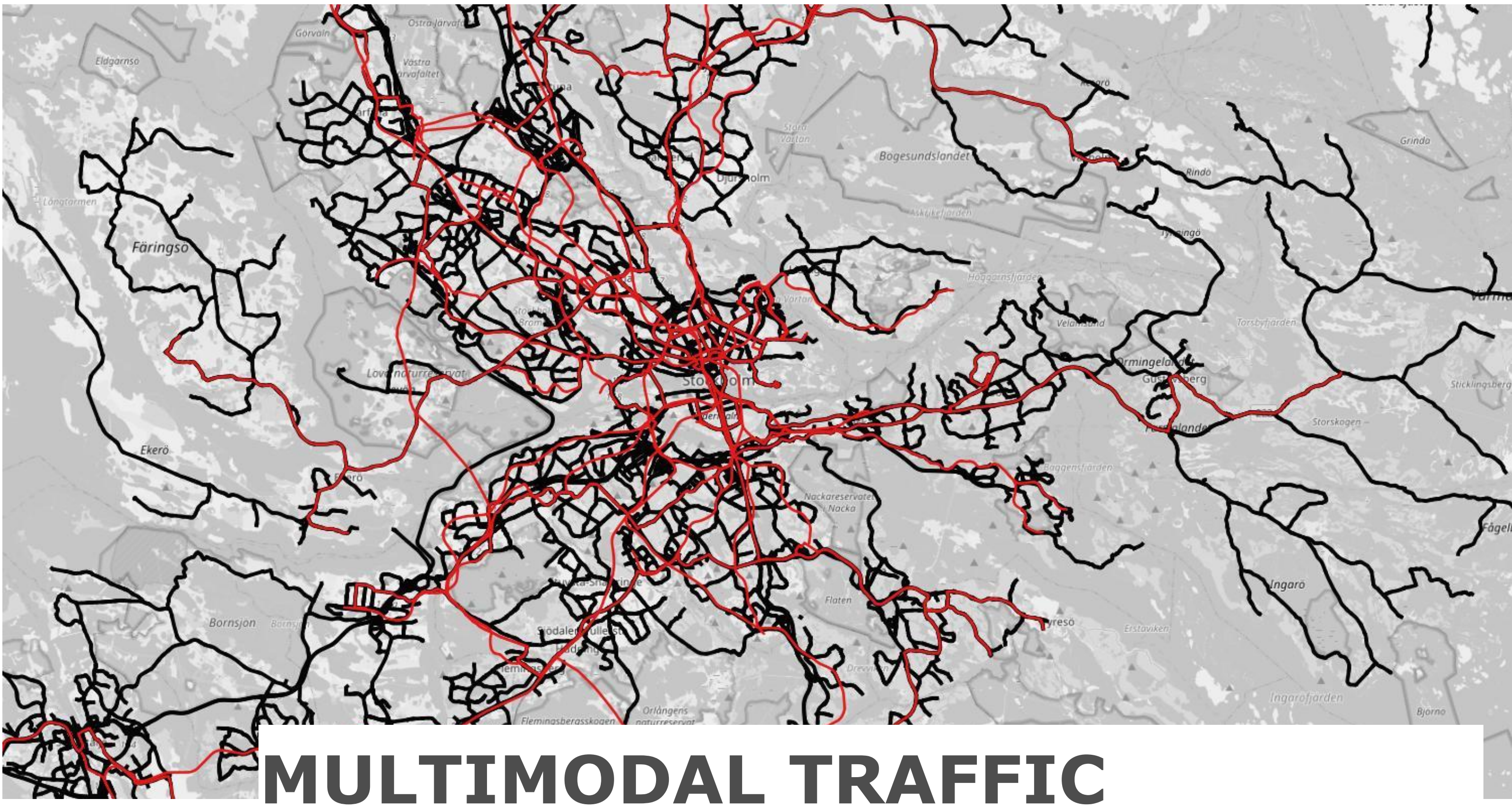

MULTIMODAL TRAFFIC MANAGEMENT

PROJECT REPORT

MULTIMODAL TRAFFIC MANAGEMENT
PROJECT REPORT

Project title	**Multimodal Traffic Management (MMTM)**
Swedish project title	**Multimodal Trafikledning (MMTL)**
Diarienummer	**TRV 2020/118663**
Authors	**Wilco Burghout, KTH** **Matej Cebecauer, KTH** **Anna Danielsson, LiU** **David Gundlegård, LiU** **Erik Jenelius, KTH** **Clas Rydergren, LiU**
Contact	**David Gundlegård (david.gundlegard@liu.se)**
Recipients	**Trafikverket** **Contact person Per-Olof Svensk**
Version	**1.1**
Datum	**2024-03-20**

SAMMANFATTNING

Nya system för att kombinera transportsätt, till exempel Mobility as a Service (MaaS), ger nya möjligheter för trafikanter att växla mellan olika färdmedel. Samtidigt ger stora mängder data från såväl kollektivtrafiknätet som vägtrafiknätet samt multimodala data från mobilnäten i kombination med nya metoder för att uppskatta resmönster uppdelat på färdmedel möjligheter till en helt ny förståelse av multimodala resmönster i en stad. Att förstå hur multimodala resmönster utvecklas över tid ger nya möjligheter att utveckla effektiva verktyg för multimodal trafikledning.

Det övergripande målet med projektet är att möjliggöra förbättrad tillgänglighet i transportsystemen genom effektivare trafikledning. Mer specifikt syftar projektet till att utveckla nya metoder för att uppskatta multimodal efterfrågan samt färdmedelsval och ruttval för multimodal trafikledning. Vidare har potentiella effekter av multimodal trafikledning analyserats.

Projektet omfattar en litteraturstudie för analys av möjligheter och utmaningar med multimodal trafikledning. En explorativ analys baserad på oövervakat lärande har utförts för att identifiera typiska nätverksövergripande mobilitetsmönster. Val av rutt och färdmedel har predikterats med hjälp av statistiska modeller. Ett multimodalt dataset för fem veckor i Stockholm med storskalig mobilitetsdata för vägnätet och biljettdata för kollektivtrafiknätet har sammanställs för den explorativa analysen samt utvärderingen av rutt- och transportsättsmodellerna i samband med trafikledning.

Baserat på litteraturstudien kan vi dra slutsatsen att koordinerad ledning av väg och kollektivtrafik har potential att minska trängseln och säkerställa effektiv förflyttning av resenärer i ett storstadsområde. Det finns flera motiv för multimodal trafikledning, där de viktigaste är potentiellt ökad efterfrågan för kollektivtrafik, förbättrad robusthet för transportsystemet och bättre prioritering av trafikledningsåtgärder. De största utmaningarna är samarbete mellan intressenter, informationsdelning och datafusion.

Resultaten av den explorativa analysen baserad på oövervakad inlärning indikerar att klustring för att ta fram typdagar kan vara användbart vid scenarioutvärdering, men också fungera som input till korttidsprediktion, vilket ger en enkel och robust predikteringsmetod för länkflöden med ett MAPE-prediktionsfel på 10-15 %.

Ruttvalsanalysen visar att en modell baserad på en ruttuppsättning med genererade rutter är mer responsiv för restidsförändringar än en modell baserad på endast observerade rutter, vilket är användbart för att förutspå effekten av olika trafikledningsåtgärder. En ruttvalsmodell med enbart restid är en vanlig förenkling att använda för att prediktera ruttval, men resultatet i denna studie visar att inkludering av fler attribut avsevärt förbättrar modellernas prestanda.

Analysen av nätverksövergripande multimodala data för 5 veckor i Stockholm indikerar att det är möjligt att uppskatta hur transportsättsandelen mellan kollektivtrafik och andra transportslag varierar i tid och rum. En bättre förståelse för spatiotemporal variation av färdmedelsvalet är en viktig input till förbättrat beslutsstöd i multimodal trafikledning.

SUMMARY

New systems for combining modes of transport, for example Mobility as a Service (MaaS), provide new opportunities for road users to switch between different means of transport. At the same time, large amounts of data from both the public transport network and the road traffic network as well as multimodal data from mobile networks in combination with new methods for estimating travel patterns divided by means of transport provide opportunities for a completely new understanding of multimodal travel patterns in a city. Understanding how multimodal travel patterns develop over time provides new opportunities to develop effective tools for multimodal traffic management.

The overall goal of the project is to enable improved accessibility in the transport systems through more efficient traffic management. More specifically, the project aims to develop new methods for estimating multimodal demand as well as mode of transport and route selection for multimodal traffic management. Furthermore, potential effects of multimodal traffic management should be analysed.

The project includes a literature survey for analysis of potential and challenges of multimodal traffic management. An explorative analysis based on unsupervised learning is performed for identification of typical network-wide mobility patterns. Route and mode choice is predicted using statistical models. A five-week multimodal dataset for Stockholm including large-scale mobility data for the road network and smartcard data for the public transport network is compiled for the explorative analysis as well as evaluation of the route and mode choice models in the context of traffic management.

Based on the literature survey, we can conclude that simultaneous management of road and public transport has the potential to reduce congestion and ensure efficient movement of travelers in an urban area. There are several motives for integrated management of multiple modes, where the most important are potential demand shifts to public transport, improved robustness for the transport system, and better prioritization of traffic management actions. The main challenges are collaboration between stakeholders, information sharing, and data fusion.

The results of the explorative analysis based on unsupervised learning indicate that day clustering can be useful in scenario evaluation, but also serve as input to short-term prediction providing a simple and robust prediction method with a MAPE prediction error of 10-15%.

The route choice analysis showed that a model based on a route set with generated routes is more responsive to travel time changes than a model based on only observed routes, which is useful for predicting the effect of traffic management actions. A route choice model with only travel time is a common simplification to use for prediction route choices. However, the result in this study shows that including more attributes significantly improves the performance of the models.

The analysis of network-wide multimodal data for 5 weeks in Stockholm indicates that it is possible to estimate how mode share between public transport and other modes of transport varies in space and time. A better understanding of spatiotemporal variation of mode share is an important input to improved decision support in multimodal traffic management.

INNEHÅLLSFÖRTECKNING

1. INTRODUCTION

1.1 Background

New systems for combining modes of transport, for example Mobility as a Service, provide new opportunities for road users to switch between different modes of transport. At the same time, large amounts of data from both the public transport network and the road traffic network as well as multimodal data from mobile networks in combination with new methods for estimating travel patterns for different modes of transport provide opportunities for a completely new understanding of multimodal travel patterns in a city. Understanding how multimodal travel patterns develop over time provides new opportunities to develop effective tools for multimodal traffic management.

Commercial companies, such as INRIX, offer large amounts of mobility data from GPS-equipped vehicles with a penetration rate of a few percent. At the same time, mobility data with lower spatiotemporal resolution but penetration rates of 20-50% and observations of all means of transport are becoming available from mobile network operators. These two large-scale mobility data sources combined with mobility data from the public transport network are enablers for multimodal traffic management.

1.2 Purpose

The overall goal of the project is to enable improved accessibility in the transport systems through more efficient traffic management. More specifically, the project aims to develop new methods for estimating multimodal demand as well as means of transport and route selection for multimodal traffic management. Furthermore, potential effects of multimodal traffic management should be analysed.

The project is documented in scientific publications and presented in academic conferences and forums targeting practitioners in the field of traffic modeling and analysis (see section 1.4). This report gives a brief summary of the results included in both scientific publications as well as conferences during the project.

1.3 Methodology

The project includes a literature survey for analysis of potential and challenges of multimodal traffic management. An explorative analysis based on unsupervised learning is performed for identification of typical network-wide mobility patterns. Route and mode choice is predicted using statistical models. A five-week multimodal dataset for Stockholm including large-scale mobility data for the road network and smartcard data for the public transport network is compiled for the explorative analysis as well as evaluation of the route and mode choice models in the context of traffic management.

1.4 Publications and presentations

This report provides a summary of the project results, for a more detailed description of the result, the reader is referred to the publications produced within the project. The project has so far resulted in one PhD thesis (Cebecauer, 2021), one journal publication (Cebecauer et. Al, 2023a), one book chapter (Danielsson, 2023), 3 conference publications (Cebecauer, et. Al, 2023b; Skoufas, et. Al, 2023; Danielsson, et. Al, 2024a) and two working papers (Cebecauer, et. Al, 2024; Danielsson, et. Al, 2024b). The project has been presented in both national conferences (Nationella transportkonferensen, Transportforum) as well as international conferences (hEART, ITSC, TRB, MFTS).

