D1.2: Form and flow dynamics at interchange, urban and regional network layers

Project acronym:	TRANS-FORM
Project title:	Smart transfers through unravelling urban form and travel flow dynamics
Funding Scheme:	ERA-NET call on Smart Cities and Communities (ENSCC)

Authors

Oded Cats	o.cats@tudelft.nl
Johanna Törnquist Krasemann	johanna.tornquist.krasemann@bth.se
Nicholas Molyneaux	nicholas.molyneaux@epfl.ch
Clas Rydergren	clas.rydergren@liu.se
Riccardo Scarinci	riccardo.scarinci@epfl.ch
Shadi Sharif Azadeh	shadi.sharifazadeh@epfl.ch
Menno Yap	M.D.Yap@tudelft.nl
Yuki Oyama	yuki.oyama@epfl.ch

Internal Reviewer

Patricia Bellver	pbellver.etraid@grupoetra.com

State:	Draft
Distribution:	[Confidential]

Deliverable History

Date	Author	Changes
15-04-2016	Riccardo Scarinci, Shadi Sharif, EPFL	Draft table of contents
30-05-2016	Riccardo Scarinci, Shadi Sharif, EPFL	Update table of contents (titles, descriptions and time frame for completion)
28-07-2016	Oded Cats, Menno Yap, TUD	Draft version chapter 4.2 (planned analysis urban network level)
19-08-2016	Menno Yap, TUD	Draft version chapter 5 (definition of multi-level PT network)
31-08-2016	Johanna Törnquist Krasemann, Clas Rydergren, BTH/LiU	Draft version chapter 4.3
21-09-2016	Johanna Törnquist Krasemann, BTH	Updated information in chapter 4.3 based on meeting with Blekingetrafiken Sept. 8th.
03-10-2016	Riccardo Scarinci, EPFL	Update time frame and scope description of Section 6
11-12-2016	Clas Rydergren, LiU	Updated Chapter 4.3
13-12-2016	Menno Yap, TUD	Update Chapter 5
10-01-2017	Nicholas Molyneaux, EPFL	Included first versions of: introduction (chap 2), review (chap 3) and descriptive analysis of tracking data chap 4.2
11-01-2017	Menno Yap, TUD	Included descriptive analysis of urban network data 4.2
12-01-2017	Clas Rydergren, LiU	Included initial analysis of regional network data 4.3
21-06-2018	Clas Rydergren, LiU	Added cross references. Added info to Chapter 4.4.
04-07-2018	Yuki Oyama, EPFL	Global review, Added summary and conclusion
28-09-2018	Patricia Bellver, ETRA	Internal review
19-10-2018	Yuki Oyama, EPFL	Addressed comments from internal review

Contents

1. Summary

The main objective of WP1 “Unravelling form and flow dynamics” of the TRANS-FORM project is to understand well the characteristics and evolution of traveller flow dynamics at three different geographical scales, namely transport hub, urban and regional network layers, and their integration. This deliverable “D1.2: Form and flow dynamics at interchange, urban and regional network layers” mainly reports the result of Task 1.2 “Multi-level analysis of traveller flows” in WP 1. In this task, we analyse the traveller dynamics that take place at each network level of above-mentioned three different levels. Descriptive analyses by using big transportation data are reported here: pedestrian flows inside a single multi-modal transport hub of Switzerland, flow dynamics on an urban public transport network of the Netherlands, and the transit flows in a region of Sweden. A number of different techniques are used and developed for the analyses. In this task, we also establish a common system representation and mathematical notations to be used throughout the project and modelling developments. Urban form plays a significant role in traveller dynamics and mobility, as topography and human-built infrastructures impact the choice and mode people choose to move around.

2. Introduction

2.1. Purpose and approach

Prior to modeling any system, understanding the dynamics which take place with a global yet precise view is necessary. Therefore, the first steps of the TRANS-FORM project cover a descriptive analysis of the dynamic processes which take place at each level and in each different system. When transportation hubs are concerned focus should be directed towards pedestrian dynamics; for urban transportation systems, dynamics of both travelers and public transport modes (buses or trams generally) need to be taken into account; finally, for the regional/national level trains are the main concern as they must satisfy the traveler demand. As these multiple different modeling scales are studied, different analysis techniques exist: some are level independent while others are scale specific. This multi-level analysis is required for understanding flow dynamics of both pedestrians and public transport. Finally, urban form also plays a significant role in mobility as topography and man built infrastructures impact the choice and mode people choose to move around.

In this deliverable, the modeling concepts are defined and a descriptive analysis specific to each level is first done, and given that, a notation system for multi-level analysis is defined. Not only does this ensure consistent notation between the different partners and the modeling levels, but it also provides an overview of the different data to all participants in the project. To fulfill this objective, a description of the planned analysis at each level (hub, urban, regional) is given. Furthermore, a preliminary list of important variables is defined. This is done for notations common to different levels and also for level-specific variables.

2.2. Report Outline

Chapter 3 provides a literature review of the main topics covered in this deliverable. Following that, the descriptive analyses at each level are detailed in Chapter 4. To unify notation, Chapter 5 presents notations common to all levels in the first sub-chapter followed by level-specific notations. To emphasize the relevance of the data for this project and gain further insight, Chapter 6 presents a descriptive analysis of the data from the case studies of each level. Finally, Chapter 7 concludes this report and provides a summary of the talk-away points.

3. Measuring Traveller Form and Flows dynamics review

The ambitious objective of improving our knowledge and understanding of travel dynamics requires thorough investigation of data collected in real case studies and strong theoretical knowledge. Not only should the data collection techniques be performed in a rigorous way, but any model should be built on strong assumptions. The present review covers some of the important topics which should be considered when analyzing passenger transfer dynamics at multiple levels; this includes passenger dynamics but also public transport dynamics and urban form.

The first section covers the link between urban form and mobility, with the implication at different levels. Then, various data collection techniques which focus either on pedestrians, urban public transport or regional rail are overviewed. Finally, some data analysis techniques and some key modeling elements are presented.

3.1. Urban form

Over the past decades, most European cities have expanded in terms of inhabitants and size. In many cases, this is accompanied by urban sprawl which impacts different aspects of a city such as job/inhabitant density, mobility and even the structure of cities themselves. The structure of urban developments has been vastly covered in literature, and diverging opinions subsist on the conclusions.

On one hand the urban form can be seen as the source of mobility and the sprawl of cities should be controlled in order to limit the increase in private vehicle dependency (Camagni, Gibelli, & Rigamonti, 2002), and on the other hand the idea that the free-market approach will allow each individual to find his preferred place in the society (Dubois-Taine & Chalas, 1997). Some examples of positive and negative government policies can be found in (Anas, 1998), along with situations where interventionism is recommended.

3.1.1. Metropolitan areas & cities

When considering metropolitan areas or large cities, understanding the interactions between housing location, job location and travel behavior is important to help policy makers make sustainable decisions. Different structures of urban form have been observed by measuring job density (Anderson & Bogart, 2001). The authors observe that only 50% of jobs are located in dense employment centers, while the rest of the jobs are spread throughout the metropolitan area. Furthermore, they see a center-specific specialization in job type. This non-uniformity of land-use is one source of travel. In (Handy, Methodologies for exploring the link between urban form and travel behavior, 1996), some recommendations for exploring and improving our understanding of the link between travel behavior and urban form are given.

Nevertheless, one needs to quantify the impact of urban expansion. To achieve this, different indicators can be used and commonly grouped under the term "urban form" (Handy, Methodologies for exploring the link between urban form and travel behavior, 1996). The authors emphasize that multiple characteristics should be used to describe the configuration, not only inhabitant or job density. Quantification of urban sprawl/compactness can be done using various metrics like size, density or polycentrality (Tsai, 2005). When considering interactions between urban form and mobility, different indicators can be used. This is done in the context of mode choice estimation and activity chain modeling (Frank, Bradley, Kavage, Chapman, & Lawton, 2008). Some quantifiable measures of urban form focused on mobility are proposed: street connectivity, land use mix or bus stop density are only a subset of meaningful measures. Another study investigates the social and environmental impact of urban structures by measuring the travel time and mode choice for trips with different purposes. Not only does the urban form have a strong impact on the competitiveness of public transport, but also on the general health of the inhabitants (Camagni, Gibelli, & Rigamonti, 2002). The compact city model is presented as a sustainable solution in (Holden & Norland, 2005). Nevertheless, the authors also mention the possibility that large compact cities lose the advantages of centralized services. This could be caused by the congestion created by the many workers converging towards the job locations. Admitting the urban morphology can impact the mode choice, one can therefore use this as a "tool" to control automobile dependency, as in (Handy, Cao, & Mokhtarian, Correlation or causality between the built environment and travel behavior? Evidence from Northern California , 2005).

3.1.2. Hub scale

In recent years, number of large transportation hubs have diversified their activities, as airports have done since many years. Today, it is common to find many shops and restaurants in large public transport hubs. Not only do hubs attract pedestrians interested in using public transportation systems, but individuals who want to shop or eat are also attracted to the hub even if they do not plan on traveling.

3.1.3. Mobility forecasting: the four-stage model

When forecasting is of interest, the four-stage model can be considered. Applied in the 1950s when computers started to become tools used by analysts, the four-step model is a general framework for modeling and forecasting land-use and travel demand. The stages are the following (McNally, 2007):

- <u>Trip generation.</u> The number of trips is generated. These trips can be inferred from survey data for example. This step transforms activity-based data into trip-based data.
- <u>Trip distribution.</u> Once individual trips have been defined, these are aggregated into an origin-destination (or production-attraction) matrix.
- <u>Mode choice.</u> Each trip is now assigned a mode, based on the selection probability from survey data.
- <u>Route choice.</u> This final stage assigns each trip to a specific route based on demand.

Although the four-stage model has been extensively applied and researched, the accuracy of the results is limited. The segregated and sequential aspects of the model mean no feedback from the mode & route choice is given to the trip generation and distribution. Furthermore, the impact on transport infrastructure on land-use is not taken into account.

3.2. Data collection

One critical step in any data analysis or modeling project is acquisition of reliable and accurate data which can be used as a case study or an industrial application. Although this is true for both pedestrian and public transport modeling, different challenges are associated with each field. For example, when considering pedestrians, one must deal with heterogeneity of the participants, whereas public transport vehicles of the same fleet generally share common characteristics.

Two significantly different approaches exist for collecting pedestrian data: surveys or computer-based methodologies. These two alternatives have different advantages and disadvantages. Even though surveys can be used to collect data which cannot be collected using external automated procedures, the distribution, collection and processing phases are time consuming and costly. Computer-based methods are efficient for collecting large-scale data with a somewhat systematic and repetitive component. When PT-focused data is concerned, localization data can be collected by GPS or accelerometer for example.

3.2.1. Pedestrian-centered

As stated previously, two different techniques can be used to collect data concerning pedestrians. When data about hidden attributes is of interest (journey purpose, route choice motivation) surveys are interesting as information concerning non-observable attributes can be collected (Sisiopiku & Akin, 2003). Other appealing reasons for using surveys are long distance motorized trips (Atasoy, Glerum, & Bierlaire, 2013), living location choice (Handy, Urban form and pedestrian choices: study of Austin neighborhoods, 1996) or pedestrian activity choice (Bierlaire & Moura, Effects of terminal planning on passenger choices, 2014).

If flow dynamics are to be studied, then computer-based technologies are efficient as camera-based tracking (detection) systems can be installed. Many different combinations of algorithms, recording technologies and applications exist; both for industrial and research purposes. The following paragraphs provide a small overview of different technologies available.

Pedestrian activity chains can be observed using WLAN traces as in (Danalet, Farooq, & Bierlaire, 2014), where traces left by smartphones are analyzed in order to determine pedestrian activities inside a campus. For modeling of pedestrian flows, detailed tracking data can be collected using different technologies such as inertial sensors (Foxlin, 2005) or networks of cameras (Alahi, Vandergheynst, Bierlaire, & Kunt, 2010). Many different flavors of these technologies exist with varying precision, reliability and focus. In some cases, the objective is to simply count the number of pedestrians who cross an imaginary line; while at the other extreme of the resolution scale is detailed space-time tracking of individual pedestrians. For an in-depth review of pedestrian detection technologies, we refer to (Bauer, Brändle, Seer, Ray, & Kitazawa, 2009).

High resolution tracking of pedestrians cannot be applied at the scale of a city or region using fixed location detection technologies as too many sensors would need to be installed, therefore an alternative approach is used: provide pedestrians with personal tracking devices. Although this seems challenging from a practical and legal aspect, the 21st century has provided the answer: smartphones. Most cellular phones nowadays have a built-in GPS which can provide localization. The service providers can also estimate the position of each cellular phones based on distance to network receivers/transmitters. These sources of large-scale localization data can be used for investigating different topics: physical activity monitoring (Anderson, et al., 2007), map matching (Bierlaire, Chen, & Newman, A probabilistic map matching method for smartphone GPS\ data , 2013), route choice for cyclists (Broach, Dill, & Gliebe, 2012), vehicle route choice (Papinski, Scott, & Doherty, 2009) or travel time analysis in public transport (Mazloumi, Currie, & Rose, 2009).

When localization data is not the main interest (as for mode choice analysis) different methodologies for data collection can be used. One recent method for collecting rich data on mode/route choice is a side-product of Automated Fare Collection (AFC). Many cities require public transport users to purchase a "smartcard" and then tap-in or tap-out at each boarding or alighting event. Such data can be integrated into frameworks for measuring reliability of public transports (Orth, Weidmann, & Dorbritz, 2012), origin-destination matrices (Munizaga, Devillaine, Navarrete, & Silva, 2014) or passenger transfer times (van den Heuvel & Hoogenraad, 2014). An extensive review of the applications from smart card data is done in (Pelletier, Trépanier, & Morency, 2011).

The last aspect of pedestrian counting systems connected to public transport is in-vehicle occupancy estimations. By using Automatic Passenger Counting (APC) systems, public transport operators can obtain an estimate of in-vehicle occupancy. This aspect is important when passenger comfort is studied as congestion inside a vehicle will significantly decrease the quality of service appreciation of passengers. To collect APC data, either manual or automatic methodologies can be used. The shortcoming of manual counting is reliability and coverage: the employment cost to obtain reliable and large-scale data is too high. In practice, most operators use either mechanical (pressure sensitive mats, turnstiles) or camera-based technologies. This kind of data can be used for arrival time prediction (Chen, Yaw, Chien, & Liu, 2007), origin-destination matrices (Zhao, Rahbee, & Wilson, 2007) or estimation of dwell times (Rajbhandari, Chien, & Daniel, 2003).

3.2.2. Light rail & bus vehicle location data

Today, many public transport operators use Automatic Vehicle Location (AVL) for locating in real time their vehicles. Over the years, multiple technologies have appeared and can even be combined to increase the amount of information which is collected and the accuracy of such information. The standard data which is transmitted from vehicles to the operator is the vehicle ID, time and position (latitude and longitude). Research towards applications of AVL for user-centered purposes became common in the 1990s. In (Lin & Zeng, 1999), the authors use GPS data combined with different prediction algorithms for estimating the arrival time of buses at their next stops. Today, this approach has been deeply investigated and the public transport operators of many large cities in European countries have such arrival and travel time predictions available for users. The impact on passenger waiting times and walking behavior of real time information at public transport stops is done in (Dziekan & Kottenhoff, 2007).

3.2.3. Heavy rail

Unlike inter-city public transport systems (tram or bus), regional train (heavy rail) operators do not systematically provide real time estimates of the travel or arrival times. At the national level in Sweden, expected arrival times of trains are openly available online. As this example shows, real-time positioning is slowly becoming standard thanks to the European Train Control System (ETCS). Thanks to rail side installations and in-vehicle transmitters, the position and velocity of each train is transmitted to the train operator. One issue with this new standard which must be dealt with is component failure, as discussed in (Zimmermann & Hommel, 2003). The authors discuss and model failure of some parts of the ETCS system and conclude that such events can lead to severe disruptions in the system.

The historical and classical approach for locating trains on the network is the block signaling system. One disadvantage of this system is that the exact position of trains is not known. The advantage on the other hand is that no further material needs to be installed, the security system which controls the lights indicating whether trains can continue at the highest velocity or must stop also provides the approximate position.

One field of heavy rail operations extensively covered by research is (re)scheduling. The data required for such problems comes down to the timetable along with spatial network and travel times. Except for the timetable which can be considered as the configurable element, the others are fixed and must accommodate the train demand (Robenek, Maknoon, Azadeh, Chen, & Bierlaire, 2016). When rescheduling is under consideration, real-time position and velocity of trains is required. This information can be provided by the ETCS system.

3.3. Big Data analysis

As stated previously, multiple alternatives exist for collecting pedestrian-centered data. All camera based technologies rely on image processing techniques to recognize pedestrians in images and calculate their positions. Today these steps are accomplished using machine learning algorithms which, once trained, can recognize and locate a pedestrian in an image. The importance of efficient algorithms is evident when considering the amount of data which can be collected today through CCTV or smartphones for example. These "Big Data" sources are common in many cities around the world.

Although collecting and processing raw data is critical for any successful enterprise, performing an efficient and meaningful analysis is important to learn the weaknesses and strengths of the data. When discussing mobile entities like pedestrians or vehicles, some measures common to different systems can be calculated to estimate the state of such system, while other measures are system-specific.
Maybe the most common and intuitive indicator for observing the state of a given transportation system is density. This notion exists in many different fields of science and is true also for transportation. Nevertheless, calculating density of public transport vehicles on a road is significantly different from the calculation of pedestrian density. Not only are the scales different, but motorized vehicles are mostly constrained to lanes while pedestrians can move freely and very rapidly change directions. This simple example emphasizes the need for mode-specific differentiation of measures, indicators and analysis techniques.

3.3.1. Congestion

The negative experience linked to density is congestion, applicable both for pedestrians and road vehicles. When the number of entities in a given space goes above a certain threshold, the movements become restricted and each entity cannot travel at its desired velocity. One common measure of congestion is done using the level-of-service (LOS) classification. This scale exists for pedestrians (Polus, Schofer, & Ushpiz, 1983) and road networks (Manual-HCM, 2010).

3.3.2. Travel time

Not only does congestion decrease the comfort experienced by users, but it also modifies journey characteristics. One measurable characteristic is travel time, applicable for pedestrians and mixed traffic. Variations will occur based on traffic situations, generally the higher the density, the higher the travel times.

3.3.3. Public transport offer frequency

Thanks to AVL, real-time positions of PT vehicles is known. With this information, estimation of punctuality, schedule adherence or time spent in congestion can be done. Naturally, this information is not directly monitored but needs to be calculated from data.

