D2.1: Passenger-focused Indicators of Service Variability

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
Menno Yap	m.d.yap@tudelft.nl
Flurin Hänseler	f.s.hanseler@tudelft.nl
Patricia Bellver Muñoz	pbellver.etraid@grupoetra.com
Clas Rydergren	clas.rydergren@liu.se
Riccardo Scarinci	riccardo.scarinci@epfl.ch
Nicholas Molyneaux	nicholas.molyneaux@epfl.ch

Internal Reviewer

Johanna Törnquist Krasemann johanna.tornquist@bth.se

State:	Complete Draft
Distribution:	[Confidential]

Deliverable History

Date	Author	Changes
12-04-2016	Oded Cats, TUD	Draft table of contents
26-05-2016	Menno Yap, TUD	Section 2.2 (draft)
31-05-2016	Menno Yap and Oded Cats, TUD	Section 3.2
10-06-2016	Patricia Bellver, ETRA	Section 2.1 and 2.3
28-06-2016	Oded Cats, TUD	Sections 2.1, 2.2
24-06-2016	TUD, ETRA	Revisions and additions to Chapter 2.3
10-08-2016	Shadi Sharif Azadeh, EPFL	Added draft of Section 3.1
01-09-2016	Clas Rydergren, LiU	Added draft of Section 3.3
28-09-2016	Menno Yap, TUD	Added draft of Section 3.2
11-10-2016	Oded Cats, TUD	Updated timeline, comments to Sections
14-12-2016	Clas Rydergren, LiU	3.1, 3.3
22-02-2017	Nicholas Molyneaux EPFL	Updated draft of Section 3.3
		First version of variability analysis at hub
23-02-2017	Patricia Bellver, ETRA	(chap 3.1 + 4.1)
01-03-2017	Flurin Hanseler, TUD	Revisions and additions to Chapter 2.1 and 2.3
05-03-2017	Clas Rydergren, LiU	Extensive proof-reading, added summary and Sec. 1.2 and 2.2
		Added text for Section 4.3
14-03-2017	Oded Cats, TUD	Formatting, proofreading, Section 5.2

Contents

Summary

This document reports the work performed in task 2.1 of the TRANS-FORM project, and is entitled "Passenger-focused Indicators of Service Variability". This task was performed as part of WP2 "Measuring and Modelling Passenger Interchange Activities".

The present document discusses and proposes methodologies for quantifying and measuring service variability in the context of public transportation. To that end, the concepts of reliability and robustness are introduced, dealing with recurrent disturbances and non-recurrent disruptions, respectively. A range of variability measures are presented that allow to analyse and evaluate passenger flows across entire public transportation networks, with a special emphasis on transferring flows. Indicators are first discussed and developed separately at the level of a transportation hub, at the urban and the regional level, before they are combined in an integrated framework. For instance, measures of travel reliability for multi-modal trips are defined that include waiting, in-vehicle and transfer times. Similarly, measures of travel and arrival time predictability as well as the likelihood of missed connections are formulated. The measures are tested and applied to three different case studies that are considered throughout the TRANS-FORM project, allowing to understand their nature at the various aggregation levels and across national borders. The wider aim of the development of service variability measures pertains to the development of a multi-layer passenger flow model that can represent flow dynamics across a public transportation network, and which allows to evaluate and steer transit operations from the perspective of individual passengers.

1. Introduction

1.1. Purpose and Approach

Service variability pertains to any deviation from planning or normal operations which results in service degradation for example in terms of prolonged travel times or a lower level of travel comfort. The public transport industry has experienced in recent years a rapid increase in the amount of data being collected and made publicly available. From the vehicle-side, automatic vehicle location (AVL) data regarding scheduled and realized departure and arrival times are available for each realized trip. From the passenger-side, since the introduction of automatic fare collection (AFC) systems – like the Oyster Card in London and the smartcard in the Netherlands – an enormous amount of passenger travelling data has become available (Pelletier et al. 2014). The growing availability of these data sources facilitates the development of novel indicators of service variability and its impacts on passenger journey. Such indicators can support decision making by contributing to the evaluation and monitoring of service performance.

This deliverable reports the review, development and application of indicators for measuring public transport service variability from passengers' perspective. The indicators developed in the course of task 2.1 in the TRANS-FORM project can be used to assess the performance of public transport systems based on empirical data as well as be used to evaluate the effects of alternative strategies aimed at improving passengers' experience. By providing means to monitor and assess measures in terms of passenger experience of service variability, this report contributes to the overarching project goal: "to better understand transferring dynamics in multi-modal public transport systems and develop insights, strategies and methods to support decision makers in transforming public transport usage to a seamless travel experience by using smart data" and with the particular emphasize on gearing "the services towards travellers rather than the infrastructure".

The term *service variability* encompasses both notions of reliability and vulnerability. The former is associated with higher probability and lower impact events while the latter pertains to lower probability and more impactful incidents. This deliverable will cover both aspects of service variability. In line with the overall approach of the TRANS-FORM project, the indicators are formulated and applied to the regional, urban and hub levels of public transport networks. Moreover, special attention is given to service performance in relation to transfer activities. The prospects of measuring service variability across network levels are discussed. The applications are designed to demonstrate how the proposed indicators can be implemented and how they can be used to identify the temporal and spatial variation in service reliability and vulnerability.

1.2. Report Outline

The remainder of this document is structured as follows: Section 2 presents a thorough review of the current state-of-the-practice and state-of-the-art of measuring service variability, outlining the most relevant performance concepts for service variability. Notably, these include reliability and robustness, useful for describing the system response in case of recurring and non-recurring disturbances and disruptions, respectively. The focus thereby lies on passenger-oriented indicators that directly capture travel experience. Section 3 introduces specific performance indicators for measuring service variability at the level of transfer stations, urban level, as well as at the regional level. These indicators are at first developed separately for each level, however keeping their interdependence in mind. Section 4 illustrates service variability at the example of three case studies from Switzerland, the Netherlands and Sweden, which are used throughout the TRANS-FORM project (see "D5.1: Case study set-up descriptions" for more details on the individual case studies). Finally, Section 5 provides an integrated perspective on service variability that encompasses entire passenger journeys. Additionally, a pathway is presented to incorporate these indicators into traffic modelling, considering both empirical and modelling prospects.

2. Measuring Service Variability

This chapter provides a review of the current state-of the-practice and state-of-the-art in sections 2.1 and 2.2, respectively, of measuring service variability with a special focus on indicators designed to capture passenger experience. Following this review, a set of requirements for service variability indicators is provided which is then used in the development of passenger-oriented indicators in the subsequent chapter.

2.1. State of the Practice

The following provides a review of research articles, policy documents and concession contracts concerning the monitoring and measuring of variability. First, this section is introduced with aspects related to quality of service in public transport, especially those related to reliability and vulnerability. Second, a review on service variability policies and mitigation strategies is presented. Finally, a description of the methods or indicators in practice is included.

2.1.1. Quality of service in public transport

In the last years, public authorities have started to integrate criteria for service quality in transport operation tenders. For instance, the QUATTRO project (Quality Approach in Tendering/contracting Urban Public Transport Operations, EU 4th framework Programme R+D, 1996-1998), was focused on how to integrate quality in urban public transport design, in order to stimulate innovations with impact on quality. More specifically, in tenders and contracts, QUATTRO makes reference to the use of an appropriate legal framework and to the choice of measures to optimise quality, reviewing different methods for sharing contractual risks among the parties, as well as possible variants and combinations, such as clauses on incentives and penalties and other risk-sharing schemes. From the review of tenders and awarded contracts in Garcia and López (2005), the following conclusions can be derived:

- Quality aspects are gradually being introduced into urban transport operation contracts. The modality of interested management for urban bus operation is the one with better possibilities to incorporate quality issues in contracts.
- The effectiveness of quality requirements and commitments in contracts can only be reached with penalties (and incentives) in the financial functioning of the contract.
- Quality aspects are generally referred to measurable produced quality (fulfilment of runs, punctuality, etc.), non-measurable produced quality (cleanliness, information, signalling, staff behaviour, etc.), and perceived quality.
- In an indirect way commercial risk in terms of patronage is intrinsically related to service quality.

Definitively, there is still a long way to go to reach a comprehensive and fair definition of quality in transport operation contracts. Definition of variables and measurement techniques are some of the weakest points that should be improved (Garcia and López, 2005).