2. LITERATURE SURVEY

This is a summary of the literature study in (Danielsson, Gundlegård & Rydergren, 2024b). All references are provided in the paper.

Urban traffic management is vital in ensuring smooth and efficient traffic flows, mitigating congestion during incidents and other disturbances in the network. Today, when urbanization is increasing and populations continue to concentrate in cities, effective traffic management is more critical than ever.

Relevant literature is studied to get an overview of how proactive traffic management is used to mitigate traffic congestion and how it can be extended to include multimodal strategies. The aim is to identify synergies and challenges of integrated traffic management of road traffic and public transport, including a general definition of traffic management and multimodality.

Literature on the topic is found by combining blocks of search terms in several databases. A block of search terms includes synonyms, different spellings, and terms related to the same topic, e.g., a traffic management search term block can be the following search string:

> *(traffic W/3 management) OR (traffic W/3 control) OR (travel?er W/3 information) OR (transport* W/3 management)*

The combined search with blocks of urban traffic management, multimodality, traffic modeling, and large-scale data has a growing trend, with up to 10 times more publications than 20 years ago indicating a growing interest in the topic.

Traffic management can have different focus and include different strategies for different modes, but in general traffic management refers to operative strategies to ensure safe and efficient movement of travelers and goods. Urban traffic management focuses on the efficiency of the urban traffic network by guiding and assisting road travelers to plan their trips in case of disturbances or incidents. It can include both freight and passenger transport and several modes. Typically, the management is unimodal, with different operators focusing on one mode of transport at a time.

The road network supply (network capacities) and demand (traffic OD flows) iteratively affect the system performance, or the traffic state, of the system. Depending on the traffic state, the traffic management center decides on traffic control strategies and what information to provide to the travelers. Traffic control includes variable speed limits, traffic signal control, and ramp metering, while traveler information influences travel behavior by informing the travelers about the traffic state, current incidents, blocked lanes etc., using radio broadcasting, variable message signs, or mobile apps. The control actions and traveler information can influence flows and the choice of route and departure time to mitigate the effects of disturbances and incidents. The static road network, together with current incidents or disturbances affects the road capacity while the choice of route and departure time affects the demand.

In general, multimodality relates to the idea of combining different modes of something, primarily for economic reasons, but also to make a system more efficient, accessible, or sustainable. The term *multimodal* is used in many different fields with different meanings. There is multimodal machine learning, with models that take multiple data sources as input multimodal literacy that refers to multiple modes of expressions, such as speech, texts and gestures and fusion of multiple data sources. In a transportation context, multimodality often relates to freight transport and public transport. In freight transport, a multimodal trip means that a sequence of different modes of transport is used to deliver goods. When it comes to multimodal public transport it is often related to the seamless transfers between different modes, such as waiting time and walking distance in terminals and stations.

A mode of transport has been defined in different ways, depending on the context. It can be infrastructure-based modes, such as road, rail, air, and water; vehicle-based modes, including private car, electrical vehicle, taxi, car pool, motorbike, bus, tram, subway, regional train, etc.; or trip-based modes, where combinations of vehicle-based modes are considered one mode, e.g. bike-bus-walk, car-bus-walk. In the context of multimodal traffic management, the focus is on urban congestion and incident management. Thus, it is reasonable to include modes that share either demand or supply within an urban area. Modes that share demand are used for the same trip purpose, for example car and public transport that both are commonly used for commuting. Modes with shared supply use the same infrastructure and are therefore affected by the same incidents, for example car and bus. For multimodal traffic management, it is not necessary to distinguish between different vehicle types, such as electrical car or car pool, since they take the same road space and contribute equally to congestion, in contrary to bus and car for example. Thus, it is reasonable to distinguish between car and public transport in the context of multimodal traffic management. It can also be interesting to include active modes of transport, e.g., bike or walk, but these are currently not included in traffic management actions.

In multimodal traffic management, the multimodality is not limited to trips that use several modes in the same trip. The idea is to use the potential of a mode shift to mitigate congestion in the road network in case of incidents or disturbances. With both the road network and public transport lines as supply, and the multimodal demand, a comprehensive overview of the state can be used as basis for traffic management strategies. The traveler information and control actions issued from the multimodal traffic management center can then be based on a holistic view of the mobility patterns in the city. Based on this and previous literature, we define multimodal traffic management as

> *Multimodal traffic management is a holistic perspective on real-time urban traffic management, utilizing the potential of the complete transport system to ease congestion and provide efficient traveler mobility by simultaneously managing road and public transport traffic.*

Identified motives and synergies of multimodal traffic management in the literature include a more robust and flexible traffic system, ensuring seamless transfers between different modes, providing travelers with a holistic view of the system so they can make informed decisions, and utilizing the potential of the transport system. The hypothesis is that joint traveler information and control actions over several modes are in favor of public transport and active modes of transport. In addition, networks are naturally multimodal, where congestion in one mode can impact the flows in other modes. Thus, an overall view of the network efficiency can help prioritize between management strategies.

Challenges of multimodal traffic management identified in the literature include collaboration and data sharing between stakeholders, competition of customers, and information standards. Multimodal traffic management also adds requirements of data and traffic models to use as decision support to evaluate multimodal traffic management actions. In addition to multimodal traffic assignment, these requirements include support for traveler information, carpooling, and connections between modes, e.g, parking.

3. MULTIMODAL NETWORK AND DATA

3.1 Network description

The project focus on multimodal traffic management with focus on road and public transport traffic. The road network used in the project is illustrated in Figure 1, whereas the public transport network is illustrated in Figure 2. Furthermore, Table 1 summarizes the public transport stops, while Table 2**Fel! Hittar inte referenskälla.** summarizes the public transport lines.

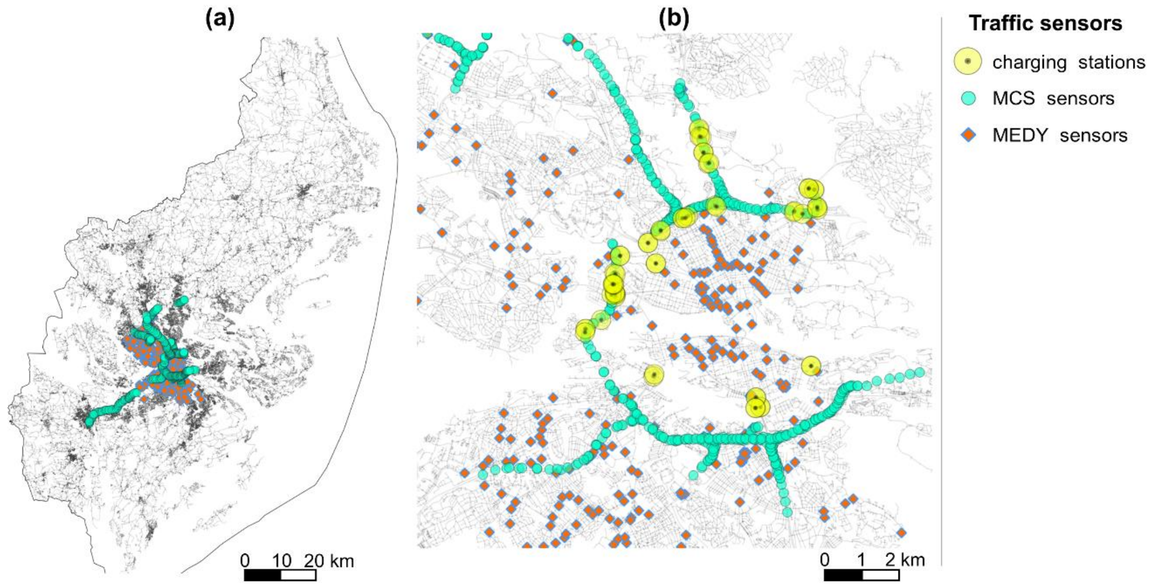

Figure 1 Traffic sensors on road network (a) Stockholm region (b) City center area

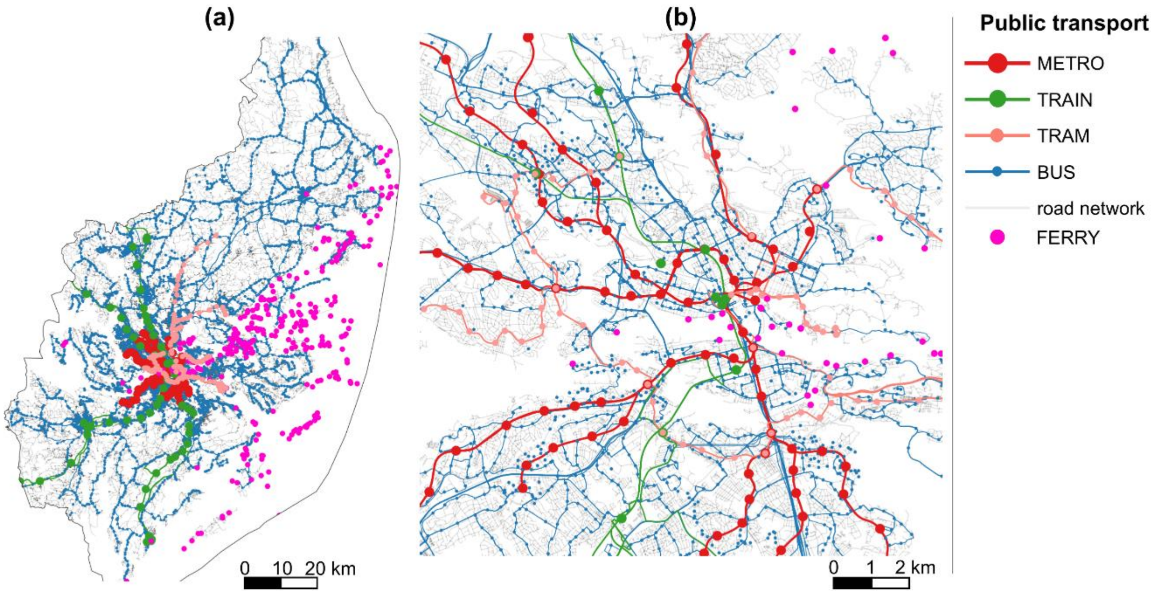

Figure 2 PT network with stops and lines (a) Stockholm region (b) City center area

Table 1 Public transport stops summary

Mode	Number of stops	Share
Bus	6754	91.28%
Metro	112	1.51%
Train	61	0.82%
Tram	130	1.76%
Ferry	342	4.62%

Table 2 Public transport lines summary

Mode	Number of lines	Share
Bus	585	86.28%
Metro	8	1.18%
Train	15	2.21%
Tram	11	1.76%
Ferry	59	8.7%

3.2 Data sources

3.2.1 Link flow measurements

The link flow measurements dataset contains data from both MCS radar stations and temporary link counts. Totally 220 000 link count observations from 700 different sensors are available. The locations of the sensors are shown in Figure 3.

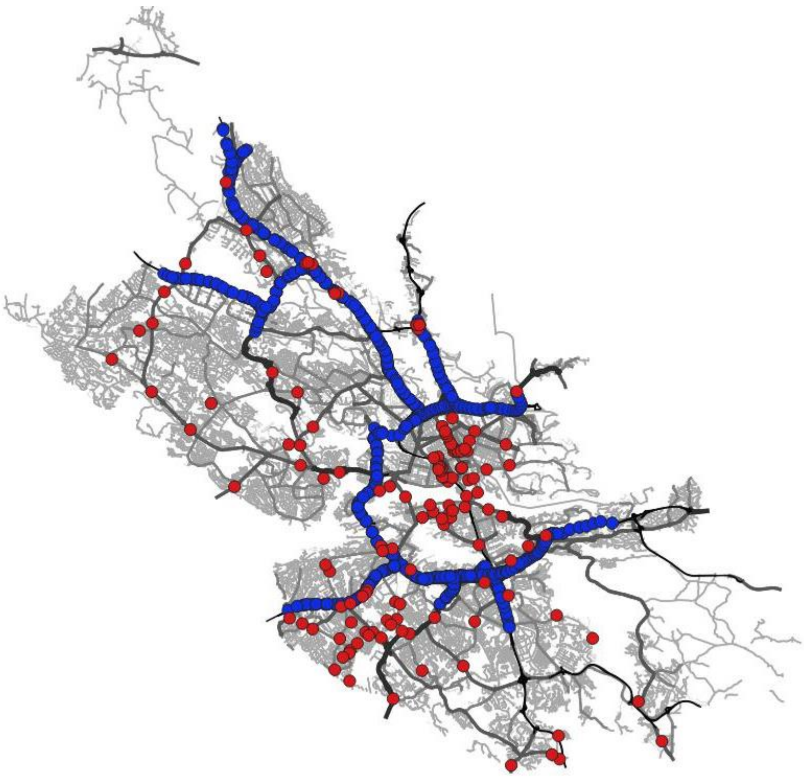

Figure 3 Sensor locations for link flow measurements used in the project.