3.3.4. Flow dynamics

As with any dynamical system, both pedestrians and public transport vehicles experience different events and situations. This can range from unrestricted daily movements up to emergency accident-induced situations. In some cases, self-organized flows can form themselves, while in other situations important control strategies must be used to limit chaos. Understanding such flows is a vital first step before any modeling or simulation can take place. In the following sections, the dynamics of pedestrians and public transport systems is discussed.

3.3.5. Daily activities patterns, week travel diary

Understanding the sources of mobility is important when planning and forecasting of urban centers is performed. Understanding the reasons which make individuals move is critical towards accurate estimation of changes in land-use, density or transportation networks when policy makers define new strategies.

As described in (Axhausen, 2005), people seem to plan their movements based on activities which are planned at different scales. Some events are systematic and take place every day while others are introduced in the activity calendar depending on "holes" in such schedule. The authors emphasize the lack of empirical studies in this direction and push researchers to investigate the way people fill their daily activity calendars.

Daily activity patterns can be deduced from mobile phone traces; including time, position and activity (Phithakkitnukoon, Horanont, Di Lorenzo, Shibasaki, & Ratti, 2010). In this study, the authors develop a methodology for inferring the time-activity couples during the day. A workday is split into temporal blocks to which an activity is assigned. Another study analyzed mobile phone traces and concluded that people systematically visit the same locations, even if the routes are different (Gonzalez, Hidalgo, & Barabasi, 2008).

3.3.6. Pedestrian dynamics

Maybe the most common event of pedestrian dynamics which is considered in scientific research is self-organization into lanes (Helbing, Molnar, Farkas, & Bolay, 2001). The phenomenon where pedestrians follow each other in lanes to avoid conflicting movements is observed in many different infrastructures and is used in many modeling and simulation projects as validation. As with most scientific work, assumptions are made to simplify the modeling challenges. One assumption which is commonly made concerns uniformity of pedestrians: many studies impose the same walking speed distributions, size or tolerance to proximity. This can be relaxed when computational cost is not a limiting factor and sufficient information is available, as in (Seyfried, Steffen, Klingsch, & Boltes, 2005), (Campanella, Hoogendoorn, & Daamen, 2009).

3.3.7. Traffic dynamics

Maybe the most famous manifestation of dynamical processes in traffic flow is shockwaves; this phenomenon has been studied since the 50' (Richards, 1956). Although it was first studied on highways, shockwaves also take place at lower speeds and impact the comfort and fluidity of traffic flow in cities (Wu & Liu, 2011).

3.4. Modelling

Three fundamental modeling paradigms are briefly presented in the following subsections: Origin-Destination matrices, fundamental diagrams and train scheduling.

3.4.1. Origin-Destination matrices

Being able to represent and model demand is vital for any transportation system. Origin-Destination (OD) matrices is one possible representation of pedestrian or vehicular demand between a given set of locations. Estimating such matrices is challenging as such information is often not available. Multiple algorithms exist to estimate such matrices and are applicable to different domains. For example, OD estimation can be performed from vehicle link counts (Yang, Sasaki, Iida, & Asakura, 1992) or more recently by using mobile phone data (Gundlegård, Rydergren, Breyer, & Rajna, 2016).

3.4.2. Fundamental diagrams

One common tool in both vehicle and pedestrian flow analysis is fundamental diagrams. They represent the link between flow, density and speed; the most common form is the flow-density graph. Although this tool can be used in both pedestrian and traffic flow, significant differences exist between them.
The fundamental diagram of traffic flow has extensively been studied and many specifications exist. Recently, research is oriented towards macroscopic fundamental diagrams (MFD) (Daganzo & Geroliminis, 2008), which consider the flow-density relationship at an urban scale. Important assumptions must be met for and MFD to exist, namely uniform flow conditions throughout the network.

Maybe the most significant difference between pedestrian and traffic fundamental diagrams is the dimensionality of the system. Traffic flow is mainly unidimensional, while pedestrian flow is bi-dimensional (Seyfried, Steffen, Klingsch, & Boltes, 2005). Pedestrian fundamental diagrams are still under active research, for example differences between cultures (Chattaraj, Seyfried, & Chakroborty, 2009) or multi-directional flows (Wong, et al., 2010) are studied.

3.4.3. Train scheduling

Scheduling trains is challenging from both the methodological and computational aspects. As summarized in (Törnquist, 2006), many different objectives and algorithms exist when heavy rail scheduling is considered. One reason for this wide range of models is the many different constraints which must be satisfied: passenger demand, infrastructure limitations or rolling stock are only a few examples.

Therefore, to address these different aspects of train scheduling multiple different categories can be defined: tactical scheduling which deals with long-term track assignment, operational scheduling covers short-term and generally non-recurrent route and track assignment or re-scheduling which takes place when a generally unforeseen modification in the schedule takes place and a new schedule needs building. These different scheduling levels can also be classified by objective. Some (re-)scheduling techniques focus on minimizing delay (of trains or passengers), maximizing profit or minimizing cost. Some 48 different approaches are reviewed by the authors.

4. Individual level data analysis

The data are analysed independently at each level. Different methodology and indexes can be used, and different dynamics can be investigated.

4.1. Hub Level

Modern transportation hubs play an important role in modern cities, both for linking different public transport services but also as important infrastructures in the heart of cities. This is particularly true for the city of Lausanne (Switzerland) as the main train station serves as a link between the northern and southern parts of the city. In this context, the data of pedestrian trajectories is analyzed for the hub level.

4.1.1. Data

The raw data available contain:

- *t* - time stamp with frequency 0.1 second
- *z* - zone key indicating the monitored area in the hub
- *x* - x-coordinate of the individual every *t*
- *y* - y-coordinate of the individual every *t*
- *i* - anonymized person ID. The ID is random, and the same person cannot be tracked in different days or in different hub zones.

Table 1 reports a sample of the row data.

Table 1 sample of data available

t	*z*	*x*	*y*	*i*
2013-02-18T07:00:00:009	PIW	9344	5836	1
2013-02-18T07:00:00:109	PIW	9353	5962	1
2013-02-18T07:00:00:209	PIW	9306	6108	1
2013-02-18T07:00:00:309	PIW	9317	6227	1
2013-02-18T07:00:00:409	PIW	9313	6309	1
2013-02-18T07:00:00:509	PIW	9321	6385	1
2013-02-18T07:00:00:609	PIW	9325	6463	1

Figure 1 shows the geographical representation of a zone in the hub. In this case, the zone is a pedestrian underpass (PU) connecting platforms and station exits in Lausanne train station. The areas marked in green indicate the field of view of the tracking sensors.

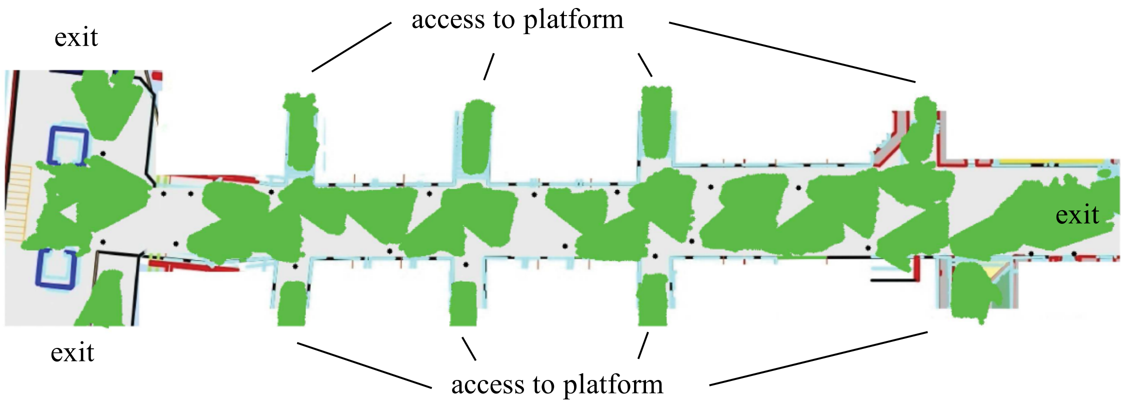

Figure 1 geographical representation of a hub zone

Two sets are available, covering Lausanne and Basel train stations. Each station has multiple days available, 10 for Lausanne and 24 for Basel. The data covers only the morning peak hour for Lausanne, whereas in Basel it covers nearly the whole day.

4.1.2. Planned analysis

Based on the above-mentioned data, we plan to analyse the followings:

- Analyse the pedestrian density and identify the critical areas inside the hub. The critical areas can be entrances/exits, the junction between multiple corridors or the areas around service points. We assess the acceptable level of density.
- Analyse the pedestrian flows inside the hub. "Origin-destination" demand is analysed by using historical data or by assuming the public transport induced flows.
- Analyse walking speed and time distributions inside train stations. The walking time distribution is critical for transfer of public transport services.

From the pedestrian trajectory data available, we can perform descriptive analysis on pedestrian flows. We analyse the current flows, velocity, travel time and pedestrian accumulation using the data, which are shown in the next Section.

4.1.3. Descriptive analysis and visualization

The following figures provide a descriptive analysis of the tracking data available for Lausanne. Some descriptive quantities are calculated and reported below, for indicators which cover reliability and variability, please see Deliverable 2.1. Emphasis is given to aggregate analyses. Representation of flows through the station, velocity distributions and pedestrian accumulation in the station are some of the aspects analyzed.

The overall demand in Lausanne's train station during the morning peak hour is summarized in Figure 2. Naturally daily variations take place and there are systematic differences between days. This was analyzed in a socio-economic point of view (Lavadinho, Alahi, & Bagnato, 2013).

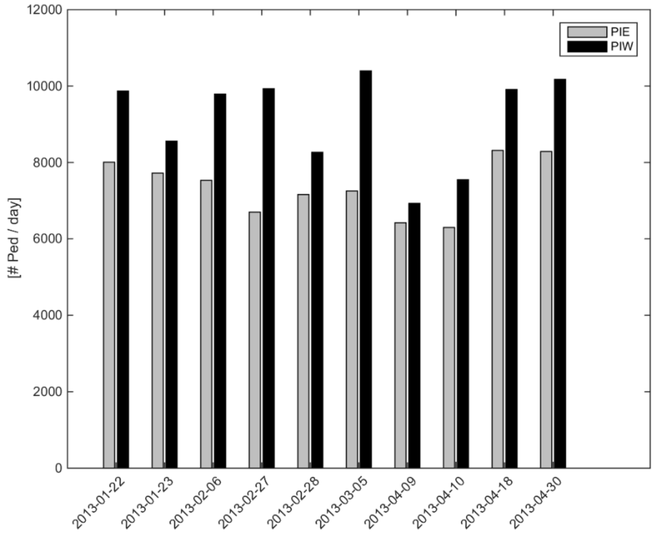

Figure 2: Daily demand in both pedestrian underpasses during the morning peak hour (7:00 to 8:30). PUW suffers from higher demand systematically.

Flows

To visualize the importance of the "transit" aspect of the station, chord diagram (Hänseler, Bierlaire, & Scarinci, 2016) is used. The flows between different zones of interest in the station are visible, naturally the train platforms attract most of the pedestrians but people do come to the station for shopping (red bands) or simply traversing the station (North to South and vice-versa).

Another possibility for visualizing the pedestrian flows in the underpasses is by plotting each individual trajectory such that higher occupied zones become darker, as in a heat map. The exercise is done in Figure 4, where both underpasses are visible. With this trajectory heat map, it is possible to locate the areas of the underpasses which are most used, and qualitatively estimate which platforms attract the most passengers. When comparing both underpasses, the eastern underpass (bottom) appears to suffer less from high densities than the western underpass. This goes in the same direction as a socio-economic qualitative survey (Lavadinho, Alahi, & Bagnato, 2013), which indicates higher utilization of the western underpass. Furthermore, the fact that the northern exit (right in images) is more heavily used than the southern one is also visible. Finally, some artificial high density zones are created from the network of cameras. In PUE (bottom image), some polygonal areas appear through high density areas.

The last analysis focusing on the flows of pedestrians through the station is the location of the entrance and exit "stamps" for each individual. This is important in assessing the reliability of the data. In Figure 5 the blue dots represent entrance points while red dots represent exits. Not only no points are found far from the physical entrance locations (this could be the result of some filtering), but very few exit stamps are located in the middle of the underpasses (indicating the tracking system lost someone). This is encouraging as the available data set appears robust (or has been filtered).

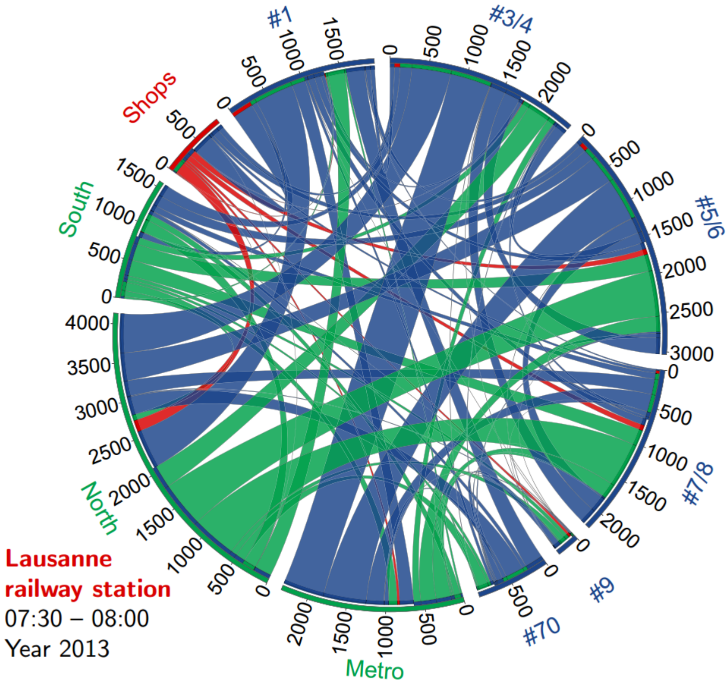

Figure 3: This chord diagram shows the distribution of pedestrian flows between different OD pairs in the station. Even though most pedestrians come to the station to use public transpot, some individuals come only for shopping or use the station to link the north and south parts of the city.

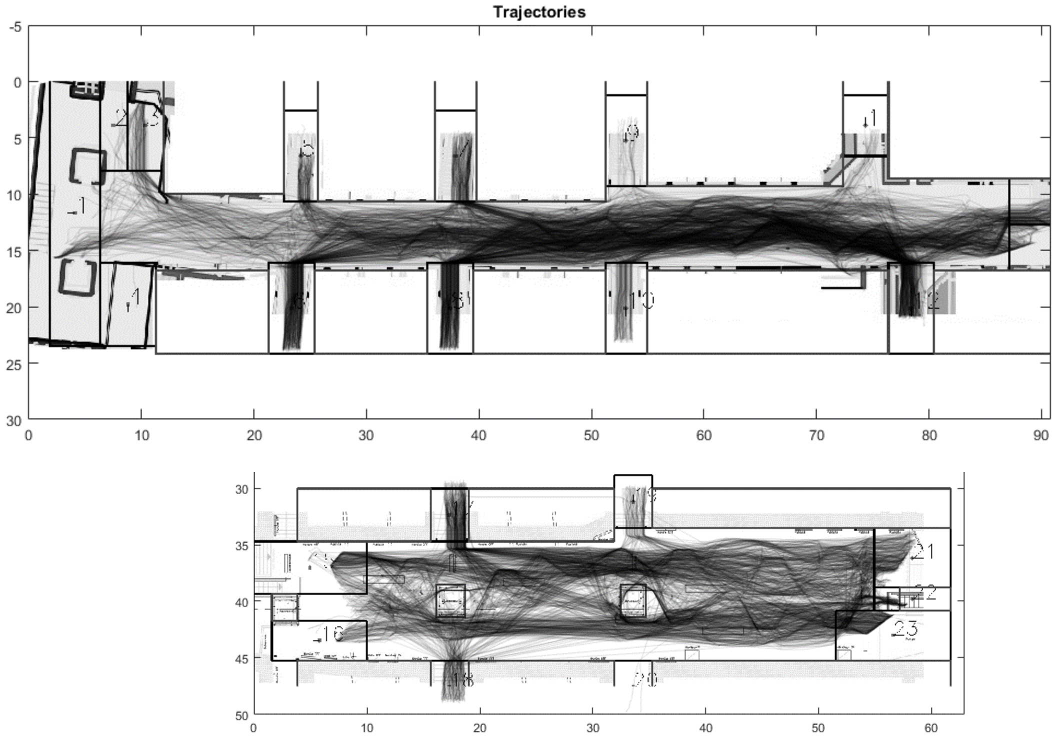

Figure 4: Subset of all trajectories for the main station in Lausanne. Only 1000 trajectories are shown and they have been smoothed using a moving-average algorithm of width 5.

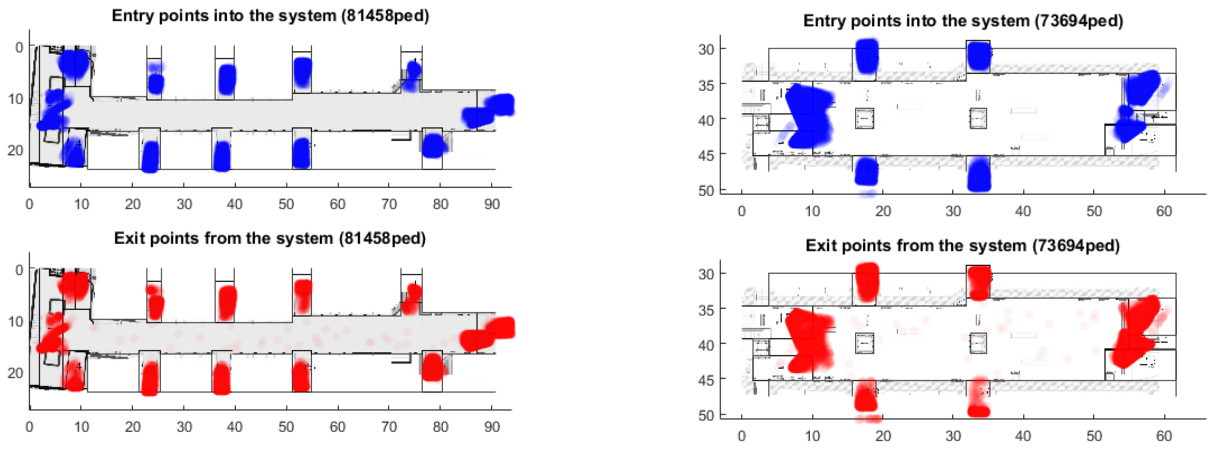

Figure 5: The entrance and exit points of each individual for all ten days are shown with colored dots. The blue points indicate entrance points and red dots represent exit locations. It is apparent that the data is reliable from this aspect: very few people « appear » far from physical entrances and very few people "disappear" in the middle of the tracking area.

