On the Computation of Accessibility Provided by Dynamic Transportation Modes

Integrating dynamic public transportation systems into public transport to improve mobility in suburban areas: A simulation-based approach

A thesis presented in part fulfilment of the requirements of the Degree of Master of Science in Transportation Systems at TUM School of Engineering and Design.

Supervisors

Univ.-Prof. Dr. Constantinos Antoniou
Chair of Transportation Systems Engineering
Technical University of Munich (TUM)

Assoc. Prof. Andrea Araldo
Institut Polytechnique de Paris (IPP)

Dr. Santa Maiti
Chair of Transportation Systems Engineering
Technical University of Munich (TUM)

Submitted by

Severin Diepolder
Weilbachweg 13
82541 Münsing

Submitted on 01.03.2023

Declaration

I hereby confirm that the presented thesis work has been done independently and using only the sources and resources as are listed. This thesis has not previously been submitted elsewhere for purposes of assessment.

Munich, 01.03.2023, Severin Diepolder, Matr. No. 03679873

Abstract

Social inequality within urban areas is reinforced by the unequal availability of public transport. Areas with high population density like city centres are served with high frequency and capacity public transport, while lower density areas are lacking such quality of service, as improvement of services is deemed uneconomic. Novel dynamic public transport modes are designed to serve large, low-density areas adapting to the demand by no longer relying on a fixed network or schedule. This can make these systems profitable as feeder lines for conventional public transport lines. The impact of these expansions can be measured in changes in accessibility, but presently no tool exists to quantify accessibility based on simulation or real-world data for intermodal trips with dynamic and line-based public transport modes. This research develops an approach to deal with the scarce and scattered data by spatio-temporal modelling performance indicators of dynamic services and converting them into a graph-based representation. This graph is used as input for the established tool City-Chrone to quantify accessibility. By comparing the status quo and scenarios with integrated dynamic transport, an improvement in public transport connectivity for low-density areas can be observed, providing a proof of concept for the novel approach.

Acknowledgement

Firstly, I want to express my special thanks to my supervisor Assoc. Prof. Andrea Araldo from Institute Polytechnique de Paris (IPP) where the research was conducted in a 6-month research internship. Andrea was an engaged, always available contact person for all scientific and organizational issues throughout my stay in Paris. His inputs, critical technical remarks, and the fruitful discussions we had decisively contributed to the success of this study. My acknowledgement also includes Andrea's team who provided for a friendly environment, favourable for the project's progress.
I would also like to thank my supervisors at Technical University of Munich (TUM), Dr. Santa Maiti and Prof. Dr. Constantinos Antoniou. The regular exchange created a productive environment, their constant feedback on the progress of the study helped me to stay on the right track while breaking fresh ground.
Last but not least, I am grateful for the successful and frictionless cooperation with the researchers Tarek Chouaki from the French graduate engineering school CentraleSupélec (CS) and Dr. Sebastian Hörl from IRT SystemX. The regular exchange on the general topics of the work and the provision of high-quality datasets for the evaluation steps has been very helpful.
The 6-month onsite research internship at IPP was funded by IPP and BayFrance (Bayerisch-Französisches Hochschulzentrum), supporting the exchange of researchers between Bavaria (Germany) and France. An additional Erasmus+ internship grant, supporting internships within countries of the European Union, was highly appreciated.

Contents

List of Figures

Glossary

C

CRS - Coordinate Reference System 30, 31
CSV - Comma Seperated Values 29, 31

D

DRT - Demand Responsive Transport 6, 7, 12, 16, 19, 21, 22, 27–32, 34–37, 44, 45, 48–50
DTS - Dynamic Transport Systems 3, 15–23, 26–28, 30, 31, 33, 34, 48–50

G

GTFS - General Transit Feed Specification 11, 12, 16, 28, 31–33, 49

L

LBT - Line Based Transit 6, 7, 14–17, 19, 21, 27, 28, 30, 34, 44–46, 48, 49

M

MIT - Motorised Individual Transport 6, 10, 11, 18
MRT - Mass Rapid Transit 14, 15, 34, 35, 44

O

OK - Ordinary Kriging 24, 50

P

PD - Pandas 29–31
PT - Public Transport 3, 6, 7, 10–20, 22, 28, 31, 35, 43–45, 48–50
PTA - Public-Transport-Analysis 31

X

XML - Extensible Markup Language 29

1. Introduction

The quality of mobility and accessibility not only is a vital point of sustainability but also a socio-political issue. This master thesis and the research it is based on gears towards the answer on how to measure the changes in accessibility of PT by integrating DRT (Demand Responsive Transport) into the existing LBT (Line Based Transit) in suburban areas and to evaluate if an equalisation of accessibility is feasible throughout the study area, and transferable beyond. Transit systems play a crucial role in the sustainable development of regions and metropolitan areas by providing commuters access to centres and hubs. The availability of services and the straightforwardness of reaching destinations of the daily need, including travel time and travel cost, considerably contributes to improving the citizens' quality of life[1]. Prior research [2] already could demonstrate this coherence by showing that the quality of mobility is correlated with the general wealth of the people. This means that the quality of services provided also is a socio-democratic issue. Especially the unavailability of high capacity PT alternatives admit to no other option than using private vehicles. This reinforces the social inequality as owning and using MIT (Motorised Individual Transport) requires a certain level of wealth and education, excluding a major share of the society [3]. Among others, specifically the aforementioned coherence of the availability of mobility options and the socio-demographic structure shows that the present situation in urban vs. suburban spaces inverts the societal needs: Wealthier people live in areas with favourable mobility options, i.e., city centres, while less privileged social classes with lower financial capabilities live in suburbs and exurbia which causes disproportionately high travel cost and time need. This not only drives up the cost of living in the suburban areas, but it also hampers the suburbanites' activities in the city, thus further reinforcing the division of society.
The lack of comparable public mobility for all people throughout all parts of the urban and suburban space and considering the long distances that consequently must be travelled, favours MIT as the mode choice. Increased demand whilst limited capacities of trunk roads cause congestion. Empirical studies show that the expansion of road capacity will just attract more traffic rather than providing a sustainable solution [4]. Consequently, a more holistic area-wide approach must be developed which integrates and strengthens the alternative public transportation modes.

Previous research clearly revealed that the capabilities of integrated PT network design are appropriate to relieve stress from the road infrastructure and to increase the effective throughput [5]. Furnishing higher capacities, likewise availability for everybody, thus equally providing mobility for the entire commuting area, are the two most significant advantages of PT. However, system design varies not only between different networks, but also within one network. The spatial variation of population density resulted in the implementation of different capacities and frequencies of public transport. On one hand, densely populated and/or business areas are hubs for high capacity and high frequency operation modes like subways. On the other hand, suburbs with lower population density are more likely served by feeder services

maintaining the connection to and from the trunk lines to the city centre. Basically, providing the same service to both, low-density areas and densely populated hubs, is not considered economically viable. The intensity of the effects varies significantly between developed and industrial countries.

To gain profound insight in the area under investigation and to compare it with other areas commonly accessibility measures are used. Such accessibility measures, considering multiple factors for the evaluation of the attractiveness of locations, are powerful tools to understand and plan measures counteracting the inequality in mobility. To this end many definitions and implementations of accessibility measures have been developed, ranging from neighbourhood mobility [6], e.g., walk and bike, to high-speed modes across entire countries [7].

In other use cases DRT, also known as flexible transport services, have been identified suitable to serve fringe areas, in which operating conventional LBT in high capacities and frequencies are not viable. Not operating on a fixed route or schedule, DRT can serve a larger area and respond to temporal requirements of the users on the first and last miles to and from high capacity and frequency services. Different operation schemes have been developed over the years of research and in different pilot phases [8], suggesting a door-to-station schema as the most promising enhancement of PT service. For impact evaluation of access and egress services, various accessibility tools have been developed and established over the last decade. Specifically, research has been conducted to calculate the change of accessibility by DRT deployments [9]. The topic of accessibility was also evaluated by analysing the changes caused by feeder transit services [10] [1]. Further, tools such as CityChrone [11] and Public Transport Analysis Tool [12] were developed to examine the impact of conventional LBT. The comparative study of the existing methodologies and tools, revealed the lack of accessibility tools for providing a holistic view of how dynamic PT modes, such as DRT, are changing the accessibility when combined with LBT. All the tools listed above are constrained to the routing on networks evaluation. However, flexible transport services, DRT, do not operate on fixed routes nor schedules, consequently, routing on networks is not available. One of the principal challenges of the research conducted in this thesis thus was, to cope with trip data obtained from simulation or real-world observations of DRT services.

To be able to create scenarios of DRT as a feeder mode to and from LBT in an integrated PT system and subsequently analysing the impact on accessibility, the apparent research gap had to be closed by developing a novel tool. The input data should be as broadly available as possible including real world and simulation data. Additionally, future extensions to other dynamic modes beyond DRT had to be considered.

The research analyses the requirements and methodology of accessibility calculation and gives deeper insight in the key parameters of dynamic transport modes. A methodology is developed to graph and theoretically analyse trip data, to deal with the scarcity of data and convert it into an output that can be used by existing accessibility analysis tools. The resulting accessibility snapshot can be used to evaluate the implemented operation scheme regarding the achievement of equality in accessibility.

Research breaking fresh ground hardly ever is a straightforward approach, rather a trial-and-

error endeavour unveiling many pitfalls. Correspondingly, a lot of effort has been applied to find a viable methodology but did not bring the desired outcome. This “barking up the wrong tree” is not described in this paper rather focusing on the targeted and positive approach to achieve the objective of this thesis. Likewise, the time-consuming encoding of the methodology’s workflow, even though an integral part of the research conducted, is not addressed in detail in this paper. An executable workflow code is available from the author and will be published as an attachment to this thesis after its defence.

2. State of the Art of Accessibility Measures in Public Transport & Shared Mobility

2.1. Accessibility

2.1.1. Definition & Quantification

Despite the extensive use of the term accessibility, no generally accepted definition is stipulated. Hence, no standard approach for accessibility calculation is defined, rather, every use case adapts the meaning and framework of accessibility to its requirements. This is well summarised by Gould [13], "Accessibility (...) It is a slippery notion, however; one of those common terms that everyone uses until faced with the problem of defining and measuring it!" To facilitate a common understanding, essential for the comprehension of the approach at hand, the short definition by Hansen [14] stating, accessibility is "the potential of opportunities for interaction" or Geurs [15] defining accessibility as "the extent to which land-use and transport systems enable (groups of) individuals to reach activities or destinations by means of a (combination of) transport mode(s)" sets a solid basis.

Basically, accessibility enables the classification of impedance to reach destination(s) from one or many origins. Existing approaches define this impedance in different manners. Location-based methodologies focus on land use and mobility, whereas individual-based methodologies focus on the individual level [16]. Further, a differentiation between infrastructure-based and utility-based accessibility can be made, focusing on mobility and economical aspects respectively [15]. Accessibility enables to classify the ease of moving from an origin to a destination, exceeding the pure graph-theoretical measure of connectivity analysing values like coverage, directness, and connectivity [17]. Accessibility allows to evaluate the mobility of people, when considering the actual movement of people within the transportation network. For this many parameters, like mode choice, emission, internal/external costs can have an impact. To measure the system benefit, mostly mobility-based methods are used [18]. In some methodologies the influence of density of destinations is considered crucial for the measurement of accessibility: The more destinations that can be reached from one location in a shorter time, the better the accessibility value. Depending on the use case, the data pool used to underpin accessibility calculations varies greatly. Thus, accessibility measurements can be compared meaningfully only when they are based on the same methodologies for calculation. Likewise, the definition of the study area influences the result, as most measures are based on relative indices, not allowing direct comparison on different location without in-depth understanding of the matter.

Basically, it is generally accepted that a valid accessibility measure should satisfy four major axioms [19]:

1. Accessibility varies from one point in space to another. “Point” in this context may mean a particular X-Y geocode, a building or parcel, or a zone centroid (traffic zone, census tract, etc.). Since they are closer to more activity opportunities, people living in denser, central urban areas, for example, are likely to have higher accessibility levels than people who live in low-density rural areas.
2. Accessibility is activity (trip purpose) specific. For example, a given location may have good access to stores for shopping, but poor access to employment opportunities.
3. Accessibility intrinsically combines a measure of the ease/difficulty involved in travelling to or interacting with different points in space (typically referred to as the disutility, generalised cost or impedance of travel) with a measure of the attractiveness and/or magnitude of opportunities (i.e. the desirability and/or number of activity opportunities) at different spatial locations.
4. The measurement of accessibility involves the integration or summation of opportunities over space, weighted by the ease of interaction; i.e. opportunities that are closer/easier to access generally will be weighted higher than those that are further away or more difficult to reach in terms of determining the overall accessibility level.

Deviated from those four axioms, Miller [18] considers the points of impedance, location attractiveness, individual taste, and the selection of locations to be included in accessibility calculations. All four aspects must be factored in when setting up an apt accessibility measure. Specifically, when considering PT networks and their accessibility impact, not only the travel time derived from distances is essential, as users are also influenced by wait times, ride comfort, etc. [20] Different kinds of accessibility measures have been developed giving various weightings to the above-mentioned points. The most relevant existing approaches emphasize distance / travel times based measures, isochrone-based measures, gravity measures, and random utility-based measures [18, 21, 22]. Within the scope of this research, the focus is on compatibility with existing accessibility tools, as addressed in 2.1.2. To lay the foundation, the relevant accessibility measure is described in the following section. Trying to categorise the various approaches and tools to calculate accessibility, most often involves the usage of travel time or distances. These distances are established on the network graph consisting of links and nodes. For MIT the road network is considered, for PT the route networks and its schedule.

A generalized summary of how to quantify accessibility A_i at the location i can be expressed as

$$A_i = \sum^{j} O_j f(C_{ij}) \tag{2.1}$$

with j being all reachable destinations and O_j the opportunities, representing locations of interest and their respective weight. E.g., supermarkets might be of higher interests than corner shops when evaluating accessibility to supply groceries. $f(c_{ij})$ represents the impedance to reach location j from location$i$. This function includes weighting regarding the covered distance, travel time, or other essential factors like costs or wait times. Overall, the accessibility A_i for location i is calculated by the sum of all accessibilities for each location j.

2.1.2. Accessibility Planning Tools for Public Transport

The principal difference of the accessibility measures developed over the past year is the calculation methodology and focus on different transportation modes. Tools like GOAT [6] focus on neighbourhood mobility and therefor only consider walking and cycling modes. Other tools consider larger study areas and hence focus on MIT. As the attractiveness of different modes varies regarding different factors, the impedance calculation varies between the modes. For active mobility modes, the distance and condition of the route are of greater importance as for MIT, for which congestion along the route and the availability of vehicle parking are more important. All these differences are reflected in the respective accessibility tools. Calculation tools capable to calculate the accessibility for multiple modes, must consider a set of impedance parameters that allows for a comparison and can be measured for all mode options. This inherently makes the tools highly complex and consequently requires a large variety of input data. To overcome the hurdles due to complexity two approaches can be chosen. Either the smallest possible intersection of impedance values is selected, at the risk of providing a poor accessibility measure, or a simulation-based approach is integrated into transport simulation tools. Two examples for agent-based simulation frameworks with the possibility of accessibility calculation are SimMobility [23] and MATSim [24]. These tools can trace the path of individual agents and their origin-destination (OD) relations. Based on the congested network, mode choice, and weight factors, these tools are capable of deliver an accessibility measure. Key differences are that SimMobility is using the observed trips of agents, whereas MATSim is generating a congested network on which the required OD relations are routed.

Focusing on public transport, some of the mentioned tools are capable to integrate PT trips into the accessibility calculation. To obtain the PT network and schedule in addition to the road network, in most cases GTFS (General Transit Feed Specification) data is used. The composition of this data format is described in detail in Section 4.2. In effect, the GTFS data is creating a secondary network on which PT vehicles travel between nodes, representing the stations, according to the set schedule. In the case of road-bound transit modes, the GTFS sub-network is mapped onto the road network and therefore vehicles are affected by congestion too.

The further evaluation of options for accessibility calculation for PT had to clarify if the tool is capable to integrate existing PT networks as a base scenario and if it allows for introducing new mobility options, such as DRT. Two promising options where reviewed.

MATSim Accessibility

This approach is executed in the agent-based simulation environment of MATSim [24]. The accessibility measure in conjunction with the agent-based simulation can offer a well-integrated solution. The principal disadvantage of the tool is that it does not allow for any data input except its own simulation scenarios. Real world data cannot be incorporated, which is essential for evaluating the true impact of a dynamic mode deployment. The accessibility calculation implemented by Ziemke [25], within an extension of MATSim, is based on the congested network resulting from iterations of simulations. In its present version it deploys a logsum approach to quantify accessibility.

$$A_i = ln(\sum^{j} e^{\beta t_j}) \tag{2.2}$$

This directly leads to Eq.2.1 with $f(t) = ln(e^{\beta t_j})$ and the equality of $\sum^j ln(x_j) = ln(\sum^j x_j)$. The description was boiled down to t as the only argument of measurable impedance. β includes the non-measurable parameters of impedance. This term can be extended by other impedance values and factors.

CityChrone

The second option is separating the accessibility calculation from the simulation framework and using a standalone tool. As not an integrated approach, this would potentially also allow for different data sources as simulation.

An apparently suitable tool for non-integrated approach was developed by [26]. The tool named CityChrone by design is laid out to calculate accessibility measures for PT in the form of velocity and sociality scores. The tool takes the graph based representation of a PT network in form of GTFS data as input for its routing. Actually, it presents how good every location of a study area is reachable, and further integrate the amount of people able to reach this location in a certain time. CityChrone utilises an isochrone accessibility definition:

$$A_i = \sum_{t_{ij} < t_{max}} n_j \tag{2.3}$$

With t_{max} being the maximum travel time, t_{ij} being the travel time from location i to j and

n_j being the number of opportunities reachable at location j. The accessibility measure A_i sums up the number of reachable opportunities within t_{max}

For sake of computational power, destinations and population are aggregated in a hexagonal grid, leading to values greater than 1 for n_j[26]. In case of the sociality score, this value is the population reachable within the time limit. Instead of a sociality score based on the amount of people in reach, other spatially distributed destinations such as worksites or leisure facilities can be surveyed. The CityChrone was developed as an open-source tool and therefore can be adapted and extended to the specific need of the particular use case. Such flexibility allows to incorporate functionalities to evaluate other performance indicators or measures based on other methodologies.