3.2.2 Mobile network data

Anonymized and aggregated mobile network data is provided by the mobile operator Telia through their service Telia Crowd Insights (Telia, 2023). The observed all-mode (walking, biking, micro-mobility, shared vehicles, car, trains, and public transport) travel data that is used is scaled to represent the full population. A more detailed description of the type of data used is available in (Ågren, Bjelkmar & Allison, 2021). Crowd Insights data is available in many different spatial resolutions, we have chosen the OD zoning illustrated in Figure 5 with 31 zones in the area. In this project we have mainly used the routed OD matrix classified as "ROAD" traffic in Crowd Insights.

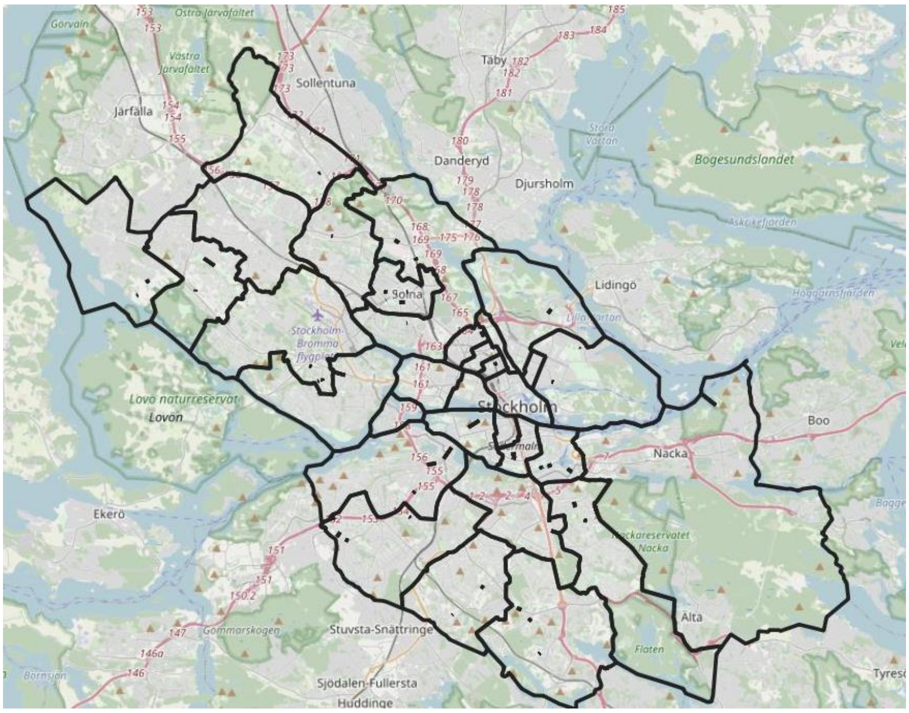

Figure 4 Zoning of the mobile network data from Telia Crowd Insights used in the project.

3.2.3 Public transport smart card data

Automated Fare Collection (AFC) data, which we also refer to as smart card data, consists of all recorded ticket validations made by PT smart cards, mobile phone apps, and debit/credit cards. This represents all the ticket validations in the Stockholm Region. These validations are anonymized. In total there is about 2 million trips per working day. The PT mode shares are shown in Figure 5.

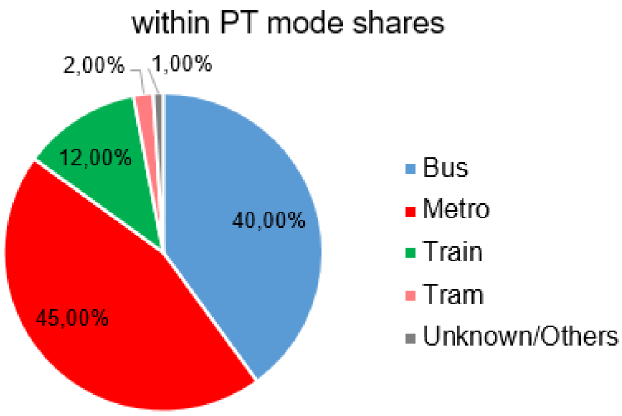

Figure 5 Within public transport mode shares

The AFC system in the Stockholm region is a tap-in only system. To have complete origin-destination travel diaries, the inference of tap-out location, destination location, line/departure/vehicle must be performed. Here, the AFC is merged with automated vehicle location (AVL) data and GTFS timetables. An illustration of a complete journey diary is shown in Figure 6.

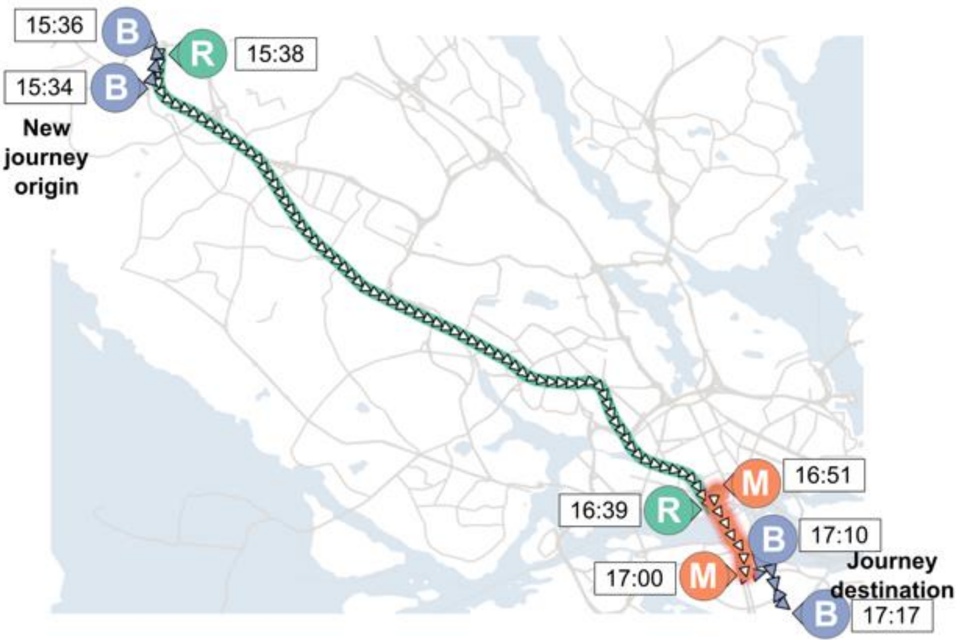

Figure 6 Example of travel diary of one jorney with sub-trips and tap-in and tap-out timestamps

R – train, M – metro, B – bus mode

The inference framework for Stockholm Region was developed in FairAccess project; thus, for more details about processing AFC data and travel diaries to origin-destinations, we refer you to the FairAccess project report (Cats et al. (2019)). The PT diaries enables extraction of routes, origin-destination flows, delays and crowding in the PT network.

3.2.4 GPS probe data

A detailed description of the detailed GPS probe data from INRIX that is used in this project can be found in (Ahlberg et. Al, 2021). The Stockholm dataset contains 421 000 trips that passes by the chosen area, where 323 000 trips start inside the area and 324 000 trips end in the area. Figure 7 shows the spatial distribution of trips in the area.

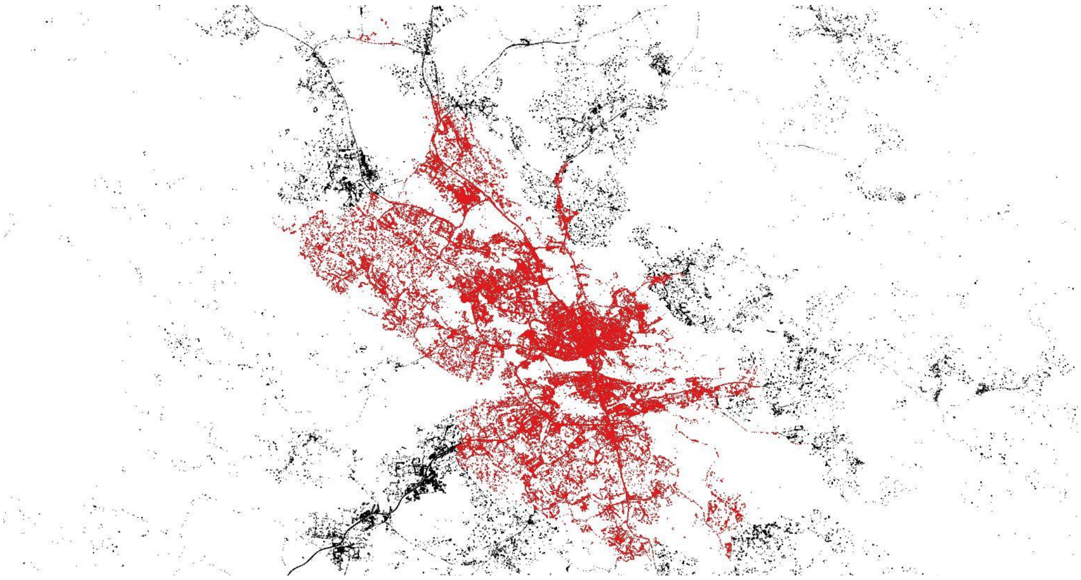

Figure 7 Start points of trips starting inside the area (red) together with start points for nearby trips (black).

4. TYPICAL DAYS

This chapter aims at understanding and identifying typical mobility patterns in multimodal traffic networks. We use clustering to identify the most representative or “typical” days, referred to in the report as day-types. We focus on day-type patterns since days represent the time period where we can assume that the travel demand of the general population restarts. The definition for the representative day-type cluster is that it groups days with similar network-wide patterns into one group to decrease the dissimilarity within the group, but increase dissimilarity towards the other groups. The detailed methodology for revealing representative day-types in transportation networks is presented in Cebecauer et al. (2023a). There are many clustering methods with various objectives that could be considered, their selection and definition is provided in Cebecauer et al. (2023a) and its supplement information file.

The first part of this chapter includes analysis of one year of sensor data from the Stockholm highway network and public transport smart card data to enable analysis of seasonal patterns. The second part includes an analysis of five weeks of data during autumn 2019 for many different sensors, including both public transport data, highway sensor data, mobile network data, congestion charging data and GPS probe data.

4.1 Yearly day-type patterns

4.1.1 Yearly MCS highway traffic day-types

Cebecauer et al. 2023a shows up-to 20 day-type clusters for Flows, and Cebecauer et al. 2024 for both flows and speeds using various clustering methods. Cebecauer et al. 2023a give a comprehensive overview of revealing representative day-types using case-study of Stockholm MCS network of sensors considering one-year 2017 observations and using these day-types as a source of predictions for 2018 to identify the most representative clusterings. Figure 8 shows 6 day-types recognized using flow and speed observations. It reflects expected weekends, day-of-week, holidays, and seasonal impacts. Importantly, it creates these groups based on their similarities in traffic observations without any calendar or expert knowledge.

In Cebecauer et al. 2024 we looked at interchangeability of speed and flow data types for prediction application in ITS. The results show that these two data-types can be substituted to some extent.

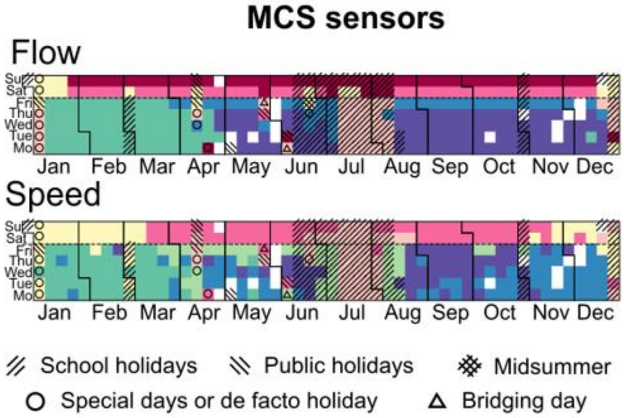

Figure 8 Yearly MCS representative day-types

In Figure 9 (a) we show profiles of day-type centroids. Day-type centroid is in this case the average vector of all days in the day-type cluster. Figure 9 (b) shows calendar visualization of day-types. Figure 9 (c) illustrates space-time values of day-type centroids and Figure 9 (d) put these centroids in the context of the day-type centroid with the highest total daily flow.