<u>Velocities & travel time</u>

Using finite difference approximations, the velocity of each pedestrian can be calculated throughout his trip in the underpasses. The distribution of the mean speed of each pedestrian is shown in Figure 6, for both underpasses. Although a small difference in the distributions is visible, they both have the same mean and globally share the same shape.

On the other hand, when considering the distributions of travel times in both underpasses, one significant difference stands out (Figure 7). For the western underpass, the travel time follows a normal-like distribution, while the eastern underpass yields a truncated like distribution. Although the left tail seems to be normal-like, the right tail appears truncated at travel times of 60 seconds. The cause for this behavior are not known for sure, one possible explanation would be the variety of length of all possible trips. In PUW, there is enough diversity in the trip lengths to create a full distribution, while PUE does not contain longer trips. This combined with the distribution of travel times would truncate the travel time distribution.

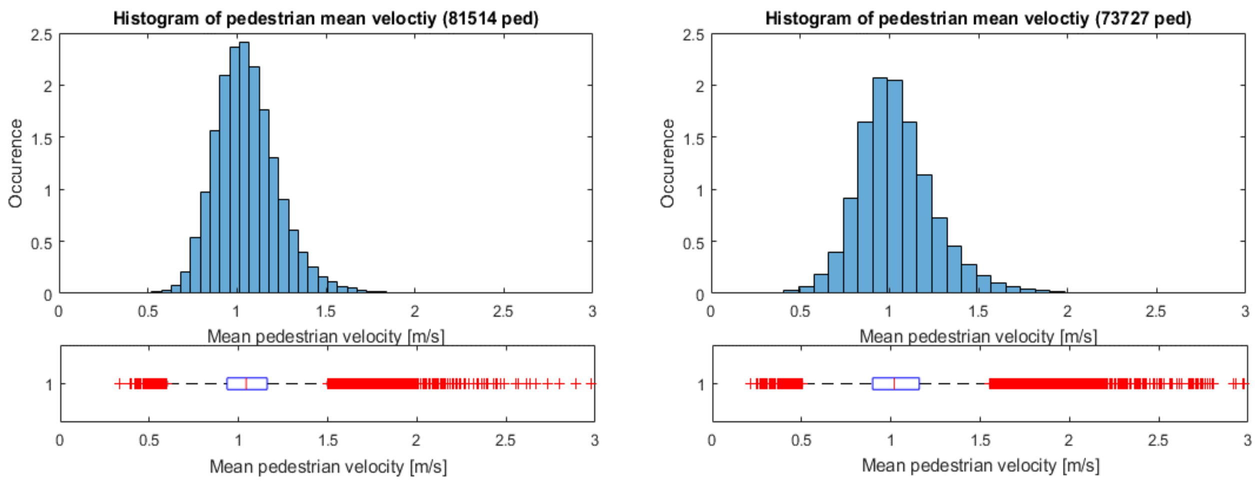

Figure 6 : The distribution of mean velocities is similar for both pedestrian underpasses (left: PUW, right: PUE). The mean is located at approximately 1.1m/s.

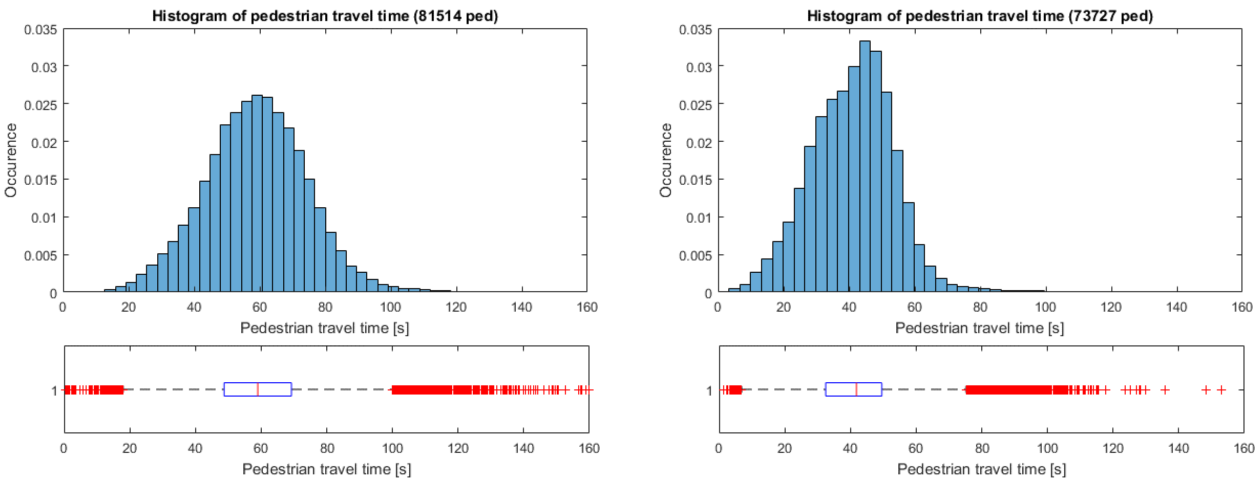

Figure 7 : When comparing the distributions of travel times in both pedestrian underpasses (left : underpass West, right: underpass East), two different distributions appear. In the West underpass (PUW), the distribution is symmetric and looks normal-like. On the other hand, the travel time distribution of PUE lacks symmetry.

Velocity & accumulation

When combining the velocity of each individual with an estimation (approximation) of density in the underpasses, an aggregate speed-density diagram can be created. To simplify the computation and obtain an aggregate estimate of density, the accumulation of pedestrians inside an underpass is used. This is simply the count of the number of pedestrians inside an underpass at a given moment. In Figure 9, the accumulation of pedestrians inside PUW is plotted for the 22nd January 2013. It is clear that very large variations induced by trains arriving in the station occur and one can easily image congestion taking place.

When considering the link between velocity and density (accumulation in fact) as in Figure 8, an interesting pattern appears. As the accumulation increases, the variability of mean pedestrian speed decreases. This converging velocity phenomenon can be explained by the pressure of the surrounding people. Individual wishing to walk slower or faster than the average speed of the crowd cannot and are forced to follow everyone else.

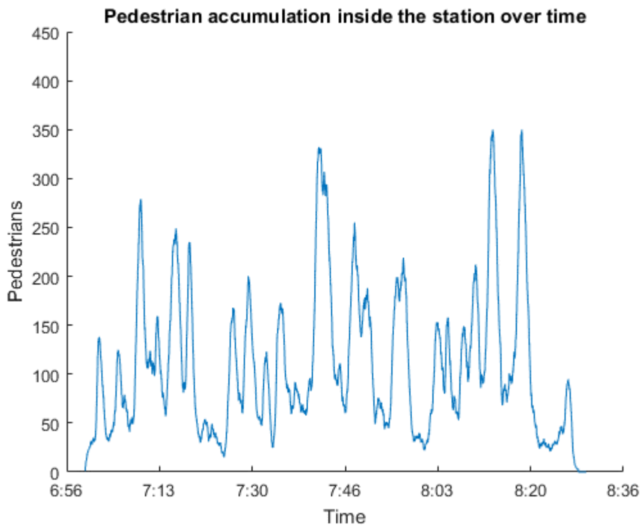

Figure 9: The accumulation of pedestrian in each underpass is highly variable. In only a couple of minutes it varies from less than 50 to over 300 people. This high demand can cause congestion in the underpasses.

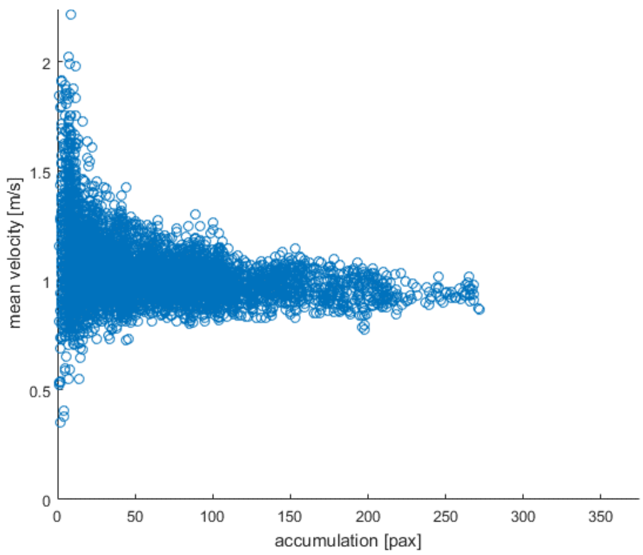

Figure 8: Each point represents the mean velocity of a pedestrian and the corresponding accumulation in the station at the time the pedestrian entered the underpass. We can observe that for high accumulations (high density), the velocity values converge towards 1m/s. The source of this convergence can come from the pressure of other users

4.2. Urban Level

In this section, the following abbreviations are used:

- AFS systems: Automatic Fare Collection systems (with spatial and/or temporal entry and/or exit information)
- AVL systems: Automatic Vehicle Location systems (real-time, GPS-based)
- APC systems: Automatic Passenger Counting systems (sensors, based on vehicle weight, or based on AFS systems)

4.2.1. Data

When travelling by tram or bus in the Netherlands, there is a closed AFC system where passengers have to tap-in and tap-out within each vehicle. This means that in theory, both the boarding location, the alighting location and the chosen service and route of each journey stage are known. It also means that data from the AFC system can be used as APC system to determine the occupancy per vehicle per line segment. When travelling by metro or train in the Netherlands, also a closed AFC system is used. The difference however is, that the tap-in and tap-out needs to be performed at the station. This means that although the boarding and alighting location of the total train or metro journey are known, transfers within this train or metro system and chosen service and route cannot be determined directly. Also, the AFC system cannot be used directly as APC system, since vehicle loads are not directly known from this data.

Since this research focuses on urban public transportation which is operated by trams and busses, we use AFC data of which in principle both boarding time and location, and alighting time and location are known for each journey stage. Table 2 below shows the format of the AFC data. Given the location of tap-in and tap-out devices within the vehicles, the AFC system is also used as APC system to infer passenger loads. There is also made use of a conversion table in which the corresponding x-y coordinates and passenger stop name are indicated for each stop-ID (see Table 3). In this table, the coordinates are expressed using the Rijksdriehoekstelsel (RD-coordinates), which is a common conversion from regular XY-coordinates in the Netherlands. Next to AFC data, there is also made use of AVL data, which is open data in the Netherlands. Table 4 shows the format of this data.

Table 2. Format of Automatic Fare Collection (AFC) data

Tap-in date+time	Tap-in stop-ID	Tap-in line	Tap-out date+time	Tap-out stop-ID	Trip -ID	Vehicle nr	Smart-card nr
4-3-2014 11:42:37	35309	6	4-3-2014 12:03:19	34997	3423	3050	81675688
4-3-2014 12:15:57	30091	18	4-3-2014 12:23:04	32857	6545	187	81675688

Table 3. Format of stop-location table

Stop-ID	RD x-coordinate	RD y-coordinate	Passenger stop name
35309	81962	450867	Dr. H. Colijnlaan
30091	82188	455213	Centraal Station

Table 4. Format of Automatic Vehicle Location (AVL) data

Stop-ID	Trip-ID	Order-nr	Nominal arr	Realized arr	Nominal dep	Realized dep
1119	4464	28	19:22:35	2016-01-06 19:23:25.000	19:22:35	2016-01-06 19:23:49.000
1119	4465	28	18:23:48	2016-01-06 18:26:26.000	18:23:48	2016-01-06 18:26:44.000

4.2.2. Planned analysis: journey inference

In the first part of this analysis, the purpose is to infer journeys from the AFC, AVL and APC data available. Given already performed studies, the focus is on the development of a new methodology for journey inference which is robust for different service conditions. This means that:

- This methodology should be able to infer journeys in undisrupted circumstances in a valid way, thereby also incorporating crowding and denied boarding.
- This methodology should be able to infer journeys in (both planned and unplanned) disrupted circumstances in a valid way. Especially the incorporation of crowding and denied boarding, and inference of journeys including the use of an intermediate public transport journey stage by another operator, on another network level or without AFC system, are of relevance here.

A better journey inference allows us to get better insight in travel flows and behaviour in both undisrupted and disrupted situations. It also allows us to better investigate and quantify network reliability and robustness, since the impact of disruptions can be determined on a journey level.

Although journey inference based on smartcard data for regular circumstances is quite advanced, there are still opportunities for improving the inference algorithm here (Gordon et al. 2013; Wilson 2014). This is especially the case, since developed methodologies are not always able to validly infer journeys in case of irregular (planned or unplanned) transit operations. Especially in case of disruptions, passengers might have to take detours which are illogical to use in case of no disruptions, or might have to use transit services of another operator in the multi-level public transport network available for passengers, or might have to additionally transfer to rail-replacement busses. This places additional challenges to journey inference algorithms, since these algorithms must be robust to infer journeys in a valid way in both disrupted and undisrupted scenarios.

Incorporating crowding and detours/transfers due to disruptions in journey inference

To infer journeys, it is important to determine for each alighting whether this is the final destination of a journey, or a transfer. A transfer is defined as a transition between two consecutive journey stages without a trip-generating activity. It should be noted here that it is still possible to perform an activity at a transfer location, as long as this activity is not the primary purpose of the presence of the traveller at this intermediate stop (Gordon et al. 2013). In other words: there should take place no purposeful activities at a transfer location (Seaborn et al. 2009). Convenience activities can take place at a transfer location as long as these activities could take place at 'any' transfer location (Nassir & Hickman, 2014). Several constraints are formulated to improve transfer inference, and to make the inference algorithm robust also for disrupted scenarios. There is distinguished between temporal, spatial and numerical constraints. Afterwards, also some special cases related to missing or different tap-in or tap-out behaviour due to system failures, human errors and fare evasion.

<u>Temporal constraints:</u>

1. If a trip is the last trip made by that card that day: alight location is destination (no transfer).

2. Did an alighting passenger take the first *reasonable* passing vehicle of the next journey stage? If yes: it is a transfer / if not: it is an activity. Many studies apply a fixed maximum transfer time as threshold to determine whether to consider the alighting stop as a transfer location or as journey destination (e.g. Jang (2010), Munigaza & Palma (2012), Seaborn et al. (2009)). However more advanced methods could be applied to prevent false positive and false negative identified transfers. Besides, some studies check whether alighting passengers boarded the first passing vehicle at a stop, to determine whether we see a transfer, or that an activity is performed at the alighting location (Gordon et al. 2013). However, the impacts of crowding and denied boarding on whether passengers are reasonably able to board this first vehicle are not considered (Wilson 2014), thereby risking the introduction of false negative transfers. This is especially relevant in disrupted scenarios, in which crowding and denied boarding is more likely to occur due to cancelled and/or irregular services. Therefore, the following aspects are introduced.

Determine walking time from alighting location to boarding location based on stop coordinates and Euclidian distance

- Determine the Euclidian distance by using 3D coordinates to incorporate transfers to different levels (e.g. tunnel, station). In most studies, only the two-dimensional Euclidian distance is used, thereby neglecting longer transfer walking times due to height differences. This could potentially lead to false negative transfers. Since for stops only x-y coordinates are known, the z-coordinate has to be estimated in a simplified way. Consider at-grade stops with height 0, and take a fixed height per level, for stops at level +1, +2, -1 or -2.
- Validate the use of Euclidian distance: compare for some transfer locations of the case study network the distance using the existing pedestrian network, and the calculated distance using 3-dimensional Euclidian distance. This is relevant, because almost all transfer inference studies base the minimum transfer walking time on the Euclidian distance. If this measure does not fit with the walking time in practice, it can potentially lead to both false positive or false negative transfers.
- Use the expected value of the walking speed distribution function of a walking speed model, based on Koutsopoulos (2014). In this distribution function, individual characteristics and varying crowding levels at platforms or transfer locations can be incorporated. If there is a systematic bias between the Euclidian transfer distance and the transfer distance over the pedestrian network, then scale the expected walking speed value accordingly to correct for this bias. Combining the Euclidian distance between stops with the expected walking speed value, it is possible to determine the first possible arrival time of an alighting passenger at the platform / station at a transfer location for boarding the next vehicle.

Count which Nth passing service a passenger boarded:

- Add a fixed buffer time to the calculated first possible arrival time at the boarding stop, which represents the possibility of performing a convenience activity at the transfer location (e.g. buying a newspaper, or getting coffee), and calculate the new arrival time (first possible arrival time + buffer time) at the boarding stop.
    - Experiment with different parameter values for this fixed convenience activity time.
- Determine all possible services a passenger can take to reach the alighting stop of this next journey stage (e.g. use hyperpaths for alighting stops served by common lines).
- Determine based on AVL data how many services to this alighting stop passed, before the service which is taken by the passenger. If this number is larger than 0, this reflects a purposeful activity instead of a transfer.

Check if there was no substantial crowding in the arriving vehicle:

- Determine the occupancy (infer occupancy based on the number of tap-in and tap-out transactions in a certain vehicle at each stop from the AFC system) of the arriving vehicle and check if the occupancy was larger than a certain threshold (e.g. the ‘norm capacity’, which typically equals the seat capacity + 50% of the standee capacity in the Netherlands).
- Determine if another service passing the same alighting stop of the next journey stage arrives at the transfer stop within a short time (e.g. within max. 2 minutes behind the quite full vehicle). If yes: do not count this vehicle as service a passenger chooses to skip. This allows the incorporation of passenger behaviour in relation to crowding: if a quite full vehicle arrives while the next vehicle is already in sight (e.g. due to bunching, especially relevant in case of disruptions), some passenger might decide to choose comfort above a short additional waiting time, and skip the first arriving vehicle on purpose. In existing methods, this would be classified as an activity instead of a transfer.
    - Experiment with parameter values of this number of minutes as threshold.

Check if there was no denied boarding, because of which the passenger could not board:

- Determine the occupancy of each arriving vehicle serving the alighting stop of the next journey stage and check if the crush capacity of this arriving vehicle has been reached. If yes: do not count this vehicle as service a passenger chooses to skip. This means that passengers were not able to board this vehicle, which should not be inferred as activity.