There is a European specific quality standard for the public passenger transport sector, CEN EN 13816-2002, which specifies the requirements for defining, targeting and measuring quality of service in public passenger transport. EN 13816 applies to passenger transport service providers and does not exclude individual passenger vehicles such as taxis, and includes the following criteria:

- Availability (network, operation time, **reliability**)
- Access (interfaces, ticketing)
- Information (travel information, regular and occasional)
- Time (travel time, punctuality, regularity)
- Customer Service (availability of personnel, competence, assistance)
- Comfort (space, driving, environment)
- Security and safety (avoidance of criminal attacks and of accidents, emergency)
- Environment (pollution, resources)

Thus, EN 13816 provides transportation with benchmark criteria to structure an approach to improve public services in the transport sector other than advancing through general quality standards.

To assess urban public transport system quality, a quality rating system should be developed, and continuous monitoring of service quality should be exercised. The most common indicators are: **reliability/punctuality** (100%), commercial speed/trip time (92%), comfort on the run (92%), service frequency/regularity (75%), cleanliness and maintenance (75%), safety (67%), trip price/fare level (58%), security (58%), trip environment (58%), transfers necessity (50%), customers contact (50%) (Pticina, 2011).

Public transport procurement is a step in the process of obtaining the most cost-effective, fit for purpose service for the transport authority – this should not be simply about procuring the lowest-cost option. A major objective must be an assessment of the appropriate quality for the services procured. What is appropriate will change over time as new technologies become available, the aspirations of passengers grow, and laws and guidance change to reflect this. It implies that the transport authority will:

- Include quality as part of its contract monitoring and reporting
- Include reporting on this in its Service Level Agreements with client departments and agree on a service quality improvement strategy
- Communicate this adequately to public transport operators in order to enable preparation for any necessary investment in vehicles, equipment and staff.

Local traffic authorities can have a major influence on the reliability, efficiency and attractiveness of bus services through their management of the road network. Either within a specific Performance (or Punctuality) Improvement Partnership (PIP) or a wider quality partnership, a framework for effective improvement of bus journey times and reliability through targeted action can be provided. This will vary according to local circumstances, but include (apart from bus priority measures, which are not always appropriate) parking control and enforcement, traffic signal settings, traffic flow changes, road and lane geometry and marking, bus stop location and layout, roadworks management, etc. This will have direct impact on tendering, where service timetables are specified by the public transport authority. Setting unrealistic timetables will, at best, create conflict with contracting operators and instability as services are adjusted to reality; at worst, it will result in no service, as no operator will be willing to risk the cost of substantial sanctions. Thus, there are both positive and negative reasons for authorities to adopt a proactive partnership approach with operators for managing the punctuality and **reliability** of bus services. They must also ensure that their tendered service timetables are appropriate and realistic, e.g. that no bus should, for a foreseeable reason, run early or more than five minutes late. Therefore, for all passenger transport, the reliability and punctuality of operation will be a core parameter (Department for Transport, 2013).

It is hard to define an ambitious yet achievable level of service. Authorities tend to set too high standards, from the perspective of operators. When a penalty is proposed or applied, much discussion starts on how the data is collected and processed. In addition, it is very important to distinguish who is responsible for which part of unreliability. As shown in Van Oort (2011), several sources together create variability and unreliability. Some of them are under the responsibility of the operator and some under the public transport authority and/or infrastructure manager (Van Oort, 2014).

The White Paper (European Commission, 2011, p. 12) remarks the relevance of quality transport services: "The quality, accessibility and **reliability** of transport services will gain increasing importance in the coming years, inter alia due to the ageing of the population and the need to promote public transport. Attractive frequencies, comfort, easy access, **reliability of services**, and intermodal integration are the main characteristics of service quality."

The aim of the Regulation (EC) No 1370/2007 of the European Parliament on public passenger transport services by rail and by road, is to define how, in accordance with the rules of Community law, competent authorities may act in the field of public passenger transport to guarantee the provision of services of general interest which are among other things more numerous, safer, of a higher quality or provided at lower cost than those that market forces alone would have allowed. More concretely, it states that in the case of a direct award of public service contracts for transport by rail, the competent authority shall make public the quality targets, such as punctuality and **reliability** and rewards and penalties applicable within one year of granting the award.

In 2050, the major part of medium distance passenger transport should be on rail. Railway needs to improve its punctuality as well as **reliability**. Punctuality and reliability, means that at least 19 trains out of 20 arrive on time (95%). Indeed, one of the main targets of ERRAC (2012) in Intelligent Mobility priority area for 2050 is nearly all customer-trips will arrive at their destination on time because 95% of all trains are punctual, arriving within 5 minutes of the planned time of arrival. Improved system performance and advanced on board and waysides technology together with full interoperability deliver reliable, available, maintainable and safe trains; moreover, new command control technology will contribute to increase the reliability of train connections and network capacity.

2.1.2. Service variability

Reliability

At present, network and service reliability are not systematically incorporated in the transport planning process and thus not reflected adequately in decision making. Even where decision-making guidelines do incorporate reliability, most of the actual project appraisals do not include monetised parameters for reliability. Reliability is rarely factored into cost-benefit analysis, the core planning tool for transport networks. There are ways to measure and value reliability that can be integrated into cost-benefit analysis that have been used on a pilot basis in some countries. These approaches provide a foundation for incorporating reliability benefits into investment appraisals and, consequently, policy frameworks. Since the demand for reliability varies markedly across users, a single monetary value for reliability will be of little use in project appraisal – a range of values is required that represents the major user groups in each case. Reliability targets and performance indicators for services and infrastructure performance enable discussions among users, operators and decision makers regarding the proper levels of reliability. Such targets present an average level of reliability not reflecting diversity in the demand for reliability. Targets should aim at reflecting both the network and the user perspective (OECD/International Transport Forum, 2010).

In OECD/International Transport Forum (2010), reliability is defined as the ability of the transport system to provide the expected level of service quality, upon which users have organised their activities. Thus, reliability can be improved either by supplying a higher level of reliability, or by changing expectations concerning the level of service quality. In other words, unpredictability of network performance is the defining characteristic of unreliability. The more random (less predictable) the performance, the harder it is for the network user to ensure against delays.

Individuals and infrastructure managers affected by changing reliability can respond in a number of ways; individuals build extra (buffer) time into their journeys to allow for the possibility of delay, while infrastructure managers provide mainly traffic flow information to reduce the impact of unreliability, among other strategies to decrease delay probability or its impact (OECD/International Transport Forum, 2010).

Service reliability and how public transport authorities deal with this important quality aspect during design, monitoring and tendering of services was assessed by Van Oort (2014). His paper presented results of two surveys (an international survey, and other with public transport authorities in The Netherlands) on the reliability practices of public transport authorities, during the design of the network and of the timetable, and in relation to concession requirements and incentive regimes. One of the main conclusions was that service variability of vehicle performance is measured and monitored frequently, whereas a focus on passenger impacts is lacking. The impacts of unreliable services on passengers are (i) average travel time extension, (ii) increased travel time variability, (iii) a lower probability of finding a seat in the vehicle. Traditional indicators focus on vehicles instead of passengers. Second, there was no consistency in the definition of service reliability. Van Oort (2014), thus recommended taking passenger interest more explicitly into account when setting indicators and objectives of service reliability. Moreover, he demonstrated that an alternative indicator, being **additional travel time** (Van Oort & Van Nes, 2009), represents the level of service reliability in a sustainable way, since it considers factors that are neglected by traditional indicators, such as driving too early and passenger boarding patterns. This indicator translates the supply-side indicators, for instance punctuality, into the additional travel time that a passenger on average needs to travel from the origin to the destination stop due to service variability. The average additional travel time may be calculated per stop or per line and it enables explicit consideration of service reliability in cost-benefit calculations, since the level of service reliability may be translated into regular travel time. The additional travel time was calculated for the tramlines in The Hague. Based on his findings, some recommendations were to improve concession requirements as well as to improve the design of networks and of timetables, both aiming at enhanced service reliability (Van Oort, 2014).

Robustness

The current focus of transportation policy in the context of **major disruptive events** is to adopt an engineering resilience-oriented approach, which focuses on returning assets to good workable order as soon as possible. There is significant adaptive capacity within society that could be better harnessed to reduce the impacts of disruptive events. OECD/International Transport Forum (2016) discussion paper has identified four areas for action to improve resilience planning and investment strategy assessment:

- Development of **Smart Resilience Strategies**: a combination of transport and non-transport responses in order to minimize the impacts of temporary infrastructure loss.
- **Improving the usefulness, impact and co-ordination of communications** with the public and businesses during disruptions, enabling social adaptation and reducing time wasted in unnecessarily risky and extended journeys.
- Further developing the capacity of travellers and businesses to adapt to different events through **greater multi-modality and an increase in smart and flexible working practices**.

- **Understanding the economic impacts of disruptive events** which extends beyond the apparent reductions in flows and increases in journey times observed on the networks and captures the societal and economic impacts in a more holistic way.