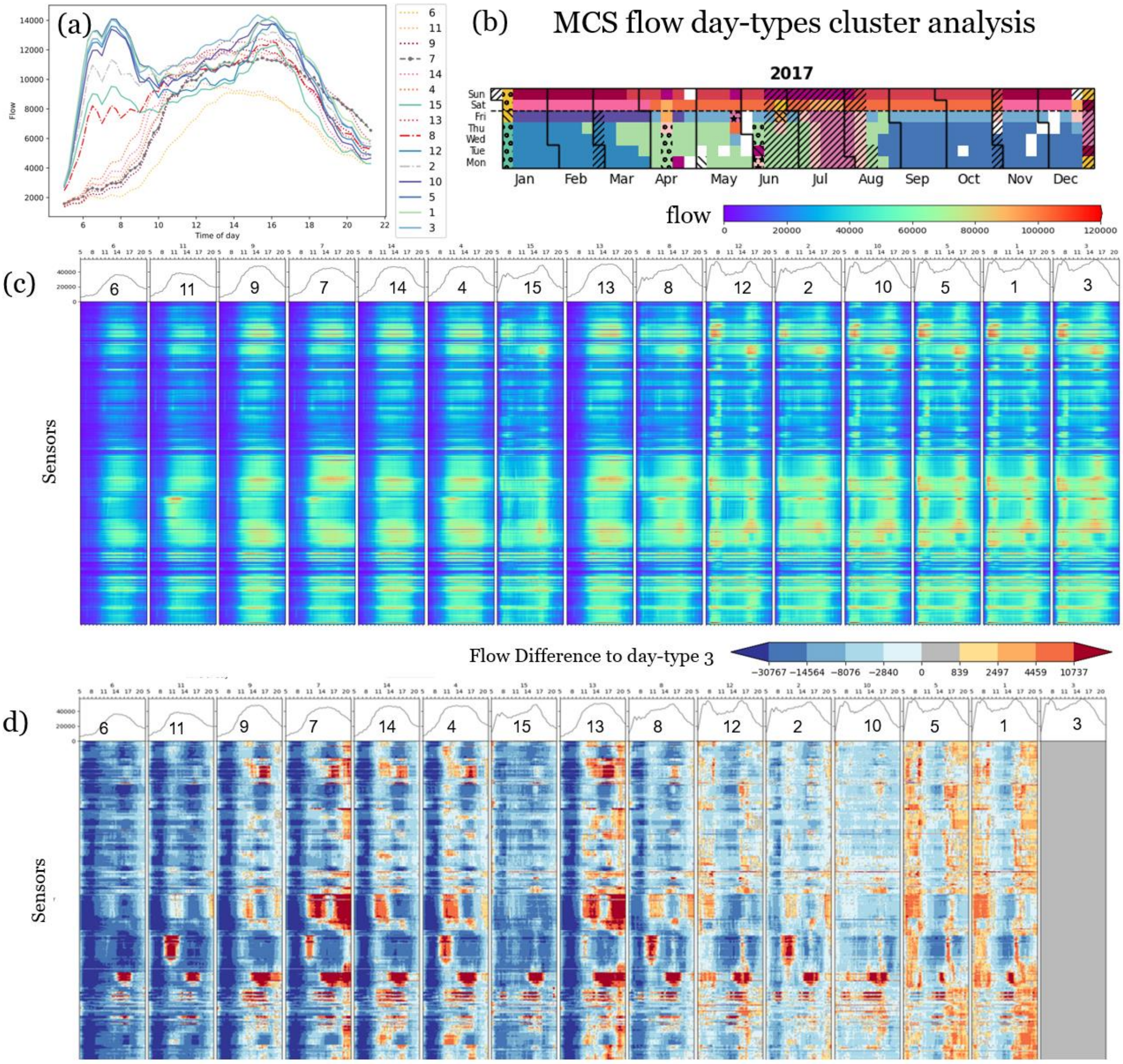

Figure 9 Day-time profiles and space-time heatmaps for yearly 15 MCS flow day-types. (a) Day-time average centroid profiles. (b) Calendar visualization of day-type clusters. (c) Space-time heatmaps of MCS flow day-types. (d) Difference of day-types to the day-type with highest total daily flow (cluster 3)

The results show that there are clear network-wide clusters for different days of the week (Monday-Thursday, Friday, Saturday and Sunday) as well as for different seasons and holidays. The results indicate that day clustering can be useful in scenario evaluation, but also as input to prediction providing a simple and robust prediction method with a MAPE prediction error of 10-15%. Several of the clustering methods evaluated in the project produce both promising and similar results, including K-means clustering and Agglomerative clustering. Clustering based on speed and flow produces similar results in terms of typical days. However, smaller changes in flow cannot be detected when the network operates in free flow conditions, which is natural given the flat speed-flow relationship for low flows. These small changes can be important for scenario evaluation during incidents.

4.1.2 Yearly OD public transport day-types

In this project, we have also looked at representative day-types in large PT dynamic OD matrices considering year 2017. To compare with the results generated by MCS datasets, Figure 10 shows 6 PT day-types. It also reflects weekends, day-of-week, holidays, and seasonal impacts comparable to MCS day-types. Thus, there is an indication of similarity in day-type level patterns on a multimodal level, which can have applications for data-driven multimodal traffic management and motivate future multimodal analysis of peaks, disruptions, and incidents.

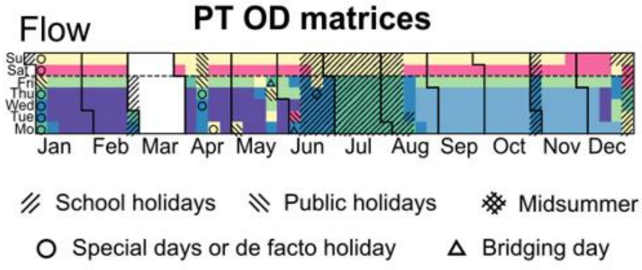

Figure 10 Yearly OD public transport day-types

4.2 5-week case-study of multimodal data

4.2.1 Case-study introduction

In this section, the representative day-types, for the case-study area (see Figure 11) and period of 5 weeks, are revealed for all data sources introduced in subsection 3.2. Table 3 summarises the case-study datasets, public transport smart card data (PT), mobile network data from Telia Crowd Insights (Telia), and GPS probe data from INRIX, where we reveal day-types considering originating flows in 36 considered zones. Other datasets (MCS, Toll, Medy) count vehicle observations using sensors located on infrastructure. We refer to various sensors or zones for which flow data-type is collected as observation entities.

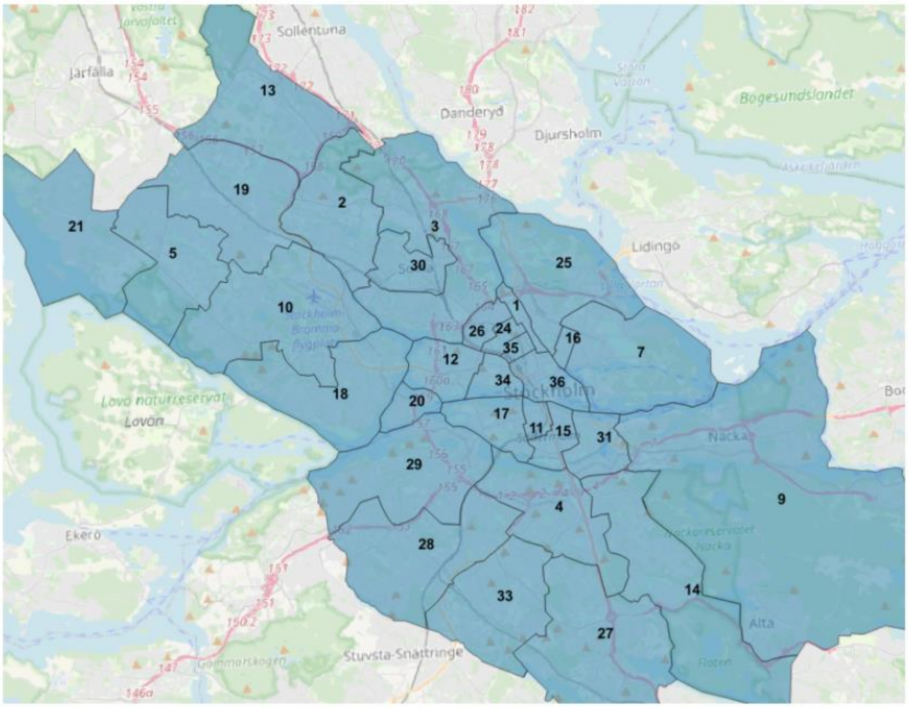

Figure 11 Case study zones: 1:Stockholms Sankt Johannes, 2:Sundbyberg, 3:Solna, 4:Enskede, 5:Vällingby, 7:Stockholms Oscar, 9:Nacka, 10:Bromma, 11:Stockholms Maria Magdalena, 12:Stockholms Sankt Göran, 13:Kista, 14:Skarpnäck, 15:Stockholms Katarina, 16:Stockholms Hedvig Eleonora, 17:Högalid, 18:Västerled, 19:Spånga, 20:Essinge, 21:Hässelby, 24:Stockholms Gustav Vasa, 25:Stockholms Engelbrekt, 26:Stockholms Sankt Matteus, 27:Farsta, 28:Brännkyrka, 29:Hägersten, 30:Råsunda, 31:Stockholms Sofia, 33:Vantör, 34:Kungsholmen, 35:Stockholms Adolf Fredrik, 36:Stockholms domkyrkodistrikt. Background map: OpenStreetMap.

Dataset	Observation entity	Aggregated entity	Time interval	Number of entities	Data type
Telia	Mobile phone	zone	1 hour	36	count (originating flow)
PT	All ticket validations	zone	1 hour	36	count (originating flow)
Inrix	GPS	zone	1 hour	36	count (originating flow)
MCS	Microwave sensor	sensor	1 hour	572	count (passing flow)
Toll	Toll portal	sensor	1 hour	57	count (passing flow)
Medy	Loop/Microwave sensor	sensor	1 hour	157	count (passing flow)

Table 3 Case study datasets summary

4.2.2 Estimating PT mode share – multimodal context

The available datasets of mobile phone data (Telia) that estimate total travel origin-destination demand and PT that capture almost complete PT origin-destination demand enable us to place PT demand in relation to all other modes and capture multimodal context. Note that this approach has certain limitations; PT data are limited to PT mode. This means that it considers PT trips without having knowledge if these are part of longer multimodal origin-destination journeys with preceding or continuing the PT trip with other modes such as walking, biking, scooters, etc. For some small zones in the city center, such as 1, 16, and 24 (see Figure 5), people can walk to the high-capacity metro service in neighboring zones, which the PT dataset cannot capture. Telia data can capture complete and complex multimodal journeys. However, estimating modes, in dense urban areas, using mobile phone data can be very challenging (Breyer et al. (2021)), and the PT dataset used in this project gives reasonably accurate PT demand.

Considering above, we investigated in Cebecauer et al. (2023b) the potential of combining mobile phone data and smart-card PT data for very cost-efficient high-resolution data-driven estimation of PT mode share by simple interpolation:

$$PT_{share} = \frac{PT}{Telia}$$

This approach, in comparison to traditional surveys, have several pros:

- Data are already being collected (no additional cost for data collection)
- These can be a large and rich historical databases (analysis of past years)
- High space and time resolution (space and time dynamics)
- Low costs

However, there are also cons:

- Anonymized data
- No additional questions
- The traveler's background information is unknown.
- Limited to observed travel.

A combination of data-driven and traditional travel survey approach have potential for more comprehensive understanding how and why people travel. According to Swedish agency for transport policy analysis (Transport Analysis (2022)) there is a continuous decline in respondent rates, from 68% in 2005 to 28% in 2021. It is also concerning whether the sample of respondents is truly representative of the population. According to Cebecauer et al. (2023b) this data-driven approach results in expected patterns comparable with travel surveys on a very aggregated level, which is a promising alternative to costly travel surveys that enable higher resolution analysis.

Cebecauer et al. (2023b) proposed the methodology for estimating background socio-economic information for this data-driven approach that would be otherwise unknown. This observed travel is enriched by merging these data with socio-economic-pt-supply statistics. Regression analysis of PT mode shares with these attributes shows high explanatory and predictive power.

For more details about data-driven PT mode share estimation, spatio-temporal and cluster analysis we refer you to Cebecauer et al. (2023b).