<u>Spatial constraints:</u>
1. Walking time between alighting stop and boarding stop should not exceed a threshold value (usually the walking radius of e.g. 500, 800 or 1000m), as investigated in several studies. If the walking time (calculated based on the Euclidian distance and (possibly corrected) walking speed distribution function exceeds this threshold, there might be an intermediate journey stage which is not registered in the AFC system. See below (in case of disruptions) for more explanations about this topic.

2. Is the next journey stage made by the same service (regardless the direction: either in same or opposite direction) & is there more than 10 minutes (or another threshold value) between the alighting time and next boarding time? If yes, it indicates that there is an activity performed instead of a transfer.

- A transfer to the same service is considered as indication for performing an activity in several studies (e.g. Gordon et al. 2013; Hofmann & O'Mahony 2005). A transfer to the same service in the other direction indicates a back-and-forth trip, whereas it makes no sense usually to transfer to the same service in the same direction, unless you perform an activity at that location.
- However, these studies do not consider the impacts of crowding and the impacts of possible rescheduling measures in case of disruptions. It is for example possible that passengers transfer from a very crowded vehicle to the next vehicle of that service, if information is provided to them that this (often emptier) vehicle is driving very close-by. Also, passengers can travel one stop in the opposite direction, where they transfer to the same service in the other direction to increase seat or boarding probabilities. It is also possible that passengers have to alight a vehicle if a rescheduling measure like short turning or deadheading (drive empty to the final destination to recover from the delay) is applied in case of disruptions. Usually this will only be applied if a next vehicle is driving close-by. These cases show, that especially in relation to crowding and disruptions, two transactions in the same service can occur, in which usually the second transaction follows the first transaction quite fast. Therefore, a threshold is applied: only if this threshold (e.g. 10 minutes) is exceeded, alighting and directly boarding the same service is considered an activity instead of a (possibly mandatory) transfer.
  - Experiment with parameter values for this threshold value.

3. If the alighting location of journey stage X+1 is less than a certain threshold value (e.g. 500m or 1 km) from the boarding location of journey stage X (to correct for detours in relation to the former trip stage), or if the alighting location of the last journey stage is less than a certain threshold value from the boarding location of the first journey stage (to correct for detours in relation to the full journey level). In Gordon et al. (2013) a correction for detours is applied by considering circuity or the detour made on the full journey length. The disadvantage of these measures is that the travelled distance is compared to the Euclidian distance between these locations. If this ratio is high, it is inferred that the made detour is due to a purposeful activity, since this can explain why passengers made such detour. However, it is not guaranteed that this detour in purposeful, since also geography or network topology might contribute to detours in transit lines compared to the Euclidian distance (e.g. in Brisbane: Nassir & Hickman 2014). Also in case of disruptions, transit passengers might have no choice than travelling by rail-replacement service or by taking remaining transit lines, which can lead to substantial detours compared to the Euclidian distance. This should however not be perceived as activity. Therefore, it is preferred to use another measure than the ratio between travelled and Euclidian distance.

4. Is there an intermediate journey stage made with another public transport mode / public transport operator / without transaction?
This is especially relevant in case of disturbances, in which passengers have to use remaining available PT services of other PT operators, on other PT network levels, or use rail-replacement services without AFC system. It can however also occur in regular circumstances, that passengers use an intermediate service of another operator if there is overlap between service areas. If this is an intermediate journey stage, the question is whether to consider the alighting stop of the last registered trip before the non-registered intermediate journey stage as destination or as transfer. Such intermediate journey stage can be made by train, tram or bus. Using trains as alternative is only possible in case of the last registered

alighting location before the ‘gap’ is near a train station, and the first registered boarding location after the ‘gap’ is also near another train station.

- If the calculated transfer time between the alighting stop and next boarding stop exceeds a threshold walking time (based on Euclidian distance and walking speed distribution function), there can be no walk transfer. This means that there are either two separate journeys (with another unknown mode in-between), or there is one total journey with an intermediate non-registered journey stage.
- Add an additional attribute to all PT stops in the vicinity (within a certain radius measured in Euclidean distance around a train station), which indicates that a train station is nearby.
- Estimate the average speed between the last registered alighting stop and first registered boarding stop. If both stops have the attribute ‘train station’, check whether the estimated average speed (transfer waiting and walking time are in fact incorporated in this time and speed) is higher than a certain threshold which corresponds to an average train speed. If yes, this probably means there is a train journey stage in between. If not, this means either a non-PT mode is used, or there is performed an activity first which reduced the estimated average speed, before starting a new journey.
    - Experiment with different threshold values for the minimum train speed.
- If not both stops have the ‘train station’ attribute, then only busses or trams of another operator, or without a (working) AFC system can be used as PT alternative. The use of private modes in the middle of a public transport chain is considered very unlikely and therefore ignored. Also for this case, calculate the average speed between the last registered alighting location before the ‘gap’ and the first registered boarding location after the ‘gap’ and determine whether the found average transfer speed is larger than a certain threshold value which represents the average speed of bus/tram (thereby incorporating transfer walk and waiting time). If the estimated average speed is lower than this threshold value, consider both parts as separate journeys separated by an activity.
    - Experiment with different threshold values for the minimum bus/tram speed.

<u>Numerical constraints:</u>
1. Apply a maximum of 4 transfers per journey. In case of the 5th journey stage, the alighting location is considered as destination. Hickman (2014) uses a maximum of 3 transfers. When applying a shortest path (e.g. Dijkstra) or logit (e.g. ZENITH) algorithm to the case study network, the number of transfers between all OD-pairs in the undisturbed situation can be determined. Often, all OD’s are connected with each other with a maximum of 2 transfers. However, here it is also important to consider the impact of disruptions. In case the intermediate part of a tram line is substituted by bus services due to planned engineering work, through passengers might be forced to make two additional transfers (tram-bus and bus-tram). In case of 2 required transfers in the undisrupted situation, this justifies the choice of a slightly higher maximum number of transfers per journey of (2+2) 4. Also, the impact of rescheduling measures in case of unplanned disruptions (e.g. deadheading), or the impact of crowding can stimulate or force passengers to make an additional transfer, which should not lead to false negative identified transfers in case of a too strict criterion here.

<u>Constraints related to different tap-in or tap-out behaviour in regular circumstances</u>
1. Is the alighting location exactly equal to the boarding location? If yes, then ignore the trip.
This might refer to a system error or a passenger who wrongly boarded a vehicle and directly alighted it at the same stop. In these cases, these trips can be ignored.

2. Are there more than 1 tap-in and/or tap-out in the same vehicle in the same trip? If yes, then only consider the first boarding location and the last alighting location (Nunes et al. 2016).
Usually, it should not happen that passengers tap-in and tap-out more than once in the same vehicle in the same trip, since this means they did not really alight the vehicle after tapping-out. However, it is possible that this occurs coincidentally in a crowded vehicle, or that passengers are uncertain whether they did tap-in, and check this by tapping-out, and tapping-in, again. It also occurs as form of fare evasion, in which passengers tap-out in the middle of the trip (e.g. if there is a long track without stops, in which no control team can board), and tap-in when approaching the next stop. In all these cases, the

intermediate tap-out and tap-in transactions should be ignored and not be counted as transfer. Only the first boarding and last alighting location are considered.

3. Is the tap-out time less than 30 seconds (or another close-by threshold value) after the tap-in time? If the alighting location is not the same (which trips are ignored in step 1), the alighting location should be considered unreliable, since it is highly unlikely that a passenger really alights after 30 seconds. It might be a case of fare evasion (e.g. entering a bus, tapping-in and directly tapping-out in the back of the bus). The alighting location can be inferred using methodologies applied for open-AFC systems, in which passengers only have to tap-in (e.g. the busses in London) (e.g. Trepanier et al. 2007; Munigaza & Palma 2012; Gordon et al. 2013).

- If this trip is the only trip registered with that card number that day & no trip is registered the next day: either ignore the trip, or keep the registered alighting location as destination.
- If this trip is the only trip registered with that card number that day & a trip is registered the next day: use origin of first trip next day as destination of this trip (assuming you stay overnight at that location).
- If this trip is the last trip registered with that card number that day & other trips are registered that day with the same: use initial origin of first trip that day as destination (assuming that you end your trip at/near the same location you started your trip that day).
- If this trip is not the last trip of that day: infer the alighting location by minimizing the Euclidean distance of all candidate stops downstream the boarding stop within a certain radius (threshold) to the boarding location of the next trip. If no candidate stops are found, check whether there might be an intermediate journey which is not registered (see below for undisrupted situations).

4. Is there a missing tap-out? If yes, then there is no alighting location and time known for that specific transaction. Apply the same four scenarios as described just above in case the tap-out time is within 30 seconds after the tap-in time.

When based on these constraints journeys are inferred, it is possible to scale these journeys given the fact that there are also transit users who do not use a smartcard, based on exogenous counts and/or surveys. Under the assumption that the distribution of destinations for a certain origin is comparable for non-smartcard users and smartcard users, journeys can be scaled accordingly, for example by using an Iterative Proportional Fitting (IPF) method (Munigaza & Palma 2012). When aggregating these journeys, an OD-matrix can be determined.

4.2.3. Planned analysis: identification and clustering

When an OD-matrix is inferred from the different data sources in a valid way, it is possible to determine the share of origins, destinations and transfers for each stop in the network. This allows us to identify the most important functions of each station, for example whether this functions as source, sink or hub. We can identify the most important hubs based on this. Therefore, it is important to define and test criteria which define a hub explicitly, for example:

- The share of transfer passengers in relation to the total number of boarding and alighting passengers should be larger than a certain threshold (experiment with threshold value).
- There has to be a substantial number of transferring passengers in absolute way (magnitude). This can for example be measured by the share of transferring passengers at that specific hub, compared to the total number of transferring passengers on the total considered network (experiment with the threshold value for this share).

Besides, an analysis will be performed in which clusters of stops are identified and classified. The identification will mainly be related to the magnitude, function and geographical location of different stops (Cats et al. 2014). The classification will be performed using hierarchical clustering as data mining technique (Trepanier 2014). Identified clusters can be classified based on their temporal patterns. The classification can also be performed based on the extent of reliability and robustness of identified clusters in relation to WP2.1 – in which measures for reliability are defined – by applying and comparing this service reliability measure on a journey level between the undisrupted and disrupted scenario.

4.2.4. Descriptive analysis and visualization

We considered travel demand data from AFC-systems of the period 7-20 November 2015, in which 10 working days, 2 Saturdays and 2 Sundays are included. For the descriptive analysis, we applied the basic transfer inference criterion in which trip stages are considered as one journey if they are made with the same smart card number and if the transfer time between alighting and the next boarding is smaller or equal to 35 minutes. We focus here on journeys on the urban network level. Interactions with other network levels (e.g. the regional train network or hub network) are not considered here.

Interchange dynamics

In the first analysis, we considered the number of transfers made in journeys on an average working day and during specific time periods of the working day (Figure 10). In general, it can be concluded that 75-80% of all journeys are direct, without transfers. About 20% of all urban network journeys contains one journey, whereas the remaining 0-5% of the journeys contains 2 or 3 transfers. When comparing the distribution for different time periods, it can be concluded that the distributions for peak- and off-peak periods are quite similar to each other. In the evening however, a higher percentage of direct journeys can be observed, at cost of a lower percentage of journeys with one or more transfers. A possible explanation for this can be the lower service frequency during evenings, making a route with a transfer less attractive due to the (risk of) relatively long transfer waiting times.

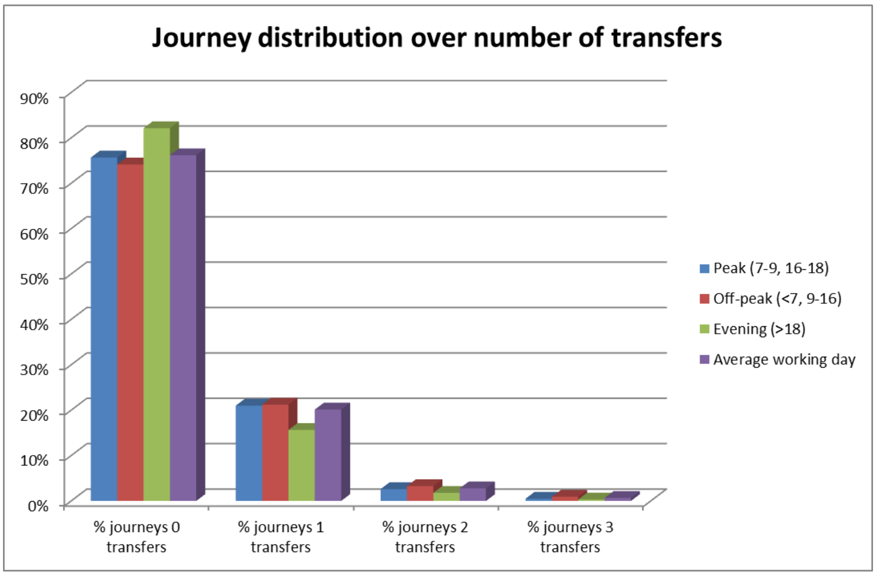

Figure 10: Journey distribution over number of transfers per time period

The observed average transfer time (transfer walking time plus transfer waiting time) on the case study network of The Hague equals 10 minutes on an average working day. During peak hours the average transfer time is shorter (8 minutes) than during other time periods, probably due to the higher peak service frequencies. Interestingly, the observed average transfer time during the evening period is no longer than during off-peak periods. Although the percentage journeys with a transfer is lower during the evening, this shows that the transfers which are actually performed take no longer than during other time periods, on average.

Table 5: Average transfer time per time period of the working day

Time period	Average transfer time (min)
Working day	10
Peak period	8
Off-peak period	10
Evening period	10

Figure 11 shows the transfer time distribution for each time period. In line with Table 5, we see that during the peak periods a higher share of transfers takes place with a relatively short transfer time. The transfer time distribution during off-peak periods has a lower share of short transfer times. During the evening period, it can be observed that a relatively high share of transfers has an intermediate to long transfer time (8-15 minutes), although on average the transfer time remains similar to the off-peak period.

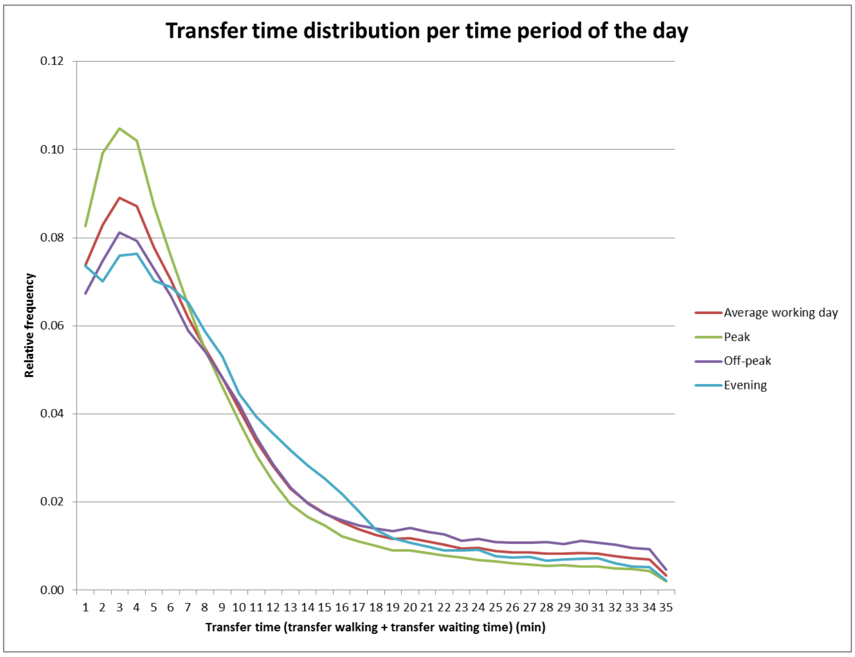

Figure 11. Transfer time distribution per time period of the working day

Temporal and spatial dynamics

Figure 12 shows the demand development over the week. When taking demand for an average *working* day as reference point, we can see that Tuesdays and Thursdays have 2% respectively 4% higher demand than an average working day. On the other hand is demand lower than average during Mondays (-6%) and Wednesdays (-2%). Mondays and Fridays are the most frequently chosen days off by part-time workers in The Netherlands, followed by Wednesdays. This results in a lower demand on Mondays and Wednesdays. The lower demand on Fridays during the day is compensated by a higher evening demand (e.g. people going out a night). Demand on Saturdays is about 60% of an average working day; while during Sunday about 40% of the working day demand remains.

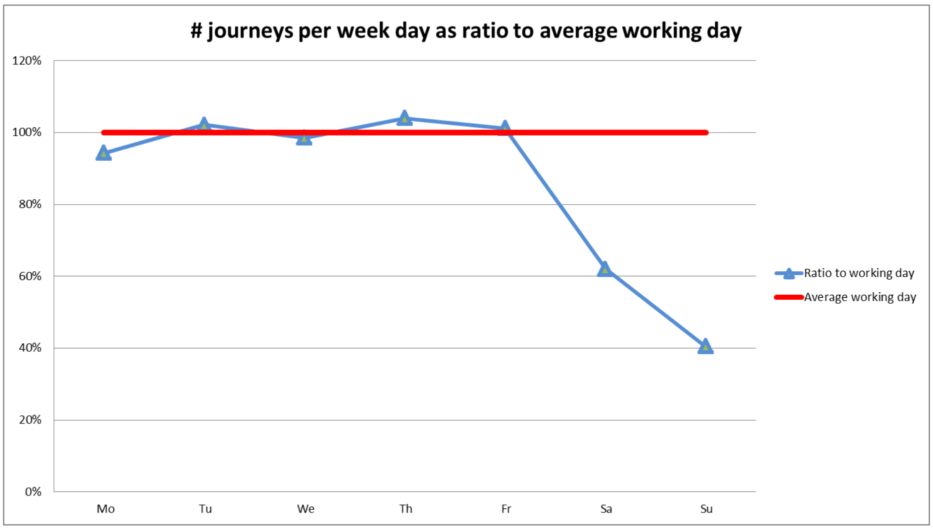

Figure 12. Demand development over days of the week

Figure 13 shows the typical demand development over the hours of a working day. The morning and evening peak periods can clearly be distinguished. Demand in the morning peak is typically more concentrated, compared to demand during the evening peak in the Netherlands. The busiest hour of the day is between 8 and 9 in the morning peak, where more than 10% of all daily journeys occurs. 36% of all daily journeys takes place in the morning and evening peak together; 42% takes place in the period between the peaks; 18% during the evening; and 3% takes place in the early morning before the morning peak.