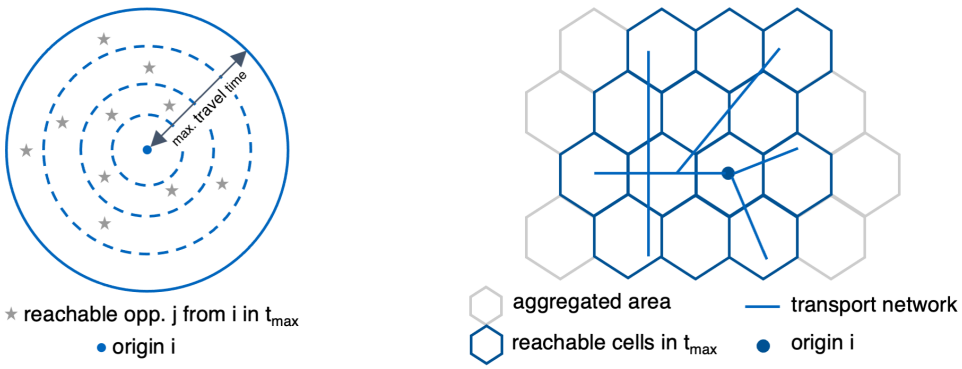

Figure 1 Schema Isochronic Accessibility Measure
Left: Opportunities that can be reached within time limit t_{max} simplified to the beeline distance
Right: Space aggregated in hexagons. The catchment area is defined by the hexagons that can be reached within t_{max} using the given transportation network.

Sociality Score

In detail, the sociality score is established by calculating the isochrone for the max travel time t_{max} summing the reachable population within this isochrone. The score represents the interplay and feedback loop of the population density and the performance of the PT system. This interplay describes the fact, that high population density areas are better served by PT than low population density areas. The measure quantifies the efficiency of PT in connecting people. The maximum travel time t_{max} represents the daily travel time budget. Further, different starting times are analysed and the average value for the metric calculated. The equation for calculating the sociality score is described in detail in [11]. To define the reachable area within the time t, CityChrone uses the Intransitive Connection Scan Algorithm (ICSA) for equally determining the routing to and from PT stations, along the PT routes, but also between PT stations in case changeovers are necessary. This routing algorithm was chosen for its high performance, as the number of reachable points is very high, e.g. $dist < 500m$ results in $\alpha 10^3$ number of journeys.

Spatial Aggregation in Accessibility Measures

Like all evaluated accessibility tools, MATSim Accessibility and CityChrone build upon spatial aggregation to reduce the amount of origins p_i and opportunities O_j. For the origin side, in most cases, infinitely many potential locations are aggregated in the centroid of a cell, mostly a hexagon, with the entire cell representing the computed accessibility from this centroid. Same applies for the opportunities side. Regarding that the weight and number of the opportunities is of particular interest for many accessibility quantifications, the aggregation process includes the aggregation of the weights, e.g. by average, or the number of opportunities for this cell. Consequently, the centroid of the destination cell, represents the number of aggregated opportunities. Another approach, as implemented in the MATSim Extension [25], is to project the aggregation of opportunities onto the closest network node. Some accessibility tools build upon a single sided aggregation, mostly aggregating origins only and calculating the accessibility for every opportunity. This is due to the finiteness of opportunities as compared to the infinite number of origins. For the accessibility quantification in CityChrone, origins are aggregated in the centroid of a hexagonal grid, whereas the aggregation of opportunities, here: population, is done by the catchment area per travel time interval. This values have been distributed over the same hexagonal grid, so that every hexagon features a particular population density. The origins of the isochrone is represented by the centroids of the hexagonal grid. Thus a isochrone is calculated for every centroid [11].

2.2. Public Transport

2.2.1. Line Based Public Transport

As opposed to private transportation, PT defines a transport mode, that can be used by the public. PT is available in most populated areas of the world, even though the quality of the services vary substantially. This makes any comparison of the different deployments virtually impossible.

Public transport in general, covers a vast variety of modes, spaces, ranges, frequencies, and needs. This research focuses on PT modes deployed in urban areas. This precondition requires the mode to offer a suitable capacity and frequency to allow passengers to cover the needed trips from origin to destination. In general, LBT is based on fixed routes and a temporal guidance for passengers on when to expect the arrival of vehicles. This either may be a fixed schedule, listing arrival and departure times, or by indicating the services' frequency. The latter is only useful at very high frequencies, e.g., indicating "every 5 minutes". The length and stop density of a LBT line varies depending on the mode of the service. For high capacity and high-speed transit systems, a larger distance between stops is standard, as frequent stops would counteract the speed advantage. These fast services are referred to as MRT (Mass Rapid Transit). In the urban area scale, MRT is serving trunk lines from outskirts to city centres or peripheral high demand hubs such as airports and ferry terminals. Areas with lower demand are served by modes featuring more stops and often with lower capacity. They are designed for collecting more dispersed passenger volumes in larger areas. For

this mode the overall travelled distance within a certain time interval is lower than for MRT. Still, these lower capacity services require a constant and medium level of demand to be run economically. To respond to the different demand over the day, peak hours vs. off-hours, likewise workdays vs. Sundays and bank holidays, the schedules are variable. This must be considered when evaluating the possibility to reach destinations[27].
PT networks, by definition, do not rely on a single line or a single mode of PT. The combination of different modes and lines favours passengers to reach more destinations by transfer at interchanges. Optimizing the connections at interchanges can significantly reduce the wait time, thus improving the convenience of PT. Further, the varying attributes of the different PT modes can be utilised to make the network economically more viable and sustainable while serving a large area [28]. E.g., low capacity modes can be deployed as feeder services to haul passengers to MRT lines, collecting the passengers on their way to the final destination. Such design of PT networks are a highly complex issue, specifically in extended networks featuring varying demand density in various time intervals. Conventional LBT modes reach their limits of viable and sustainable operation when demand gets too low. To provide acceptable services also in low demand areas and/or low demand hours and to integrate them in the overall PT network, novel mobility modes named DTS have emerged.

2.2.2. Dynamic Transportation Systems

New technologies have enabled the deployment of novel mobility modes. Still being available to the general public, these modes equally count towards PT. The main difference between LBT options and the dynamic services is the removal of a fixed schedule or frequency and no fixed route of the vehicles. New technology enable an online communication between the users and the PT service. As the PT operator now has knowledge about the exact origin-destination relation, a users wants to cover, PT vehicles can be set on flexible routes, that picks up the passenger at the origin and drops him off at or close to the destination location. To make this economically feasible by serving as many ride requests with a small number of dynamic vehicles. passengers that what to cover similar origin-destination relations are pooled, thus served by the same vehicle. The route is chosen so passengers are subsequently picked up and dropped off. This leads to the fact, that during a pooled ride, the passenger is not traveling the direct route on the network but is taking has to accept detours. The dispatching algorithm have to be laid out in a way, so this detour is minimised for each individual passenger. On one hand, as more, but lower capacity vehicles are used for this flexible routing and the pooling can lead to large deviations from the direct route, the speed and overall capacity of dynamic transport systems are lower as the ones presented in Section 2.2. On the other hand, the dynamic services can serve a larger area and therefore feature a larger catchment area [8].
Continuing with the concept from Section 2.2, DTS is capable of efficiently serving even lower demand areas, where it becomes unsustainable to operate low capacity LBT. Consequently, dynamic modes are ideally suitable as feeder services to transport passengers to and from MRT stations. A well calibrated system in terms of fleet size, service area and coordination between dynamic and line based modes is expected to deliver great advantages for users but

also the operator, as it is more economical. Apart from this feeder towards LBT, the presented modes can also serve door to door trips. For this research. This is not relevant and will not be analysed. For future systems, it is expected, that vehicles are driving autonomously, without a driver. This will make the mode even more attractive for PT operators, as it further reduces the cost of operation [29].
Other that dynamic modes pooling passengers to get them to their destinations, other modes can be viewed as dynamic/flexible too. E.g. Looking at bike, scooter and car sharing, a similar pattern can be observed. The passengers have to accept detours to find a an available vehicle and/or wait for a vehicle to become available. Once again, the overall route does not follow the direct route on the network between the origin and destination. Sticking with the analysis of dynamic pooled modes in PT, these options have to be kept in the back of the head. To also account for them, dynamic modes including these other options will be referred to as Dynamic Transport Systems DTS.

2.3. Research Gap

Literature research and building up a knowledge base on the state of the art in accessibility quantifications for different fields of application, then constraining it to the promising tools for application in an integrated DTS in LBT scenario, was the conceptual phase of finding general solutions for the problems in calculation of accessibility in a specific use case. To date, no data format or application has been established that can process both, dynamic and line based PT modes. Neither pure simulation-based accessibility tools, like the one integrated into MATSim, nor tools specifically designed for PT, like CityChrone, are by default capable of integrating DTS trips in their analysis. With regard to the accessibility calculation of DTS, only approaches are defined that consider dynamic and line-based modes separately. E.g., Chandra 2013 [10] compares line- based feeder services with DRT, but only considers the feeder part, disregarding the holistic effects on the entire network, specifically intermodal PT trips. Same applies for Nahmias 2021 [9] calculating activity-based accessibility measures for DRT. In summary, so far, no tool was published for calculating accessibility of DTS integrated LBT. Alternatively, research conducted by Hasif et al. 2022 [30] illustrates a promising approach representing DRT and PT as graph model. The LBT data originates from GTFS data, whereas the dynamic mode, DRT, is modelled analytically by continuous approximation. As this approach does not include analytical DRT data, rather aims at using simulation-based trips, it cannot satisfy all requirements of this research's questions. Summarising the findings of Section 2.2, the difference between DTS trips and LBT trips is obvious. Hence, if following the approach of Hasif et al. a supplementing methodology must be developed, that allows for the translation of DTS trips originating form simulation, or any other data source with the same format, into a representation like LBT. According to [31] the GTFS data format allows for the representation of PT networks as time dependent graphs. Consequently, the setup of a suitable translations tool was the necessary next step, aiming at the representation of dynamic mobility modes based on observations as representative schedule base PT modes, that can be used by accessibility analysis tools like CityChrone.

3. Methodology

The approach to close the research gap identified in Section 2.3 consists of three principal steps:

- The possibility of representing PT networks as time dependent graphs is set out, which also lays the foundation for the accessibility calculation using these graphs.
- A method is proposed to integrate DTS into such graphs. This unified representation of LBT and DTS in the same time dependent graph allows executing the same accessibility analysis for both.
- Arcs representing DTS trips are added into the time-dependent graph, with time labels inferred from as collected dataset of DTS trips.

The aforementioned inference of time labels should be performed for the entire study area and all times of day, which is challenging when dealing with scarce data. This challenge is addressed in Section 3.3. Implementation of these three steps allows to set up a workflow for importing raw trip data, modelling it in space and time, estimating respective values at needed locations, and representing DTS as time dependent graphs. Formatting these outcomes in line with existing LBT schemes resp. standards enable accessibility calculations with the obtained graphs. In the following, the methodology of the consecutive steps will be presented. As the input requirements for the subsequent step define the requirements on the outcomes of the previous step the explanation of the steps is given in reverse order, starting with the layout of the accessibility measure on a time-dependent graph.

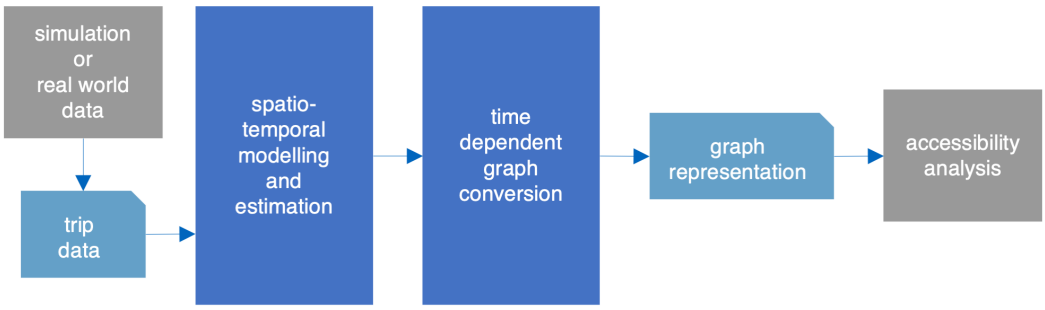

Figure 2 Schema of Methodological Workflow - Methodological workflow to prepare DTS trip data for accessibility analysis on time dependent graphs

3.1. Accessibility Calculation on Time-dependent Graphs for PT

Commonly, PT networks are either represented in a schedule based or a frequency based manner[32]. Both approaches have in common, that services serve stops v along a route r. So, each stop along route r can be denoted as s_r. One/single stop/s can also be served

by multiple routes. Each route r has a predefined sequence of stops, with each stop being served at specific times t_v. The travel time between two consecutive stops $v1$ and v_2 is the difference of its time of departure at v_1 and the time of arrival at v_2. To generalise the structure, the PT network can be modeled as a time-dependent graph $\mathscr{G}^{PT}$. Vertices v represent the stops of a PT service, edges e represent the path in between the stops without any intermediary stop. The sets of vertices and edges are denoted with V and E, respectively.A PT route is represented as a sequence of vertices $v_1, v_2, ..., v_k \in V$ with edges $e_{n,n+1}$ for $n = 1, \ldots,$connecting vertices v_n and v_{n+1}. The overall travel time from vertex v_1 to v_k is given by the difference of the departure time at v_1 and the arrival time at v_k. This can be expanded to any pair of stations i, j with $i < j$. Expanding the graph to accommodate multiple trips along the same route and further, multiple routes serving one stop, a representation of a PT network is achieved. One trip is composed out of a consecutive series of stops v. The shortest path along such time-dependent graph can be found by minimising the travel time from the initial stop v_s to the final stop v_k. This is done by finding the sequence of vertices v_i with $i \in V$ and $i \neq s, k$. Each departure time $t_{d,k}$ from a vertex must be greater than all previous departure and arrival times in the sequence. Every arrival time $t_{a,k}$ at a vertex is derived from the lines travel time along the connecting edge e. Transfers between stops can be integrated by adding edges between lines representing such interchanges.

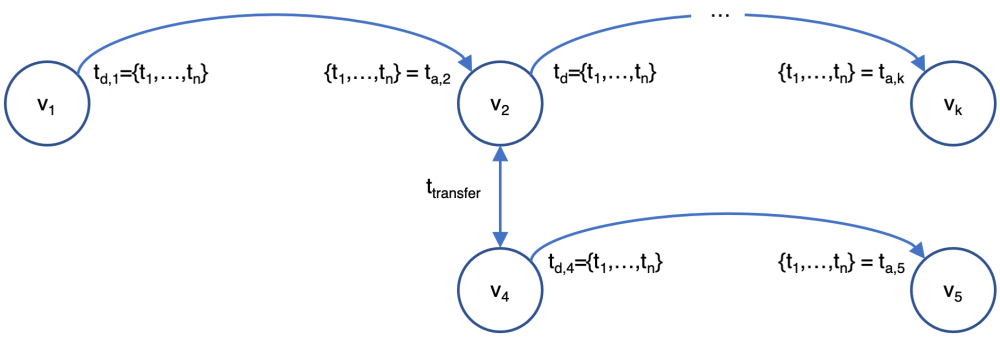

Figure 3 Time-Dependent Graph for Public Transport - Schema of a pt time-dependent graph representation of a simple pt network. Each vertex v holds arrival t_a and departure times t_d of each trip i passing trough the stop. t_d, i indicates the departure of a trip from a stop and t_a, i indicates the arrival of the same trip at the consecutive stop. Transfer times are set for changing between stops.

Expanding from here, having established a graph representation of the PT network, approaches to measure accessibility can be examined. The graph $\mathscr{G}^{PT}$ can be utilised to calculate the shortest path between every pair of stops. Further, each stop has a catchment area. At this point only walking is considered as feeder mode, hence a maximum of 15 min walking at a speed of 5km/h results in an approximated circular catchment area of 1.25 km. The catchment area can be further refined by introducing routing on the street network. Loading is not required, as DTS is approximately assumed to be not capacity constrained as MIT. After the catchment area of each stop is discretised, to exclude areas, that are not served by the DTS of one hub, and the shortest travel time between every station is known, an isochrone accessibility measure as described in Section 2.1 can be utilised to establish accessibility quantifications.

3.2. Time-dependent graphs for DTS

Having introduced a method of obtaining a graph representation of PT in Section 3.1 on which accessibility can be calculates as described in Section 2.1.2, the subsequent step is the integration of DTS into the graph $\mathscr{G}$.DTS is assumed to provide a feeder service to traditional PT. This includes the transport op passengers to, access and from, egress, LBT stations. The key information related to an access trip of DTS can be expressed as (the information related to an egress trip is similar, *mutatis mutandis*):

- Time of day $t \in \mathbb{R}$ when the user requested a trip to the flexible service
- Location $\mathbf{x} \in \mathbb{R}^2$ where the user is at time t
- Station s where the user want to arrive via the DTS feeder service
- Duration w indicating the wait time before the service starts: it can be the time passed between the time of request and the time of pickup from a vehicle, in case of ride sharing, DRT or carpooling; it can be the time to wait until a vehicle is available at the docks in a car-sharing or bike-sharing system.
- In-vehicle travel time y (or, shortly, travel time): time spent in the vehicle to arrive at s after the service starts.

The arrival of users requesting access trips via DTS feeder service can be modelled as a marked spatial-temporal random process [33], i.e., a random set of events located each in a position $\mathbf{x} \in \mathbb{R}^2$, in a time instant $t \in \mathbb{R}$ and associated with two marks, i.e., w and y. The intensity $\lambda(\mathbf{x}, t)$ is the expected amount of arrivals per unit of time and space in an infinitesimal area and an infinitesimal time interval around $\mathbf{x}$ and t.

The observed marks y and w are assumed to be realization of respective spatial-temporal random fields [34] W^s, Y^s, i.e., for each time of day $t \in \mathbb{R}$ and physical location $\mathbf{x} \in \mathbb{R}^2$, marks $W^s(\mathbf{x}, t), Y^s(\mathbf{x}, t)$ that would be associated to a user appearing in t and $\mathbf{x}$ is a random variable. Expressing the distribution of y and w as spatio-temporal random processes is of great importance for estimating values at specific instances of space and time in Section 3.3. Observe that each stop served by a feeder DTS service induces two random fields W^s and Y^s and that in the same $\mathbf{x}$ and t multiple random fields, from multiple stops, may insist. It is assumed that estimations $\hat{W}^s(\mathbf{x}, t)$ and $\hat{Y}^s(\mathbf{x}, t)$ are available and that their expected value $E[\hat{W}^s(\mathbf{x}, t)], E[\hat{Y}^s(\mathbf{x}, t)]$ can be obtained at any centroid $\mathbf{x} \in C$ (centroids are defined in Section 2.1.2. The centroids for this methodology are calculated by CityChrone to ensure compatibility in later steps) in which the feeder DTS is deployed and at any time t. Section 3.3 will be devoted to constructing such estimations.