4.2.3 Representative day-types

In this report and Figure 12 we show the representative day-types revealed by using agglomerative clustering with ward-linkage and Euclidean distance as similarity/dissimilarity measure. Figure 12 shows the evaluation of day-types based on their significance. The first row shows the three most dissimilar groups, the second row adds two following dissimilar groups, and so on. Note that agglomerative clustering starts with clustering when each day is one group and combines the two most similar/closest groups in each step and creates hierarchies of clusters in a relationship. Thus, going from 10 to 9 clusters would mean merging two existing groups; in other words, 10 clusters are created as a split of one cluster from 9 clusters. These attributes of hierarchical agglomerative clustering make it a good candidate for cluster analysis of day-type patterns.

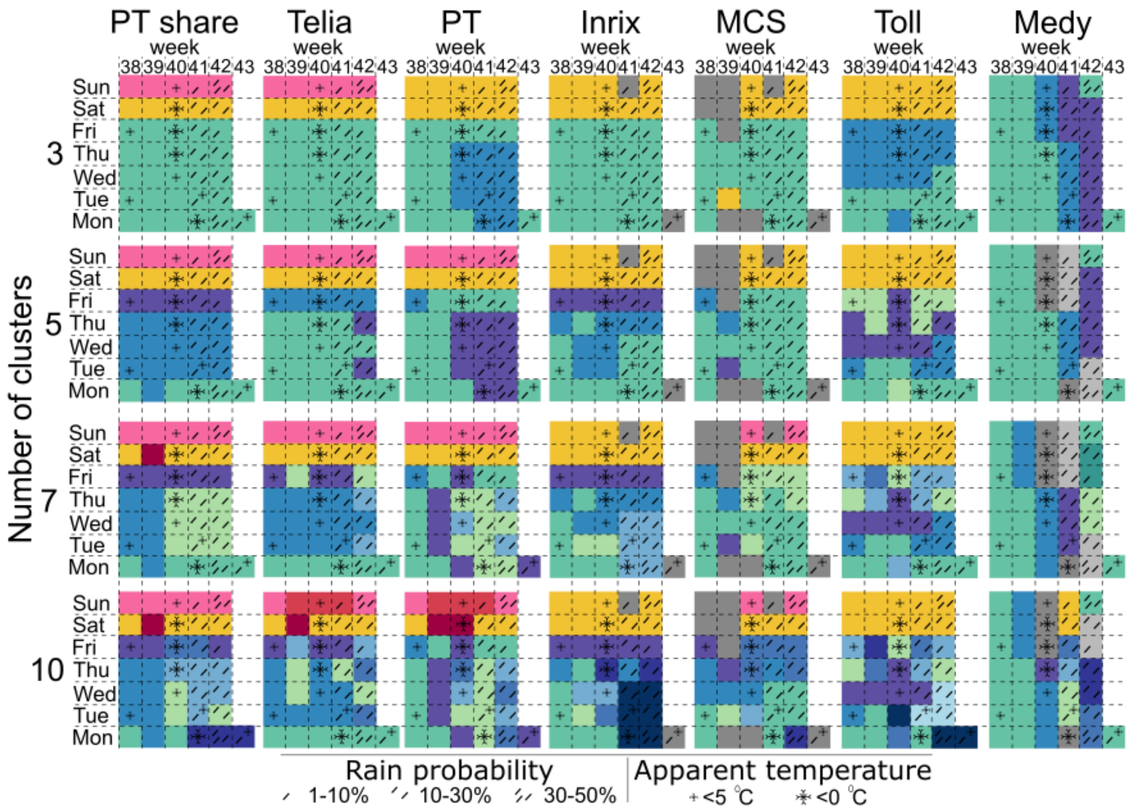

Figure 12 Case-study period representative day-types. Each color is one day-type cluster.

The most distant-looking day-types among datasets are Medy data; it is shown in Figure 13 that day-types reflects activation and deactivation of sensors. Medy sensors are temporally located and relocated in certain intervals, thus hardly comparable to other datasets. The most apparent day-type groups are weekdays and weekends. For certain datasets such as INRIX and MCS data, gray color clusters are outlier days with missing observations grouped in one cluster. Interestingly, for PT mode share and Telia data, the split of weekend group to Saturdays and Sundays is of higher importance compared to other weekdays. For the PT dataset, the rainy and cold weather during weekdays is of higher importance compared to Saturdays and Sundays. The weather pattern also gets recognition by reserving one day-type in PT mode share and INRIX day-types if there are more than 7 clusters. In both PT share and Telia data, we can see one day (28th September) cluster (dark red cell in Figure 12) when there was a large memorial ceremony related to the passenger ferry M/S Estonia ship disaster.

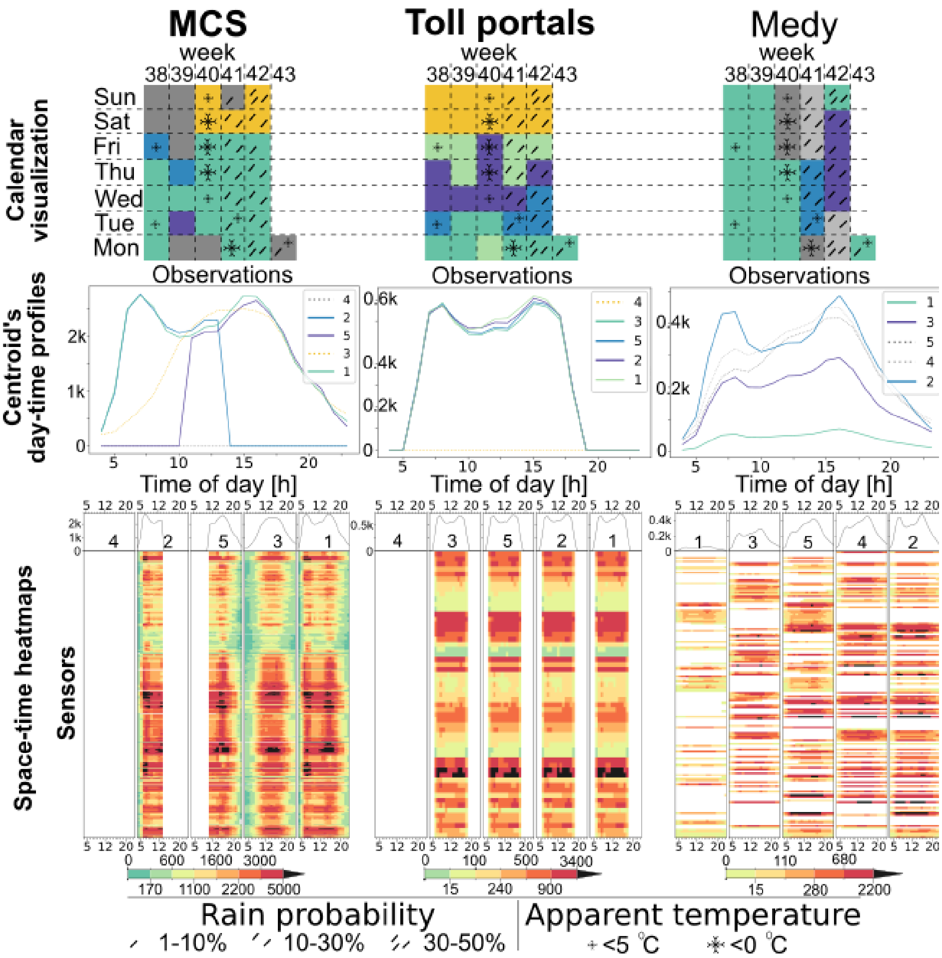

Figure 13 Sensor day-type day-time profiles and space-time heatmaps

There are also other groups of outlier days, such as cluster 2 (blue group of two days) for MCS sensors (see Figure 13), with no observation collected on any of the sensors in the afternoon. Cluster 5 of one day (purple group) has no observation collected in the morning. Weekends for Toll data are without observations as the system is not collecting tolls during the weekend. For each dataset in Figure 13 the spatiotemporal differences among the different day-type observations are apparent. However, note that only the day-time dynamics of these datasets can be compared since the spatial perspective in heatmaps is unique for each dataset.

In Figure 14 heatmaps, the Y-axis is the same for all the datasets and thus enables space-time comparison of originating flows in the zones and time. Zones on Y-axis for each dataset are ordered based on their distance to the central zone 36. When it comes to outliers, only in the INRIX dataset there are two days (gray cluster 3) with almost no OD flow observations. For most datasets, in Figure 13 and Figure 14, we can see in day-time profiles and space-time heatmaps the presence of morning and afternoon peaks and no morning peak during weekends. PT, INRIX, and especially Telia show a lunch peak. Interestingly, the lunch peak is absent in the PT mode share that combines Telia and PT datasets. PT mode also has higher ridership during cold and rainy days (purple cluster 3 in Figure 8), attracting most likely people who would bike or walk.

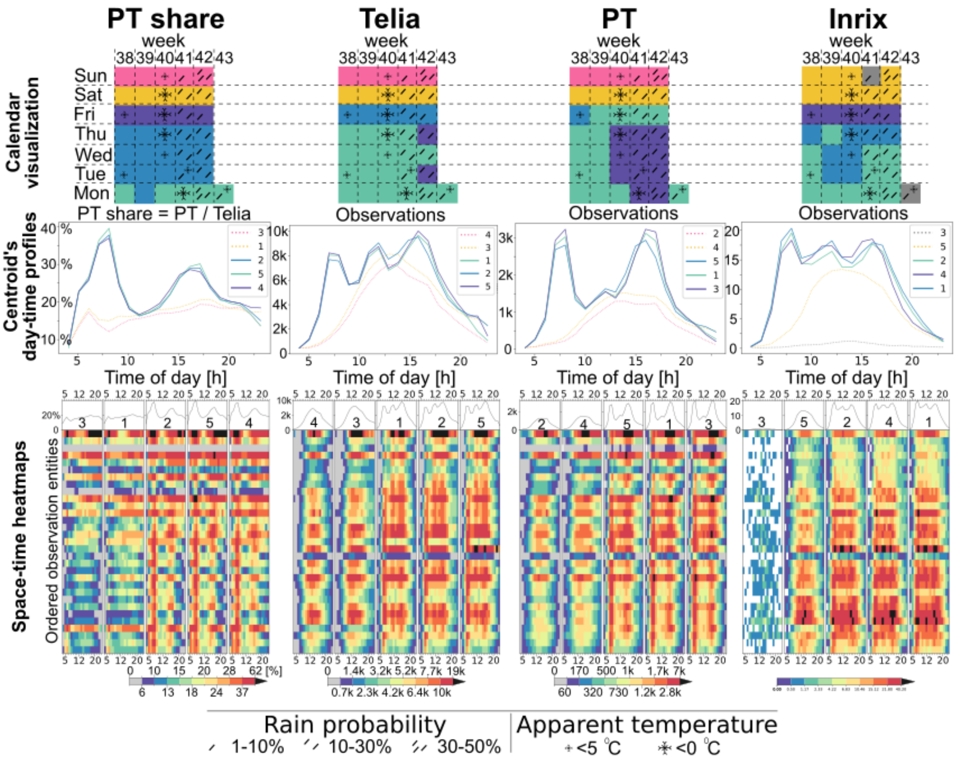

Figure 14 Originating flows per zone day-type day-time profiles and space-time heatmaps

5. ROUTE CHOICE ANALYSIS

This chapter presents the route choice analysis, both for public transport and road network, including route set generation and route choice modeling.

5.1 Route choice analysis for public transport

One part of the project includes generation and analysis of route sets for public transport travelers. Smart-card data travel diaries allow us to infer observed routes, their capacity and flows. Figure 15 is an illustration of observed multimodal public transport route alternatives traveled by travelers between two selected areas. However, the limitation of considering only observed routes is that other non-observed but viable alternatives for disruption events that travelers do not usually take will be absent in these route sets. For this reason, we investigated the generation of logical-feasible route alternatives in the public transport multimodal network. A more comprehensive description of this work is available in (Skoufas et. Al, 2023).