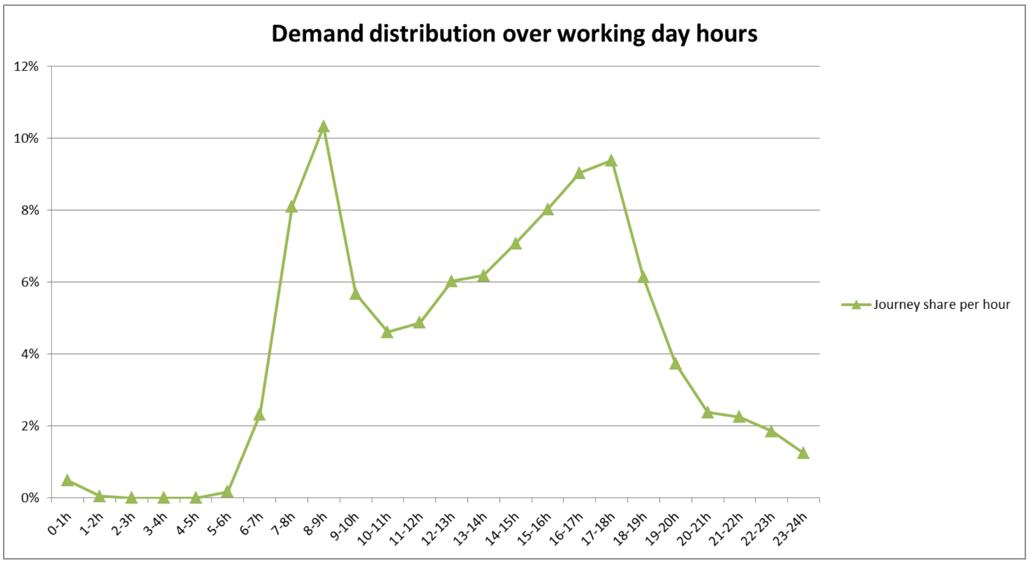

Figure 13. Demand distribution over the hours of a working day

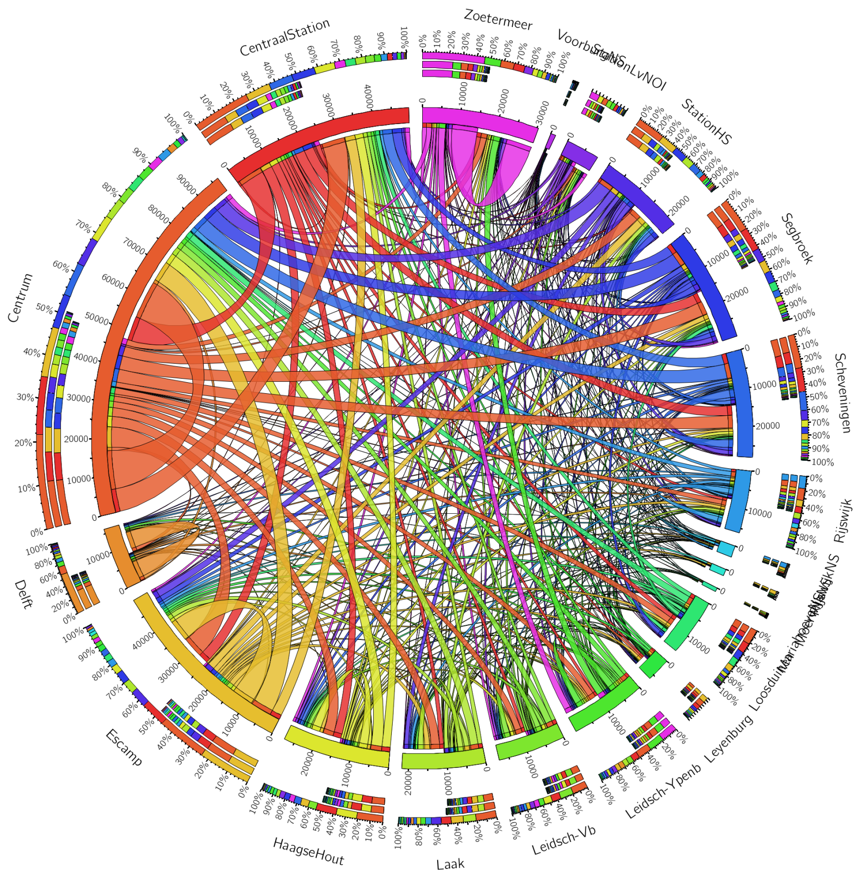

Figure 14. Chord diagram showing spatial travel patterns over the urban case study network in The Hague

Figure 14 shows a chord diagram which visualizes the number of journeys between different clusters of the case study network. All stops in the public transport network of The Hague are categorized within 20 clusters. All stations and most important transfer points are considered as separate clusters (8). The remaining stops within the municipality of The Hague are divided within 8 neighborhoods, according to the classification applied by the municipality of The Hague. At last, 4 small municipalities around The Hague are separately considered. Figure 14 shows the number of journeys between these 20 clusters for an average working day. The total number of journeys per average working day (about 240.000) is divided over 20*20 relations. It can be seen that most journeys are made from/to/between the center ('Centrum') and the central station of The Hague ('Centraal Station'). Besides, a substantial number of intra-cluster movements can be observed specifically for the city center of The Hague and for the satellite city of Zoetermeer.

<u>Stop classification</u>

Based on the AFC data combined with the applied transfer inference rules, hubs can be inferred. In line with definitions applied for airports, hubs are defined as locations where a high relative and high absolute number of transfers takes place. The relative importance can be calculated by dividing the number of transferring passengers at a stop by the total number of boarding, alighting and transferring passengers at that same stop. The magnitude or absolute importance can be calculated by dividing the number of transferring passengers at a specific stop by the total number of transferring passengers on the considered network as a whole. The AFC data allows the formulation of a ranking of stops based on these two criteria, to infer the most important hubs in the network.

Table 6 shows the top 10-ranked hubs from the case study network according to the percentage transferring passengers, the magnitude of transferring passengers, and the product of these two criteria. It can be seen that the highest transfer percentage equals about 35-36% for the Wouwermanstraat stop: an important transfer location between tram line 9 on the one hand, and tram lines 11 and 12 on the other hand. It should be noted that only transfers within the urban network level are incorporated in this data, since the AFC data is not integrated with AFC data of other operators (e.g. the train operator). This means that important multi-level transfer locations (e.g. stations) where transfers between the different network levels take place, do not show up in this data set as transfers, but as boarding/alighting instead. This explains why no train station for example occurs relatively high in the first column of relative hub importance. When considering the top 10 of most important hubs based on the absolute importance, it can be noted that more than 20% of all urban network level transfers occurs at Central Station, followed by 12% at the city center stop bundle 'Centrum', 'Spui' and 'Kalvermarkt-Stadhuis'. This criterion results mostly in a different top 10. Only the 'Wouwermanstraat', 'Leyenburg' and 'Leyweg' transfer stops also appear within the top 10 here, given the large absolute transfer volume as well. The combined criterion shows the product of the percentages shown in the columns before. The resulting top 10 of most important hubs is more similar to the top 10 based on the second hub criterion. Central Station appears as most important hub using this combined criterion.

Table 6: Top-10 hubs based on relative, absolute and combined transfer volume per station

Top 10 most important hubs based on relative importance		**Top 10 most important hubs based on absolute importance**		**Top 10 most important hubs based on combined criteria**	
Stop	*%trans-ferring pax*	*Stop*	*% transfer magnitude*	*Stop*	*Combined criterion*
Wouwermanstraat	35.7%	Centraal Station	21.16%	Centraal Station	0.0442
Muurbloemweg/Balsemienlaan	35.6%	Centrum/Spui/ Kalvermarkt-Stadhuis	12.09%	Centrum/Spui/ Kalvermarkt-Stadhuis	0.0259
Herenstraat	34.0%	Station Hollands Spoor	8.37%	Station Hollands Spoor	0.0164
Leyweg/Hengelolaan	32.0%	Leyenburg	3.26%	Leyenburg	0.0084
Nieuwe Plantage	28.4%	Hobbemaplein	2.20%	Wouwermanstraat	0.0073
Goudenregenstraat	26.7%	Brouwersgracht	2.16%	Leyweg/Hengelolaan	0.0056
Leyenburg	25.8%	Wouwermanstraat	2.03%	Brouwersgracht	0.0048
Loevesteinlaan/Melis Stokelaan	25.4%	Grote Markt	1.90%	Herenstraat	0.0042
Oostinje/Juliana van Stolberglaan	24.5%	Leyweg/Hengelolaan	1.76%	Hobbemaplein	0.0041
Monstersestraat/Loosduinseweg	24.5%	Leidschenveen	1.75%	Voorweg	0.0036

Table 7 shows the top 10 of busiest stops in the case study network, based on the total number of boarding, alighting and transferring passengers per average working day. Also here, the busiest stops are the two largest train stations of the city (Central Station and station Hollands Spoor), and the central stop clustering within the city center Centrum/Spui/Kalvermarkt-Stadhuis.

Table 7: Top-10 busiest stops in the The Hague case study network

Stop
Centraal Station
Centrum/Spui/Kalvermarkt-Stadhuis
Station Hollands Spoor
Grote Markt
Leidschenveen
Laan van NOI
Leyenburg
Hobbemaplein
Centrum West
Brouwersgracht

4.3. Regional Level

The data analysis on the regional level focuses on the traveler flows, using the available public transport services (bus and train) in the South-Eastern part of Sweden (mainly Blekinge Län). The analysis is based on the combination of multiple sources of data. This chapter presents available and potential sources of data and specifies how these can be used in different ways to enable the planned analysis.

In Sweden, we can distinguish between three different types of public transport services:

- Local/regional transport services which are organized and provided by the regional public transport authority, RKM, (but operated by various train operators, bus operators and also possibly taxi operators and boat operators).
- Inter-regional transport services which are services jointly organized by two or more regional public transport authorities.
- Commercial long-distance public transport services provided by and operated by private rail companies and bus companies.

In this context, it is relevant to make a distinction between data concerning a) the transport services and b) the passenger flows. Data regarding the different planned transport services and data about the realization of the planned service, can be found via the organization responsible for each service. This is also the organisation that has the rights to use the data. Data on passenger flows is, however, more challenging to find since many passengers move between different service networks and buy their tickets via different organisations. Hence, the data is rather scattered.

A detailed presentation of the case study connected to the regional level can be found in project deliverable *D5.1 Case study set-up descriptions*.

4.3.1. Data

Data sources for transport service data

Information about the transport services, that is vehicle movements, can be categorized into two types, static and dynamic data. The static data consists of planned services, and the data is normally structured as timetables. The dynamic data is the resulting realization of the timetables, i.e. data revealing actual and forecasted arrival and departure times of the planned services including information about delays and cancellations.

Static data, e.g. timetables, is available from Samtrafiken and includes both planned train services and other public transport services connecting the nodes/stations. This data is available in the GTFS-format[1]. The timetable includes information on the main service connections and minimum transfer time for guaranteed connections is set to 10 minutes. Samtrafiken is a Swedish, independent organization with the main purpose to coordinate certain transport activities and provide services for the stakeholders. It is owned by 38 different transport companies. RESPLUS-data is not available to us.

Dynamic data concerning trains operated within the Swedish railway network are published as public data by Trafikverket via an open API[2]. This data gives the actual state of the running trains and their ETA (Estimated Time of Arrival) when the trains are delayed. The system does not provide any historic data. This dynamic data does not include any information on actual connections or transfer times. Scheduled connections can be identified via the timetables instead.

[1] https://www.trafiklab.se/api/gtfs-sverige-2
[2] http://api.trafikinfo.trafikverket.se/API/Model

Dynamic data for local and regional bus services belongs to the responsible regional public transport authority (i.e. Blekingetrafiken) and the dedicated bus operator (i.e. Bergkvara buss). Via the SQL-based planning system REBUS all buses are monitored in real-time including cycles and rolling-stock schedules. This data is currently not owned by Blekingetrafiken and consequently we do not know whether it will be accessible. Historic data about certain bus services may be available via the smart card data and the timetables, to conclude about services' on-time performance and other performance metrics.

Data sources for passenger flow data

In order to perform analysis on the passenger flows, several data sources need to be combined. Typically, information from ticket data provide information about the number of travelers on trip legs, but does not contain information about the intended destination and planned connections. In addition, each region (in Swedish: *Län*) and associated public transport service provider (in Swedish: Regional kollektivtrafikmyndighet, RKM) have their own ticketing system and commuter smart card system, but between neighboring regions (e.g. Blekinge Län and Skåne Län) the different smart cards work.

Many travelers have a smart card but it is registered differently depending on which public transport service the traveler is using. If the traveler buys a ticket with the card in a ticket machine for a certain stretch, the origin and destination is known while it is not known which service he/she takes and at what exact time. If the traveler enters a bus, or a specific train, ("tap in") the following information is registered (at least):

- Card-id
- Time of tapping in
- Zone when tap in was made
- Station/Stop where tap in was made
- Line/service where tap in was made (e.g. Linje 1)
- Direction of the service where tap in was made
- Price (if applicable), which indicates type of traveler and how far he/she intends to travel.

The destination is, however, consequently not known, but it is possible to deduce likely destinations based on the traveler's past and following trips.

The smart card data connected to "Blekingekortet" is stored in a system called BUS-POS, which is based on an Oracle database. Since the data is connected to a specific card-id, and some travelers have connected their name and social security number to their card-id via the web/app interface, there are some legal issues when using the card data for those travelers. Blekingetrafiken has therefore consulted their legal counselor and agreed to provide anonymized smart card data for our use. The anonymized card-id will, however, remain unique for the data instance so that connected trips can be identified and linked over the given time period. BUS-POS also includes single trips purchased in the ticket machines, but it does currently not include trips bought via the apps. Those purchases correspond, however, to less than 5% of all purchases. A screen shot of the smart card data in BUS-POS is shown in Figure 10.

Information required to connect the smart card data with the service network and its unique stops with geographical positions, will be exported from REBUS.

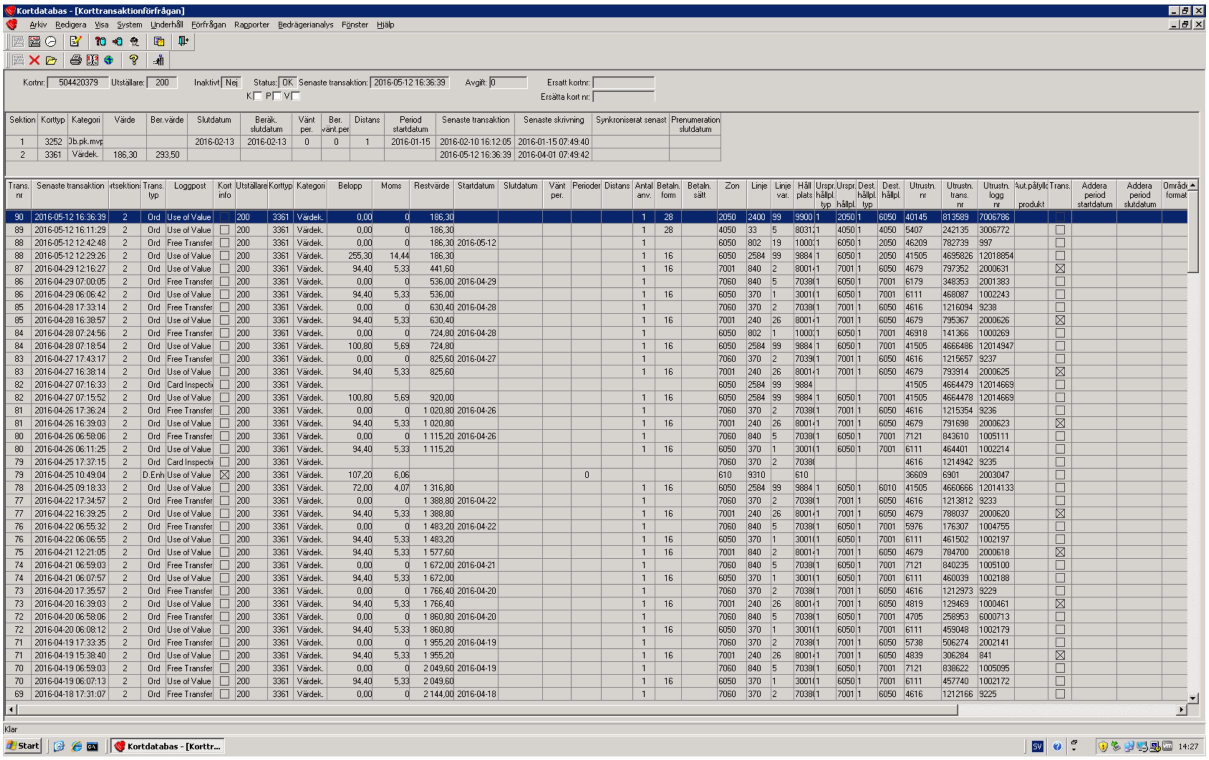

Figure 10: A snapshot of the smart card data view in BUS-POS. Source: Lars Karlinius, Blekingetrafiken.

Other potential data sources

New systems to count passengers on the trains will be implemented during 2016/2017, but for now manual counts are done three times per year to reflect the seasonal changes. This data is available to us already in an aggregated form showing which station people steps on and steps off (i.e. an OD-matrix which is not time dependent in the aggregated form).

When a travel includes a connection involving two (or more) services provided by different transport service providers which do not have connected commuter card systems (e.g. regional train service and a commercial long-distance train service), the passenger can purchase a so-called ”RESPLUS”-ticket. This ticket often costs more than buying the tickets for the individual stretches separately, but then he/she has a travel guarantee if the connection is missed (and thus can take the next train or bus from the same service provider without additional charges). These RESPLUS-tickets are issued via a service provided by Samtrafiken.

A potential data source for an estimate of the number of travelers in relations where train is the main mode, can be constructed from course mobile phone data (Gundlegård et al. 2016). This type of data is normally used in aggregated format, such as time-sliced OD-matrices. If such data become available for the region around Blekinge, an evaluation of the benefits of fusing the ticketing data with mobile phone data will be made as part of the regional level analysis. However, the availability of mobile phone data for the specific case study which is matching the ticketing data, is not guaranteed.

4.3.2. Planned analysis

Based on the data we know we have access to, we intend to analyse the following:

- Analyse robustness and punctuality of selected important train services and assess selected train-train and train-bus connections.
- Analyse the passenger flow in certain important OD-relations connecting the region of Blekinge with the commercial trains running on the Southern Mainline (main transfer stations are then Hässleholm and Alvesta) and evaluate the connection success rates.