As stated in Cats (2016, p. 1), "**investments in transport are increasingly motivated by the need to improve reliability and robustness** and not merely travel time savings under normal operations (Mackie et al., 2014). Transport systems are subject to recurrent disruptions that may have substantial implications for network performance and society at large. On the other hand, disruptions in the public transport service could result from various reasons including mechanical and technical failures, planned maintenance works or targeted attacks. The impact of disruptions on network performance may vary from one link to the other, depending on the interaction between network topology and travel demand. In addition, the consequences in case of a disruption on a given link may change as a result of changes in network topology or operations such as the addition or removal of links or the increase or decrease of link capacity (Cats and Jenelius, 2015b). The relatively low connectivity of public transport systems makes them particularly vulnerable in case of disruptions. Urban rail-bound systems are particularly vulnerable to link failures because of their restricted capability to bypass link closures, and the low network density. However, these systems often comprise of several separate railway systems which could potentially offer redundancy and thus alternative routes in case of disruptions. Even though robustness of critical infrastructures such as mass metropolitan public transport networks is high on the planning and policy agenda (Homeland Security, 2010), there is thus a growing need to develop techniques and indicators to assess the consequences of alternative network developments in terms of robustness".

2.1.3. Methods/indicators in practice

The first step in recognising the importance of reliability is to monitor it. Two distinctive activities are involved: Monitoring service reliability, and setting targets to compare the service provider's actual performance. Most of the existing reliability targets can be found in the rail sector due to their strict working timetables. To the extent that the provider is perceived to be a monopoly, governments usually oversee supply standards by monitoring and setting performance standards; the target provides a degree of accountability in service quality. However, according to OECD/International Transport Forum (2010), there are several **shortcomings** in the reliability indicators currently available:

- **Aggregation across users**. Most existing reliability indicators monitor the performance characteristics of the whole system rather than satisfaction of diverse users' needs.
- **Aggregation across time**. The indicators normally show annual or monthly averages, and therefore mask shorter-term variations in service standards.
- **Design oriented to stakeholders**. Most of the existing indicators were originally designed to provide feedback to network managers, rather than to measure reliability as perceived by end-users.

A number of countries are moving towards incorporating travel time reliability into their cost-benefit analysis methods. The Netherlands, Sweden, and Norway appear to be the furthest along, but the United Kingdom is studying reliability as well. Both Australia and the New Zealand already include travel time reliability in cost-benefit analysis. Japan is just starting to look at methods. Some project appraisals do incorporate reliability, although, the values used are typically assumed the same for all users. To appraise reliability effects in cost-benefit analysis it is important to measure both average travel time and travel time variability. Incorporating reliability requires three sets of data:

- Existing travel time reliability, defined in minutes.
- Anticipated reliability level, e.g. in minutes.

- Monetary values of reliability, disaggregated at the appropriate level of granularity.

In order to be able to take into account service variability in policy impact evaluation, only a cost-benefit assessment framework provides consistency in assessing the societal pros and cons of policy interventions in terms of their positive, or negative, effects on variability. For **policy making, it is important to measure and report on both network operator and user perspectives of variability, as shown in the below figure**:

- For a network provider or operator, the focus is on:

  \- **System robustness/vulnerability**: link and network performance indicators, under changing conditions.

  \- **System operating performance**: indicators to describe the performance of a system in terms of deviations from expected, or agreed, levels of service.

- For a network user, the focus is on:

  \- **Variability of travel times experienced** by the user: indicators to describe issues regarding general variability of travel times, and issues regarding the elimination of extreme, unexpected, travel times.

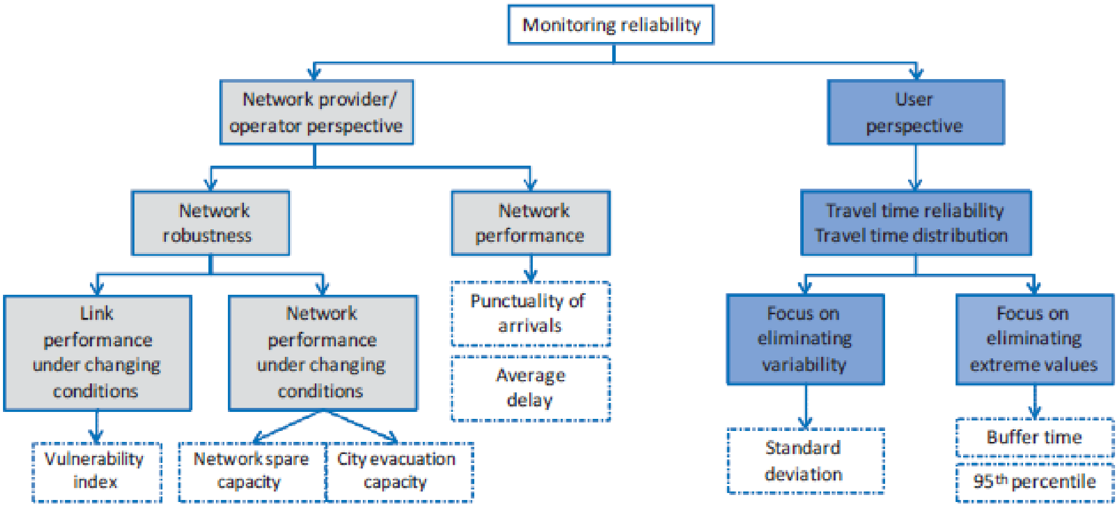


Figure 1. Network and user perspective of reliability (OECD/International Transport Forum, 2010)

Reliability measures are used within performance regimes to evaluate the quality of service of public transport providers. Most of these regimes are based on the percentage of services arriving on time, where the notion of being on time and the penalty structure associated with not adhering to this differences between cities. For instance, the UK rail industry tends to use the Public Performance Measure (PPM) within its incentive regime, with differing tolerance to late running depending on the distance of the total service: lower tolerance for shorter services that are likely to be more frequent (Vincent, 2008).

Today most countries measure reliability and punctuality of their transportation systems on a daily basis. Common to both of these service characteristics is that they are measured after the operation has taken place. Most applied measures of reliability focus on **departure time deviations**: early and late vehicles are treated the same. The departure time deviations may be calculated per line or network, including all or only a few stops. However, from a passenger perspective, late vehicles tend to affect travel times less than early vehicles. With regard to penalising unpunctual vehicles, in 18% of the investigated tender documents, a penalty regime was applied (Van Oort, 2014).

Another way to express schedule adherence is the **percentage of vehicles' schedule deviations within a certain bandwidth**. This method is very common in heavy railways. 74% of public transport systems use a bandwidth to quantify and analyse schedule adherence, while 21% use the average punctuality (the extent to which the scheduled departure times are met). Only Transport for London uses indicators showing the effects of unreliability for passengers, being **excess journey time**. Furthermore, the boundaries of the bandwidth are not uniform among public transport systems. If a broad bandwidth is set, excess variability will be small for instance. These differences in bandwidth obviously have a large impact on the percentage of on-time vehicles (Van Oort, 2014).

Besides differences in indicators and boundaries, locations of measuring service reliability differ among public transport systems. In the Dutch survey, results showed that in 18% of the investigated concessions punctuality was measured at the first stop, in 27% at the last stop and in 23% at the main transfer points (Van Oort, 2014).

Service reliability of high-frequency services is determined by **headway variation**. Improved service regularity yields shorter passengers waiting times, lower levels of experienced crowding, better capacity utilization and higher operational certainty. Based on the results of Cats (2014), the public transport agency in Stockholm requires when tendering the inner-city trunk lines 1–4 to operate these lines based on regularity. Moreover, the bus operator decided to shift to regularity operations even before the official requirement went into effect (Cats, 2014).

Finally, specific examples of service reliability figures around Europe and the USA are mentioned below.

EU
The Dutch Railways used to periodically present the number of trains departed not later than 3 min from 32 main stations in the Netherlands until 2010. Most heavy railway companies in Europe use 5 min as a maximum (Landex & Kaas, 2009; Schittenhelm & Landex, 2009), as the Dutch railways currently does.

In the United Kingdom, the mean-lateness approach was proposed for passenger rail by the Association of Train Operating Companies. It consists of two components: schedule time (the travel time between actual departure time and scheduled arrival time of the train), and the mean-lateness at the destination (the mean of the travel time between scheduled arrival and actual arrival) (Batley and Ibanez, 2009).

USA
For the railway companies in U.S., even 30 min delay is considered being on time (Bush, 2007).

2.2. State of the Art

As stated previously, the term service variability encompasses both notions of reliability and vulnerability. The former is associated with higher probability and lower impact events while the latter pertains to lower probability and more impactful incidents. In the following, commonly used definitions are outlined and discussed.