To integrate DTS into the PT graph model, DTS is represented as a set of “virtual” PT lines, each virtual line running in the feeder area around stop s, from any centroid $\mathbf{x}$ within that area

to s and back. As for the travel time on this virtual line, it will be assumed to be $E[\hat{Y}^s(\mathbf{x},t)]$. The schedule-based description of PT (Section 3.1) requires for each line a set of departure times. Such departure times are assumed to be spaced along the time axis and separated by an interval $H^s(\mathbf{x},t)$, which may evolve over time of day and may be different from a centroid to another. The value of such an interval in $\mathbf{x}$ and t is also a spatial-temporal random field. To compute $H^s(\mathbf{x},t)$, it is relied on the relation between wait time and headway defined by Larson and Odoni [27]. To apply the method, the precondition of a random arrival of passengers must be determined. To apply the intended method, requests by passengers are assumed to be random in space and time. It is reasonable to assume DTS complies to this prerequisite as its requests are modeled as a spatio-temporal random field. Based on that relation, it is assumed that:

$$\mathbb{E}[W^s(\mathbf{x},t)] = \frac{\mathbb{E}^2[H^s(\mathbf{x},t)] + \sigma(\mathbf{x},t)^2}{2\mathbb{E}[H^s(\mathbf{x},t)]} = \frac{\mathbb{E}[H^s(\mathbf{x},t)]}{2} + \frac{\sigma(\mathbf{x},t)^2}{2\mathbb{E}[H^s(\mathbf{x},t)]}, \tag{3.1}$$

where $\sigma(\mathbf{x},t)$ is the variance of the interarrival time between one vehicle and the next. For ease of tractability, such a variance is assumed to be 0. Such a strong assumption could be eliminated by estimating such variance from the observations collected in the dataset. This is left outside of this work. Under the made assumptions, specifically the constant $\mathbb{E}[w]$ for the viewed time interval, σ^2 is 0 and therefore the term reduces to a simple expression only depending on $\mathbb{E}[h]$ and $\mathbb{E}[w]$. The previous formula thus become:

$$\mathbb{E}[H^s(\mathbf{x},t)] = 2\mathbb{E}[W^s(\mathbf{x},t)], \forall \text{ time of day } t \text{ and centroid } \mathbf{x}. \tag{3.2}$$

This reduced equation to assess the headway out of the wait time is equally defined by Dial [35]. It states that the average wait time is equal to half of the headway. Since only an approximation $\mathbb{E}[\hat{W}^s(\mathbf{x},t)]$ of $\mathbb{E}[W^s(\mathbf{x},t)]$ is available, only an approximation $\mathbb{E}[\hat{H}^s(\mathbf{x},t)]$ of $\mathbb{E}[\hat{H}^s(\mathbf{x},t)]$ is available:

$$\mathbb{E}[\hat{H}^s(\mathbf{x},t)] = 2\mathbb{E}[\hat{W}^s(\mathbf{x},t)], \forall \text{ time of day } t \text{ and centroid } \mathbf{x}. \tag{3.3}$$

The set of departure times of the DTS feeder service at location $\mathbf{x}$ and time of day t is based on $\mathbb{E}[\hat{H}^s(\mathbf{x},t)]$. First a random time of day t_0 is fixed. Then, edges corresponding to access DTS feeder connection from centroid $\mathbf{x}$ to station s are added at the following times of day:

$$t_0, \tag{3.4}$$

$$t_j = t_{j-1} + \mathbb{E}[\hat{H}^s(\mathbf{x},t_{j-1})] \qquad \text{for } j = 1,2, \text{ until 11:59 pm}, \tag{3.5}$$

$$t_j = t_{j+1} - \mathbb{E}[\hat{H}^s(\mathbf{x},t_{j+1})] \qquad \text{for } j = -1,-2, \text{ until 00:00 am}, \tag{3.6}$$

Applying best practice out of the domain of accessibility calculation, not every $x \in \mathbb{R}^2$ can be modelled, thus, a spatial aggregation is necessary, with the centroid of each aggregation

area representing the corresponding area. Using the methodology described in the next Section 3.3, information for this location is estimated from its surroundings. The centroid of each cell provides the location for the virtual stop s_v of the DTS service. The service will then operate between this virtual stop and the stop of the LBT, henceforth referred to as hub s_h. The temporal aggregation is constrained by time-slots during the overall operation time interval $[t_{start}, t_{end}]$. For each temporal boundary t_k of every time-slot, $t_{start} \leq t_k \leq t_{end}$ and $t_k < t_{k+1}$ holds true. Thus, every time slot is confined in the interval $[t_k, t_{k+1}[$, with a constant h for each time-slot.

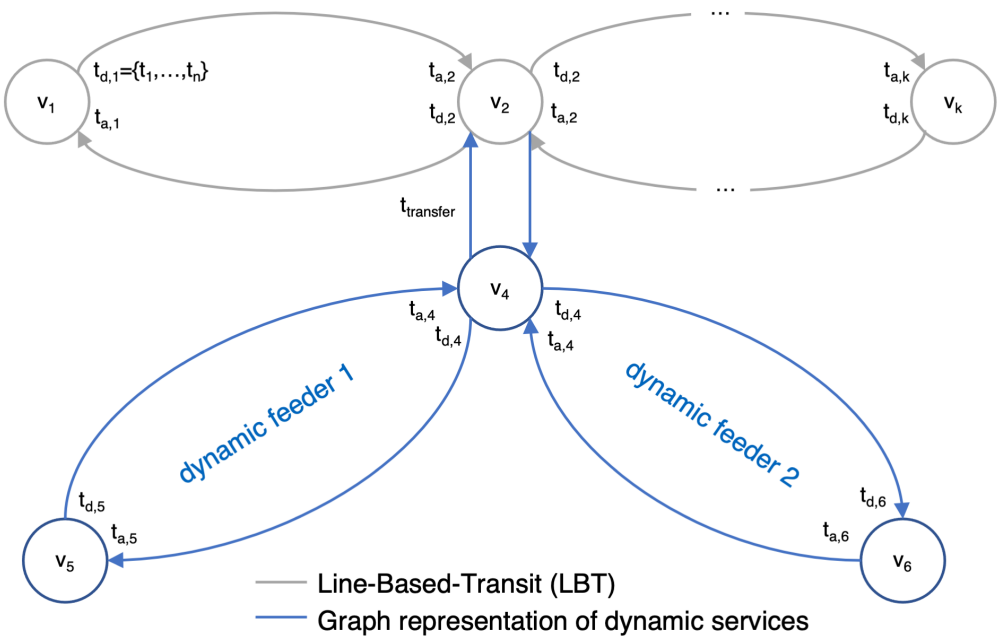

Figure 4 Time-Dependent Graph for Dynamic Feeder Service - The conversion of DTS as an extension of the graph in Figure 3. Information of dynamic services is used to set fixed departure and arrival times for all locations in the study area (here: v_5 and v_6) and at the drop-off / pick-up point close to the station (here v_4). The transfer between the LBT stop and the DTS stop is bridged by a transfer leg adding a temporal delay of $t_{transfer}$

Considering the travel time between two vertices v_n, a simple routing on the road network between s_v and s_h only yields the pure travel time. This measurement does only hold true in a few cases. DTS includes additional attributes that reduce or increase the travel time. DRT for example deviates from the direct travel route due to collection or drop off of other passengers. Only during unshared rides, the direct route is taken. As discussed, equal to w, y can be expressed as a spatio-temporal random field as well, and consequently an estimator can add values to other points in space and time. Once the location of all instances of s_v at the centroid of the aggregated areas and the corresponding locations at the hubs s_h are determined, the spatial relation of the graph to the real world is set. To comply with the requirement in terms of providing t_d and d_a for each v, the mentioned time intervals combined with the calculation for h_k during time slot k is applied. During each interval, trips depart at v with a temporal spacing of h_k. The same trips arrive at the next stop at the hub or virtual stop

site after y_k. Again, y remains constant during the time interval k.

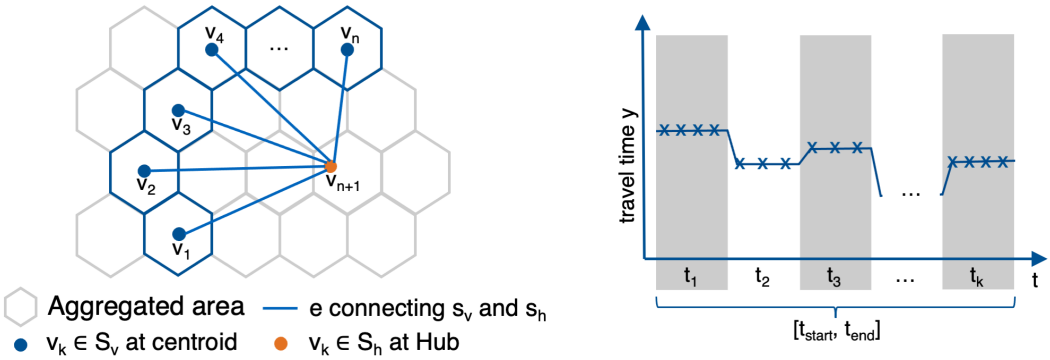

Figure 5 Spatial and Temporal Aggregation and Representation
Left: Schema of connecting each centroid to the hub. Only the graphs for the blue cells are depicted for clarity. Having n cells, n vertices $v_1, ..., v_n$ is created at the centroids. Each vertex is connected to the vertex $v_n + 1$ at the hub location.
Right: The study time interval $[t_{start}, t_{end}]$ is broken down into time slots t_k. For each time slot and for every centroid, $y_{n,k}$ and $w_{n,k}$ is estimated. y and w are constant during each t_k

3.3. Processing of scarce and scattered data

To allow for modelling DTS in terms of a time dependent graph $\mathscr{G}^{PT}$ as described in Section 3.1 a travel time on the graph as well as departure times, gained from the respective headway of the service, need to be known for every graph. Considering the feeder services, providing access for users from a location within the catchment area to a station, v.v., likewise in 3.2, the following parameters are known from the input data:

- Time interval T of interest with limits $t_{\text{start}}, t_{\text{end}}$
- Time of day $t \in \mathbb{R}$ when the user requested a trip to the flexible service
- Location $\mathbf{x} \in \mathbb{R}^2$ where the user is at time t
- Duration w indicating the wait time before the service starts: it can be the wait time before the vehicle arrives and leaves, in case of ride sharing or DRT or carpooling; it can be the time to wait until a vehicle is available at the docks in a car-sharing or bike-sharing system.
- In-vehicle travel time y: time spent in the vehicle to arrive at s after the service starts.

Operating as a feeder service, each trip of the DTS, either starts or ends at a hub station, allowing for interchange between feeder and PT, denoted as s_i. The other end of the trips is at the start or final location of passenger/trip i at location x_i. Taking t_0 as the departure time at a location, with $t_0 = t + w$. Consequently, egress trips $x(t_0) \in S$ with all $s_i \in S$. On the other hand all access trips have $x(t_0 + y) \in S$. The respective other side of the trip $x(t_0 + y) \in C_s$ and $x(t_0) \in C_s$ with C_s being the catchment area of the feeder service of station s. The non-station end of each trip is randomly distributed throughout the catchment area C_s. Further, the departure times t_0 of DTS are randomly distributed as they are according to the request of passengers.

As described in Section 3.2, both the observed t and w are realizations of respective spatial-temporal random fields W^s, Y^s, i.e., for each time of day $t \in \mathbb{R}$ and physical location $\mathbf{x} \in \mathbb{R}^2$, marks $W^s(\mathbf{x}, t), Y^s(\mathbf{x}, t)$ that would be associated to a user appearing in t and $\mathbf{x}$ are random variables.

To represent DTS as a time dependent graph $\mathscr{G}^{PT}$ as described in Section 3.2, information is needed at specific times of the day t_j with $t_j \in [t_{start}, t_{end}]$. Further, the graph $\mathscr{G}^{PT}$ has a discrete start and end position in $\mathbb{R}^2$. At this moment, information is only present at locations x_i at the time t. The time point, as well as the location of the information is part of the same time frame $[t_{start}, t_{end}]$ and the same catchment area C, but as information is randomly distributed in C, the locations do not necessarily match with the endpoint of our graphs. A methodology has to be developed to model information for DTS at locations and points in time, at which no information is present. The modeled information has to be able to statistically represent the observations made in the input data.

To estimate information for travel time y and wait time w at any given location in C, and at any given time t in time interval T, two main issues have to be overcome. Spatial distribution of data points is not even and gaps within the data have to be filled with knowledge gained using the available observations. Same holds for the temporal distribution. Distributed in time, not every possible time $t \in T$ can be assumed to be given in the dataset. By demand of DTS following the magnitude of the overall temporal and spacial travel demand, the scarcity of data is worsened by the accumulation of data points during peak hours and in high demand locations. Off peak hours and low demand areas therefore show a lower information density.

The following approach aims at building a framework to estimate y and w in space, to obtain the needed information at location and times as described in Section 3.2. The approach consists out of to main steps. First, a temporal aggregation, assuming similar performance of DTS during certain time-slots. Each time slots covers the temporal interval $[t_k, t_{k+1}[$ with $t_k \in T$ and [$t_{k=0}, t_{k=n}[\in [t_{start}, t_{end}[$

For each time-slot $[t_k, t_{k+1}[$, the marked spatial point process that generates the locations of users appearing to the system in $[t_k, t_{k+1}[$ is considered, together with the experienced waiting and in-vehicle travel times. Such times can be intended as realizations of spatial random fields $W^s_{t_k}(\mathbf{x}), Y^s_{t_k}(\mathbf{x})$. The simplifying assumption that $W^s(\mathbf{x}, t)$ is very close to $W^s_{t_k}(\mathbf{x})$ and $Y^s(\mathbf{x}, t)$ is very close to $Y^s_{t_k}(\mathbf{x})$ for all $t \in [t_k, t_{k+1}[$ is taken. Therefore, instead of estimating $W^s(\mathbf{x}, t)$ and $Y^s(\mathbf{x}, t)$, $W^s_{t_k}(\mathbf{x})$ and $Y^s_{t_k}(\mathbf{x})$ is estimated.

Thus two spatial random fields $\hat{W}^s_{t_k}(\mathbf{x}), \hat{Y}^s_{t_k}(\mathbf{x})$ to estimate $W^s_{t_k}(\mathbf{x})$ and $Y^s_{t_k}(\mathbf{x})$, respectively are constructed. First the observations $\mathscr{O}$ are projected onto time-slot $[t_k, t_{k+1}]$ to remove their reference to time of day:

$$\mathscr{O}[t_k, t_{k+1}[= \{(\mathbf{x}, w, y) | (\mathbf{x}, t, w, y) \in \mathscr{O}, t \in [t_k, t_{k+1}[\} \tag{3.7}$$

$\hat{W}^s_{t_k}(\mathbf{x}), \hat{Y}^s_{t_k}(\mathbf{x})$ is computed by Kriging [36] on the elements of $\mathscr{O}[t_k, t_{k+1}[$. Kriging is a method of inter- and extrapolation originating from geostatistics led by the covariance of observations. Inter- and extrapolation will from here on be mentioned as estimation. The method of kriging is known for delivering a very good estimate for values at unsampled locations [37]. Kriging offers multiple variations such as OK (Ordinary Kriging) or universal kriging. First is very widely used and applicable to a large variety of use cases. Being the default kriging method, its fundamental point is, that the constant mean is unknown. Talking about OK, is especially useful for variables that can not be represented by deterministic functions and consequently is random functions are used. More precise, kriging is done in an order-2 stationary random function model. The rationalized marker,in this case $W^s_{t_k}(\mathbf{x})$ and $Y^s_{t_k}(\mathbf{x})$, is the realisation of the stationary random function with x being a point in $\mathbb{R}^2$. Going forward, the kriging process is explained using $W^s_{t_k}(\mathbf{x})$. The process for $Y^s_{t_k}(\mathbf{x})$ is identical. The number of available data points for time slots t_k is given by N. Each at location x_α with $\alpha = 1, 2, ..., N$. Further the values are represented by $W^s_{t_k,\alpha}(\mathbf{x})$. The kriging estimator realises as follows, with $W^s_{t_k,0}(\mathbf{x})$ being the true marker at location x and $W^s_{t_k,*}(\mathbf{x})$ the modeled value:

$$W^s_{t_k,*}(\mathbf{x}) = \sum_{a=1}^{N} \lambda_\alpha W^s_{t_k,\alpha}(\mathbf{x}) \tag{3.8}$$

All weights λ_α sum to 1. The weights are selected, so overall the variance of the estimator error is minimised. This error can be quantified by $W^s_{t_k,*}(\mathbf{x}) - W^s_{t_k,0}(\mathbf{x})$. To minimise the variance of the estimation error, an equation system is constructed with $N + 1$ equations.

$$\sum_\beta \lambda_\beta \sigma_{\alpha\beta} + \mu = \sigma_{\alpha 0} \text{ with } \alpha = 1, 2, .., N \ \sum_\beta \lambda_\beta = 1 \tag{3.9}$$

where $\sigma_{\alpha\beta}$ is the covariance of the observations Z_α and Z_beta and $\sigma_{\alpha 0}$ the covariance of Z_α and the target Z_0. These attributes assemble the OK system, which variance can be expressed as:

$$\sigma^2_{OK} = E(W^s_{t_k,*}(\mathbf{x}) - W^s_{t_k,0}(\mathbf{x}))^2 = \sigma_{00} - \sum_\alpha \lambda_\alpha \sigma_{\alpha 0} - \mu \tag{3.10}$$

with μ_{00} being the variance of $W^s_{t_k,*}(\mathbf{x})$ [38] In contrast to interpolation methods like Inverse Distance Weight (IDW), the weight λ_α not only depends on the distance between the known points and the unsampled point of interest. OK is utilising the arrangement in space $\mathbb{R}^2$. To model this relation between points, spatial-autocorrelation is utilised.

Spatial autocorrelation is a possibility to indicate the relation of values of close locations in space. The key point of spatial-correlation states, that the values of data points that are in close spatial proximity have higher dependencies than values of data points in greater distance [39]. A well established measure to quantify the degree of similarity of values in proximity is Moron's index:

$$I_W = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(y_i - \overline{y})(y_j - \overline{y})}{\sum_i (y_i - \overline{y})^2} i \neq j \tag{3.11}$$

H_0: No co-variation of neighbors $I_W > 0$: Positive spatial auto-correlation

In this case, modelling the values for headway and wait time at the centroids with surrounding values. Spatial auto-correlation by definition is expected to vary with the distance to the modelling location [40, 39].