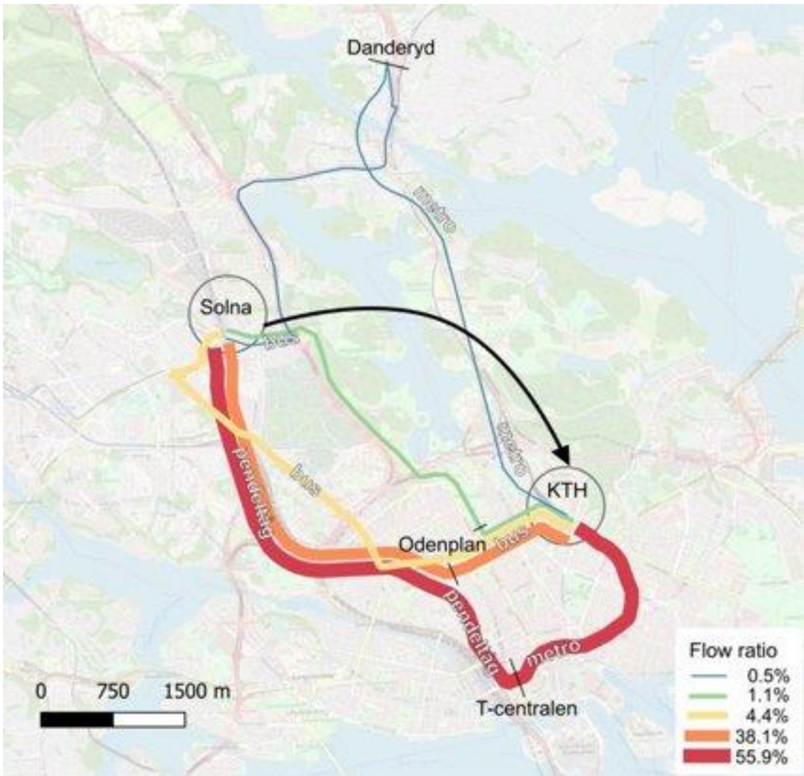

Figure 15 - Example of public transport OD-pair traveled routes

5.2 Route choice analysis for road transport

5.2.1 Route set generation

Two approaches are used for route set generation in the road network, described in more details in (Danielsson, Gundlegård & Rydergren, 2024a). One data-driven path identification (DDPI) choice set, constituting observed routes, and one where the DDPI choice set is augmented with routes from a shortest path search with link penalty (network-based path augmentation, NBPA). Figure 16 shows an example OD-pair with DDPI routes in blue and NBPA routes in orange.

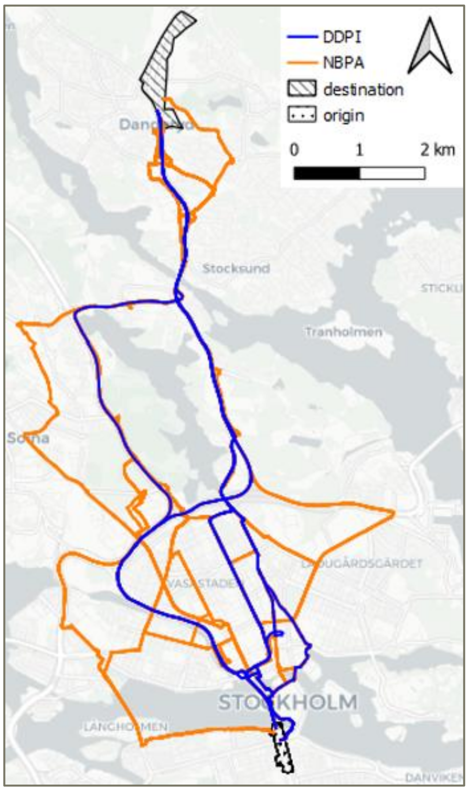

Figure 16 - Example OD-pair with DDPI routes in blue and NBPA routes in orange.

The DDPI choice set is constructed from trajectories in the GPS probe data from INRIX. The routes from up to 200 vehicle trajectories are extracted in each OD-pair. The first two weeks of the data set constitute a training data set that models are estimated based on and the following two weeks a test data set that is used for evaluation. The number of routes in the training data in 127 OD-pairs is compared to the number of routes in the training data in the left part of Figure 17. Since there is a correlation, the number of routes seems to be similar over the two time periods. The right part of Figure 17 presents the number of routes in the training data that are also observed in the test data with different exclusion thresholds. With exclusion threshold 0, all routes are included in the comparison, with exclusion threshold 1, all routes that only were observed once are excluded. Accordingly, with exclusion threshold 2, all routes that only were observed twice were excluded from the comparison. With exclusion threshold 1, over 90 % of the training routes are observed also in the test data, thus the route set can be considered stable over the four weeks.

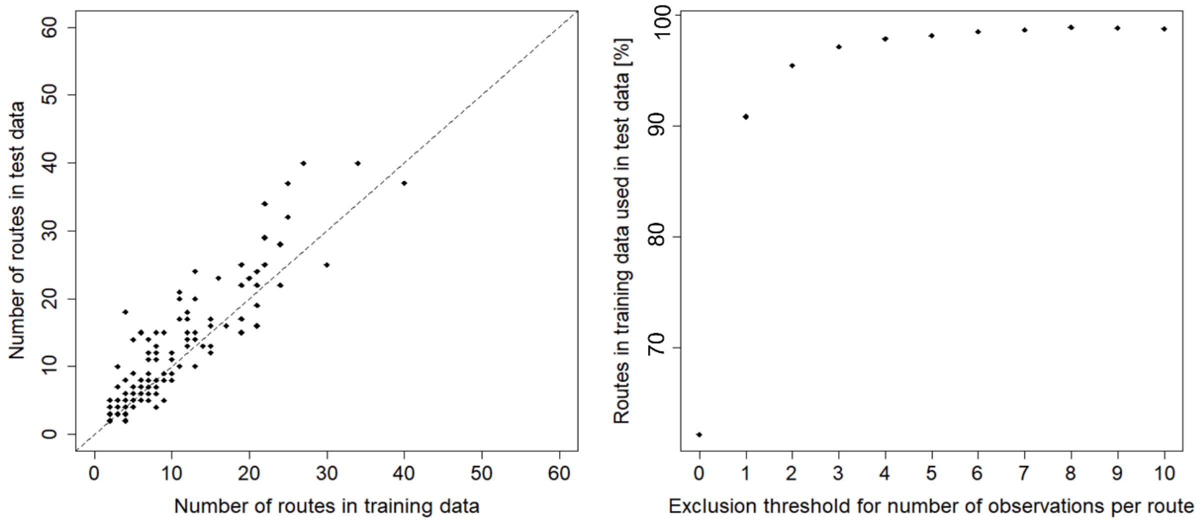

Figure 17 - Number of routes per OD-pair in training vs test data to the left and percentage of routes in training data that are observed in test data as well with different exclusion thresholds to the right.

In the DDPI route set, the routes are filtered out if i) the route is more than 50 % longer than the calculated shortest path, ii) if the route is observed less than 2 times or by less than 3 % of the travelers in the data set, or iii) if the route is too similar to another, already added route in the choice set. The last condition is to avoid alternatives being too similar in the route selection estimation. Instead, the routes are considered the same if the commonality factor for any route is larger than -0.1. The commonality factor is calculated as

$$CF_{k,l} = ln\frac{L_{l,k}}{\sqrt{L_l L_k}}\ \forall k \in K$$

L_{ik} is the length of the links that route i and k have in common, and L_i and L_k are the lengths of route i and k respectively. K is the choice set in the current OD-pair.

The DDPI route set constitutes historic routes, thus where there are few observations or when there are disruptions in the network, it can be questioned if all routes are observed. Another drawback of the method is that the routes in the choice set have all been used at least twice. The routes are therefore all reasonably good and there are no routes in the choice set that were considered but not chosen. That all routes in the choice set have been chosen can have unexpected consequences in the model. For example, if all routes in an OD-pair have the same traveltime, the traveltime will not have an impact on the choice. Intuitively, we know that is not the case in reality though. To account for these drawbacks, the NBPA choice set generation method was implemented.

In the NBPA choice set, the DDPI choice set is augmented with routes from an iterative shortest path search with a link penalty added each iteration, calculated as

$$\boldsymbol{p(a) = \frac{l_a}{\mu L} ln \sum_{j \in C_w} \delta_{aj}}$$

a is the current link, l_a is the free flow traveltime of link a, μ is a parameter regulating the magnitude of the penalty (0.4 was is in this study), L is the minimum traveltime between origin and destination, C_w is the set of generated routes in current OD-pair at the end of iteration w and δ_{aj} equals 1 if link a is in route j, 0 otherwise. Routes were added to the route set until they were more than 80% longer than the shortest path in the OD-pair. In average the NBPA choice set has 43.6 routes per OD-pair, while the DDPI choice set has 6.8 routes per OD-pair.

5.2.2 Route choice modeling

The road network route choice is modeled using logit-based discrete choice models, where the probability of alternative $i \in C$ (choice set) being chosen by traveler $t \in T$ (set of travelers in the OD-pair) is calculated as

$$\boldsymbol{P_{it}}(\boldsymbol{\beta}) = \frac{e^{V_{it}}}{\sum_{j \in C} e^{V_{jt}}}$$

Each alternative is associated with a utility function, describing the alternative. The utility function, V_{it}, for alternative $i \in C$ for traveler $t \in T$ is based on a number of attributes, x_{ikt}, $k \in K$ (set of attributes) is formulated as

$$\boldsymbol{V_{it}} = \sum_{k \in K} \boldsymbol{\beta_k x_{ikt}}$$

The weights of the attributes, β, are estimated using maximum likelihood estimation as

$$\boldsymbol{\beta}^* = \mathbf{argmax}_{\beta} \sum_{t \in T} \sum_{i \in C} \boldsymbol{\delta_{it} ln} \left(\boldsymbol{P_{it}}(\boldsymbol{\beta})\right)$$

where δ_{it} is 1 if traveler t chose alternative i, otherwise 0.Attributes included in the utility function in this study are inspired by route choice modeling literature, examples of common and less common attributes seen in the literature are presented in Table 4.

Table 4 – Attributes used in route choice literature based on passive data with example references.

Common attributes	Traveltime Route length Road type Number of turns or crossings Overlap variable (path size or commonality factor)	e.g. Bovy et al. 2008 Montini et al. 2017 Deng et al. 2021
Less common attributes	Speed Route delay Euclidean distance in OD-pair Percentage of route within city center Variance of traveltime	e.g. Yao & Bekhor 2020 Dhakar & Srinivasan 2014 Mai et al. 2021

In this study, attributes are constructed from the GPS and network data, presented in Table 5. Thus, there is no socioeconomic information about the travelers.

Table 5 – Attributes constructed in this study divided into five categories.

Traveltime	Mean traveltime over 4 time periods Free flow traveltime 0/1 variable long traveltime
Route length	Route length in meter 0/1 variable long distance
Delay	Relative delay Traveltime variance
Route simplicity	Percentage major roads Number of intersections Number of left turns Congestion charge
Overlap	Path size factor to account for overlap between alternatives

The attributes are divided into four main categories, i.e., traveltime, route length, delay, and route simplicity. An overlap attribute accounting for similarities among the alternatives is also included. The traveltime or route length are considered "long" if they are more than 50 % longer than the shortest path. The relative delay is the mean traveltime subtracted by the free flow traveltime, divided by the free flow traveltime. The roads considered "major" are shown in Figure 18, together with congestion charging stations and the 147 OD-pairs used in the choice estimation. The mean traveltime, the delay and the congestion charge vary depending on when the trip starts.

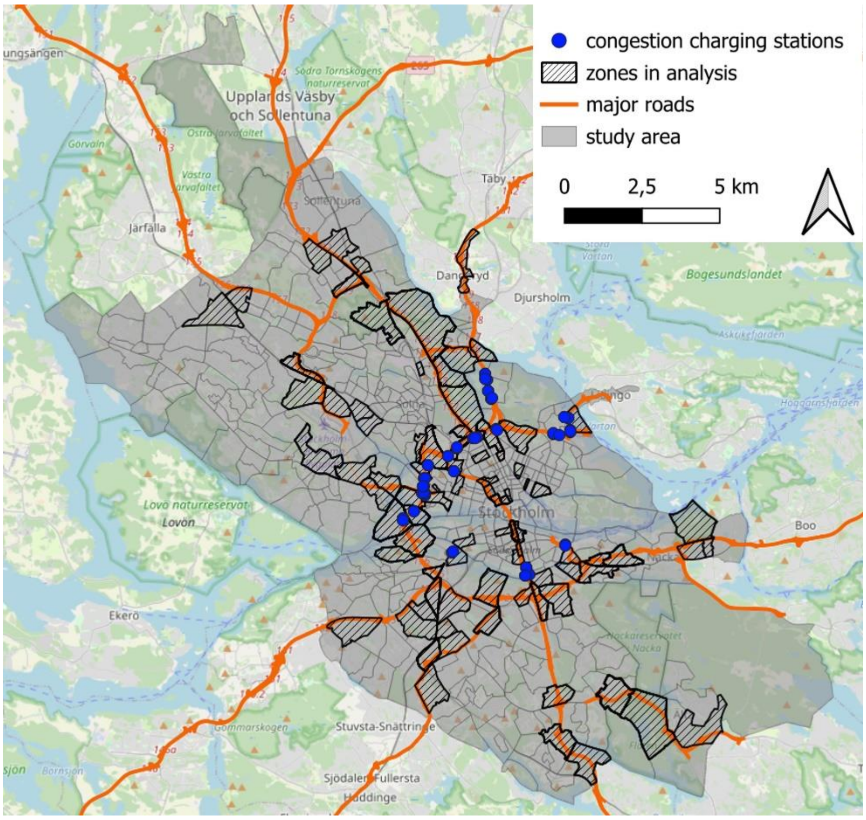

Figure 18 – Study area with major roads, congestion charging stations and zones in the included OD-pairs.