2.2.1 Definitions and terminology

In the scientific literature, a variety of definitions is used for reliability and vulnerability of road networks and public transport networks. Oliveira et al. (2016) define *reliability* as "the susceptibility of a network to everyday perturbations in supply or demand", whereas *vulnerability* is defined as "the susceptibility to severe disruptions". The distinction between these two concepts is based on the question whether one considers recurrent events or non-recurrent events. *Recurrent events* – related to the concept of reliability – are stochastic fluctuations in system demand and/or supply under normal everyday conditions, in which all network links remain available. In contrast, *non-recurrent events* – related to robustness – are non-random events which reduce capacity on a (group of) network link(s), usually having a larger impact (Oliveira et al. 2014, Mattson and Jenelius 2015). Vulnerability is related to the consequences of disruptions on network functionalities, once these disruptions occur. Some authors prefer using robustness instead of vulnerability (Scott et al. 2006; Sullivan et al. 2010). *Robustness* is considered as the antonym of vulnerability, which also is only related to non-recurrent events: the capability of a system to absorb shocks and withstand disruptions (Snelder et al. 2012; Reggiani et al. 2015). In line with this, Jenelius et al. (2006) and Jenelius and Mattson (2012) define robustness of road networks as "network ability to maintain an acceptable performance after an event impairs one or more links". This definition is also applied for public transport networks by Cats and Jenelius (2015b) and Cats et al. (2016b).

Network *resilience* requires both robustness and a rapid recovery to normal operations in case of non-recurrent events (Barker et al. 2013). Most studies only consider robustness: the impact of non-recurrent events (for example Taylor et al. 2006). However, there are also studies in which exposure to disruptions and the impact of disruptions are integrated to a concept of *risk* for road networks (Jenelius 2006; Dehganisanij 2013) and recently also for public transport networks (Yap et al. 2015; Cats et al. 2016b). Incorporating exposure can be especially of relevance when considering multi-level public transport networks, with usually different link failure probabilities and passenger flows on different network levels. Not considering exposure differences between links can therefore lead to wrong prioritizing of network links of which robustness should be improved from the perspective of the passenger. While robustness and risk are network characteristics, on a link level a similar distinction can be made between link importance (the impact of a disruption on a link), link weakness (the probability of the occurrence of a disruption on a link: exposure) and link criticality (accounting for both exposure to, and impact of disruptions on a link) (Cats et al. 2016b).

There are also studies in which reliability is considered in relation to *connectivity*: the probability that network nodes remain connected and can be reached by passengers (Iida and Wakabayashi 1989; Asakura et al. 2001; Berdica 2002; Kurauchi et al. 2004; Chen et al. 2002). However, given the abovementioned definitions used in other studies, this definition is more relevant in case of non-recurrent events resulting in link unavailability, instead of recurrent events where link availability is not affected. This definition is therefore more related to robustness. Some studies consider reliability and robustness as parts of a common notion, in which robustness is the more extreme subset of reliability in case of severe disruptions (e.g. Chen et al. 2002). Other studies explicitly distinguish between these two concepts, and consider reliability and robustness as separate, non-overlapping concepts (e.g. Taylor et al. 2006; Li 2008).

Based on the aforementioned literature review, in this research the following definitions of reliability and robustness of public transport networks are proposed, based on Cats et al. (2016b) and Oliveira et al. (2016):

Reliability: the ability of a public transport network to withstand recurrent, everyday stochastic fluctuations in supply and/or demand which do not result with link unavailability.

Robustness (with its antonym vulnerability): the ability of a public transport network to withstand non-recurrent events with a partial or full unavailability of one or more links.

Disturbances and disruptions that lead to reduction in system performance could be caused by either planned (e.g. maintenance works) or unplanned (e.g. accident, technical or manual failure) events and by either random (e.g. adverse weather conditions) or targeted (e.g. terror attacks) incidents. The supply- and demand-oriented measures described in the following can be applied to any service layer (e.g. urban, regional or national service network; road, rail, air or waterbone systems). To the best of our knowledge, no such measures have been proposed in the context of transport hub operations.

2.2.2 Supply-oriented measures of reliability and robustness

Traditionally, reliability measures for public transport networks have been strongly supply-oriented, thereby using vehicle-based reliability measures. Mostly, the service characteristics of consideration measure transportation operators' performance rather than passengers' convenience. This may seem counterintuitive since the focus of a well-functioning transportation hub operating system mainly should be on moving passengers from point A to point B as easily as possible according to a predefined schedule. This stems especially from a lack of accurate data of public transport passenger travel patterns in the past, while vehicle performance could be determined objectively and without knowledge about passenger flows and patterns.

Parbo et al. (2013) provide an overview of how the three service characteristics (Robustness, Reliability and Punctuality) have been measured and defined. The way of measuring service characteristics is of great interest because the chosen approach may bias the results obtained when the performance of transportation networks is assessed. It is considered how these measures could be defined to better reflect the provided level-of-service seen from a passenger's perspective. They refer to reliability as time-dependent measure of the performance of public transport network.

Recently, the integration of passenger satisfaction within a utility framework coupled with operational planning of transportation systems has been addressed in several papers especially in the context of railway networks. For instance, Robenek et al. (2016) and Binder et al (2016) have presented the passenger satisfaction by presenting a utility function consisting of "traveling time", "in-vehicle waiting time", and the "waiting time in the platforms" as well as "scheduled delays" (earlier or later).

In the following, the state-of-the-art concerning reliability indicators is discussed.

Reliability indicators on the urban network level

Punctuality

In scientific literature, various vehicle-based measures of reliability can be found in urban public transport. With the availability of Automatic Vehicle Location (AVL) data in recent years, it is possible to determine these indicators based on all realized trips. An often used category of indicators is *punctuality*. Punctuality P_l of public transport line l can be defined in terms of the average deviation of the actual departure (or arrival) time d^a from the scheduled departure (or arrival) time d^s of the number of trips n_i of line l measured at one, multiple or all stops j (see formula 1).

$$P_l = \frac{\sum_j^{n_{l,j}} \sum_i^{n_{l,i}} |d_{l,i,j}^a - d_{l,i,j}^s|}{n_{l,j} * n_{l,i}} \quad (1)$$

Punctuality can also be measured as the percentage p_l of vehicles of line l of which d^a is realized within a certain bandwidth between a lower bound schedule deviation δ^{min} and upper bound schedule deviation δ^{max} from d^s, measured for all stops j of line l (see formula 2). The latter indicator is typically coined on-time performance.

$$P_l = \frac{\sum_j^{n_{l,j}} \sum_i^{n_{l,i}} p_{l,i,j}\left(\delta^{min} < d_{l,i,j}^a - d_{l,i,j}^s < \delta^{max}\right)}{n_{l,j} * n_{l,i}} \quad (2)$$

An important disadvantage of the abovementioned punctuality indicators is that a distinction between early and late departures is lacking, whereas the impact of an early or a late departure on passengers is different (Van Oort 2011). An early departure increases passenger waiting time for passengers arriving at stop j at $t_{a,j}$ between d^a and d^s with $\left(d^s - t_{a,j}\right) + h^s$, where h^s is the planned service headway. This means that the waiting time increases with more than one headway h^s (assuming that the next vehicle operates according to schedule). For passengers with $t_{a,j} < d^a$, the waiting time is reduced by $d^s - d^a$ in case of early departures. On the other hand, a late departure increases the waiting time for all passengers by $d^a - d^s$. These different passenger impacts are not distinguished when applying the vehicle-punctuality measures from formulas (1) and (2).