Consequently, the weights λ_{alpha} are dependent on the distance to the estimated point, the fitted model to the estimated point and the spatial relation of the values around the estimated point. The spatial relation between known points is modeled in a semivariogram. It is used to estimate the spacial dependency of points. For all pairs of points, semivariance is calculated. As for large datasets, large quantities of markers can have equal distances, pairs with equal distances are grouped into bins. The average of the squared difference of the pairs in one bin are giving the semivariance for this distance.

$$\text{semivar}(d_j) = \frac{1}{2} * \text{avg}((W^s_{t_k,\alpha}(\mathbf{x}) - W^s_{t_k,alpha}(\mathbf{x}))^2) \tag{3.12}$$

Further aggregation is done by extending the bins to intervals $[d_j, d_{j+1}[$. These bins are represented by legs l_j with $l_j = d_{j+1} - d_j$.

Kriging gives means $\mathbb{E}\hat{W}^s_{t_k}(\mathbf{x}), \mathbb{E}\hat{Y}^s_{t_k}(\mathbf{x})$ of the estimators, for each location $\mathbf{x}$, as a linear combination of the wait and in-vehicle travel times measured at observations close to $\mathbf{x}$, i.e., as a linear combination of the following values

$$\{(w, y) | (\mathbf{x}, w, y) \in \mathscr{O}[t_k, t_{k+1}[, \mathbf{x} \in \mathscr{X}_{\mathbf{x}}\}, \tag{3.13}$$

where $\mathscr{X}_{\mathbf{x}} \subseteq \mathbb{R}^2$ is the set of locations "close" to $\mathbf{x}$.

Kriging also gives the variances Var$\hat{W}^s_{t_k}(\mathbf{x})$, Var$\hat{Y}^s_{t_k}(\mathbf{x})$ of the estimators, for each location $\mathbf{x}$. Note that if the variance is high in a certain location $\mathbf{x}$, it means in simple terms that no trust can be put into the estimation at $\mathbf{x}$. In the numerical results, in addition to the mean of the estimator, the variance will be specified. The variance is computed based on the *semivariogram*, which "can be regarded as a scaled distance" [36] between $\mathbf{x}$ the observed points. The more observed points close to $\mathbf{x}$, the less the variance, the more reliably $\mathbb{E}\hat{W}^s_{t_k}(\mathbf{x}), \mathbb{E}\hat{Y}^s_{t_k}(\mathbf{x})$ represent the "true" random fields $W^s_{t_k}(\mathbf{x}), Y^s_{t_k}(\mathbf{x})$.

$\hat{W}^s(\mathbf{x}, t) = \hat{W}^s_{t_k}(\mathbf{x}), \forall \mathbf{x} \in \mathbb{R}^2, t \in [t_k, t_{k+1}[$ is set, for all time slots $[t_k, t_{k+1}[$.

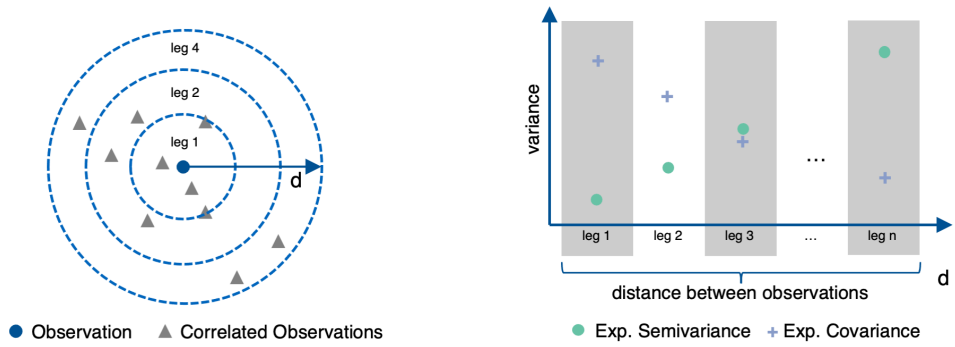

Figure 6 Schema of a Semivariogram - Separating the correlated observations into legs depending on the distance to the viewed observation an calculating the co and semivariance for each leg. Covariance is a indicator of similarity whereas semivariance indicates the dissimilarity of values.

At a last step the the confidence in the estimator was viewed. As described, the data is scares in time and space. Taking it into account as spatio-temporal random fields as every trip request can be viewed as random in those three dimensions. Under these circumstances, the confidence in the estimator can be formulated as follows. It is well-known that Kriging converges to the true expected value of the random field which wants to be estimated, as the number of observations increases [41]. To express this important property more formally, let us first notice that all the estimators discussed so far depend on the observations $\mathscr{O}$. Let us build set $\mathscr{O}_0 = \emptyset$. Then, let us obtain $\mathscr{O}_N$ by adding to $\mathscr{O}_{N-1}$ a point extracted by the spatial-temporal marked point process representing the arrival of users in DTS, $\forall N \in \mathbb{N}$. Let us denote with $\hat{W}^s_N, \hat{Y}^s_N, \hat{W}^s_{t_k,N}, \hat{Y}^s_{t_k,N}$ the estimators obtained on the set of observations $\mathscr{O}_N$.

Theorem 2.3 of [41] allows us to state that:

$$\lim_{N\to\infty} \mathbb{E}\hat{W}^s_{t_k,N}(\mathbf{x}) = \mathbb{E}W^s_{t_k}(\mathbf{x}) \tag{3.14}$$

$$\lim_{N\to\infty} \mathbb{E}\hat{Y}^s_{t_k,N}(\mathbf{x}) = \mathbb{E}Y^s_{t_k}(\mathbf{x}) \tag{3.15}$$

Such result is of great importance for us as it ensures that our estimators are unbiased. On the other side, an infinite number of observations is needed in order for our estimators to reach the “true” means of wait and in-vehicle travel times.

Collecting a big number of observations would require observing the system (either real or simulated) a big number of days from t_{start} to t_{end}. In anticipation that in the numerical results are considering that observations from only one day of operation are available. As a consequence, the numerical values reported therein must not be considered as reliable estimates, but only as a showcase of the methodology. To obtain more reliable estimates, one would just need to run more simulation runs and perform kriging on all the collected observations, which is however outside of the scope of our work, focused more on the methodological contribution.

4. Implementation

After establishing an understanding of both the domain of accessibility, it's quantification methods and the differences in LBT and DTS in Section 2 and consequently developed a methodology in, Section 3, to bring all three components together, the holistic process of converting DTS trip data to schedule based data has to be implemented. This Implementation step consists of three major steps. First, the identification and outlining of input datasets, such as simulation and real world. Second, the formulation of requirements for the output to be be compatible with existing accessibility calculation tools. And third, the implementation of the described methodology to convert the data between the formats obtained from step one and two. This final step represents a major part of the conducted research as it implements the a novel tool. The implementation is described in a conceptual manner without presenting the code itself. The code will be made available for further development after this research. The implementation was realized in a python environment. Further, the implementation is designed for DRT data.

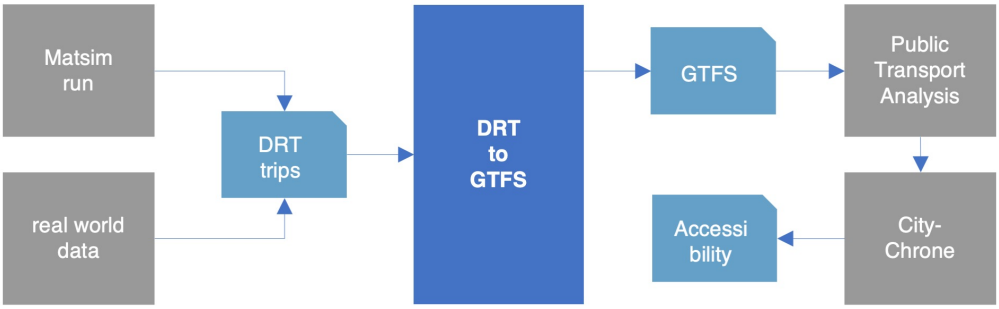

Figure 7 Schema of Implementation for DRT Services - Other input modes are possible.

4.1. Input Data Format

The input data can either be provided by simulation data, e.g. MATSim agent based simulation, or real world data, collected by DTS systems that are already deployed. The data has to map individual trips and must not aggregate any information. The implementation is designed to require some mandatory parameters for each individual trip. As described in Section 3, a spatio-temporal random field for wait time and travel time is established. For this, the minimum requirement for the input data can be summarized to:

- origin coordinates, defining at which location the passenger entered the transport system.
- destination coordinates, defining at which location the passenger left the transport system. *One condition, resulting from the consideration of feeder services, is that either the origin coordinate or the destination coordinate has to be at a hub location, allowing for*

interchange between the DTS and LBT.

- departure time, indicating at what time the passenger left the origin location in the transportation system
- wait time or pre-trip delay. Any sort of delay between the decision of choosing the transportation and the actual departure in the transport system. This can be either waiting after requesting a pickup or delay caused by the search for a vehicle.
- travel time or in vehicle time, indicating the time spent in the vehicle between the origin and destination location. This time includes any kind of delays such as detour tool create pooled rides or delays caused by capacity constrains of the transport network.

The collection of theses data points are providing the foundation for the presented methodology. As the tool is meant to be easily transferable between different dynamic transport modes, attention is being payed to use parameters that are not mode specific. As one example, the routing of DRT can be mentioned. Most datasets, including DRT trips, include in-depth information about detours due to pooling or separate the wait time into delay in regards to the passengers request and in regards to the vehicle being stuck in traffic. By generalizing this information to only using the overall travel time or one delay / wait time, a universal applicability is simplified. In some cases a previous preparation of the data to conform to the set requirements in needed.

4.2. Output Data Format

The output format of the presented tool claims to be as universally applicable as possible. Therefore an output format is chosen that enables conforms to the most general known standards for publishing PT schedule information. The format is called General Transit Feed Specification, GTFS in short, and was invented and is maintained by Google. [42]. The format is of special interest, as established accessibility tools for PT, like CityChrone, take this format as input. This was already described in Section 2.1 [11]. GTFS data is published by the local PT operator as a zip archive. This archive has to contain a number of files to describe the operation. Many datapoints can be added as an option to enable further functionalities. As CityChrone is relying on the basic information of GTFS data, complying with the minimum requirements set by Google is sufficient to ensure a valid accessibility analysis. GTFS data can be seen as a representation like it was described in Section 3 [31].

Summarizing the structure of GTFS data, each stop of the network is represented in the *stop.txt* file with an id and its location. Each stop is assigned stop times in *stop_times.txt* stating which trip passes though the stop at what time, arrival and departure. Further it is stated which number in the sequence of stops of one trip the current stop is. Each of the mentioned trips is assigned to a service id and a route id in the *trips.txt* file. The service id allows to define the dates of operation in either the *calendar.txt* or the *calendar_dates.txt* file. On the other hand, the route id links links to the PT operator. For representing DTS services

as schedule based trips, the location in *stops.txt* and the time details given in *stop_times.txt* are of most interest. When those are defined, all further information can be deviated from the first two.

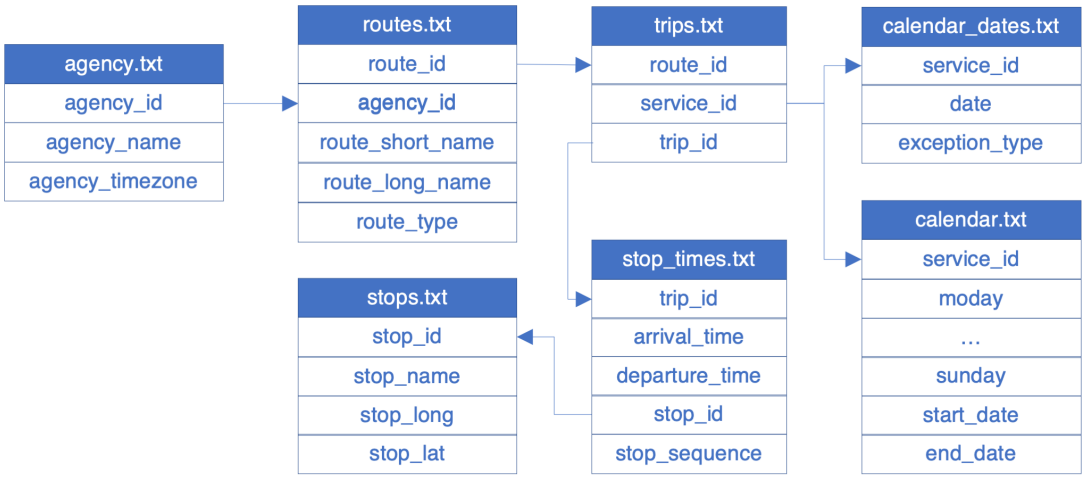

Figure 8 GTFS File Structure - Lists all required .txt files and attributes within the files

4.3. Novel Approach: DRT2GTFS Pipeline

4.3.1. Trip Input Processing

Referring to *Legs2df.py*

In the presented case, MATSim output files are used. The implementation can easily be altered to accept deviating input data. The used MATSim setup produces as variety of output files containing large numbers of plan, trip, mode or simulation run related information. Some information is aggregated, so one information block contains multiple steps of the individual agents journey. Focusing on the task at hand, and the intention to represent as detailed information of DRT as possible, we need the most in detail information as possible. The simulation output file has to contain data about the separate legs of the agents plans, holding information such as the origin and destination, the time and duration. In the case of DRT, the wait time is a key value needed for the conversion into a schedule. If this methodology is transferred to other operation schemes or other dynamic modes, e.g. car sharing or shared e-scooters, this wait time parameter can be replaced by the times a costumer has to search/walk to the closest vehicle. Generally, two output files were considered, both holding the needed information, described in the previous section. First, the *plans.xml* file. Holding all this information in a multi level XML (Extensible Markup Language) structure defined by MATSim, a converter is needed to parse through the given information and filter out the relevant parts. Given this slow, not optimized, parsing process, a second source was found in the *output_legs.csv* file. Here every single leg done done by each agents in the selected plan of the last iteration is listed in a very clear manner. A simple importing of the CSV (Comma Seperated Values) file into a PD (Pandas) data frame and filtering for DRT legs provides the base data to perform the calculation on. This initial filtering step only leaves the data points in the set, that have rel-

evance for further calculations. The size of the obtained set can give an initial warning about the accuracy of the output. Too little points will lead to a high statistical uncertainty and variance in the results. The import process also ensures the consistency of the trip data with all other required data by converting time values to consistent format and projecting all provided coordinates into the same coordinate reference system. To guarantee universal applicability of the tool, the working CRS (Coordinate Reference System) needs to be given by the user as well as the coordinate system the user is providing the raw data in. The working CRS needs to be metric to enable upcoming calculations of distances. One major data preparation step is the cleanup of the time values. As given by the *output_legs.csv* file, the travel time contains the wait time. By subtracting the wait time of the travel time, the later needed of the pure travel time on the network becomes available. The output of the *importLegs* function is a PD data frame containing the listed items. The pointer to this data frame is made available to the main run method.

4.3.2. Hub Detection

Referring to *ExtractHubs.py* and *assignHub2Trip.py*

Not limiting the data input to one source, holds the possibility of having no preliminary knowledge of the amount and location of DRT -Hubs. The location at which passengers are able to transfer from DTS to LBT systems. To obtain the knowledge of hub locations, the *hubDetection* function is exploiting the boundary condition of this tool to process only door-to-hub trips and especially excluding door-to-door trips. This consequently means, that either the origin or destination of each trip is located at a hub location, thus that the number of occurrence of the hub locations H_j equaling the number of trips T_i in the dataset.

$$\sum^{j} \#H_j = \#T_i \tag{4.1}$$

By grouping the trip data frame once by origin and once by destination coordinates and calculating the number of occurrences of each coordinate, a ranking can be created. Joining the to lists by Coordinate, summing the number of occurrences in origin an destination and finally sorting the list in descending order by the number of occurrence of each location, the locations of hubs will appear at the top of the list. The number of Hubs can be detected by extracting the entries of the list from top to bottom and summing up the entries in the number of occurrence field. The breaking condition is, when the sum equals the number of total trips. The extracted locations are the locations of the hubs H_j. The locations are written into a separate data frame, and Hubs are named with integers following the schema $Hub_0, Hub_1, ..., Hub_j$.

Using the knowledge of Hubs location, each trip is checked to which hub it leads or from which hub it departs. At the same time, the information, weather it is a access (leading to hub), or egress (departing from hub) trip is evaluated. Both information-points are added to the trips data frame.

4.3.3. Study Area and Grid Preparation

Referring to *GTFSpreprocessing.py* and *importCells.py*

At this part of the pipeline, the overall requirement of archiving compatibility with the CityChrone tool and therefore with the PTA (Public-Transport-Analysis) tools as the preprocessing unit of CityChrone, becomes relevant. As mentioned in Section 2.1, the accessibility analysis is done in a aggregated manner, in which origins and destinations are aggregated in gird cells. These cells are in shape of hexagons to keep the computational demands low, achieve full coverage of the study area on one hand and stay close to circular shape to only aggregate areas with similar initial properties. As the goal of this pipeline is to place virtual stops representing DTS trip options to and from hubs, the stations locations at the centroid of the cells have to aline with the analysis centroids. The PTA tool starts the trips to establish a travel time from the centroids. If this pipeline ensures, that the virtual stations are at the centroids, no additional walk time to the station is included and therefore the true door-to-hub scenario is analysed. To achieve this, a part of PTA is executed at this point with the base-case, no DTS. It is assumed, that the extend of the GTFS file exceeds the extend of the DTS study area, so the bounding box of the GTFS file is the same with and without the virtual DRT stops. The PTA is aborted after the hexagonal grid was calculated. The only change in the default process is that no filtering for cells that contain PT stations as the pipeline needs a continuous hexagonal grid for its analysis. the PTA tool outputs a CSV file containing all cells and their centroids. This CSV file is then read in by the pipeline and made available in a PD data frame. Throughout this process, the continuity of the CRS needs to be guaranteed to obtain a valid result.