From the smart card data provided by Blekingetrafiken, we intend to analyse passenger flows over time as well as passenger delays compared to train and bus delays. Furthermore, we intend to assess how much transfers (and service frequencies) contribute to increased (or decreased) passenger delays.

From the anonymous card-ID for each ticket transaction, it is possible to map the traveler to several stops (or vehicles or possibly by ticket machine) during a specific time period. For example, one ID can be mapped to a tap-in at a specific stop on a specific line at time t_1, and later mapped by a tap-in at another stop on another specific line at time t_2, see Figure 11.

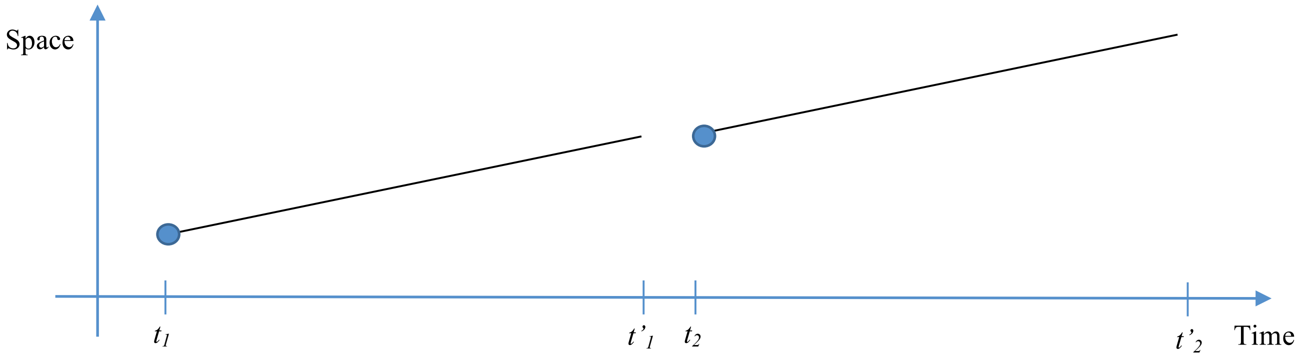

Figure 11: Time/space diagram for one anonymous ID with two tap-ins.

A connection between two public transport journeys can be detected by analysis of the timestamp of the transaction, in combination with knowledge of the timetable, or the realized time table, and an assumption about a maximal connection time. In the case in Figure 1, and we assume that the journey departed from the origin at t_1 arrived at time t'_1 to a stop where the passenger alighted. If the time span $w_1=t_2-t'_1$ is shorter than some time threshold, say w_t, we have identified a trip consisting of two journeys where the traveler has waited at location s_1 for a time w_1. If w_1 is larger than w_t, the traveler is assumed to have stopped at location s_1 with some other purpose than waiting for the connecting public transport service. In order to compute w_1, information of the realized time table, is needed, which in cases where this data is not available, be approximated by the data in the planned schedule.

The alighting station of the passengers has to be inferred from the next connection, if such is available in the ticket data set. If no connection is registered in the ticket data, some default distribution of travelers on the destination stops has to be assumed. This default distribution may be based on analysis of the return trips for the passengers, that is, the boarding stop for a trip later the same day for the specific passenger. This is illustrated in Figure 12, where it is assumed that the time between $t_{1'}$ and t_2 is larger than the threshold time w_t.

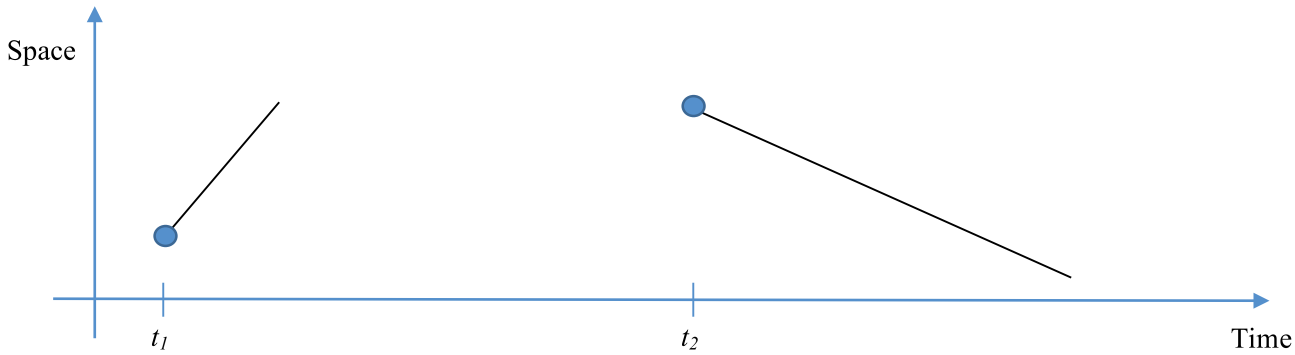

Figure 12: Return trip for a passenger, that is, the boarding stop for a trip later the same day for a specific passenger

4.3.3. Descriptive analysis and visualization

Different type of visualizations of the regional level data can be considered. The focus of the analysis will be on the service time variability and passenger travel time variability:

- Visualizations of the origins, destinations and transfer stations over time.
- Visualization of passenger flows over some selected routes over time.
- Visualization of passenger flow and travel time variability for some selected routes over time.

Some of the visualization will be made using diagrams, others on maps. Passengers arriving at a train stop in the area of analysis can a) continue with the same train, b) alight at the station and get to theirs planned destination without the one an additional public transport leg, or c) alight at the station and continue the trip by boarding anther train, leaving later, or boarding a bus, leaving later.

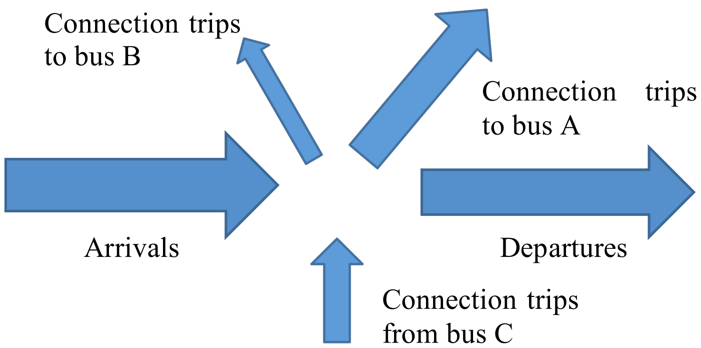

Figure 13: Travelers' exchanges made at a specific station for a specific train number.

Processing of the source ticket data by the procedures given in Section 4.3.2, will produce input that can be used for visualizing the proportions of the trips continuing on other public transport legs from a specific stop, for a specific train number.

This data can also be visualized in a more standard format of a vehicle specific origin-destination (OD) matrix. The visualization of such OD matrix requires that each of connecting vehicles (bus A, bus B and bus C in Figure 13) are defined as origins and destination for the OD matrix. However, the true destination for the last leg of the trip needs to be estimated, and for the purpose of the analysis, it might be more appropriate to define the boarding of a specific bus line as the destination.

Given the visualization of the exchanges made at a specific stop, for a specific train, the resulting wait times can be visualized, see Figure 14. Such visualization can be made in an aggregate format, where the waiting times for the different arrivals and departures are, for example, weighted by the actual number of travelers making the exchange. The visualization can also be made for each combination of vehicles, that is, for the specific train to and from a specific bus line.

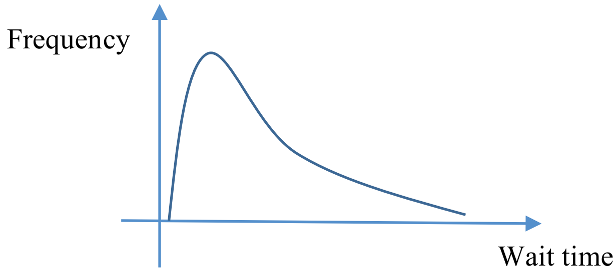

Figure 14: Passenger wait time.

Initial Tap-in analysis

This subsection includes results from the initial analysis of the Blekingetrafiken. Some of the measures described in the previous subsection are shown for a subset of the available data. Further analysis, related to the case, will be presented in Section 5.1. Indicators related to reliability, computable from this type of data combined with other data sources (planned and realized time tables), is presented in Deliverable 2.1.

The ticket data provides, as shown in 4.3.3, information of the location of the boarding stop of each traveler, and the route that the boarded vehicle is serving. The number of boarding for the month of October 2016 are shown in Figure 15. From the figure, it can be noted that for weekdays, about 25.000 boarding events are recorded, and for Saturday and Sunday less than 10.000 tap-in events.

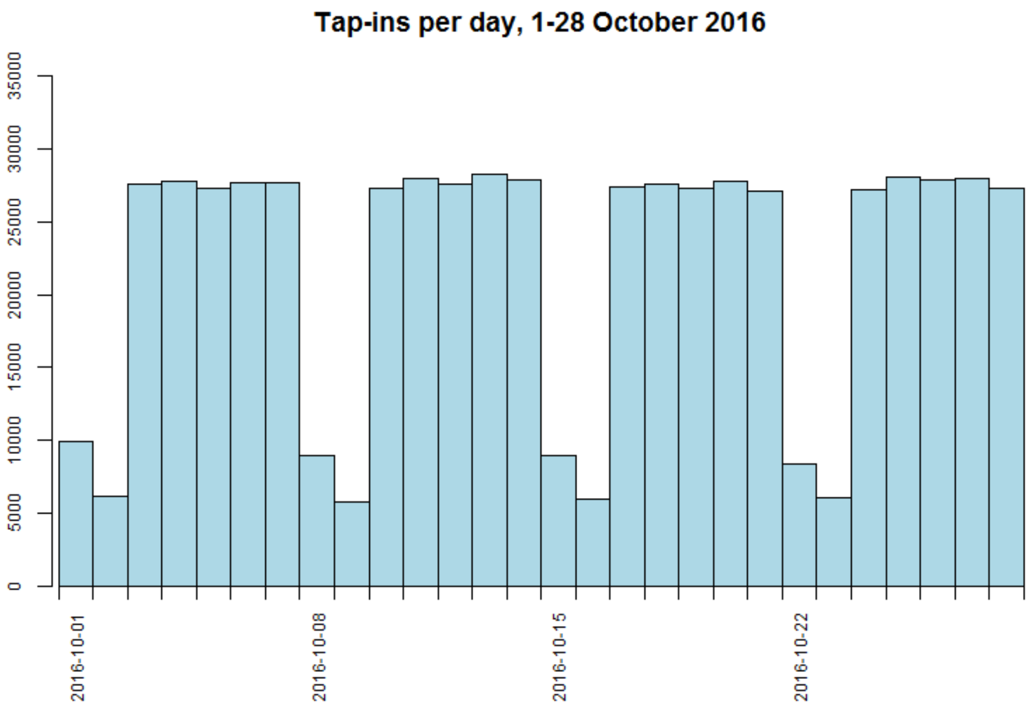

Figure 15 Number of tap-in events for one month of data.

Figure 16 shows the number of tap-in events for the first Monday in October (October the third). The number of tap-ins shows typical travelling peaks in the morning and in the afternoon, the morning peak sharper than the afternoon peak.

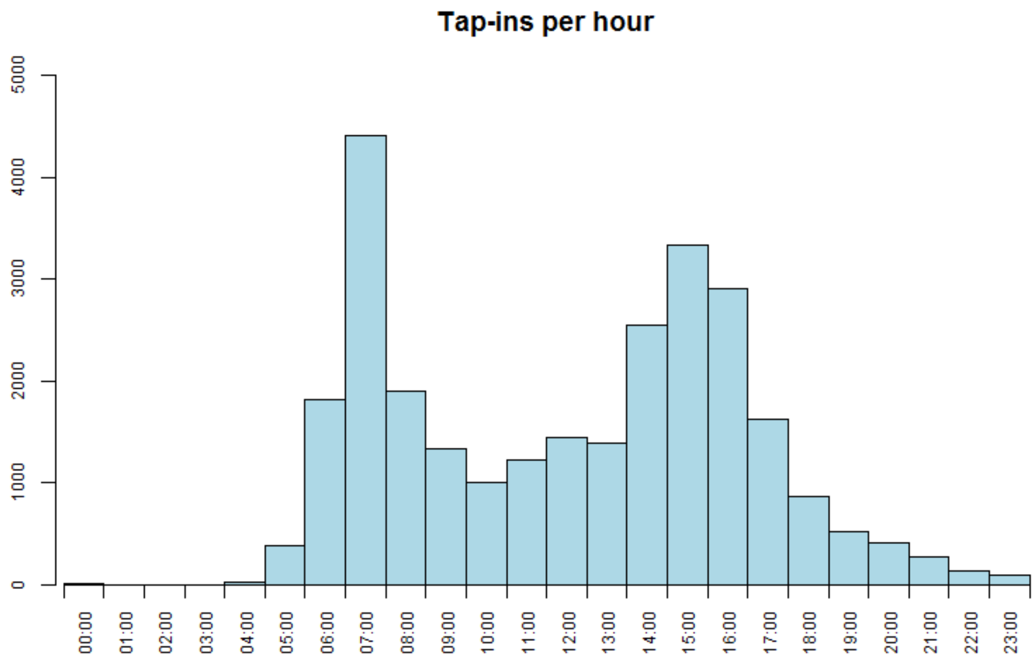

Figure 16: Tap-ins per hour for one day of tap-ins.

Figure 17 shows the number of tap-in events made between 07:00 and 08:00 the first Monday in October for a large part of the region covered by the tap-ins registered in the Blekingetrafiken ticketing system. In the figure, the radiuses of the red circles are proportional to the number of tap-in at the different bus stops.

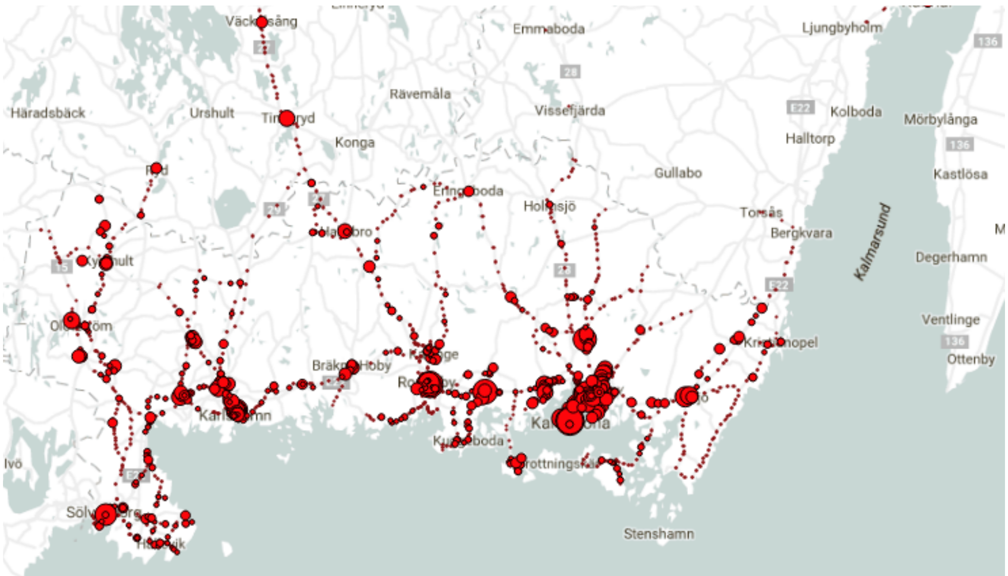

Figure 17: Circles with size proportional to the number of tap-in at each stop or station.

Initial flow analysis

Since the ticket data only includes time stamp of tap-ins, connected to stop locations and service route number, the alighting stop needs to be inferred from analysis of sequences of boardings. The technique presented in Section 4.3.5 has been applied to infer the alighting stops. Many of the boarding sequences for a given day, consists of only two tap-in at two different stop, and registered to the same route. In such situation, the procedure would estimate the alighting stop of the first trip as the boarding stop of the second tap-in, and the alighting stop for the second trip as the boarding stop of the first trip. Applying this technique for all data for one specific weekday, give us the estimated trips. In Figure 18 is all estimated trips, for bus, train and boat trips, shown. In the figure, a red line represents a trip, with a start point and an endpoint. The width of the line is proportional to the logarithm of the estimated number of travelers in the same start and endpoint relation. Each small black dot represents a bus stop or train station.

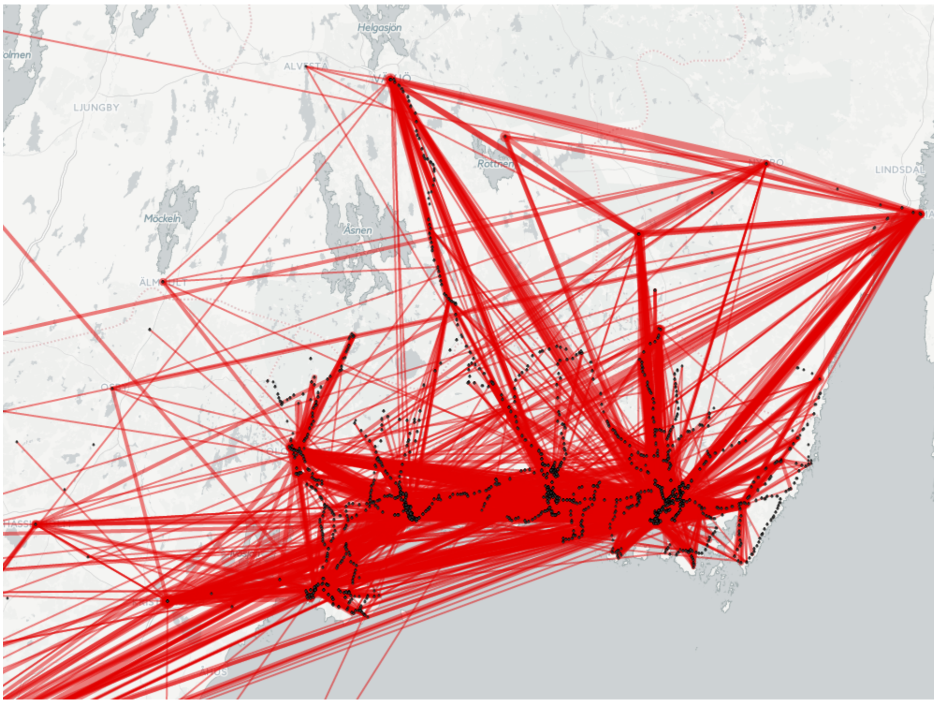

Figure 18: Lines representing trips with the public transport modes, bus, train and boat.