When used as a single attribute in the model, increasing route variance increases the utility of the route, which is not in line with the expected effect of the attribute. Therefore, the route variance is excluded. In addition, Figure 19 shows that the free flow traveltime, the route length and the number of links all correlate strongly with the mean traveltime. They are therefore also excluded from the further analysis of which attributes to include in the model.

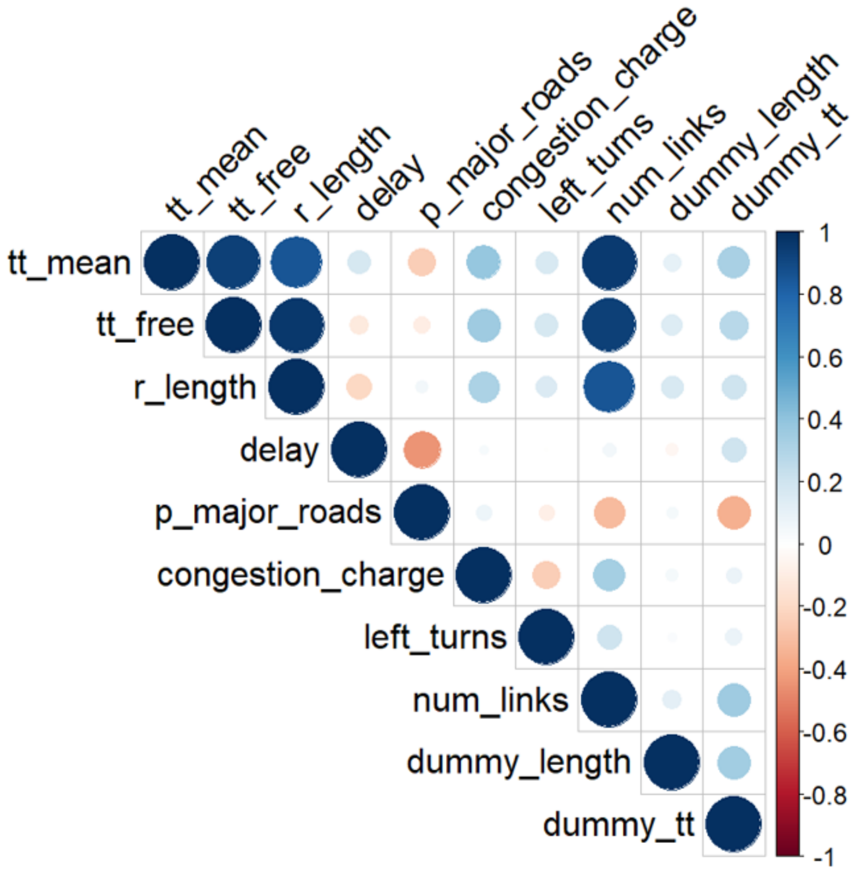

Figure 19 – Correlation matrix over the attributes. A larger dot indicates higher correation between the attributes.

A forward stepwise selection process, where one attribute at a time is added to the model, is used to identify the attributes with the largest impact on the route choice. The attribute improving the log-likelihood the most is added each iteration. The log-likelihood of each iteration is shown in Figure 20 for both the NBPA and the DDPI choice set.

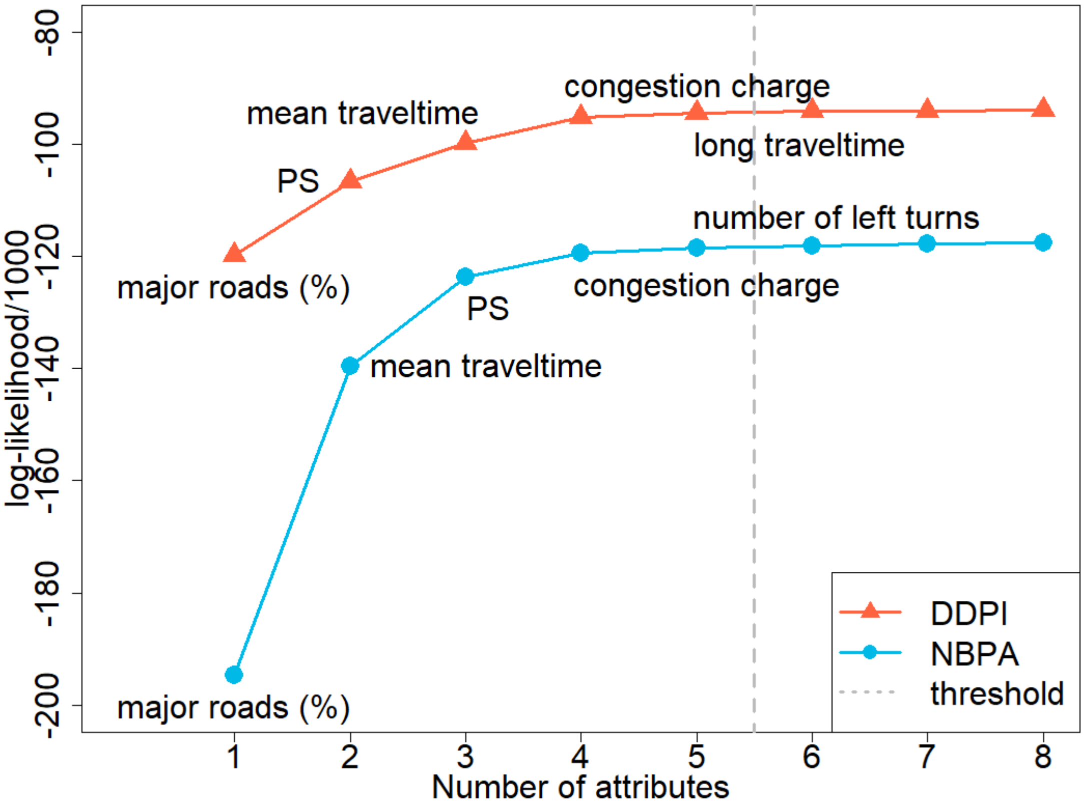

Figure 20 – Log-likelihood of models with limited number of attributes

Major roads, mean traveltime, cost of congestion charge and number of left turns are identified as important attributes to make the route choice. It can be noted that the mean traveltime is added earlier when the models are estimated on the NBPA choice set, which suggests that traveltime plays a larger role in those models compared to the DDPI models. This is expected since that is one of the motives of adding generating routes to the choice set.

A model based on the forward stepwise selection was compared with two naïve models; one model with free flow traveltime and one with no attributes. The models are presented in Table 6 with the attribute weights, β, training metrics and test metrics. The evaluation metrics include

- log-likelighood that is maximized in the estimation (larger values shows a better)
- BIC value (bayesian information criterion) that normalizes the log-likelihood with number of attributes and number of observations (the lower the better)
- Hitrate that shows the share of correctly estimated route choices
- The r^2 of observed and estimated route shares
- The r^2 of observed and estimated linkflow within one OD-pair

Table 6 - Estimated models and evaluation metrics.

	NBPA			DDPI		
	mNull	mTraveltime	mFSS	mNull	mTraveltime	mFSS
tt_free		-0.34^{***}			-0.21^{***}	
p_major_roads			4.26^{***}			2.66^{***}
tt_mean			-0.25^{***}			-0.15^{***}
PS		0.82^{***}	0.16^{***}		-2.51^{***}	-1.92^{***}
congestion_charge			-0.01^{***}			-0.01^{***}
left_turns			-0.38^{***}			
dummy_tt						-0.41^{***}
Training						
Log Likelihood	−161,842.30	−140,101.20	−99,380.71	−98,596.54	−87,128.20	−76,728.46
BIC	323694.4	280222.1	198810.5	197202.9	174276.0	153506.0
Hit rate [%]	2.46	12.70	49.40	15.08	46.57	57.95
Route probabilities [r^2]	0.06	0.09	0.40	0.01	0.31	0.52
OD specific link flow [r^2]	0.31	0.54	0.90	0.76	0.77	0.89
Number of observations	42977	42977	42977	42977	42977	42977
Test						
Hit rate [%]	2.43	10.01	50.23	15.40	44.17	55.67
Route probabilities [r^2]	0.06	0.10	0.42	0.02	0.29	0.51
OD specific link flow [r^2]	0.47	0.68	0.86	0.89	0.89	0.93
Number of observations	42422	42422	42422	42422	42422	42422

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table created with the Stargazer package in R

The model coefficients all have the expected sign. The magnitude of the tt_mean coefficient is larger for the NBPA model, which, as the forward stepwise selection, suggests that the mean traveltime has a larger impact on the choice in the model with the NBPA route set. The models show a significantly better fit than the naive models on both datasets in all evaluation metrics. The NBPA model predicts the correct route for around 50% of the travelers, and the DDPI model predicts around 56% of the travelers correctly. This can be compared to the naïve models, that predicts 2-10% correctly for the NBPA models and 15-44% correctly for the DDPI models.

The results of the study together with an anticipated large increase in available GPS-data indicates a promising future for improved data-driven route choice models. Improved route choice models are an important component of efficient traffic management with targeted traveler information and improved state estimation during incidents. More details about the route choice modeling on the road network are presented in (Danielsson et al., 2024a).

6. SCENARIO EVALUATION

6.1 Analysis of highway incidents in Dynameq

Within the project an overview of different modeling tools for analysis of road network incidents is made together with an analysis of how the modeling tool Dynameq can be used for modeling incidents. An analysis of 5 incidents in Stockholm is made using incident data from the Traffic Management Centre in Stockholm and GPS data from INRIX. The results shows promising results in using Dynameq for offline analysis of incidents, but further work is needed to get a more detailed understanding of potential and limitations of using Dynameq for analysis of incidents and traffic management actions. The 5 incidents are illustrated in Figure 21. A more detailed description of the analysis and the literature survey is available in (Rek, 2022).

Figure 21 Incidents in the Stockholm highway network analysed using Dynameq and INRIX GPS data.

6.2 Multimodal analysis of example incident

An example incident on the 1st of October 2019 can illustrate the effect on the multimodal network. The incident, a collision between a private car and a truck in the morning peak, was reported 6.15 – 8.30. A change in speed around the incident (in Figure 22) and flow on nearby links (see Figure 23) identified the incident as major with an observed change in route choice.

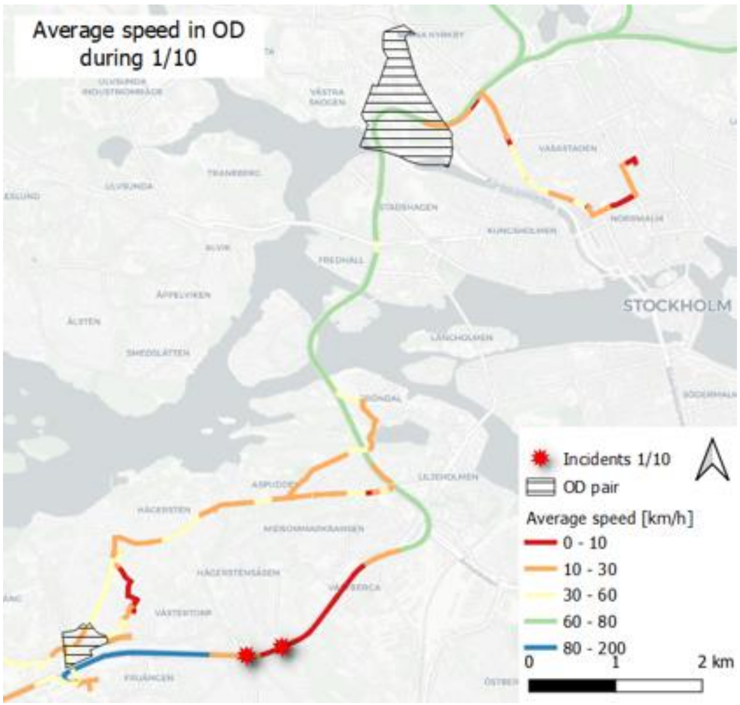

Figure 22 - Average speeds during the incident. Around the incident, the average speed is 0-10 km/h on Essingeleden with speed limits of 80 km/h.