4.3.4. Temporal Aggregation and Spatial Modelling

Referring to *InterpolateTrips.py*

In the following the processing of access trips is described in detail. The same procedure is performed for egress trips with the difference, that instead of the origin location, the destination location is used as reference for the spatial estimation. Otherwise all reference locations are at the same location and no spatial patterns can be picked up by the autocorrelation. The given trip data only represents one single trip at one time instance. One of the fundamental assumptions of this research is, that in the case of DTS trips, relations between one spacial region of the study area to one hub show similarities to a certain extent. This similarity and therefore the standard deviation for this set of values is further reduced by grouping the data into time intervals. This assumption is based on data showing longer wait and travel times during peak hours than during low demand times. Attention needs to be payed towards the size of the intervals. Regarding the size of the initial dataset broken down to each DTS hub and further reducing the size of the set by grouping by time interval, can lead to a high spacial scarcity of the data points and thus, breaking with the first assumption, no longer having similar values in one region. This introduces uncertainty to the prediction of values in the locations of interest. The time intervals therefore have to be chosen carefully. Having this broken down dataset, the challenge is to predict values at each location at which a virtual DTS stop should be introduced. As defined before, these locations are the centroids of the

hexagons. The selected approach is the usage of spacial interpolation implemented by the python library *Pyinterpolate* [43]. To realise the targeted GTFS conversion, two values are needed at the centroid location of each cell. On one hand the wait time and on the other the travel time from origin to destination. For each time interval, and each hub, the access and egress trips are analysed for spatial autocorrelation by the pyInterpolate tool. After establishing a semivariogram, representing the autocorrelation, the values at the centroids are estimated. Estimation is only performed for centroids that are within the catchment area of one hub. This catchment area is defined by the extend of the input data. This step also ensures, that locations far away from any actual data are not taken into account and that the overall estimation process in this work does not influence the catchment area of the DRT services, but only fills the gaps in areas where service is feasible. After this this step being finished, every centroid in the catchment area of one hub has an estimated wait time and an estimated travel time. This enables to construct a time-dependent graph represented in a GTFS file in the next step.

4.3.5. Schedule Creation and Output

Having established travel and wait times for every time interval and every centroid serving a hub, graph based representation can be developed. This step is done for both access trips and egress trips. As mentioned in Section 3, the headway of a service is approximated to double the wait time. Setting an overall analysis time interval for one day and having the time intervals during the day, for each interval the departure times at each virtual stop at the centroids are established. Each time interval is filled with as many departures as the headway allows for. The first departure is set to a random time between 0 and the first headway. If, caused by scarce data, for some time intervals during the day, no estimation was possible, the interval is skipped by setting the headway to the duration of one interval. The stop locations

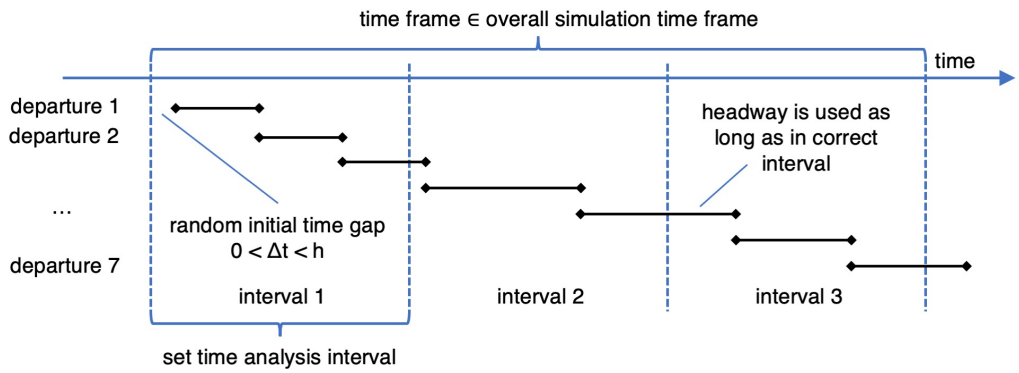

Figure 9 Schema of Schedule Creation - The duration of the headway (length of each black line) stays the same during each time interval until the next time interval is reached, then the next headway is applied

are the centroids and the hub locations, respective departure and arrival times are defined as mentioned by the headway and the addition of the travel time. Each cell-hub and hub-cell relation is modeled as a stand alone line. As they are virtual lines no loops or other operational factors are taken into account. For access lines, the sequence of stops starts at the cell and ends at the hub, for egress lines this order in inverse. The selection of ids is

chosen in a way, so when inspecting the GTFS data, direction, included cell and hub become are included in the name. This format is directly compatible with GTFS. A final step includes the joining of existing GTFS files with the schedule representing the DTS lines. The out

5. Results

The methodology presented in Chapter 3 and the consequent implementation described in Chapter 4 have been applied to a simulation scenario of the Île-de-France region in France. It is a scenario for DRT feeder services to and from MRT lines in the Paris Saclay area. The scenario and its output is described in further detail in Section 5.1. A verification of whether the estimation of wait and travel times follow the temporal and spatial pattern of the area under study is performed in Section 5.3. The improvement in accessibility brought by DTS is measured in Section 5.4. Based on the mentioned scenario the sociality score is calculated as one of many possible option for accessibility measures. The population used for this score is the same synthetic population that was used in the simulation itself.

5.1. Data Source: Simulation of DRT Feeder in the Paris Saclay Area

It is of great importance to ensure good quality input data for the evaluation of the novel accessibility calculation approach. The setup, calibration and execution of large scale, realistic simulation scenarios requires high attention to detail and thus, creating a simulation scenario for the pure purpose of this research would exceed the scope of this work. For this reason, the input data used in this Section comes from an existing scenario, which also showcases the compatibility of the proposed approach with external data sources. In particular, the data consists of the output of a simulation of a DRT feeder service in the Paris Saclay area. This data integrates DRT as a feeder service toward main MRT lines in the Paris Saclay area but simulates mobility behavior in the entire Île-de-France region. In particular, this means a door to station and v.v. operation. No service DRT service is provided for pure door-to-door DRT trips. The scenario was setup and simulated by Tarek Chouaki and Sebastian Hörl, researchers at System X. MATSim is the agent-based simulation environment of choice. The scenario is described in great detail in their publication [44]. In the following, the most relevant points regarding the DRT integration are summarized. Within the regarded study area, DRT serves the last-mile around 16 MRT stations. The scenario allows for DRT rides from and to all MRT stops within the Saclay area. Such stops will be called hubs from this point on. The routing module deciding to which MRT station an agents wants to travel by DRT, bases its decision on the euclidean distance between the non-hub location and the station. Always the closest station is being chosen, even though this might not provide the quickest route to the final destination of the agent. Agents choosing public transport, can either route via conventional LBT and walk or include a DRT leg in their first or last mile. DRT trips are exclusively requested and assigned online. No pre-booking is possible. If agents plan to use DRT, they request it at the time of departure and the request is processed by a centralized dispatcher. The wait time of the agents therefore is the time passed between the time instant in which the user requests a trip and the time instant in which the DRT vehicle arrives at the user's loca-

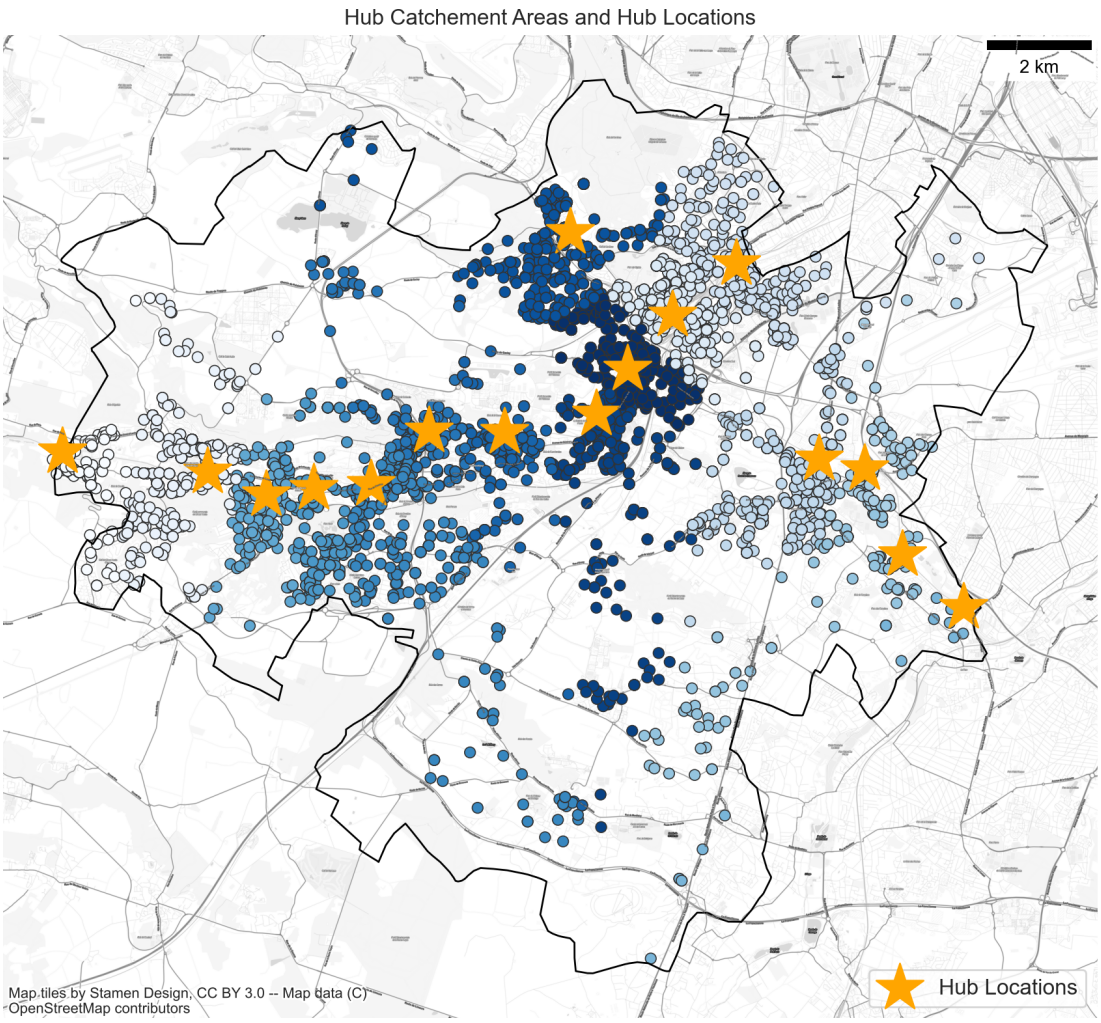

Figure 10 Hub Catchment Areas and Hub Locations - North direction always at the top. Each dot corresponds to the origin of one trip observed during the simulation. The differentiation in color of the observed trip origins indicates the catchment by different hubs

tion. Within the simulation, the wait time of all agents is limited to 10 minutes (600 seconds) in the DRT planning phase. Exceeding this, the trip request is rejected. Further, the dataset only contains DRT trips that were executed. Rejections are not included. The fleet consists of 400 vehicles, which serve 14700 trips during the single-day simulation. Among these trips, 5289 are *access trips* to MRT stations and 9412 are *egress trips* from MRT stations. Not every hub is serving equal amounts of trips. Due to the spatial distribution of hub locations and the variation in population density troughout the study area, the total amount of served trips, access and egress, per hub are in the range of 162 to 1856 trips. Looking at Figure 11, one can observe, that the amount of DRT trips have an equal share of the total number of trips at every time over the day. Considering that mainly home-work trips are simulated, there peaks in demand are to expect. Further, the simulation sets a fixed share of passengers to use PT, resulting in this almost mirrored trend. By looking at all trips with either their origin or their destination inside the Paris Saclay study area and comparing it to the number of DRT trips, the share of DRT trips in the overall number of trips is depicted in Figure 11. The rate of DRT trips follows the usual peak / off-peak hour patterns.

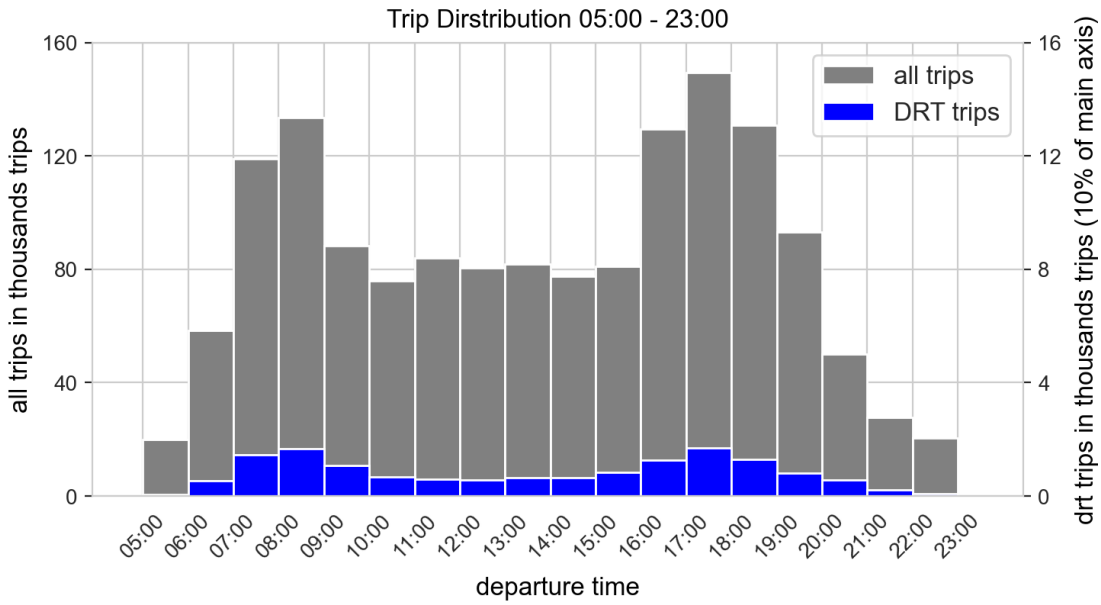

Figure 11 Histogram Total Trips (DRT + MIT) and DRT Trips - Non composite (blue bars overlay the grey bars) representation of the number of drt trips in relation to the number of all performed trips by MIT and DRT. Note: The number of DRT trips is scaled by the factor of 10. Refer to secondary axis. One trip is defined by the departure within the study area of Paris Saclay. A trip consisting out of multiple legs (e.g. walk + drt + pt is considered as one trip)

The estimation of the random fields corresponding to wait and in-vehicle travel times resulting from the DRT (see Section 3) is calculated on the dataset containing the observed DRT trip. Each observation i is a DRT trip. Considering an access trip, each observation i would have the following features:[1]

- location $c^{(i)}$ of origin and destination of the trip
- departure time $t^{(i)}$ of the trip (time of pickup)
- wait time $wt^{(i)}$ of the trip (difference between the time of request and time of pickup)
- travel time $T^{(i)}$
- traveled distance $td^{(i)}$ by vehicle (includes pooling)
- direct route distance $td^{(i)}_{\text{dir}}$ (if no pooling had occurred)

In addition to the previous features, other derived features can be easily computed:

- beeline distance $d^{(i)}_{\text{beeline}}$ between $c^{(i)}$ and the respective hub
- virtual speed $v^{(i)}_{\text{virtual}} = \frac{d^{(i)}_{\text{beeline}}}{T^{(i)}}$
- detour factor $df^{(i)} = td^{(i)}/d^{(i)}_{\text{beeline}}$

[1] A similar description, *mutatis mutandis*, would hold for the egress trips

The difference between the traveled distanced by vehicle and the possible direct route distance is that the former includes all detours caused by the pooling of passengers and the deviation from the beeline as the route is along the provided network and the latter only includes the detour due to the routing on the network. It is therefore equally long for a pooling detour factor of one. Implying no pooling is happening. Otherwise, it is always shorter than the travelled distance.

5.2. Analysis of Temporal and Spatial Patterns of DRT trips

The first step of the estimation of wait and in-vehicle travel time is to aggregate the observations over time and over space. As for the temporal aggregation, the day is split into 3 different parts. As visible in Figure 11, demand peaks are present in the morning and the evening. During those peaks, a change in performance of both the road network, affecting mainly the travel time, and the amount of DRT request, affecting the wait time, has to be taken into account. To be able to analyse these different preconditions, the morning peak (07:00 - 10:00) and the evening peak (16:00 - 19:00) are viewed. All other times are marked as off peak and not further differentiated. By comparing the observed travel times with the the actually traveled distance, the direct route on the network and the beeline distance between the pickup location and the hub, it can be identified if a representation as a virtual link between the origin and the hub is feasible. Only if similar correlation between the travel time and the three distance measures exist, confidence can be build into only relying on the beeline distance for further analysis. This requires the detour factor to not have a too high variance for the observations.

In what follows, with slight abuse of notation, observation i will be said to be "close" to centroid c and departure time t if $c^{(i)}$ is close to c and $t^{(i)}$ is close to t. The term "close" is left intentionally undefined, as the actual criterion for considering observations close or not results from the estimation method. When estimating the travel time $T(c,t)$ at a certain time of day t between a centroid c and the respective hub (and vice-versa), one could use a "direct" or "indirect" estimation. The direct estimation consists in aggregating travel time measurements of observations i close to t and c. The indirect estimation consists instead in (i) estimating virtual speed $v_{\text{virtual}}(c,t)$ at t and c from virtual speeds $v^{(i)}_{\text{virtual}}$ of the observations close to t and c and (ii) estimating $T(c,s)$, based on the distance $d_{\text{beeline}}(c)$ between c and the respective hub as $T(c,s) = d_{\text{beeline}}(c)/v_{\text{virtual}}(c,t)$.

The advantage of the indirect method is that one can hope that virtual speeds are an invariant of the dataset, i.e., they do not vary too much with respect to the considered times of day and locations, or at least they vary less than travel times. The estimation of virtual speeds would thus suffer less variance then the direct estimation of travel times. This could consent to use more observations when estimating quantities at c and t, which would be a great advantage, due to the scarcity of data available. Similarly, another indirect estimation would rely on estimating detour factors.

Observe that in case of indirect estimation, one would estimate travel time $T(c,t)$ as a function of distance $d_{\text{beeline}}(c)$. To evaluate the feasibility of using indirect estimations, it is important to check the information contained in $d_{\text{beeline}}(c)$ is sufficient to keep a satisfying correlation with the travel time.

Figure 12 shows correlations ranging from 0.83 to 0.7. The highest correlation is observed between the actual traveled distance and the travel time. By simplifying the routes down to a beeline, information about the detour and the circuity is lost. Only the travel time is representing the detours taken by vehicles. This suggests that using indirect estimation might require additional treatment, which is outside of the scope of this work. This justifies the fact the choice in this work to only use direct estimation.

In effect, the virtual speed v_{virtual} reduces with the overall distance, indicated by the trend of slope larger than 1 of the beeline vs travel time, whereas the actual speed of the vehicles are more constant along the traveled route, trend of slope of traveled distance vs travel time closer to 1. $v_{virtual} = \frac{d_{beeline}}{t_{traveltime}}$. Concluding from the decrease in $v_{virtual}$ with higher distances $d_{beeline}$, by traveling longer times, longer detours and higher circuity are observed. Figure 12 also allows to verify if the expected evolution of travel times over time of day is reflected in the dataset. Comparing the off peak to peak hour trips, the increase in travel time for same distances can be observed. The actual travelled distance and the connected travel times are of special interest. For same distances, more time is needed during peak hours as the higher number of trip requests allow for efficient pooling, causing a higher detour factor. On the other side, ruling out detour and circuity, the effect of representing this difference in travel time over the beeline distance is preserved.