Regularity

There are also several studies in which vehicle-based service *regularity* measures are used for high-frequency service reliability, instead of or besides punctuality measures (Cats 2014). In some studies, service regularity is directly based on statistical measures. For example, the standard deviation or percentiles of headways between vehicles i on line l can be used (Van Oort 2011). Often, the coefficient of variation CoV of actual headways – the standard deviation of actual vehicle headways h^a between successive vehicles divided by the expected vehicle headway of line l at stop j – is used as indicator for reliability (see formula 3) (Asakura and Kashiwadani 1991; Bates et al. 2001; Borjesson et al. 2002; Clark and Watling 2005; Cham and Wilson 2006; Engelson and Fosgerau 2011).

$$CoV\left(h_{l,j}^a\right) = \frac{\sigma\left(h_{l,j}^a\right)}{E\left(h_{l,j}^a\right)} \quad (3)$$

Hakkesteegt and Muller (1981) use the Percentage Regularity Deviation Mean (PRDM) at a certain stop j served by vehicles i from line l as measure for service regularity, expressing the average deviation of the actual headway h^a over all vehicles i from the scheduled headwayh^s as percentage of h^s (see formula 4).

$$PRDM_{j,l} = \frac{\sum_i \frac{h_{l,i}^s - h_{l,i,j}^a}{h_{l,i}^s}}{n_{l,j}} \quad (4)$$

All metrics as discussed until now depend merely on vehicle-related metrics, without explicitly considering the impact of service variability on passengers. In case of a delay of a vehicle i on high-frequency line l where the passenger arrival pattern at each stop j can be considered random, regularity measures can be advantageous compared to punctuality measures. In case of a delayed vehicle, the average waiting time and average crowding level will be lower in case of equal headways between successive vehicles (headway regularity), compared to a focus on punctuality (thus with unequal headways in case of a delay). Applying vehicle punctuality as the governing, monitoring and operation criterion in case of recurrent, stochastic supply fluctuations, may lead to counterproductive results because passengers will be distributed unevenly over vehicles. This means that a relatively large group of passengers will experience longer waiting times and busier vehicles, whereas a relatively small group of passengers will experience shorter waiting times and less crowded vehicles. This effect will start oscillating when the vehicle dwell time is also affected due to varying passenger loads at stops, often indicated as 'bunching'. When a regularity measure would be applied, passengers will be distributed more equally over successive vehicles, which means that average waiting times and crowding levels will be reduced, and that required dwell times will be more similar for successive vehicles (Van Oort 2011, Cats 2014).

Capacity and connectivity reliability

Other reliability measures found in literature are related to residual capacity or connectivity reliability. Chen et al. (2002) use the available network capacity to accommodate demand as indicator for capacity reliability. This measure can also be important in case of non-recurrent events in relation to robustness. Connectivity reliability is often quantified by using the number of trips which could be performed over a longer period of time as indicator (Bell and Iida 1997; Al Deek and Ben Emam 2006). To scale reliability measures to another aggregation level, Eliasson (2007) and Engelson and Fosgerau (2011) state that reliability can additively be calculated, meaning that the network reliability equals the sum of the link reliability of all network links.

Reliability indicators on the interregional train network level

Vehicle punctuality

On the interregional train network, punctuality is often used as the primary reliability indicator. This is usually measured as the percentage of trains departing or arriving with a delay of a certain maximal number of minutes at a set of predefined stations. In the Netherlands and several other European countries, a threshold of 5 minutes is used hereby (Vromans 2005). Because early departures are very rare on the train network, the problem of this indicator to distinguish between early and late departures on the urban public transport network, is not of relevance for train services. However, similarly to the on-time performance indicator expressed in formula 2, the disadvantage of this indicator is that a delay of 6 minutes is equally weighted as a delay of 20 minutes, whereas the passenger impact can differ substantially between these cases. Therefore, it might be better to consider instead the mean delay of trains. However, a problem with this measure in its turn is that long delays (i.e. the tail of the distribution) can influence this indicator quite heavily. Another disadvantage of the punctuality measure is that cancelled train services are not discarded (Vromans 2005). Therefore, this punctuality indicator could stimulate a perverse response from an operator by cancelling train services with (heavy) delays. It is therefore of relevance to use the number or percentages of cancelled train services as indicator of connectivity reliability as well, next to punctuality.

Passenger-weighted vehicle punctuality

The abovementioned reliability measures for the train network do not consider the impact of delayed / cancelled train services on passengers explicitly. These vehicle-based metrics are used as an approximation of passenger impacts. However, one should be aware that especially the impact on transferring passengers is not captured by these indicators. When an operator focuses on improving vehicle punctuality, this means that synchronized transfers between two train services can be lost even in case of a small delay of one of these services. Although vehicle punctuality might be improved, this delays transferring passengers by one headway. Adding the relatively high disutility as perceived by passengers for waiting time compared to in-vehicle time, there can be a substantial misalignment between this vehicle-based metric and experienced passenger delays, especially in case of transfers to low-frequent train services (Balcombe et al. 2004). Therefore, the percentage of planned train transfers which is maintained during operations is sometimes used as indicator related to transfer reliability (Vromans 2005). One should however realize that all transfers, at all locations and at all times of the day, are weighted equally when applying this measure. Maintaining a very busy transfer in the morning peak is equally weighted as a transfer which is only used by a few passengers during a late evening. On the train network level, the vehicle arrival punctuality can be adjusted to better approximate passenger impacts. For example, the weighted train arrival punctuality (in which the train arrival punctuality is weighted according to the expected number of passengers in each train) can be used (Vromans 2005). Other reliability indicators developed for the train network are the mean difference between realized and scheduled train arrival times, the standard deviation of train arrival times, and the adjusted standard deviation of arrival times, in which early arrivals are not incorporated (Rietveld et al. 2001). In general, most indicators for the train network relate to schedule adherence: the relation between scheduled and realized train travel times. Regularity-focused indicators are hardly applied to the train network.

Vulnerability / robustness indicators

In scientific literature, different service-oriented vulnerability indicators can be found. For complex real life networks, it is often computationally demanding to quantify link importance or link *criticality* for all network links. Therefore, often an indicator is used as proxy to determine a subset of most critical links, for which criticality can then be quantified. Especially in this phase of identifying the most important or critical links, the identification process can be based on network/supply characteristics instead of passenger characteristics. For example, topological network centrality indicators such as node degree (the number of links incident to a node) or node betweenness centrality (the number of shortest paths between all nodes passing a specific node) are used in the identification phase by Angeloudis and Fisk (2006), Criado et al. (2007), Von Ferber et al. (2009), Derrible and Kennedy (2010) and Von Ferber et al. (2012).

Studies concerning the measurement of vulnerability of public transport networks are limited. However, for road networks many studies on this topic are already performed. Therefore, we also incorporate vulnerability measures developed for road networks. Vulnerability is often measured by the difference in network performance between the undisturbed situation, and the situation after removal of a link (Jenelius et al. 2006; Scott et al. 2006; Taylor et al. 2006). This approach is extended by Knoop et al. (2008), Sullivan et al. (2010) and Snelder et al. (2012), by also considering partial link unavailability next to full link unavailability. Different studies use different assignment models to determine the difference in network performance. A dynamic assignment model is for example be used by Knoop et al. (2008) and Snelder et al. (2012), whereas others use a static equilibrium assignment model, or a mixture between static and dynamic assignment models (Tampere et al. 2007). Dynamic models are in general more detailed, but computationally demanding for complex, large, real networks. The Network Robustness Index (NRI) expresses how critical a link is according to its importance to network robustness, by calculating the difference in link travel time (or costs) between the undisturbed scenario and the scenario after removal of a link.

Most studies on public transport network vulnerability focus on topological indicators and how the degradation of physical links in a specific sub-network - the metro network - affects network connectivity. Using graph theory measures of link importance, previous studies investigated the impact of random and targeted attacks on network vulnerability for the world largest metro systems (Angeloudis and Fisk 2006), 32 metro systems worldwide (Derrible and Kennedy 2010), Shanghai (Zhang et al. 2011), London and Paris (von Ferber et al. 2012) and Nanjing (Deng et al. 2013) analysed network vulnerability in strictly topological terms, implying uniform link labels and attributing equal importance to connections between each origin-destination. However, the impact of link failure depends not only on the availability of travel connection alternatives, but also on their attractiveness and the number of travellers that relied on the disrupted link. Indeed, Dupuy (2013) argues that by overlooking key network characteristics as well as the urban planning context, studies performed by scientists from other disciplines in the field of network geometry and urban railway systems provide very limited recommendations to network planners and thus obstruct potential implementations.

2.2.3 Passenger-oriented measures of reliability and robustness

Reliability indicators at the urban and interregional network level

Route travel time variation

Passenger-focused measures of reliability attract an increasing attention in the scientific literature. First, a literature overview is given for indicators used for urban public transport. From a passenger perspective, one may consider statistics of the total route travel time in these indicators. For example, the standard deviation of route travel time, coefficient of variation of route travel time, and the difference between the 80th or 90th and 50th percentile of route travel time are mentioned for recurrent delays (Turnquist and Bowman 1980; Rietveld et al. 2001; Tseng 2008; Vincent and Hamilton 2008). However, to this end, very detailed data is required regarding OD-relations in time and space, which was not available until recent years. This made it difficult to adequately quantify these indicators. Earlier, only the number of boardings and alightings and line occupancy were known, based on which passenger-oriented reliability measures were therefore developed (Van Oort 2011).