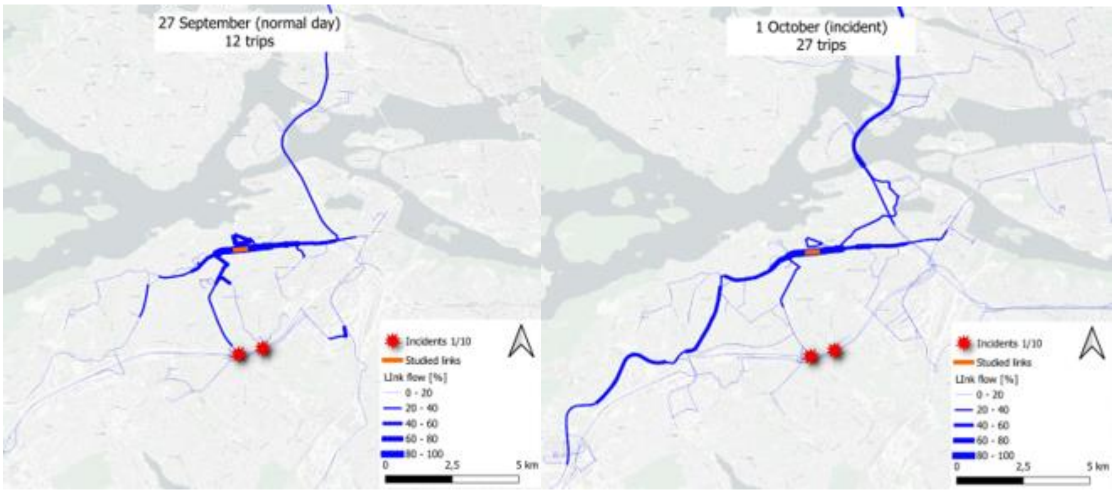

Figure 23 - Spider plot of link parallel to the incidents. Normal conditions to the left and during incident to the right, where flow comes from the main street rather than from the neighborhood as during the normal conditions.

Observed effects of the incident include increased traffic on non-motorway roads, delays on buses, increased public transport ridership, change in in- and outflows on toll portals, changed route choice and route traveltimes in an affected OD-pair. Figure 24 shows the difference between the road route choice during normal conditions and during the incident day in one example OD-pair. Figure 25 shows the only observed public transport route in the OD-pair together with the ridership over the week of the incident.

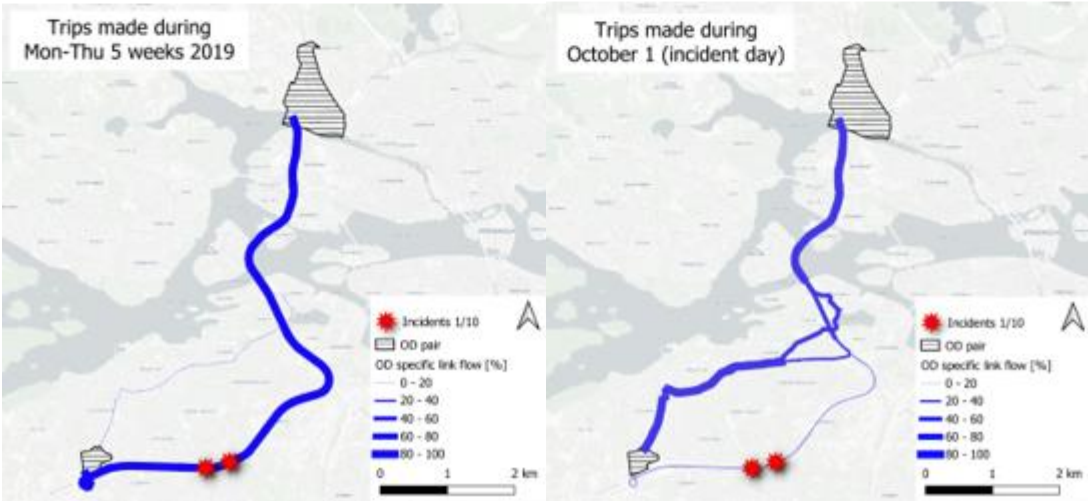

Figure 24 - Example OD-pair affected by the incident. The route choice is expressed as link flow in percentage of total OD flow. The left map shows the route choice during normal conditions and the right map shows the route choice during the incident day.

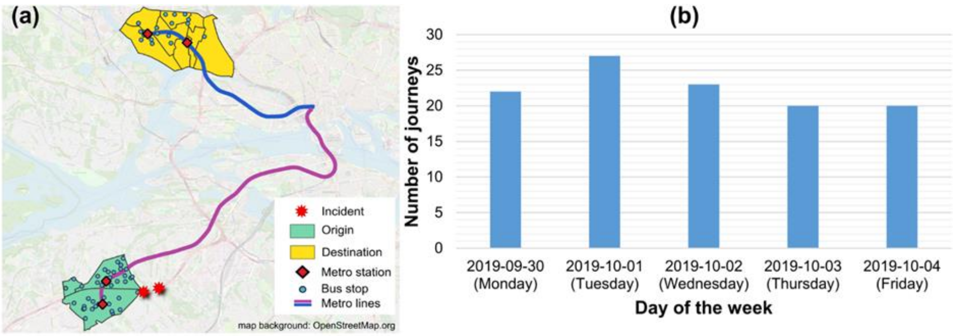

Figure 25 - Observed public transport route in the example OD-pair to the left and number of journeys over the week to the right.

The effect of the incident on the route choice could be captured by the developed road route choice model with a traveltime increase of 150% on the affected routes compared to normal conditions. The observed traveltime increase for road traffic was up to 190% compared to normal conditions.

7. CONCLUSIONS AND FUTURE WORK

Multimodal traffic management is in this project defined as a holistic perspective on real-time traffic management, utilizing the potential of several modes of transport to provide efficient traveler mobility. Based on the literature survey, we can conclude that simultaneous management of road and public transport has the potential to reduce congestion and ensure efficient movement of travelers in an urban area. There are several motives for integrated management of multiple modes. Information is assumed to favor public transport with the potential to improve urban sustainability. It can also be argued that multimodal traffic management is assisting travelers better by offering more mobility options, with the potential to reduce time and cost spent on traveling. There is also an argument that the transport network is naturally multimodal, implying that congestion in one mode has an effect on other modes. Thus, managing them together can help prioritize between management strategies. The main challenges are collaboration between stakeholders, information sharing, and data fusion. The increasing attention to multimodal traffic management in the literature reflects a growing need for holistic solutions to urban transportation challenges. Despite the increasing focus on multimodality in the traffic management domain, there is a gap in the existing body of research on quantitative assessments of the potential of multimodal traffic management strategies. There is a need for models and data to be used for the evaluation of such strategies.

The results of the explorative analysis based on unsupervised learning indicate that day clustering can be useful in scenario evaluation, but also serve as input to short-term link flow prediction providing a simple and robust prediction method with a MAPE prediction error of 10-15%. Several of the clustering methods evaluated in the project produce both promising and similar results, including K-means clustering and Agglomerative clustering. Clustering based on speed and flow produces similar results in terms of typical days. However, smaller changes in flow cannot be detected when the network operates in free flow conditions, which is natural given the flat speed-flow relationship for low flows. These small changes can be important for scenario evaluation during incidents.

The analysis of network-wide multimodal data for 5 weeks in Stockholm indicates that it is possible to estimate how mode share between public transport and other modes of transport varies in space and time. A better understanding of spatiotemporal variation of mode share is an important input to improved decision support in multimodal traffic management.

The route choice analysis showed that a model based on a route set with generated routes, on top of observed routes, is more responsive to travel time changes than a model based on only observed routes. Considering travel times varies depending on congestion and increase when there are incidents in the network, a model responsive to travel times is assumed to be useful for predicting the effect of traffic management actions. A route choice model with only travel time is a common simplification to use for prediction route choices. However, the result of this project shows that including more attributes significantly improves the performance of the models. The analysis indicated that simplicity of the route and a dynamic travel time (that changes depending on the starting time of the trip) are important attributes to include in a route choice model to match observed choices.

Important further work includes a compilation of a larger multimodal incident dataset to enable more extensive analysis of multimodal effects of incidents. Furthermore, based on estimated mode shares, a mode choice model adapted for traffic management should be developed. Efficient fusion of different data sources for multimodal demand estimation together with more detailed scenario evaluation based on a network-wide mesoscopic traffic model are also important topics for future work.

8. REFERENCES

Ahlberg, J., Danielsson, A., Drageryd, L., Gundlegård, D., Ramsey, J., Sjöholm, A., & Sjöstrand, S. (2021). Probedata: förstudie kring användning av gps-baserad probedata för skattning av hastigheter, länkflöden och ruttval (No. TRV 2019/98384). Retrieved from https://fudinfo.trafikverket.se/fudinfoexternwebb/Publikationer/Publikationer_004501_004600/

Breyer, N., Gundlegård, D. and Rydergren, C., "Travel mode classification of intercity trips using cellular network data," Transportation Research Procedia, vol. 52, pp. 211–218, 2021

Cats, O., Rubensson, I., Cebecauer, M., Kholodov, Y., Vermeulen, A., Jenelius, E., & Susilo, Y. (2019). How fair is the fare? Estimating travel patterns and the impacts of fare schemes for different user groups in Stockholm based on smartcard data : Final report for Trafik och Region 2018 SLL-KTH research project. Retrieved from KTH Royal Institute of Technology website: https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-333090

Cebecauer, M. (2021), "Enhancing Short-Term Traffic Prediction for Large-Scale Transport Networks by Spatio-Temporal Clustering," Doctoral thesis Stockholm, Sweden : KTH Royal Institute of Technology, TRITA-ABE-DLT, 2143.

Cebecauer, M., Jenelius, E., Gundlegård, D., & Burghout, W. (2023a). Revealing representative day-types in transport networks using traffic data clustering. *Journal of Intelligent Transportation Systems*, 1-24.

Cebecauer M., Gundlegård D., Jenelius E., Burghout W. (2023b). Spatio-Temporal Public Transport Mode Share Estimation and Analysis Using Mobile Network and Smart Card Data. In 2023 26th International Conference on Intelligent Transportation Systems (ITSC), IEEE.

Cebecauer M., Gundlegård D., Jenelius E., Burghout W. (2024). Interchangeability of speed and flow data-types and its internal validation. Working paper.

Danielsson, A., Gundlegård, D., Rydergren, C., and Tsanakas, N. (2023). 10: Transition towards more efficient road transports: insights from mobility analytics. In Handbook on Climate Change and Technology, Cheltenham, UK: Edward Elgar Publishing. https://doi.org/10.4337/9781800882119.00019

Danielsson, A., Gundlegård, D., & Rydergren, C. (2024a). Analysis of Route Sets and Attributes in Route Choice Estimation for Urban Traffic Management Using GPS-data. In Proceedings of the 103rd Transportation Research Board Annual Meeting. Washington DC, USA. January 7–11.

Danielsson, A., Gundlegård, D., & Rydergren, C. (2024b). Extending traffic management to a multimodal perspective: a survey of potential and challenges. Working paper.

Rek, R. (2022), Analysis of model-based decision support systems for traffic management, Master's thesis, ISRN: LiU-ITN-TEK-A--22/058-SE. Retrieved from https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1714221&dswid=3821

Skoufas, A., Cebecauer, M., Burghout, W. and Jenelius, E (2023). Generating and Evaluating Route Choice Sets for Large Multimodal Public Transport Networks: A Case Study for Stockholm Region, In 2023 26th International Conference on Intelligent Transportation Systems (ITSC), IEEE.

Telia Sverige AB. (2023) Telia crowd insights. [Online]. Available: https://business.teliacompany.com/crowd-insights

Transport Analysis: a Swedish government agency for transport policy analysis (2022). Travel survey. online: https://www.trafa.se/en/travel-survey/travel-survey/

Ågren, K, Bjelkmar, P and Allison, E, "The use of anonymized and aggregated telecom mobility data by a public health agency during the COVID-19 pandemic: Learnings from both the operator and agency perspective," Data & Policy, vol. 3, p. e17, 2021