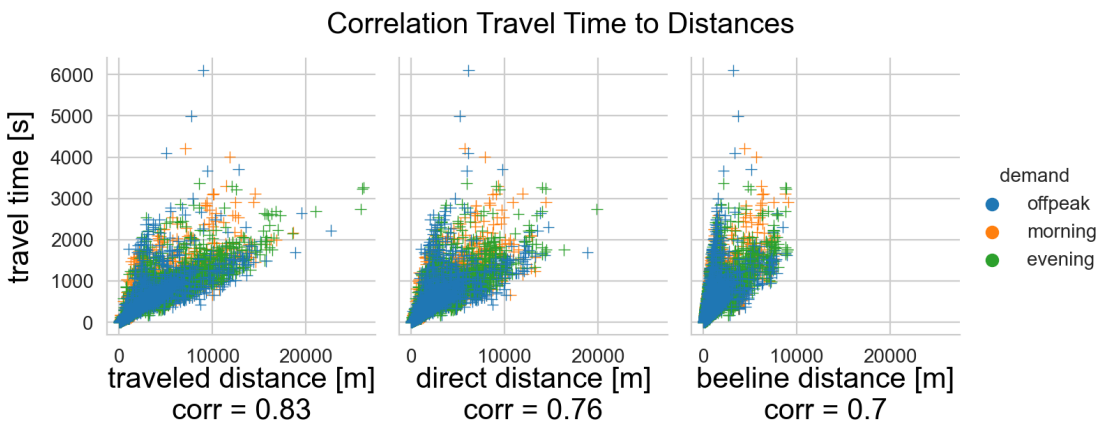

Figure 12 Pair-plot Travel Time - Access and Egress Trips - Traveled Distance, Direct Distance, Beeline Distance - Morning peak hours from 07:00 until 10:00 evening peak hours from 16:00 - 21:00 deviated from Figure 11.

The same analysis can be performed for wait time. Insight into the correlation of the wait time and the three distance measures are giving indications about the dependencies of service availability in close to hub versus remote regions of the study area. Visualised in Figure 13, it becomes obvious that no correlation between wait time and any of the observed distances exists. Thus, the theorem is formulated, that the wait time is independent in space, and only

varies in time in correlation to the overall amount of requests and state of traffic. This needs to be confirmed by further analysis.

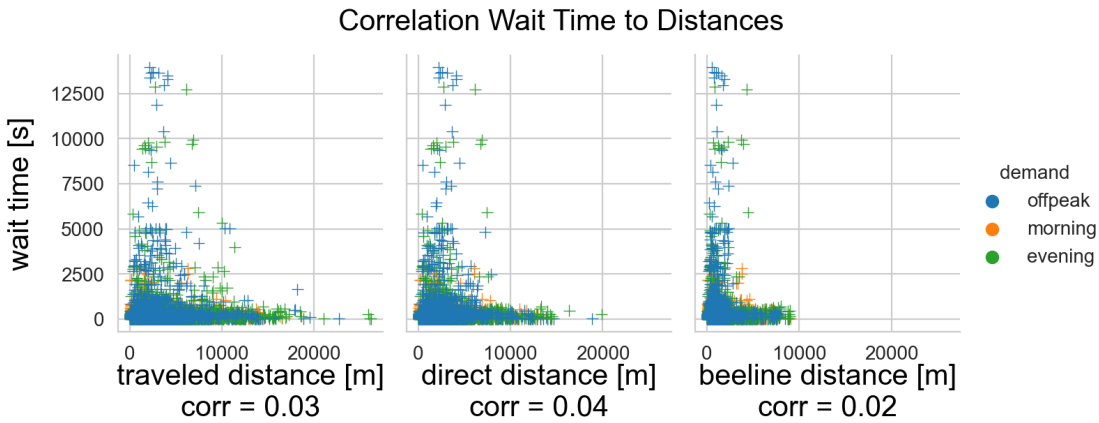

Figure 13 Pair-plot Wait Time - Traveled Distance, Direct Distance, Beeline Distance - Access and Egress Trips - Morning peak hours from 07:00 until 10:00 evening peak hours from 16:00 - 21:00 deviated from Figure 11.

Analysing the development of travel time over the course of one day using Figure 14, the upper limit of 600 seconds during the dispatching phase becomes obvious. All trips with wait times exceeding this limit experience delay of any other kind, like unexpected traffic. Even with the confidence interval during the entire day spanning a large range, the set limit of 10 minutes ensures a certain degree of confidence into the mean value. A very important factor not visualized in this plot is the spatial and temporal distribution. The spread of values is drastically reduced by only taking into account values of one spatial region or one time interval, further increasing the confidence in the mean value. To showcase this narrowing of the confidence interval, Figure 14 illustrated the wait times for access trips only. This replicates the differentiation as it is used in the implementation. Further, wait times longer that the predefined time limit of 10 minutes are observable. As initially mentioned, this limit applies to the planning phase, when the dispatcher decides to serve one request. If one vehicle is dispatched to serve a request, it will definitely serve the request, even though the route to the location takes longer and therefore the wait time limit is exceeded.

5.3. Quality of Aggregation and Estimation Model

After having established an overview of the input data, the introduces methodology is applied, including temporal aggregation and spatial estimation. Starting with the temporal aggregation. As described, a suitable time slot size has to be found to ensure, that global trends of wait and travel time are not generalized but also as many datasets are included in one time interval as possible to increase confidence in the spatial estimation model.

The three boxplots in Figure 15 compares three different temporal aggregations. 1h, 30min and 15min intervals. Once again, only the access trips are observed to include the same differentiation as the implementation is using. Looking at the coarsest aggregation of 1h,

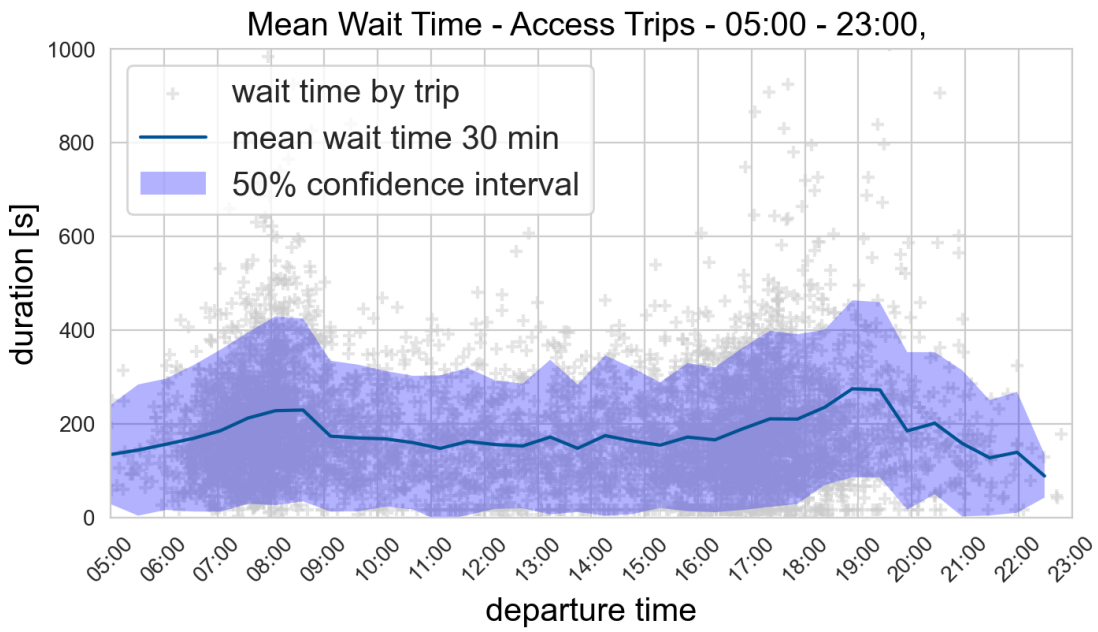

Figure 14 Mean Wait Time - Moving average of wait time during one day

the trends over one day, as presented in Figure 14 are clearly visible. Looking into the two shorter time interval aggregations, no further information is revealed. Thus, shortening the time intervals to less than 1h does not benefit the quality of the model, but only reduces the amount of trips within one interval and therefore increases the risk of uncertainty in the following step. Similar to all other analysis steps, only the holistic picture is presented, not differentiating between hubs. Taking into account, the identification of hubs only serving a low number of trips over the course of one day, it needs to be ensured, that the time intervals are as long as possible. Settling for 1h is identified as a suitable middle ground to go forward. The selection was equally confirmed with egress trips. No additional conclusions were established from it.

Having established and confirmed a suitable temporal aggregation, spatial estimation model needs to be applied to the provided data. As described in the methodology, spatial autocorrelation is used to model both travel and wait time in space. To verify, the feasibility of this step, the auto-correlation of the values in space is reviewed. Figure 16 illustrates the travel time in the study area during the morning peak hour. Equivalently, off peak ad evening peaks were examined to confirm the findings. The plot includes all available hubs, so the differentiation of estimating travel time for only one hub is not yet included. Nevertheless, a clear trend is observable. Travel times close to hub locations are homogeneously short and further away locations have higher travel times. It is of special importance for utilising spatial autocorrelation, that values close to each other show a similarity. This can be clearly observed in the plot.

This impression does not apply as significantly for wait time, as can be seen in Figure 17. Overall, the range of values is lower, only reaching to a maximum of 350 seconds (approx. 6

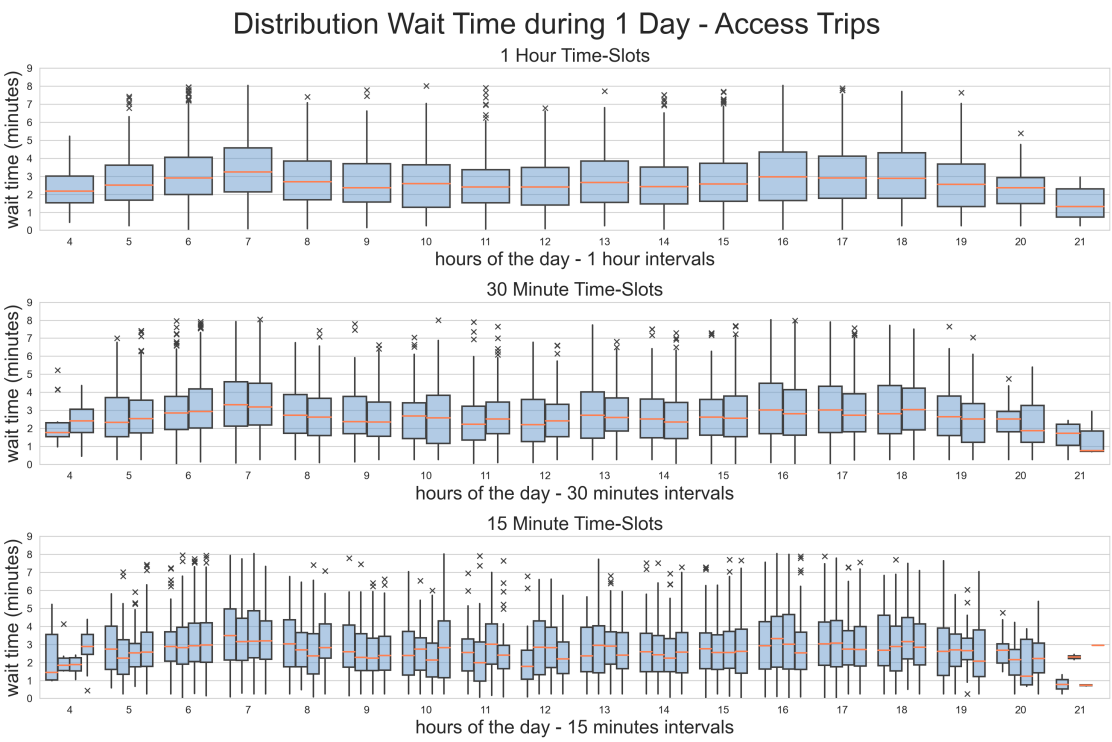

Figure 15 Confidence and Preservation of Detail for Time Intervals - Comparing 1h, 30min and 15min time intervals to illustrate that the overall trend in the data is not lost by aggregation. Breakout values were not considered, as they result from irregularities in the simulation, further emphasising the temporal pattern, focusing the observation on wait times in the targeted 0-10 minute range.

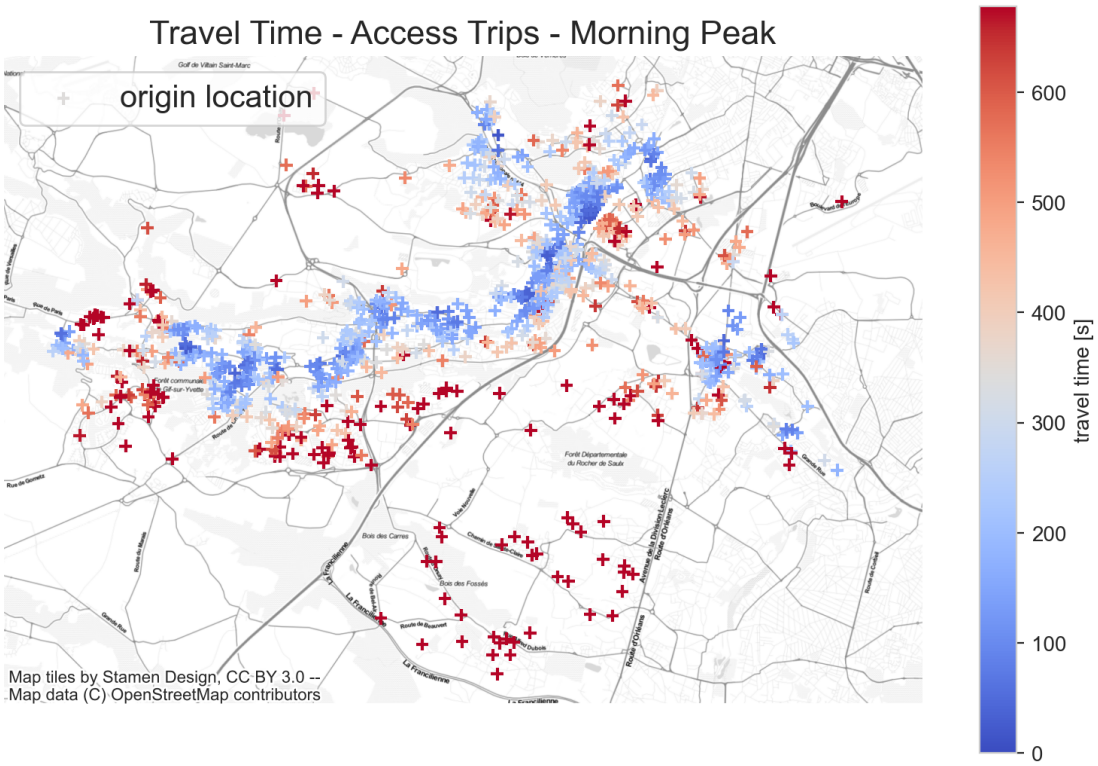

Figure 16 Spacial Trend Travel Time - North always at the top. Pattern can be identified, indicating high spatial autocorrelation

minutes). But values within one area do not show a very clear pattern of homogeneity. This can challenge the the spatial estimation. As discussed before, by relying on a mean value, constructed from surrounding values. local fluctuations are compensated. Thus, the spatial estimator will intentionally flatten these highest and lowest wait times. It can be expected, that the variance error of the estimation will be higher for wait time than for travel time as later shows a clear pattern that can be picked up by the model.

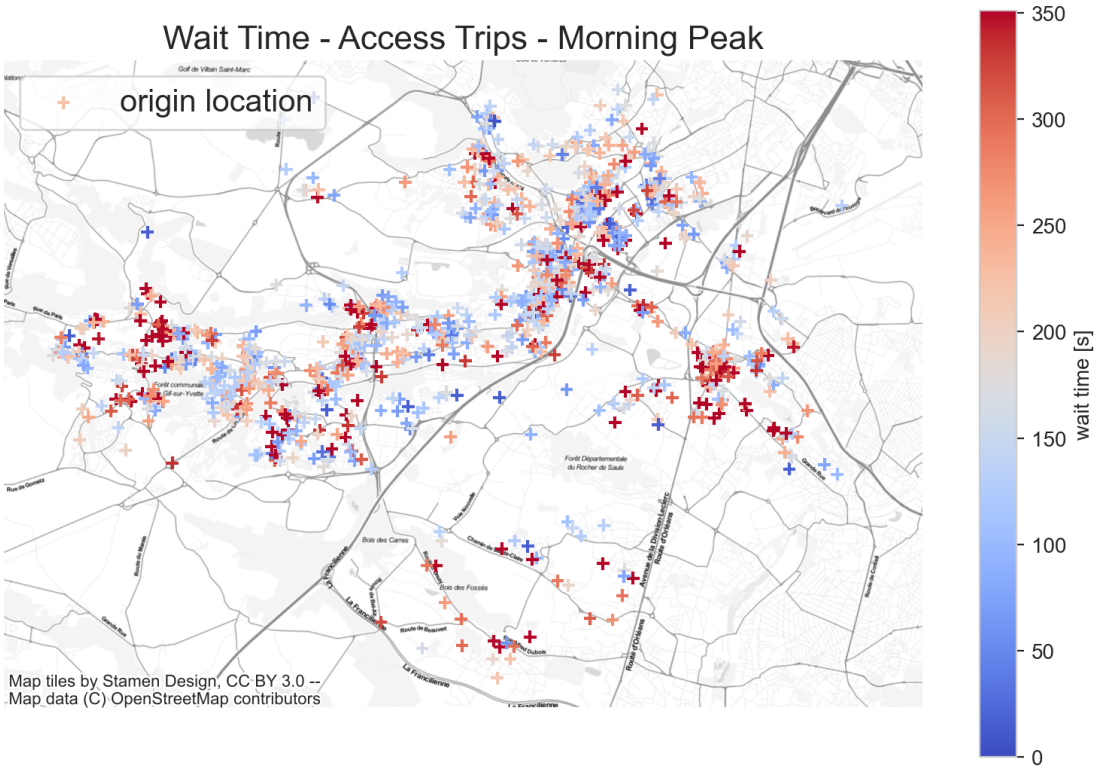

Figure 17 Spacial Trend Wait Time - North always at the top. No clear pattern can be identified, indicating low spatial autocorrelation

To showcase the spatial autocorrelation, semivariograms for both travel time (Figure 19)and wait time (Figure 19) are plotted. The made observations in Figure 16 and 17 in regards to correlation for travel time and low correlation for wait times can be confirmed. The measure for similarity, covariance, starts off higher than the measure for dissimilarity, semivariance, and reduces with increasing distance. The same can be observed for the wait time with the important difference of the semivariance dominating even at low distance between observations. Taking this information, the following conclusion can be drawn. Ordinary Kriging for travel time is picking up the spatial auto-correlation. Travel time is clearly related to the location in space. On the other hand, wait time does not correlate with the position. Estimation of values at unknown location with surrounding values nevertheless will provide a estimation that represents the service level at that location at the given time. As a result of these circumstances, the estimation is expected to have a large variance error for wait time estimations, even when the location is well surrounded by observations. This variance error will be lower for estimation of travel time, as clear trends are observed that can be picked up. The error will increase towards the edges of the observations.