Additional passenger travel time

Van Oort and Van Nes (2004; 2006) developed the average additional travel time for passengers on a line (leg) level as an indicator of unreliability. The additional passenger waiting time is the component of the total route travel time which is mostly influenced due to unreliability. In case of bunching, the average additional passenger in-vehicle time also increases in case of unreliability. In case of random passenger arrivals at a stop, the average additional waiting time $E\left(t_{l,j}^{add,wait}\right)$ at stop j of line l can be calculated using the coefficient of variation CoV of realized headways h^a (formula 5).

$$E\left(t_{l,j}^{add,wait}\right) = \frac{E(h_{l,j}^a)}{2} * \left(CoV^2(h_{l,j}^a)\right) \tag{5}$$

In case of low-frequency public transport services in which passengers arrive according to schedule, $E\left(t_{l,j}^{add,wait}\right)$ is calculated according to formula 6 based on the assumption that passengers arrive randomly in a bandwidth $\left[\tau_{early}, \tau_{late}\right]$ around the scheduled departure time.

$$E\left(t_{l,j}^{add,wait}\right) = \frac{\sum_i E\left(t_{l,i,j}^{add,wait}\right)}{n_{l,i}} \tag{6}$$

Where

$$t_{l,i,j}^{add,wait} = \begin{cases} h_l^s & \delta_{l,i,j}^d \leq -\tau_{early} \\ 0 & -\tau_{early} < \delta_{l,i,j}^d < \tau_{late} \\ \delta_{l,i,j}^d & \delta_{l,i,j}^d \geq \tau_{late} \end{cases}$$

Contrary to vehicle-based punctuality measures, this measure explicitly distinguishes between the situation in which there is an early or a late departure $\delta_{l,i,j}^d$ of vehicle i of line l from stop j. A disadvantage of this measure is that it only focuses on a single-leg trip, in which the impact of transfers on the expected additional travel time are not considered. Lee et al. (2014) extended the passenger perspective of reliability by considering service reliability of multi-leg journeys including transfers.

Predictability of passenger travel time

In other studies, developed reliability measures are related to *predictability*. Uniman et al. (2010) use the Reliability Buffer Time (RBT): the buffer time passengers should allocate to arrive in time at their destination with a defined uncertainty. The difference between the 50' percentile travel time and 95' percentile travel time t^{95} is often considered as acceptable for recurrent delays. For non-recurrent delays, the difference between the 95' percentile and nominal travel time is often used, defined as the Excess Reliability Buffer Time (ERBT). Other studies use passenger-oriented reliability measures related to punctuality or schedule adherence. The Excess Journey Time (EJT) is defined as the difference between the experienced and scheduled passenger arrival time (Hendren et al. 2015). Also for the train network, the average passenger delay can be used as measure for unreliability (Vromans 2005). Ideally, for all these measures again the total journey travel time is used. However, this requires detailed passenger origin-destination and transfer information. Due to the shortage of passenger-related data until recent years, those measures were often calculated on a single-leg level based on passenger waiting time and in-vehicle time only and often derived and approximated based on vehicle-based metrics. Formulas 7 and 8 show the calculation of the RBT due to recurrent variability in waiting time $RBT_{l,j}^{wait}$ and in-vehicle time $RBT_{l,j}^{-veh}$ respectively at stop j of line l.

$$RBT_{l,j}^{wait} = t_{l,j}^{wait,95} - E\left(t_{l,j}^{wait}\right) \quad (7)$$

$$RBT_{l,j}^{-veh} = t_{l,j}^{-veh,95} - E\left(t_{l,j}^{-veh}\right) \quad (8)$$

The impact of service variability and capacity constraints was modelled by Cats et al. (2016a) in a dynamic and stochastic assignment model, The model is capable of assessing the costs of congestion (denied-boarding; longer dwell times) and crowding in case of recurrent as well as non-recurrent events.

Journey excess time

For urban public transport in The Netherlands, it is required for passengers to tap-in and tap-out in each vehicle, also when making a transfer. This means that the total OD-relation (on a stop-to-stop level), including trip legs and transfer locations, can directly be derived from the smartcard transactions. For the train network, tap-in and tap-out need to be performed at the stations. At this point, it is not necessary to tap-out and tap-in again when making a transfer between two trains of the same operator. When transferring between trains of different operators, it is however necessary to tap-out from the upstream train operator and tap-in for the downstream train operator. For train travelling, the total OD-relations (station-to-station level) are available from smartcard data. However, the chosen route, including possible transfer stations, cannot be derived directly from the transactions. With the availability of data from AFC systems, smartcard data is applied in several public transport studies ranging from strategic to operational applications (e.g. Paul 2014; Sun 2015; Van Oort et al. 2015).

This also allows the further development of passenger-focused reliability measures. For example, since in the Netherlands the total OD-matrix can be inferred based on smartcard data, it is possible to quantify already developed reliability measures like travel time variability, the reliability buffer time or excess journey time based on the total route travel times (Uniman et al. 2010; Wood 2015). The excess journey time on a route level was applied to the London Overground network based on smartcard data (Zhao et al. 2013; Hendren et al. 2015). Bagherian et al. (2016) developed two measures for passenger-oriented urban public transport reliability, based on smartcard data. These measures are fully based on smartcard data and therefore consider the total passenger OD-relations including transfers on the urban public transport network level. The measures can directly be quantified based on realized smartcard transactions. A tool is developed where smartcard transactions can directly be imported, based on which both measures are directly quantified for OD-pairs i, j of (part of) the public transport network during time period τ. One measure is related to passenger travel time predictability: the daily variability measure $DV_{i,j}(\tau)$ is the ratio between an upper percentile travel time $t^{upp}$and the travel time in normal conditions t^n (see formula 9). For example, the 95th and 50th percentile could be used as upper and typical percentile travel time (Uniman et al. 2010; Wood 2015).

$$DV_{i,j}(\tau) = \frac{t_{i,j}^{upp} - t_{i,j}^{n}}{t_{i,j}^{n}} \qquad (9)$$

The second measure focuses on punctuality and schedule adherence. The Schedule Deviation measure $SD_{i,j}(\tau)$ reflects the ratio of the excess delay of the actual travel time t^a to the scheduled travel time t^sat a certain percentile (see formula 10).

$$SD_{i,j}(\tau) = \frac{\left(t_{i,j}^{a} - t_{i,j}^{s}\right)^{upp}}{\left(t_{i,j}^{a}\right)^{upp}} \qquad (10)$$

Both measures can be scaled and expressed at the stop, line or network level. A limitation of these developed measures is that robustness is in fact considered as subset of reliability, and cannot be quantified as a separate concept. Besides, only nominal travel times are used here: travel time perception (for example in relation to waiting, transferring or crowding) is not incorporated. Although these measures consider the total passenger trip on the urban public transport network, interactions with other network levels are not considered. This means that these measures do not consider passenger reliability on the entire passenger journey if it extends beyond the urban level.

Vulnerability / robustness indicators

Studies pertaining to passenger-focused vulnerability measures for public transport networks are still limited. In the identification phase of important links, Cats and Jenelius (2014) use the passenger betweenness centrality (passenger flow on a link). When identifying critical links Cats et al. (2016b) used the expected passenger-exposure (the product of the expected link exposure and passenger betweenness centrality) as indicator for link criticality, thereby incorporating exposure to disruptions as well. The advantage of this last measure is that especially in multi-level public transport networks, differences in exposure to disruptions on different network levels are incorporated in the identification process of critical links. The applicability of this measure is however dependent on the availability of historical disruption data, which is not universally available.

When quantifying vulnerability of public transport networks, the difference in total travel time (over all OD-pairs and all passengers) between the disrupted and undisrupted scenario is used as passenger-oriented indicator for the impact of a non-recurrent event (robustness / link importance). This measure is extended in several ways. For example, Cats et al. (2016b) use the probability weighted average impact, in which link failure probabilities are incorporated next to disruption impact (risk / link criticality). The impact of disruptions might be larger than the direct time losses. Borjesson et al. (2012) use a delay time multiplier, to express the highly negatively perceived waiting time by passengers in case of denied boarding during disruptions. Incorporating the effect of real-time information during disruptions on additional passenger travel time was performed by Cats and Jenelius (2014). Going beyond the analysis of full link breakdowns, Cats and Jenelius (2015a) distinguished between partial and complete link unavailability for public transport networks by considering partial capacity reductions.

2.3. Requirements

This section proposes criteria for developing service variability measures, and discusses data requirements and scalability, in a way that supports the subsequent developments of indicators at the three different levels in Section 3. The aim is to lay down the requirements for developing useful, relevant and meaningful indicators for transit authorities and operators across Europe.