Figure 19 shows the trips that are registered to routes that are serviced by train. As in the previous figures, the width of the lines in the figure is proportional to the logarithm of the estimated number of travelers, and the black dot represents bus stops or train stations.

Figure 20 shows the trips that are registered to one specific train route and operator. The width of the line is proportional to the number of passengers. Figure 21 shows the journeys mapped to the sequence of stations stops. The width of the lines is proportional to the number of passengers between each pair of stations.

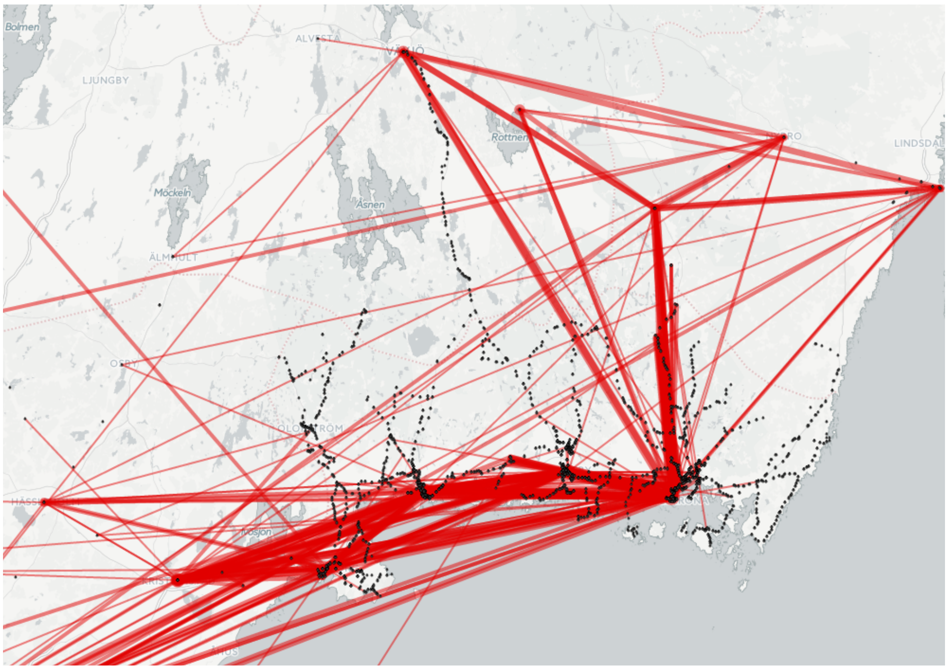

Figure 19: Lines representing trips using train lines.

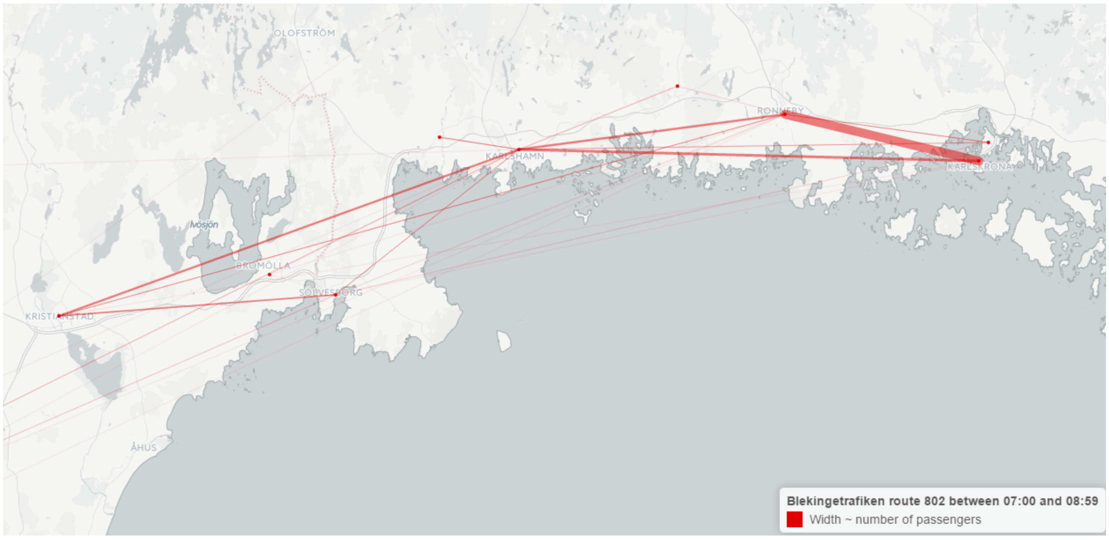

Figure 20: Start and ends of trips with Öresundståg between 7 and 9 for one day.

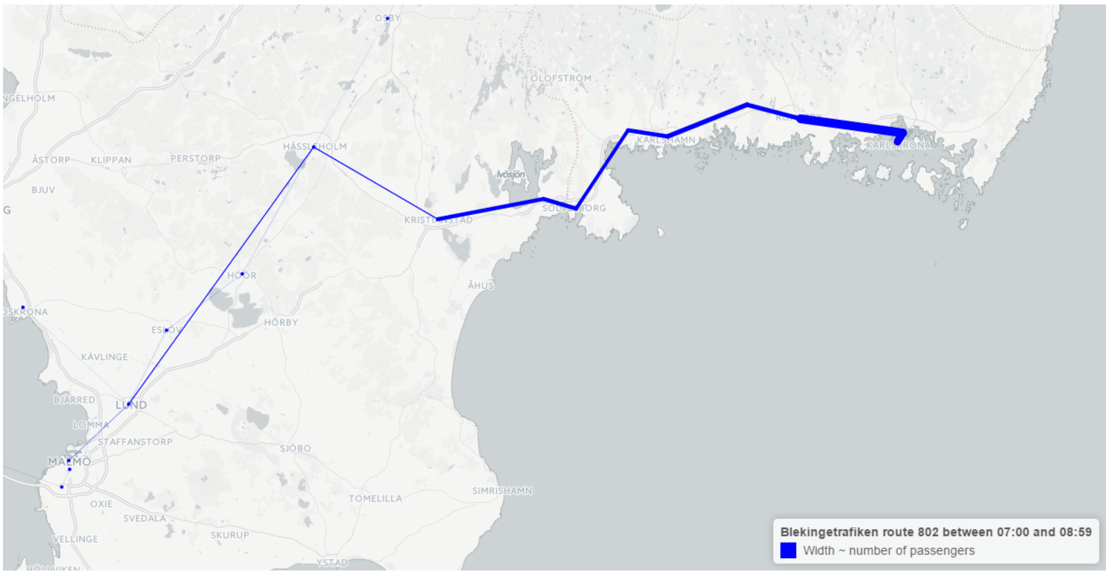

Figure 21: Route segment load for one specific train route and operator between 7 and 9 for one day.

4.4. Data visualization and analysis with common visualization tool

The visualizations of travel demand data shown in the above sections can also be done using the common visualization tool. This tool is documented in Deliverable 1.3, Graphical analysis interface of travel demand flows, where examples of visualizations of the demand data for the different levels are provided. The same tool can also be used to visualize the output from simulation runs. That is, the travel demand resulting from changes in the time table, platform allocation etc. can be visualized. Such visualizations will be exemplified in Deliverable 4.1.

5. Definition of multi-level data analysis and modelling concepts

This chapter defines the multi-level data analysis and modelling concepts. First, indices which are common among network levels are defined (Section 5.1). Then, the modelling scale and geographical scale of each level of the multi-level public transport network are defined in Section 5.2 and Section 5.3, respectively. Dynamics and interactions between network levels are described in Section 5.4.

5.1. Definition of indices of the multi-level public transport network

This section provides an overview of common notations used for the representation of the different public transport network levels. These notations are derived from Gentile and Noekel (2016) and Hänseler (2016).

- Each network level is represented by a simple directed graph $G(V, A)$ where V is the node set $v \in V$ and $A \subseteq VxV$ represents the set of arcs $a \in A$. The number of nodes and links are denoted as $|V|$ and $|A|$ respectively.
- A geographical location where a pedestrian can alight or board a tram/bus/train is called a stop, denoted s. S represents the total set of public transport stops $s \in S \subseteq V$. The number of stops is denoted as $|S|$.
- A platform $p \in P$ is an area where pedestrians alight or board a public transport vehicle (buses or railway vehicles alike).

- The set of public transport lines is denoted by L. Each public transport line $l \in L$ is defined as an ordered sequence of stops $S_l = (s_{l,1}, s_{l,2} .. s_{l,|l|})$. The first stop of a line is denoted as $S_l^- \in S_l$, whereas the last stop of a line is denoted as $S_l^+ \in S_l$.
- Each line $l \in L$ is operated by an ordered set of runs (run sequence), denoted by R_l. A run $r \in R_l$ is performed by one vehicle serving the ordered stop sequence S_l. The first run and last run of a line $l \in L$ are indicated by $R_l^- \in R_l$ and $R_l^+ \in R_l$, respectively. For each run $r \in R_l$ there exists a schedule with scheduled arrival and departure times for each stop $s_{l,j} \in S_l$.
- The total time period T of interest can be considered continuous or can be discretized into n time intervals $\tau_i, i = 0, .., n-1$ with each interval having a duration $h_i = \tau_{i+1} - \tau_i$.
- Each mode which can be used in a public transport network is denoted by $m \subset M$.
- The different distinguished public transport network levels are indicated by r (regional network level), u (urban network level) and h (hub level) in superscript.
- A sub-network, composed of nodes and links, is denoted $G_\alpha(V_\alpha, A_\alpha)$. Such a sub-network can be motivated by pedestrian activities or data limitations.
- A path k is an ordered sequence of transit links and walking links (for the urban network level) or an ordered sequence of cells (for the hub level).
- A trajectory $\eta(x, y, t)$ is the continuous position of vehicles, transit passengers or pedestrians in an area.
- Travel demand between origin $o \subset O \subseteq V$ and destination $d \subset D \subseteq V$ during time interval τ_i is denoted by d_i^{od}. The travel demand between a given set of origin nodes O and destination nodes D for time interval τ_i can be grouped into an origin-destination matrix OD_i of size $|O| \times |D|$. The passenger demand between a certain OD-pair in a certain time period d_i^{od} is distributed over a set of available and relevant path alternatives K^{od}. The number of passengers between $o \subset O$ and $d \subset D$ using path $k \subset K^{od}$ is indicated as the path flow $q_{k,\tau}^{od}$.

5.2. Definition of modelling scale of multi-level public transport networks

5.2.1. Modelling the urban public transport network level

- When modelling the urban public transport network level, each stop $s \subset S$ is represented as a node. The arc set $A_S \subseteq SxS$ represents direct connections between public transport stops. Each public transport stop $s \in S$ is connected to the public transport network by one or more arcs $a \in A$.
- A path $k = (a_1, a_2, a_3, \dots)$ over the urban public transport network is defined by a unique, ordered sequence of transit links and walking links.
- Trajectories are defined as the continuous position a passenger or vehicle follows in a geographical area. For the urban network level, trajectories can be considered from a passenger perspective and from a vehicle perspective.
- Vehicle occupancy of each run $r \in R$ of each public transport line $l \in L$ between each stop $s_{l,j}$ and $s_{l,j+1}$ is reflected by $q_{s_{l,j}-s_{l,j+1}}^r$. The link load q_τ^a equals summation of $q_{l,\tau}^r$ over all runs made by all lines operating over this link within a certain time period.
- For each run $r_l \subset R_l$ at each stop $s_{l,j} \in S_l$ the scheduled arrival and departure times are denoted by $\check{t}_{r,l,s}^a$ and $\check{t}_{r,l,s}^d$. The realized arrival and departure times are denoted by $\tilde{t}_{r,l,s}^a$ and $\tilde{t}_{r,l,s}^d$.

5.2.2. Modelling the hub network level

- In the hub level network, the walkable space is represented by a directed graph $G(V, A)$ and/or a space discretization composed of cells $\omega \in \Omega$.
- Entry/exit points of the hub are represented by nodes for the network representation. These entry/exit points in fact connect the hub network level to the network of other modes, like bus/tram/metro/train, pedestrian or bicycle network. For the space discretization, these entry/exit nodes become cells.

- For the network representation, the areas where pedestrians move, like corridors, elevators, stairs, escalators and platforms are represented by links. In the space discretization representation, all the walkable area is discretized by cells.
- Platforms can contain multiple entry/exit nodes-or cells- representing the different doors to vehicles.
- The nodes where passengers perform strategic activities are denoted as OD nodes, such that $V^{od} \subseteq V$. The set of OD nodes V^{od} is chosen based on the geographical characteristics of the hub, e.g. platforms, entrances, exits stops, shops, ticket machines, waiting areas. Train arrivals/departures could be considered as "time-dynamic" areas, given that they induce passenger flows. Operational activity nodes can be defined if relevant.
- The definition of a pedestrian's route is similar for both the network and discrete representations. For the network representation, the path $k = (a_1, a_2, a_3, \dots)$ is the ordered set of links the pedestrian used whilst when a space discretization is used, the path $k = (\omega_1, \omega_2, \omega_3, \dots)$ is the sequence of cells the pedestrian touched. Measurement variables like density k and flow q can be inferred link-wise or cell-wise. Aggregation can be done over time (by aggregating time intervals τ to larger periods) and over space (by aggregating nodes $v \in V$ – or cells – to sub regions of a hub α).
- A trajectory $\eta(x, y, t)$ is the continuous position a pedestrian follows in the walkable area.
- The set of public transport runs at a hub is denoted by $r \in R$. Each vehicle has an actual arrival time $\tilde{t}_r^a$ and actual departure time $\tilde{t}_r^d$. The platform where the vehicle from run r dwells is denoted by p_r.

5.2.3. Modelling the regional train network level

- The node set V represents public transport stops $s \subset S$, infrastructure junctions and borders between signaling blocks. Thus, the set of all stops on this network level $S \subseteq V$.
- Because the tap-in and tap-out only occurs at the first and last train station of the journey for train travelling (and not within each vehicle, as is the case for the urban network level), path enumeration and path choice cannot be observed directly, but need to be inferred.
- Vehicle occupancy of each run $r \in R$ of each public transport line $l \in L$ between each stop $s_{l,j}$ and $s_{l,j+1}$ is reflected by $q^r_{s_{l,j}-s_{l,j+1}}$. Since tap-in and tap-out for train travelling takes place on the platform, q^r_l cannot be observed directly but needs to be inferred.

5.3. Definition of geographical scale of multi-level public transport networks

We adopt a multi-level approach in this study. In this multi-level approach, there is a hierarchical structure with different functional network levels on different spatial scales, usually with different typical means of transportation (see Table 8: Relation between the studied network level, and its spatial structure.).

Table 8: Relation between the studied network level, and its spatial structure.

Network level	Intra / inter station	Intra / inter urban	Transportation means
(Inter)Regional	Inter-station	Inter-urban	Intercity train service Regional train service Local train service
Urban	Inter-station	Intra-urban	Light rail Tram Bus
Hub	Intra-station	-	Pedestrian

The definition of a hub is derived from definitions applied for airports (Rodriguez-Déniz et al. 2013). Two criteria are used for this (equation 1 and 2). The first criterion (formula 1) expresses the share of transferring passengers q^{trans}, which equals the ratio between the number of transferring passengers at the considered stop $s \in S$ and the sum of passengers originating q^o, terminating q^d and transferring

q^{trans} at that same stop. A higher ratio indicates a more dominant role as transfer location between public transport services on the same or between different network levels, instead of having a dominant role as source or sink. We do not define an explicit threshold for this ratio, in order to define a stop as a hub. We mainly consider this as a continuous scale with a decreasing importance for stops as transfer location.
The second criterion (formula 2) equals the share of transferring passengers at stop $s \in S$ in relation to all transferring movements of all stops S on the considered network. This equals the ratio between the number of transferring people at stop $s \in S$ and the total number of transferring people for S. This criterion reflects the magnitude of a specific transfer stop compared to the rest of the network. A higher ratio indicates that a larger volume of transferring passengers uses this specific stop to make this transfer. Also here, we do not define an explicit threshold for this criterion in order to define a stop as a hub or not. We consider values on this criterion as a continuous scale, where a higher value reflects a more important role as transfer location within the considered network. Scores on criteria 1 and 2 together give insight in the extent that a certain stop $s \in S$ can be considered as hub.

$$Hub\ criterion\ 1 = \frac{q_s^{trans}}{q_s^o + q_s^d + q_s^{trans}} \quad \forall\ s \in S \tag{1}$$

$$Hub\ criterion\ 2 = \frac{q_s^{trans}}{\sum_{s=1}^{|S|} q_s^{trans}} \quad \forall\ s \in S \tag{2}$$

5.4. Description of dynamics and interactions between public transport network levels

When passengers perform a door-to-door public transport journey using the multi-level public transport network, they transfer between different network levels using the hub level as intermediate connection. In general, the urban and regional public transport network levels are not directly connected with each other, but are connected by means of the hub level. This means that multi-level transfers can be considered as three-level transfers, with the hub level being the intermediate level (Figure 22). While a geographical stop on the regional level is usually modelled as one regional stop $s^r \in S^r$, this same stop can actually be a hub and contain many hub-level stops $s^h \in S^h$. These hub-level stops reflect the different platform sections, entries/exits and activity locations within the hub. On the urban level a hub can be modelled by a few stops $s^u \in S^u$, reflecting the urban stops of one or multiple urban public transport lines and platforms at this hub.

If passengers choose another path $k \subset K^{od}$ for a given OD-relation on the urban network level, this can influence the OD-matrix on the hub level. If the chosen urban path $k^{od,u}$ results in an arrival with another line and/or at another location at the hub, the OD-matrix on the hub level will be affected. The other way around, congestion on the hub level will influence the paths pedestrians choose and therefore affect the path choice on the urban level. Due to overcrowded vehicles, passengers might not make their preferred connection on the urban public transport network, leading to the choice for another urban public transport line and thus another urban path $k^{od,u}$. Delays on the urban public transport network can also affect the chosen path on the regional level. If due to delays on the urban level the preferred connection on the regional level cannot be realized anymore, passengers might decide to use another vehicle run or line on the regional level. Depending whether this vehicle departs from the same platform, or passengers decide to wait for their next vehicle at another location at the hub (e.g. at a shop or in the waiting room), this urban network level delay thus also influences the OD-matrix and chosen hub path k. These interactions also hold the other way around. In case of arrival delays of the regional train at the hub, passengers might decide to change their path k in order to catch their planned run on the urban level (e.g. run directly to the urban public transport platform, instead of first visiting a shop at the hub). Passengers also might (have to) choose another path on the urban level due to late arrival of a service on the regional network level in case the planned urban connection cannot be realized anymore. Depending whether the departure location of the chosen urban public transport service differs from the originally planned departure location at the specific hub, this arrival delay on the regional level might also change the hub-level OD-matrix. Besides, changes in arrival or departure platform for the regional train services (e.g. due to a delay, disruptions etc.) will also change the OD-matrix on the hub level.