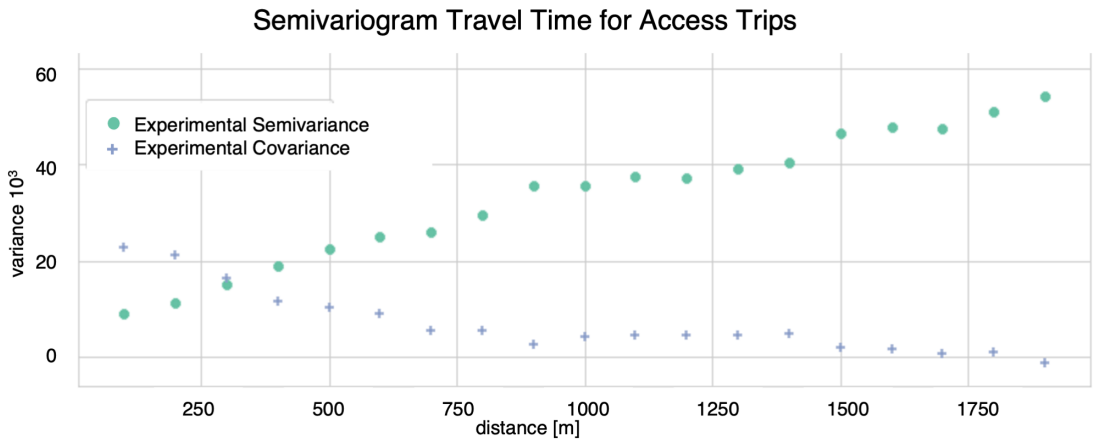

Figure 18 Semivariogram Travel Time of Access Trip Observations - Covariance indicating similarities in close proximity whereas with increasing distance between observations, the semivariance, dissimilarity, is the ruling factor. This is indicating spatial autocorrelation between observations

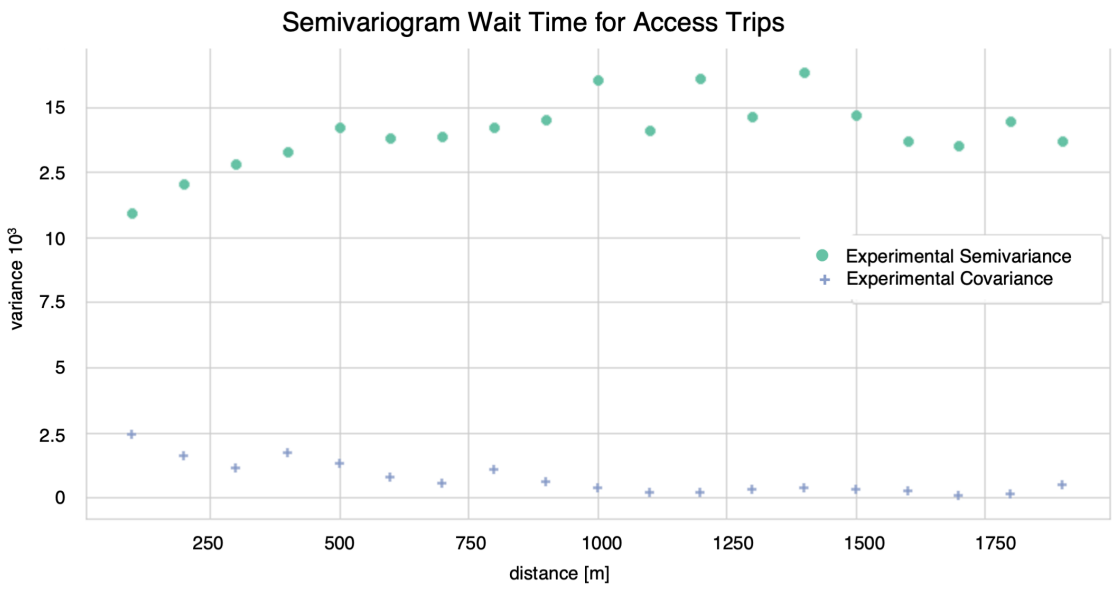

Figure 19 Semivariogram Wait Time of Access Trip Observations - High dissimilarity is observable, semivariance higher that covariance, even in areas with high spatial proximity, indicating low spatial autocorrelation between observations.

Having analysed the quality of every step of the estimation process, it is now possible to estimate values at every needed location in space. For every time slot in which sufficient data is present, the wait time and travel time is estimated. As described in Section 4. The next step involves the establishment of the virtual PT lines. As laid out in Section 3, during every time interval, a constant headway and travel time is assumed. Figure 20 shows the resulting departure times for one day for one access cell-hub relation. For demonstration purposes relations from cells with high frequency, many departures during one time interval, and low frequency, less departures during one time interval were visualised. The overall fluctuation trough out of headway, space between each cross, and the travel time, value on y axis, become obvious. For low frequency hubs, the fluctuations are less significant than for high frequency hubs.

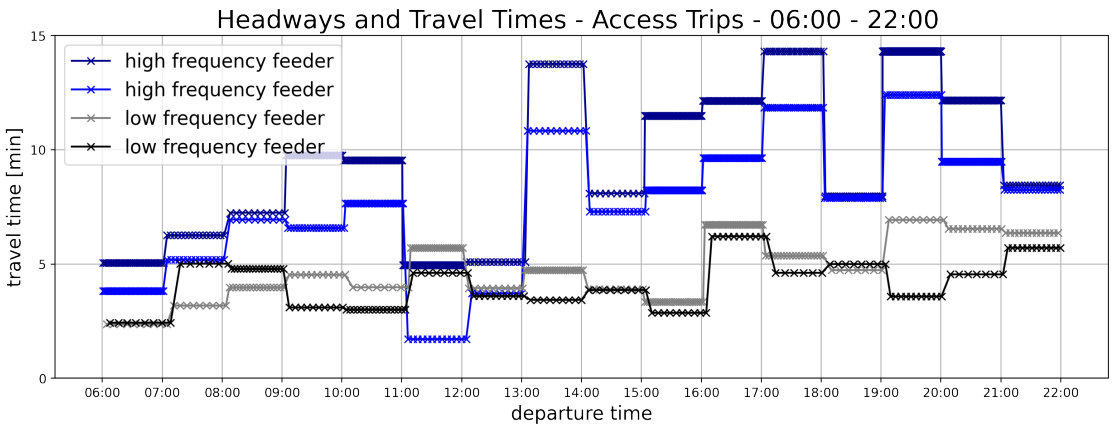

Figure 20 Headway and Travel Times in output GTFS - The variation of headway, density of crosses on each line, and the difference in travel time, values on y-axis, are visualised. The blue lines represent cell-hub relations with a very high number of departures, showing great fluctuations in travel time. The grey lines have lower number of trips and a more constant travel time.

5.4. Improvement of Accessibility Brought by DRT Integration

Having analysed the quality of steps towards the graph bases representation of dynamic mobility modes, the actual accessibility analysis is conducted using CityChrone. As described in Section 4, the output of the novel tool is a combined GTFS dataset in which both LBT and the representation of DRT are included. Using the synthetic population of the simulation scenario and the street network from open street maps, enables to calculate the sociality score. To be able to analyse the differences between the base case, without any dynamic services, and the scenario integrating dynamic services as feeder, for both cases the accessibility scores are calculated. As described in Section 2.1, the sociality score indicates how many people can be reached within a certain time. For the used setup of CityChrone, this time limit is set to 1h. Furthermore, the tool is calculating the score for every hour allowing for differentiation between different times of the day. Using this fact, the change of accessibility score will be analysed, similar to the previous results, for the different demands of the day, e.g. peak and off peak hours. Even though different times intervals of one day are considered, the scale remains constant as in every case the average of each time interval is taken. To explain the effect of the newly integrated DRT services better, at first a system with pure access services is observed. The DRT representation therefore only delivers passengers to the hubs but not provides egress trips. This enhances the the contrast of the result. Figure 21 shows the morning peak hour compared in the base case and with dirt access trips. Two main observations can be made. First the extension of the catchment area, especially in the south of the study area. By implementing virtual DRT stations in areas without any PT service, new connectivity is created. In areas with preexisting service, the quality of service is enhanced by offering higher frequency and faster access to rail lines. After the DRT implementation most areas with low sociality score within the study area are improved, indicating a equalisation within the area. As dynamic feeders still require changeovers including wait times at the station, a contrast to locations with direct walk access to MRT stops in visible.

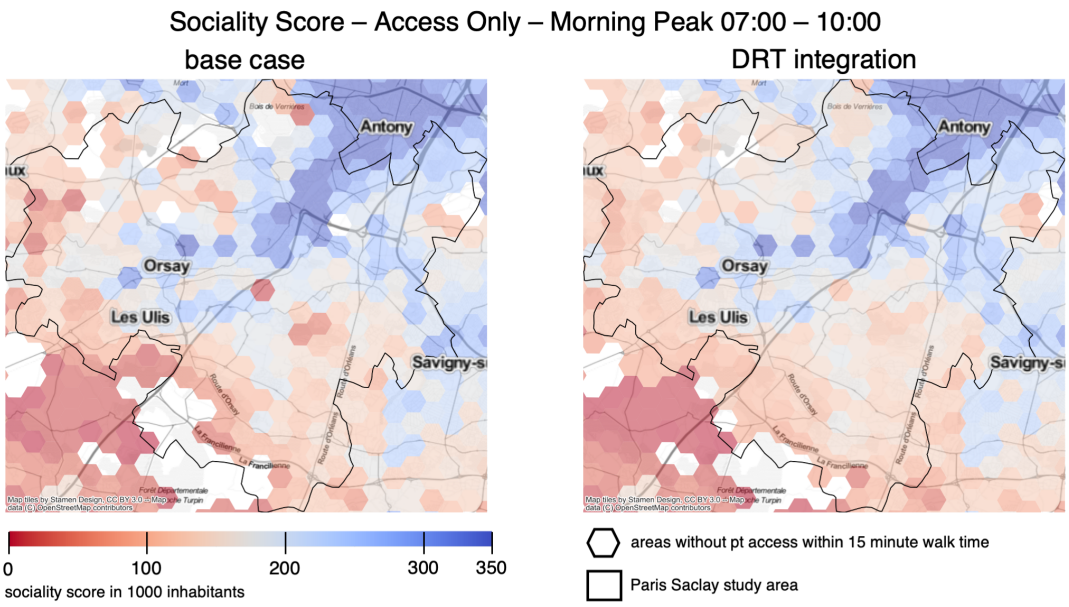

Figure 21 Sociality Score - Access Only - Morning Peak 07:00 - 10:00

The made observations can be verified by analysing Figure 22. Here, only the improvement of the sociality score is plotted. Areas with initially low service quality experience the largest improvements and areas along the existing rail line the least change in accessibility, as relying on direct walk access remains more flexible. As only access services are observed, the areas outside of the study area do not show any changes. The areas with slides improvements in the Île-de-France area can be explained with unconventional changeover connection that are made possible by the dynamic modes.

Looking at other time intervals, like the off peak hours during noon and afternoon, more significant improvements can be observed both in Figure 23 and Figure 24. This can be explained with the variation of PT schedules trough out the course of one day. LBT, especially buses, offer higher frequencies during peak hours and reduce their service during the rest of the day. By integrating DRT, this difference in service can be compensated and a higher service quality can be offered in many areas. So during off peak hours, the new dynamic mode not only enlarges the catchment area ans slightly improves the service in already served areas, but also compensates for the reduction of frequency of LBT.

Lastly, the difference of sociality score for one entire day is analysed for a system integrating both access and egress trips. This opens the possibility for users from outside the study area to enter by rail and access population within the area using DRT. Consequently the improvement outside of the study area will be more identifiable than in the so far analysed cases. By looking at Figure 25 exactly this becomes visible. Furthermore, the connectivity within the study are in drastically increased. As the main improvement is once again confined by the boundaries of the study area, the conclusion can be drawn, that users are now able to reach population within the the study area by choosing inter modal chain with two DRT legs.

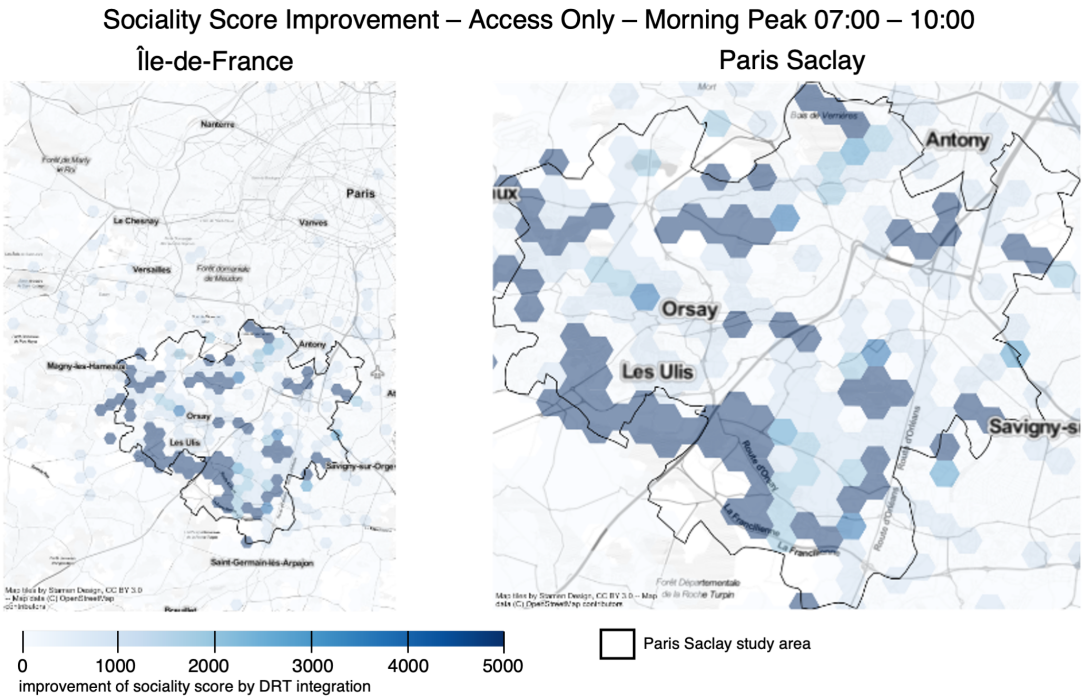

Figure 22 Sociality Score Improvement - Access Only - Morning Peak 07:00 - 10:00

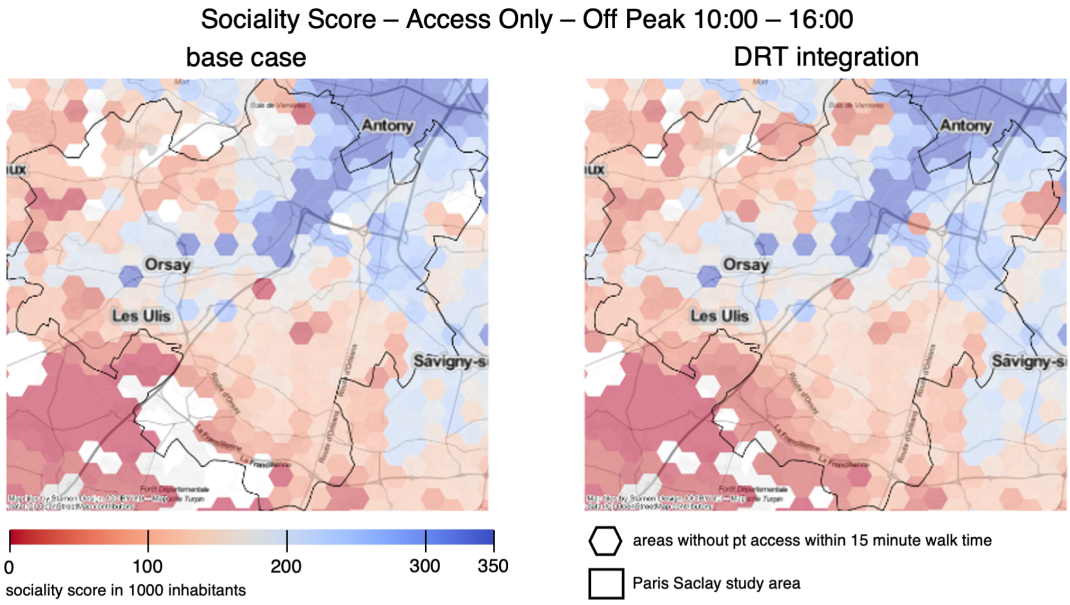

Figure 23 Sociality Score - Access Only - Off Peak 10:00 - 16:00

As a significant increase in the sociality score can be observed, a many new connectivities in comparison to the base case were created that weren't at all served by LBT. This uncovers potential of the dynamic modes, that have not been taken into account so far and that have occurred as a side product of the main intention. Other time intervals, that the visualized ones, were analyzed to confirm above made statements. No additional insights were obtained from this.