The ability of transit operators to understand and improve reliability relies on their ability to measure it. Until recently, efforts to quantify this attribute of service from the perspective of passengers were limited by the small sample sizes obtained from manual surveys, or the use of supply-side data to indirectly capture the passenger experience. With the emergence of data from automated fare collection, it becomes possible under certain conditions to directly observe travel times experienced by passengers and obtain improved estimates of the reliability of transit service. More specifically, the proliferation of smartcards has opened up a range of applications benefiting from the high resolution of this source of data, capturing information at the individual passenger level over time. In this context, measuring service variability can be an effective management tool for assessing system performance using simple figures that have straightforward interpretations (Chan, 2007). One important application of this source of data is in the area of service quality monitoring, where smartcards have the inherent advantage over traditional methods based on vehicle-location data of directly capturing the performance of the system as experienced by passengers. (Uniman, 2009)

2.3.1 Guidelines for Developing Measures of Variability

The fundamental principle is to classify performance into recurring and non-recurring system conditions, so that both reliability and robustness aspects of the service will be measured separately. Moreover, this distinction can be used to gain insight into the causes of unreliability.

In Uniman (2009) five general criteria are used to develop a set of reliability measures. These criteria balance the need to accurately represent the passenger experience with the desire to produce straightforward measures that can be applied by transit agencies as part of their routine monitoring of performance. These criteria build on the work by Chan (2007), and consist of:

- Representative of the Passenger Perspective – One of the primary drawbacks of existing measures of reliability is their focus on operational quality as opposed to the passenger experience. An appropriate set of reliability measures should take into account the effects of travel time variability on passengers by focusing on extreme values (as opposed to average performance), and should be sensitive to the way passengers build their expectations of service. The measure should also be representative of the experience of most passengers, and not be overly susceptible to the effects of individual traveller behaviour or day-specific events that could skew the characterization of performance.
- Straightforward Estimation and Interpretation – A successful measure should balance between technical/mathematical complexity and ease of use and implementation. That is, the results should be easily interpreted not only by analysts but also by decision-makers and passengers alike. On the other hand, the reliability measure should not be so simple as to lose its value for analysts and transit operators. Data availability and estimation costs should also be factored in as part of the degree to which estimation of a particular measure is straightforward, and the likelihood that it is implementable from the perspective of the transit agency.
- Comparability and Aggregation of Results – It is important that the output of any particular measure be comparable across different spatial and temporal levels. The ability to estimate results for particular portions of the system is useful for both comparative analyses, and the aggregation of results into higher spatial units. The ability to estimate performance across smaller time intervals allows for different types of analysis and makes the aggregation of results into time periods appropriate for reporting performance.
- Useful as input to the Evaluation of Performance – There are three aspects that should be covered by the set of reliability measures in order for them to be useful in the context of monitoring and evaluation of performance. First, the information provided by the measure must be useful in setting specific performance goals (i.e. setting a standard for performance) and evaluating progress made towards them (i.e. determining the proportion of passengers receiving good and bad service). The second aspect is precision and accuracy. Namely, the results must be both repeatable and take into account sampling errors. The third aspect that is important to keep in mind is the compatibility of the proposed measures with the existing service quality measures in place at the transit agency. The proposed measures should be sensitive to the characteristics of current measures (e.g. units of measurement) and resource availability (e.g. staff and data related constraints) already in place at the transit agency.
- Provides Insight into Causes of Unreliability – The measure should ideally help analysts identify and quantify the contribution of different causes of unreliability. Through meaningful feedback obtained from the measures of reliability strategies can be selected to target the specific causes of unreliability and improve performance.

2.3.2 Data requirements for the Measures of Variability

The use of passenger-oriented indicators of service variability is inherently based on the exploitation of available data from automated fare collection and vehicle location traces. The use of large continuous streams of data is thus required to support analyses during short time periods, which subsequently need to be stored. Therefore, big data techniques will be required to process, filter and analyse detailed travel demand data. The data requirements and formats for travel demand analysis will be defined in WP1 Unravelling Form and Flow Dynamics, and more concretely in Task 1.1 "Big data analytics for travel demand."

Although the main input data required for the implementation of service variability measures will come from the automated fare collection, there are other key sources of data that could be used for the calculation of the passenger-focused indicators of service variability (defined for each spatial level in the next section of this deliverable).

The sources of data will be used in each level depending on their availability, namely:

- Data on the planned services of public transport connecting the nodes/stations where it is of interest to apply the model. This data is available in GTFS format.
- Continuous data on public transport operations and disruption logfile
- Raw mobile phone data (aggregated data)
- Pedestrian counts in hubs
- Public transport smartcard transactions
- Public transport ticketing data
- Public transport vehicle positions (GPS)
- Passenger survey data

2.3.3 Scalability for the Measures of Variability

Considering TRANS-FORM's overall approach, the methodology for variability measurement should be very flexible, since the indicators should be formulated and applied to the regional, urban and hub levels of public transport networks. In addition, several time periods will be taken into account to perform the variability analysis. The defined methodology should thus be scalable with respect to both temporal and spatial dimensions.

Spatially, the ability to estimate results for particular portions of the system that are meaningful when placed side-by-side is useful for both comparative analyses, and the aggregation of results into higher spatial units (e.g. moving from the O-D pair level to line level estimates of performance). Regarding temporal scalability, the ability to estimate performance across smaller time intervals allows for different types of analysis (e.g. study of crowding within the morning peak), and makes the aggregation of results into time periods appropriate for reporting performance (e.g. 3-hour morning peak period).

3. Passenger-oriented Service Variability Indicators

The following section discusses passenger-oriented indicators of service variability from the perspective of a transportation hub, as well as from the urban and regional level.

3.1. Hub Level

The existing methodologies of measuring service characteristics are evaluated in terms of their ability to reflect the actual (provided) level-of-service from a passenger's perspective at the hub level. When considering the variability and reliability of the performance of a hub, many different metrics and indicators of different complexity can be defined. The challenge lies in the identification of meaningful indicators which can be consistently evaluated and for which data is available. The following hub-level indicators are put forward:

Travel time: The time pedestrians take to perform the origin-destination trip inside the hub is maybe the most critical indicator which can be identified. As an individual plans ahead the time required for reaching the platform from the entrance of the hub (for example), the variability of travel time is critical in assessing the performance of a hub. The distribution of travel times can change depending on different factors like congestion, construction work or even the users themselves.

Travel distance: Similarly to travel time, the distance travelled by pedestrians inside the hub can provide valuable information on the variability of the system. If many individuals dwell in the highly used areas of the hub, then the passengers must extend their walking distance to avoid these “obstacles”. Likewise, many different situations can make a user divert from his ideal path to his objective, therefore inducing variability in his trip.

Pedestrian mean velocity: Combining both previous quantities can be done to calculate mean velocity. The simple division of travel distance by travel time is defined as the pedestrian mean velocity. Pedestrians can be classified into different groups based on variables like age or trip purpose and these different behaviours will imply different free flow velocities.

Transfer reliability: Missing a connecting train often leads to long travel time extensions. The transfer time between two platforms is critical when considering this aspect and any extra walking time between platforms can be problematic for pedestrians with short connection times. The probability to miss a connection is defined as follows:

Arrival time of feeder train: t_r^a
Departure time of connecting train:$t_{r'}^d$
Transfer time between feeder train and connecting train: $t_{sch.transfer} = t_{r'}^d - t_r^a$
Time of transfer for the passenger: $t_{transfer}$
Travel time distribution (distribution of $t_{transfer}$) :f
Probability of missed connection: $\Pr[t_{transfer} > t_{sch.transfer}] = 1 - F(t_{sch.transfer})$

This probability of missing the connection can also be seen as the number of people who will miss the connection out of the number of people who wish to take the connecting train. This approach allows a convenient measure of the global transfer reliability of the hub: the mean of this measure can be considered as the average number of people who will miss their connection in the considered time interval. A second approach to measure the transfer reliability of the hub is through the distribution of the probabilities to miss the connection. After excluding the cases where the connection time is significantly larger than the walking transfer time, such a distribution will represent the overall situation of the hub and the large (hence problematic) probability of missing a connection can be identified.

By using the travel time, travel distance and velocity, insight about daily or hourly variations of level-of-service inside the hub can be acquired. Moreover, bottlenecks can be identified for a specific geographical place and a specific time inside the hub. This information can help planners to come up with better resolutions in case of a disruption or disturbances which may cause problems in the hub.

Naturally, these indicators depend on the availability of the data. In some cases, all the indicators discussed previously cannot be evaluated as sufficient information is lacking.

3.2. Urban Level

This section defines passenger oriented service reliability indicators for the urban public transport network level. Indicators are defined for each origin-destination (OD) pair, for each origin, for each destination and each hub, respectively. The newly developed indicators can be considered as extension of the journey excess time indicators proposed by Uniman et al. (2010) and the journey-based predictability indicator and schedule adherence indicator proposed by Bagherian et al. (2016).