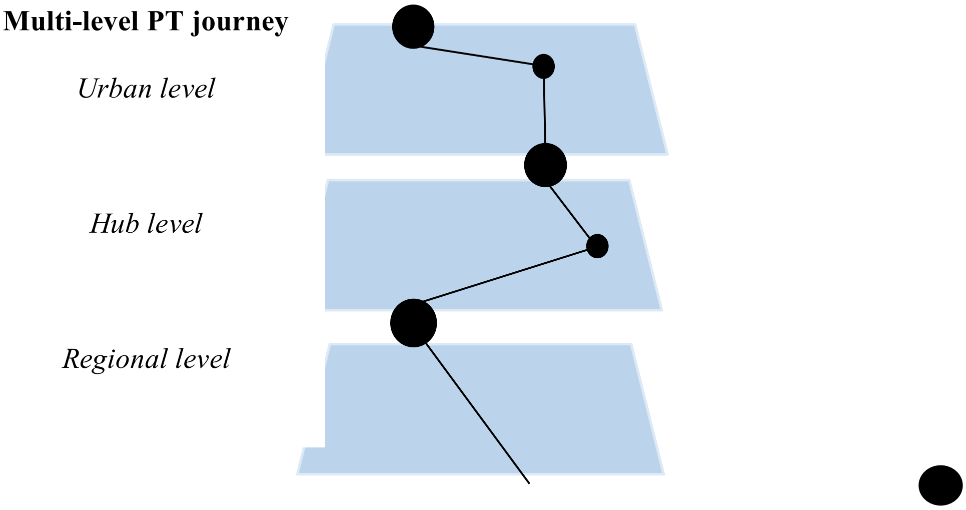

Figure 22: Multi-level transfers with the hub level connecting the urban and regional network level.

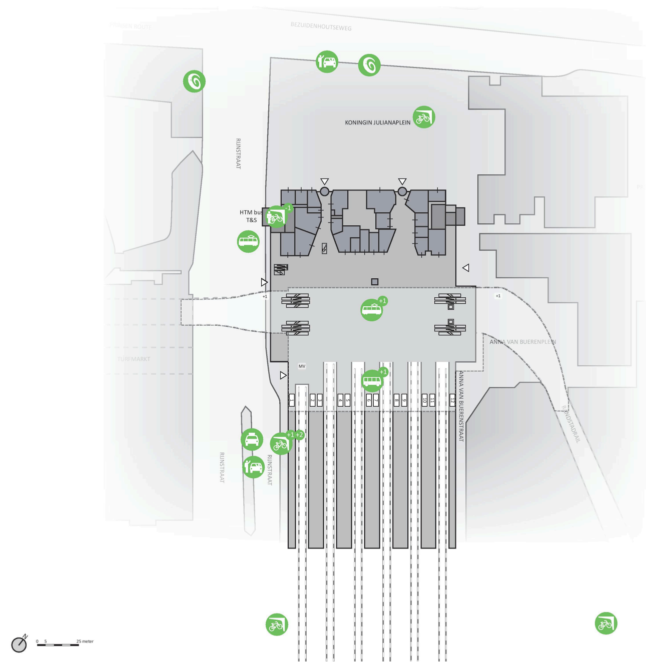

Figure 23: Illustration of multi-level transfer at The Hague Central Station, the Netherlands

We illustrate the multi-level transfer concept with the central station of The Hague, the Netherlands (Figure 23). Trains depart and arrive at level 0. Passengers transferring to the urban public transport system first use the hub level as connection between the (inter)regional train and urban network. Trams depart and arrive at level 0 and level 1; busses depart and arrive at level 1; metros depart and arrive at level 2. Other access/egress modes for train passengers are for example bicycle (with paid and unpaid bicycle storage at level 0 and level -1), taxi or car (using Kiss&Ride locations).

6. Conclusions

In this task, we have analysed the traveller dynamics that take place at each network level of three different levels. Descriptive analyses were reported: pedestrian flows inside a single multi-modal transport hub of Switzerland, flow dynamics on an urban public transport network of the Netherlands, and the transit flows in a region of Sweden. These analyses by using big transportation data revealed and visualized the flow dynamics at each level, which gave the well understanding. We also established a common system representation and mathematical notations to be used throughout the project and modelling developments. This is important for the subsequent tasks in the project.

7. References

Alahi, A., Vandergheynst, P., Bierlaire, M., & Kunt, M. (2010). Cascade of descriptors to detect and track objects across any network of cameras . *Computer Vision and Image Understanding , 114*, 624-640. doi:http://dx.doi.org/10.1016/j.cviu.2010.01.004

Anas, A. a. (1998). Urban Spatial Structure. *Journal of Economic Literature, 36*, 1426-1464. Retrieved from http://www.jstor.org/stable/2564805

Anderson, I., Maitland, J., Sherwood, S., Barkhuus, L., Chalmers, M., Hall, M., . . . Muller, H. (2007). Shakra: tracking and sharing daily activity levels with unaugmented mobile phones. *Mobile Networks and Applications, 12*, 185-199.

Anderson, N. B., & Bogart, W. T. (2001). The structure of sprawl: Identifying and characterizing employment centers in polycentric metropolitan areas. *American Journal of Economics and Sociology, 60*, 147-169.

Atasoy, B., Glerum, A., & Bierlaire, M. (2013). Attitudes towards mode choice in Switzerland. *disP-The Planning Review, 49*, 101-117.

Axhausen, K. W. (2005). A dynamic understanding of travel demand: A sketch. *Integrated Land-Use and Transportation Models: Behavioural Foundations*, 1-20.

Bauer, D., Brändle, N., Seer, S., Ray, M., & Kitazawa, K. (2009). Measurement of pedestrian movements: A comparative study on various existing systems. 325-344. Emerald Group Publishing.

Bierlaire, M., & Moura, E. P. (2014). Effects of terminal planning on passenger choices.

Bierlaire, M., Chen, J., & Newman, J. (2013). A probabilistic map matching method for smartphone GPS\ data . *Transportation Research Part C: Emerging Technologies , 26*, 78-98. doi:http://dx.doi.org/10.1016/j.trc.2012.08.001

Broach, J., Dill, J., & Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS\ data . *Transportation Research Part A: Policy and Practice , 46*, 1730-1740. doi:http://dx.doi.org/10.1016/j.tra.2012.07.005

Camagni, R., Gibelli, M. C., & Rigamonti, P. (2002). Urban mobility and urban form: the social and environmental costs of different patterns of urban expansion . *Ecological Economics , 40*, 199-216. doi:http://dx.doi.org/10.1016/S0921-8009(01)00254-3

Campanella, M., Hoogendoorn, S., & Daamen, W. (2009). Effects of heterogeneity on self-organized pedestrian flows. *Transportation Research Record: Journal of the Transportation Research Board*, 148-156.

Chattaraj, U., Seyfried, A., & Chakroborty, P. (2009). Comparison of pedestrian fundamental diagram across cultures. *Advances in complex systems, 12*, 393-405.

Chen, M., Yaw, J., Chien, S. I., & Liu, X. (2007). Using automatic passenger counter data in bus arrival time prediction. *Journal of Advanced Transportation, 41*, 267-283.

Crisostomi, E., Kirkland, S., & Shorten, R. (2011). A Google-Like Model of Road Netwrok Dynamics and its Application to Regulation and Control. *International Journal of Control, 84*(3), 633-651.

Csillery, K., Blum, M. G., Gaggiotti, O. E., & Francois, O. (2010). Approximate Bayesian Computation (ABC) in Practice. *Trends in Ecology and Evolution, 25*(7), 410-418.

Daganzo, C. F., & Geroliminis, N. (2008). An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transportation Research Part B: Methodological, 42*, 771-781.

Danalet, A., Farooq, B., & Bierlaire, M. (2014). A Bayesian approach to detect pedestrian destination-sequences from WiFi signatures. *Transportation Research Part C: Emerging Technologies, 44*, 146-170.

De Freitas, A., Septier, F., & Mihaylova, L. (2015). Sequential Markov Chain Monte Carlo for Bayesian Filtering with Massive Data. *Submitted to IEEE Transactions oon SIgnal Processing*.

Dubois-Taine, G., & Chalas, Y. (1997). La ville émergente.

Dziekan, K., & Kottenhoff, K. (2007). Dynamic at-stop real-time information displays for public transport: effects on customers . *Transportation Research Part A: Policy and Practice , 41*, 489-501. doi:http://dx.doi.org/10.1016/j.tra.2006.11.006

Fearnhead, P., & D., P. (2012). Constructing Summary Statistics for Approximate Bayesian Computation: Semi-Automatic Approximate Bayesian COmputation. *Journal of the Royal Statistical Society Series B, 74*(3), 419-474.

Foxlin, E. (2005). Pedestrian tracking with shoe-mounted inertial sensors. *IEEE Computer graphics and applications, 25*, 38-46.

Frank, L., Bradley, M., Kavage, S., Chapman, J., & Lawton, T. K. (2008). Urban form, travel time, and cost relationships with tour complexity and mode choice. *Transportation, 35*, 37-54. doi:10.1007/s11116-007-9136-6

Gning, A., Mihaylova, L., & Boel, R. (2011). Interval Macroscopic Models for Traffic Networks. *IEEE Transactions on Intellignet Trasnportation Systems, 12*(2), 523-536.

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *Nature, 453*, 779-782.

Gundlegård, D., Rydergren, C., Breyer, N., & Rajna, B. (2016). Travel demand estimation and network assignment based on cellular network data. *Computer Communications*.

Hänseler, F. S., Bierlaire, M., & Scarinci, R. (2016). Assessing the usage and level-of-service of pedestrian facilities in train stations: A Swiss case study. *Transportation Research Part A: Policy and Practice*, 106-123.

Handy, S. (1996). Methodologies for exploring the link between urban form and travel behavior. *Transportation Research Part D: Transport and Environment, 1*, 151-165.

Handy, S. (1996). Urban form and pedestrian choices: study of Austin neighborhoods. *Transportation Research Record: Journal of the Transportation Research Board*, 135-144.

Handy, S., Cao, X., & Mokhtarian, P. (2005). Correlation or causality between the built environment and travel behavior? Evidence from Northern California . *Transportation Research Part D: Transport and Environment , 10*, 427-444. doi:http://dx.doi.org/10.1016/j.trd.2005.05.002

Helbing, D., Molnar, P., Farkas, I. J., & Bolay, K. (2001). Self-organizing pedestrian movement. *Environment and planning B: planning and design, 28*, 361-383.

Holden, E., & Norland, I. T. (2005). Three challenges for the compact city as a sustainable urban form: household consumption of energy and transport in eight residential areas in the greater Oslo region. *Urban studies, 42*, 2145-2166.

Hoogedoorn, S., Ossen, S., Blosseville, J. M., & Hadj-Salem, H. (2006). Unscented Particle Filter for Delayed Car-Following Models Estimation. in procc. IEEE Intelligent Trasnportation Conference.

Jaulin, L. (2009). Robust Set Membership State Estimation. *Automatica, 45*(1), 202-226.

Jaulin, L., & Walter, E. (2002). Guaranteed Robust Nonlinear Minimax Estimation. *IEEE Transactions on AUtomatic Control, 47*(11), 1857-1864.

Jaulin, L., Kieffer, M., Didrit, O., & Walter, E. (2001). *Applied Interval Analysis.* New York: Springer-Verlag.

Kotecha, J. H., & Djuri, P. M. (2001). Gaussian Sum Particle Filtering for Dynamic State Space Models. Slat Lake City: in Procc. International Conference on Acoustics, Speech and Signal Processing.

Kotecha, J. H., & Djuri, P. M. (2003). Gaussian Particle Filtering. *IEEE Transactions on Signal Processing, 51*(10), 2592-2601.

Kotecha, J. H., & Djuri, P. M. (2003). Gaussian Sum Particle Filtering . *IEEE Transactions on Signal Processing, 51*(10), 2602-2612.

Lavadinho, S., Alahi, A., & Bagnato, L. (2013). *Analysis of Pedestrian Flows: Underground pedestrian walkways of Lausanne train station.* VisioSafe SA.

Lin, W.-H., & Zeng, J. (1999). Experimental study of real-time bus arrival time prediction with GPS data. *Transportation Research Record: Journal of the Transportation Research Board*, 101-109.

Manual-HCM, H. C. (2010). Transportation Research Board. *National Research*.

Marjoram, P., Molitor, J., Plagnol, V., & Tavare, S. (2003). Markov Chain Monte Carlo Without Likelihoods. *Proceedings of the National Academy of Sciences, 100*(26), 15324-15328.

Mazloumi, E., Currie, G., & Rose, G. (2009). Using GPS data to gain insight into public transport travel time variability. *Journal of Transportation Engineering, 136*, 623-631.

McNally, M. G. (2007). The four step model. *Handbook of transport modelling, 1*, 35-41.

Mihaylova, L., Boel, R., & Hegyi, A. (2007). Freeway Traffic Estimation Within Particle Filtering Framework. *Automatica, 43*, 290-300.

Mihaylova, L., Hegyi, A., Gning, A., & Boel, R. (2012). Parallelized Particle and Gaussian Sum Particle Filters for Large-Scale Freeway Traffic Systems. *IEEE Transactions on Intelligent Trasnportation Systems, 13*(1), 36-48.

Munizaga, M., Devillaine, F., Navarrete, C., & Silva, D. (2014). Validating travel behavior estimated from smartcard data . *Transportation Research Part C: Emerging Technologies , 44*, 70-79. doi:http://dx.doi.org/10.1016/j.trc.2014.03.008

Orth, H., Weidmann, U., & Dorbritz, R. (2012). Development of measurement system for public transport performance. *Transportation Research Record: Journal of the Transportation Research Board*, 135-143.

Papinski, D., Scott, D. M., & Doherty, S. T. (2009). Exploring the route choice decision-making process: A comparison of planned and observed routes obtained using person-based GPS\ . *Transportation Research Part F: Traffic Psychology and Behaviour , 12*, 347-358. doi:http://dx.doi.org/10.1016/j.trf.2009.04.001

Pelletier, M.-P., Trépanier, M., & Morency, C. (2011). Smart card data use in public transit: A literature review . *Transportation Research Part C: Emerging Technologies , 19*, 557-568. doi:http://dx.doi.org/10.1016/j.trc.2010.12.003

Phithakkitnukoon, S., Horanont, T., Di Lorenzo, G., Shibasaki, R., & Ratti, C. (2010). Activity-aware map: Identifying human daily activity pattern using mobile phone data. *International Workshop on Human Behavior Understanding*, (pp. 14-25).

Polus, A., Schofer, J. L., & Ushpiz, A. (1983). Pedestrian flow and level of service. *Journal of transportation engineering, 109*, 46-56.

Rajbhandari, R., Chien, S., & Daniel, J. (2003). Estimation of bus dwell times with automatic passenger counter information. *Transportation Research Record: Journal of the Transportation Research Board*, 120-127.

Richards, P. I. (1956). Shock Waves on the Highway. *Operations Research, 4*, 42-51. Retrieved from http://www.jstor.org/stable/167515

Robenek, T., Maknoon, Y., Azadeh, S. S., Chen, J., & Bierlaire, M. (2016). Passenger centric train timetabling problem. *Transportation Research Part B: Methodological, 89*, 107-126.

Seyfried, A., Steffen, B., Klingsch, W., & Boltes, M. (2005). The fundamental diagram of pedestrian movement revisited. *Journal of Statistical Mechanics: Theory and Experiment, 2005*, P10002.

Sisiopiku, V. P., & Akin, D. (2003). Pedestrian behaviors at and perceptions towards various pedestrian facilities: an examination based on observation and survey data. *Transportation Research Part F: Traffic Psychology and Behaviour*, 249-274.

Törnquist, J. (2006). Computer-based decision support for railway traffic scheduling and dispatching: A review of models and algorithms. *OASIcs-OpenAccess Series in Informatics*, *2.*

Tahmasbi, R., & Hashemi, S. M. (2014). Modeling and Forecasting the Urban Volume Using Stochastic Differential Equations. *IEEE Transactions on Intelligent Trasnportation Systems, 15*(1), 250-259.

Tossavainen, O.-P., & Work, D. (2013). Markov Chain Monte Carlo Based Inverse Modeling of Traffic Flows Using GPS data. *Networks and Heterogeneous Media, 8*(3), 803-824.

Tsai, Y.-H. (2005). Quantifying urban form: compactness versus' sprawl'. *Urban studies, 42*, 141-161.

van den Heuvel, J. P., & Hoogenraad, J. H. (2014). Monitoring the Performance of the Pedestrian Transfer Function of Train Stations Using Automatic Fare Collection Data. *Transportation Research Procedia, 2*, 642-650. doi:http://dx.doi.org/10.1016/j.trpro.2014.09.107

Wang, R., Work, D. B., & Sowers, R. (2014). Multiple Model Particle Filter for Traffic Estimation and Indicident Detection. *Submitted to IEEE Transactions on Intelligent Transportation Systems*.

Wong, S. C., Leung, W. L., Chan, S. H., Lam, W. H., Yung, N. H., Liu, C. Y., & Zhang, P. (2010). Bidirectional pedestrian stream model with oblique intersecting angle. *Journal of transportation Engineering, 136*, 234-242.

Wu, X., & Liu, H. X. (2011). A shockwave profile model for traffic flow on congested urban arterials. *Transportation Research Part B: Methodological, 45*, 1768-1786.

Yang, H., Sasaki, T., Iida, Y., & Asakura, Y. (1992). Estimation of origin-destination matrices from link traffic counts on congested networks. *Transportation Research Part B: Methodological, 26*, 417-434.

Zhao, J., Rahbee, A., & Wilson, N. H. (2007). Estimating a Rail Passenger Trip Origin-Destination Matrix Using Automatic Data Collection Systems. *Computer-Aided Civil and Infrastructure Engineering, 22*, 376-387.

Zimmermann, A., & Hommel, G. (2003). A train control system case study in model-based real time system design. *Parallel and Distributed Processing Symposium, 2003. Proceedings. International*, (pp. 8--pp).