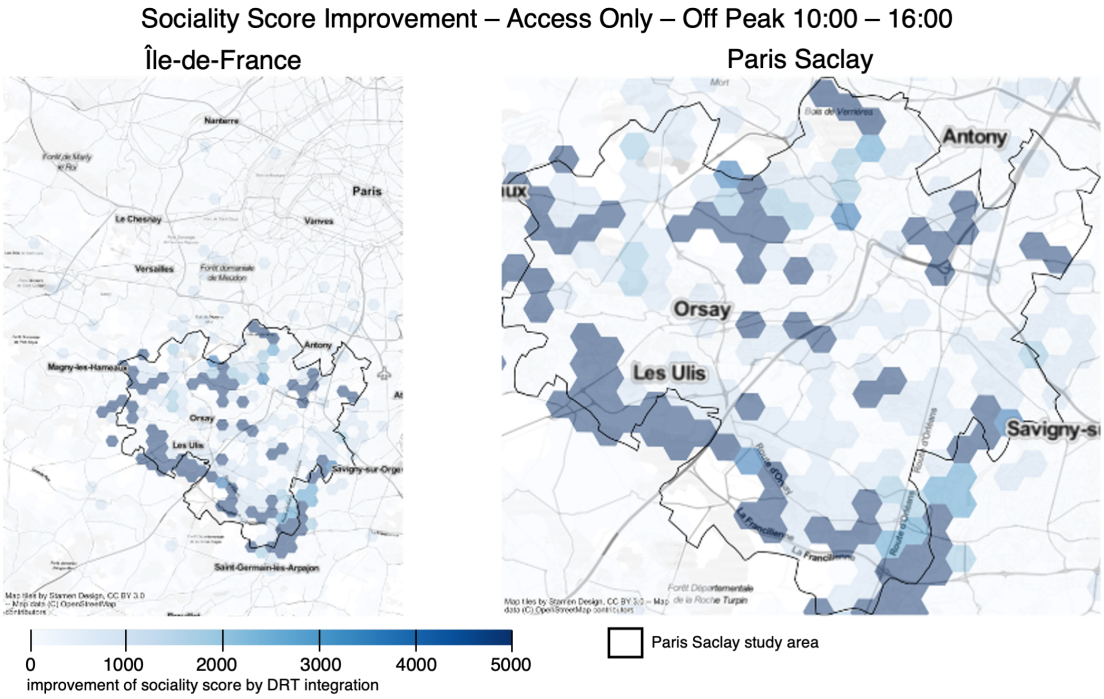

Figure 24 Sociality Score Improvement - Access Only - Off Peak 10:00 - 16:00

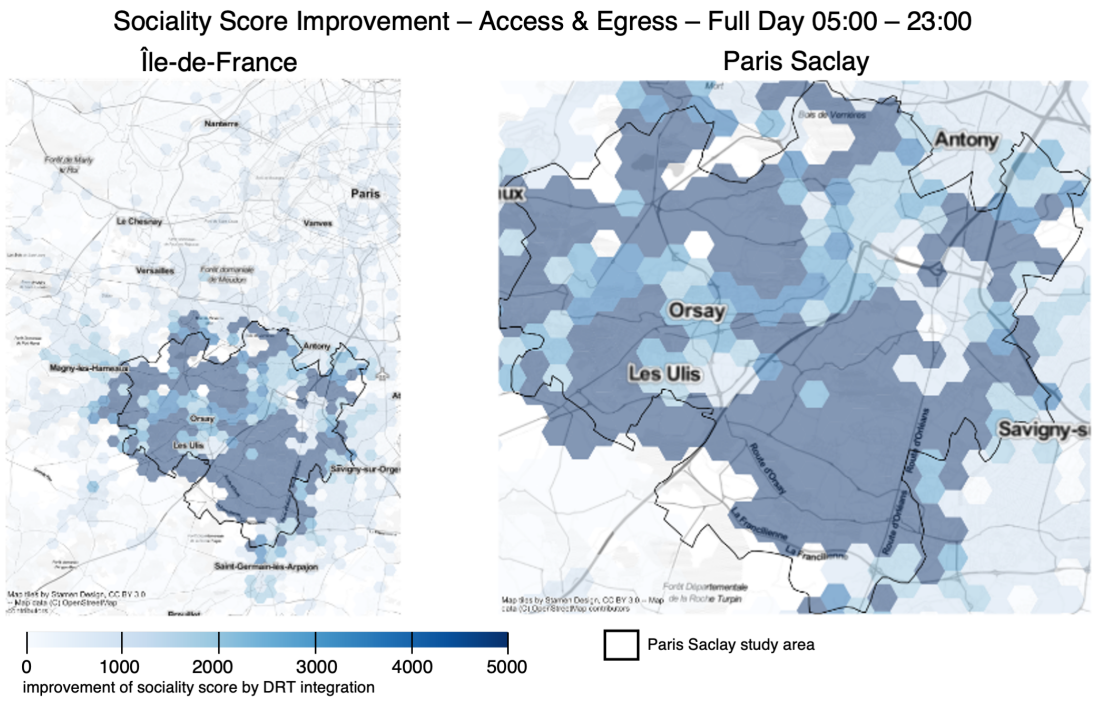

Figure 25 Sociality Score Improvement - Access & Egress - Full Day 05:00 - 23:00

6. Discussion & Conclusion

Layout drafting of the entire process for determining the gaps earlier research left, developing a methodology to bridge this shortcoming, implement a methodology as a novel conversion pipeline and finally applying the tool to existing datasets, are the most essential of the many steps this thesis incorporates, all having an important share of the outcome of the work worth reflecting. In the following, this work will be summarised to give an overview of the achievements made during the conduction of the research project.

6.1. Summary of Contribution

By developing a novel tool to aggregate and estimate trip data in space and time, the approach is breaking fresh ground that was fairly underestimated in previous research on DTS. One of the key motivations to follow this approach is its universal applicability. Not being tailored to only one dynamic mode but providing a reliable tool for many dynamic PT systems, is one of its outstanding feature. By establishing a framework within which the assumptions for spatial estimation of trips parameters are made, can be considered a solid foundation to build upon. The implementation then picks this up to provide a pipeline to the end user to exploit the methodological findings. In addition, the estimation part, a suit of functionalities is implemented to prepare the standardised data and to reduce the risk of delivering indeterminate or uncertain results by only providing solutions in areas with dense data coverage. For the input, the developed tool only builds upon on widely available information such as simulation and real-world data. The required information is limited to the bar minimum to ensure compatibility with the widest range of datasets possible. Further, the output of the pipeline is compatible with the industry standard for PT schedule distribution, enabling the use of the results in many different fields of application. Regarding the research topic to be solved, to ensure compatibility with an established accessibility tool was accomplished by not developing a new tool or significantly modifying the existing ones. To obtain a prove of concept, the developed tool is applied to a dataset for dynamic PT feeder service. A program test of each methodological step verifies assumptions made in the theoretical part. The resulting pattern clearly show improvements of accessibility by deploying DRT.

Summarizing the work carried out, the achieved goals are evaluated:

- Altogether a tools has been developed to obtain an accessibility measure for DTS modes integrated into LBT. The tool passed it initial tests for the prove of concept, but for now lacks an in-depth verification.
- The developed methodology allows the usage for multiple dynamic transport modes. The tool implementing the methodology has been designed for the use of DRT data but effort-

less can be adapted to account for different modes.

- The tool operates for different modes but also for different data sources. As only a few requirements had to be met to comply with the set standards, data sources such as simulation or real worlds data can be considered.
- The tool is universally applicable to every study area needed. The user only must provide the needed datasets, mainly the DTS trip data, spatial information about the opportunity of interest for the accessibility analysis and the existing PT and road network. Except for the trip data for DTS, these datasets are available as open-source datasets.
- Before for the final accessibility calculation step, the methodology outputs a graph-based representation of DTS. In terms of implementation, this graph representation is in the format of GTFS data. This standard is widely accepted and known. Many analysis tools for pt rely on format. Complying to this format allows a manifold use of these tools.

6.2. Concluding on Research Question

Referring to the research topic and thus, the main goal of the conducted work, a process to conduct accessibility analysis for DTS, especially DRT, needed to be developed. One approach for this process was successfully derived including the above-mentioned attributes. The resulting tool can be utilised to analyse the changes in accessibility by deploying dynamic transport modes in suburban areas, hence it can also analyse the possible equalisation of accessibility. As the initial literature research revealed the lack of a viable tool to analyse given scenarios for changes in accessibility, the focus of the research was shifted from the generation of different DRT integration scenarios to the development of the described tool. Nevertheless, the concluding application of the tool to an existing scenario already shows an improvement of accessibility in areas with DRT deployment. Further development of scenarios can further improve this result.

6.3. Future Outlook

The presented work acts as a prove of concept and establishes a solid basis for future developments of accessibility tools for DTS integration into LBT. Nevertheless, the series of evaluation steps and the implementation of the tool revealed some shortcomings that can be addressed by future research. Additionally, to fully be able to rely on the novel approach, verification steps must be accomplished. For further analysis of the suitability of the methodology developed, the results need to be compared to some benchmark. This comparison can be done by using the output graph-based description the PT network to rerun the simulation. Presently, no DRT is deployed but the service is simulated by the line based representation. The results, especially the mode choice of the agents, can give insights on how well the graph conversion of the dynamic service was performed. One parameter, that was not considered directly in this research are capacity constraints of the dynamic services. At this point, these

are aimed to represent by the performance indicators of wait time. If capacity is at the limit, wait times are higher and thus, the graph provides less service. Other solutions can be evaluated to incorporates this.
To improve the conversion pipeline additional research can be conducted on the aggregation and estimation using OK. It can be examined if the temporal component can be included directly into the kriging process by applying a 3-dimensional kriging method. This would further help to deal with scarcity of data.
In general, the inclusion of global knowledge, transferred between scenarios can be investigated. Counteracting the easy applicability of the tool independent of any non-scenario data, as it was developed, the inclusion of this information has the potential of improving the result significantly. On a high level, the computational optimisation of the implementation should be considered if large scale DTS deployments need to be analysed. For the presented scenario, the current implementation is sufficiently efficient and can be run on a standard personal computer. But definitely, there is room for improvement regarding the performance of the implementation.

To finalize the thesis, a general perspective regarding DTS, and DRT in particular, as feeder services, seems appropriate. The relevance of dynamic PT options is continuously increasing. The economic and environmental advantages in addition to the potential increase in comfort for the passenger, thus more demanded, makes any improvement surly viable.PT operators, public stakeholders and potential users need to be convinced that these novel transportation systems can provide widely accepted solutions for the requirements each party has. An accessibility tool such as the presented one or similar, can help to support decision making for deployment of such systems. Given the interest in the resulting product, also research must further deepen its knowledge in this field to cope with questions and problems looming ahead. DTS and its impact on the travel behaviour of the population will continue to be a research field of high interest and importance.

Bibliography

[1] M. G. Boarnet, G. Giuliano, Y. Hou, and E. J. Shin, "First/last mile transit access as an equity planning issue," *Transportation Research Part A: Policy and Practice*, vol. 103, pp. 296–310, 2017.

[2] S. Kaufman, M. L. Moss, J. Tyndall, and J. Hernandez, "Mobility, economic opportunity and new york city neighborhoods," *NYU Wagner Research Paper*, no. 2598566, 2014.

[3] C. Vitrano and L. Mellquist, "Spatiotemporal accessibility by public transport and time wealth: Insights from two peripheral neighbourhoods in malmö, sweden," *Time & Society*, p. 0961463X221112305, 2022.

[4] P. B. Goodwin, "Empirical evidence on induced traffic: A review and synthesis," *Transportation*, vol. 23, pp. 35–54, 1996.

[5] L. J. Davison and R. D. Knowles, "Bus quality partnerships, modal shift and traffic decongestion," *Journal of Transport Geography*, vol. 14, no. 3, pp. 177–194, 2006.

[6] E. Pajares, U. Jehle, and E. B. Hassine, *Geo Open Accessibility Tool (GOAT)-Schlussbericht: Entwicklung eines offenen und interaktiven Erreichbarkeitstools für den Fuß-und Radverkehr auf Basis von Open Data: Laufzeit: 11/2019-02/2021*. Technische Universität München, 2021.

[7] G. Wulfhorst, B. Büttner, and C. Ji, "The tum accessibility atlas as a tool for supporting policies of sustainable mobility in metropolitan regions," *Transportation Research Part A: Policy and Practice*, vol. 104, pp. 121–136, 2017.

[8] C. Mulley and J. D. Nelson, "Flexible transport services: A new market opportunity for public transport," *Research in Transportation Economics*, vol. 25, no. 1, pp. 39–45, 2009.

[9] B.-h. Nahmias-Biran, J. B. Oke, N. Kumar, C. Lima Azevedo, and M. Ben-Akiva, "Evaluating the impacts of shared automated mobility on-demand services: An activity-based accessibility approach," *Transportation*, vol. 48, pp. 1613–1638, 2021.

[10] S. Chandra, M. E. Bari, P. C. Devarasetty, and S. Vadali, "Accessibility evaluations of feeder transit services," *Transportation Research Part A: Policy and Practice*, vol. 52, pp. 47–63, 2013.

[11] I. Biazzo, "Citychrone: an interactive platform for transport network analysis and planning in urban systems," in *Complex Networks & Their Applications X: Volume 2, Proceed-*

ings of the Tenth International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2021 10, pp. 780–791, Springer, 2022.

[12] A. Badeanlou, A. Araldo, and M. Diana, "Assessing transportation accessibility equity via open data," *arXiv*, no. 2206.09037.

[13] P. R. Gould, "Spatial diffusion, resource paper no. 4.," 1969.

[14] W. G. Hansen, "How accessibility shapes land use," *Journal of the American Institute of planners*, vol. 25, no. 2, pp. 73–76, 1959.

[15] K. T. Geurs and B. Van Wee, "Accessibility evaluation of land-use and transport strategies: review and research directions," *Journal of Transport geography*, vol. 12, no. 2, pp. 127–140, 2004.

[16] H. J. Miller, "Place-based versus people-based accessibility," in *Access to destinations*, Emerald Group Publishing Limited, 2005.

[17] S. Derrible and C. Kennedy, "Network analysis of world subway systems using updated graph theory," *Transportation Research Record*, vol. 2112, no. 1, pp. 17–25, 2009.

[18] E. J. Miller, "Measuring accessibility: Methods and issues," International Transport Forum Discussion Paper.

[19] E. J. Miller, "Accessibility: measurement and application in transportation planning," *Transport Reviews*, vol. 38, no. 5, pp. 551–555, 2018.

[20] D. O'Sullivan, A. Morrison, and J. Shearer, "Using desktop gis for the investigation of accessibility by public transport: an isochrone approach," *International Journal of Geographical Information Science*, vol. 14, no. 1, pp. 85–104, 2000.

[21] S. L. Handy and D. A. Niemeier, "Measuring accessibility: an exploration of issues and alternatives," *Environment and planning A*, vol. 29, no. 7, pp. 1175–1194, 1997.

[22] M.-P. Kwan, "Space-time and integral measures of individual accessibility: a comparative analysis using a point-based framework," *Geographical analysis*, vol. 30, no. 3, pp. 191–216, 1998.

[23] M. Adnan, F. C. Pereira, C. M. L. Azevedo, K. Basak, M. Lovric, S. Raveau, Y. Zhu, J. Ferreira, C. Zegras, and M. Ben-Akiva, "Simmobility: A multi-scale integrated agent-based simulation platform," in *95th Annual Meeting of the Transportation Research Board Forth-*

coming in Transportation Research Record, vol. 2, The National Academies of Sciences, Engineering, and Medicine Washington, DC, 2016.

[24] K. W Axhausen, A. Horni, and K. Nagel, *The multi-agent transport simulation MATSim*. Ubiquity Press, 2016.

[25] D. Ziemke, *Accessibility. In: Horni, A et al (eds.), Multi-Agent Transport Simulation MATSim*. Ubiquity Press, 2016.

[26] I. Biazzo, B. Monechi, and V. Loreto, "General scores for accessibility and inequality measures in urban areas," *Royal Society open science*, vol. 6, no. 8, 2019.

[27] R. C. Larson and A. R. Odoni, *Urban Operations Research*. No. Monograph, Prentice-Hall, 1981.

[28] C. A. Nash, *Economics of public transport*. No. Monograph, 1982.

[29] R. Gomes, J. P. de Sousa, and T. G. Dias, "Design and operation of demand responsive transportation systems," in *EWGT 2012 International Scientific Conference*, 2012.

[30] C. L. Hasif, A. Araldo, S. Dumbrava, and D. Watel, "A graph-database approach to assess the impact of demand-responsive services on public transit accessibility," in *Proceedings of the 15th ACM SIGSPATIAL International Workshop on Computational Transportation Science*, pp. 1–4, 2022.

[31] P. Fortin, C. Morency, and M. Trépanier, "Innovative gtfs data application for transit network analysis using a graph-oriented method," *Journal of Public Transportation*, vol. 19, no. 4, pp. 18–37, 2016.

[32] A. Nuzzolo and A. Comi, "A run-based optimal strategy search method in intelligent stochastic transit networks," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 2633–2638, IEEE, 2019.

[33] F. R. P. Schoenberg, "Point process, spatial-temporal." 2023.

[34] M. S. Handcock and J. R. Wallis, "An approach to statistical spatial-temporal modeling of meteorological fields," *Journal of the American Statistical Association*, vol. 89, no. 426, pp. 368–378, 1994.

[35] R. B. Dial, "Transit pathfinder algorithm," *Highway Research Record*, no. 205, 1967.

[36] AA.VV., *Advances in Sensitivity Analysis of Uncertainty to Changes in Sampling Density When Modeling Spatially Correlated Attributes*, ch. 19. Springer Open, 2018.

[37] S. Y. Chung, S. Venkatramanan, H. E. Elzain, S. Selvam, and M. Prasanna, "Supplement of missing data in groundwater-level variations of peak type using geostatistical methods," *GIS and geostatistical techniques for groundwater science*, pp. 33–41, 2019.

[38] J.-P. Chilès and N. Desassis, *Fifty Years of Kriging*, pp. 589–612. Cham: Springer International Publishing, 2018.

[39] B. A. Salima and M. Bellefon, "Spatial autocorrelation indices," *Handbook of Spatial Analysis: Theory Aplication with R*, pp. 51–68, 2018.

[40] A. Getis, "Reflections on spatial autocorrelation," *Regional Science and Urban Economics*, vol. 37, no. 4, pp. 491–496, 2007.

[41] S. J. Yakowitz and F. Szidarovsky, "A Comparison of Kriging with Nonparametric Regression Methods," *Journal of Multivariate Analysis*, 1985.

[42] "Reference document, static transit, https://developers.google.com/transit/gtfs/reference?hl=en, google transit feed reference document, 01.03.2023."

[43] S. Moliński, "Pyinterpolate: Spatial interpolation in python for point measurements and aggregated datasets," *Journal of Open Source Software*, vol. 7, no. 70, p. 2869, 2022.

[44] T. Chouaki and J. Puchinger, "Agent based simulation for the design of a mobility service in the paris-saclay area," *Transportation Research Procedia*, vol. 52, pp. 677–683, 2021.

Appendix

The appendix is provided digitally in a LRZ Sync+Share folder. The folder contains:

- All code files described in the implementation section
- A digital pdf version of this thesis
- All figures presented in this thesis

The folder can be accessed using the following short url:
shorturl.at/KPZ07
or the following link:
https://syncandshare.lrz.de/getlink/fi4r8b95MGsAVunx75xwk8/

To simplify the handling of code, it is additionally pushed to the Git Repository:
https://github.com/severindiepolder/AccessibilityOfDynamicTransport