The first indicator applied measures passenger travel time predictability: the *daily variability* measure $DV_{i,j}(\tau)$ is the ratio between an upper percentile travel time t^{upp} and the travel time in normal conditions t^{n} (see formula 11). The 95th and 50th percentile are used as upper and typical percentile travel time (Uniman et al. 2010; Wood 2015).

$$DV_{i,j}(\tau) = \frac{t_{i,j}^{upp} - t_{i,j}^{n}}{t_{i,j}^{n}} \qquad (11)$$

The second journey-level indicator we apply in this study is the *perceived journey access time* $T_{p,acc}^{od}$ as passenger-oriented indicator for urban network reliability (formula 12). The scheduled perceived journey travel time $T_{p,sch}^{od}$ is compared to the realized perceived journey travel time $T_{p,rel}^{od}$. To make the indicator independent from the journey length, the indicator is expressed as a ratio of the perceived journey excess time compared to the scheduled perceived journey travel time. In line with Bagherian et al. (2016), this indicator compares the realized travel time with the scheduled travel time, instead of the expected travel time (50' percentile travel time) often applied in other studies. Using the scheduled travel time as reference leads to an objective reference travel time, independent from the reliability performance of the public transport network. This makes the indicator more easily applicable by public transport operators and authorities, and also eases the communication and interpretation of the indicator. In line with the criteria for service reliability measures as proposed by Uniman et al. (2010), an indicator should be easy to understand and interpret. The indicator differs from the work of Bagherian et al. (2016), since not only the nominal travel times are incorporated. This is because disruptions can have severe impacts on travel time perception as well. For example, bunching effects caused by unreliability on a high frequency line might have limited effects on the waiting times, but can substantially increase the average crowding level and, therefore, the perceived in-vehicle time. During incidents, crowding levels on alternative routes often increase even more. Not incorporating the perception of passengers to crowding and the uncertainty related to unreliability or non-robustness can underestimate the passenger impact of service variability, based on which we emphasis to incorporate travel time perception in the developed indicators.

$$T_{p,acc}^{od} = \frac{T_{p,rel}^{od} - T_{p,sch}^{od}}{T_{p,sch}^{od}} \qquad (12)$$

Formula (13) shows the generic calculation of the perceived travel time. Scheduled travel times are determined using GTFS data, given the chosen path in case of reliability. The difference between realized and expected crowding levels is determined by comparing the hourly expected crowding level per line segment (based on long-term historical data) and the realized crowding level, correcting this by applying a crowding multiplier to the nominal in-vehicle time.

$$T_{p}^{k\in od} = \alpha_{walk} * T_{walk}^{k\in od} + \alpha_{wait} * T_{wait}^{k\in od} + \sum_{s_{l,1}-s_{l,2}}^{s_{l,n-1}-s_{l,n}} \alpha_{crowding} * T_{ivt}^{s-s} + \alpha_{transfer} * N_{transfer}^{k\in od} \qquad (13)$$

The perceived journey excess time measure, as defined by formula (12), can also be aggregated in space and time. Instead of measuring this for each OD-pair, one can also quantify reliability in a similar manner for a specific origin (apply the formula to all destination conditionally to this origin) or a specific destination (apply the formula from all origins conditionally to this destination). By aggregating over both all origins and all destinations, a measure for network wide reliability can be obtained. Also a temporal aggregation is possible, for example when only a specific period of the day is incorporated. Since the scheduled perceived travel time can be quantified based on GTFS data for each period of the day, different time periods can easily be aggregated.

Next to the developed measure for perceived journey excess time to measure reliability on an OD-level (per OD-pair, or aggregated over a group of origins and/or destinations), it is important to explicitly consider reliability of hubs. This becomes even more relevant when considering multi-level public transport networks, with hubs as transfer points interconnecting the different network levels. Differences in hub reliability might affect passenger route choice. Hub reliability is considered in this section in the context of a multi-leg public transport trip rather than the pedestrian perspective taken in the previous section.

Based on existing indicators for trip or journey reliability, we develop two new indicators for hub reliability. A first indicator is the *percentage of passengers missing a connection*, Q_{mc} , at the considered hub (Equation 14). Based on AVL data we can determine for each combination of arriving and departing trips $r \in R$ of line $l \in L$ whether a connection could be made, comparing the minimum required transfer time $\tilde{t}_t$ and the available transfer time based on the realized arrival and departure time $\tilde{t}_a$ and $\tilde{t}_d$. Based on the transfer volume $q_{t,r_{l1}-r_{l2}}$, inferred from AFC data, we quantify the number of transferring passengers which missed their connection. By aggregating this over all trips and transfer combinations, we compare hub reliability between hubs and between different transfer combinations. The second hub reliability indicator is an extension of the previous indicator. This indicator (Equation 15) considers the *perceived journey excess time due to a missed connection at the hub*, $PJET_{mc}$. In case of a missed connection, we compare the resulting actual perceived travel time for the total passenger journey $\widetilde{T_p}$ with the scheduled perceived journey travel time $\overline{T_p}$ in case the connection would be made, as ratio to the scheduled perceived travel time. This allows incorporating the impact of a missed connection on the total remainder of the journey. For example, missing a high volume transfer connection to a high-frequent service might have a smaller net impact on the perceived journey time, compared to missing a lower volume connection to a low-frequent service.

$$Q_{mc} = \frac{\sum_l^L \sum_r^R q_{t,r_{l1}-r_{l2}} * MC_{q_{r_{l1}-r_{l2}}}}{\sum_l^L \sum_r^R q_{t,r_{l1}-r_{l2}}} \quad \forall\, s_{hub} \in S_{hub} \tag{14}$$

$$PJET_{mc} = \frac{\sum_l^L \sum_r^R q_{t,r_{l1}-r_{l2}} * MC_{q_{r_{l1}-r_{l2}}} * (\widetilde{T_p} - \overline{T_p})}{\sum_l^L \sum_r^R q_{t,r_{l1}-r_{l2}} * \overline{T_p}} \quad \forall\, s_{hub} \in S_{hub} \tag{15}$$

$$\text{with MC} \begin{cases} 1 \; if \; \tilde{t}_{drl_2} - \tilde{t}_{arl_1} < \tilde{t}_{t,rl_1-rl_2} \\ 0 \; if \; \tilde{t}_{drl_2} - \tilde{t}_{arl_1} \geq \tilde{t}_{t,rl_1-rl_2} \end{cases}$$

3.3. Regional Level

The focus of this section is on the variability in passenger travel time in the context of public transport at the regional level. Other types of variability related to the passenger experience, like variability in quality of the traffic information or in the travel comfort, is not discussed.

The variability in regional level train and bus services has some direct effects on the travel times perceived by the travellers, but also affects the possibility to make, or break, connections necessary for the traveller to get to the destination. The variability, in the case of a cyclic timetable, and typically with a high frequency, is related to the headway regularity of the service. In this case, the regularity is normally measured as the standard deviation of the headway of the service (see D1.2, Section 2.2.2).

The variability is also affected by the timetable, specifically by the robustness of the timetable. A timetable is said to be robust if it the timetable remains feasible even when delays occur. Feasibility of a timetable states that the train paths do not conflict with each other, and that the trains do not need a capacity larger than allocated. For timetables also the notions of stability (the ability to absorb initial and primary delays), resilience (the flexibility of the timetable to prevent and reduce secondary delays by re-timing, re-ordering and re-routing) are related to the concept of robustness.

The perceived reliability of the service is related to the variability. A route with consistent travel time, close to the times stated by the timetable, is perceived as reliable, and a route where the travel times vary more is considered less reliable. The reliability for one individual leg is normally measured by the average delay for a leg, and the width of the distribution, describing the variability of the travel time at the stop.

Two kinds of travel time variability can be identified. First, variations that occur frequently, and are more or less predictable. This type of variability has limited effect on the perceived reliability. Second, variability that stems from disruptions in the service. Such disruptions may be caused by, for example, incidents or bad weather. The variability of the travel time for a specific leg in the train or bus service is normally computed based on a set of differences between the travel time compared to the travel time stated by the timetable for the service.

The punctuality of a service is the difference in the arrival time at stops, compared to the stated times at the same stops in the timetable. Mathematically, punctuality P_l of public transport line l can be defined as in (1) (see Section 2) or, with the definition that a train is "on-time" if the delay (or, punctuality) compared to the timetable is less than five minutes as in (2), see Section 2. An important disadvantage of the abovementioned punctuality indicators is that a distinction between early and late departures is lacking, whereas the impact of an early or a late departure on passengers is different (Van Oort 2011).

A small variability does not necessarily imply a good punctuality, but it is often the case in practice that a large variability implies a low punctuality. The punctuality is often used to monitor the performance of a train service. This performance measure can, for example, be presented as the percentage of the trains that has arrived at a specific station and has a delay which is less than five minutes. Note that the punctuality and variability is normally presented for the train, or the vehicle, and not for the traveller. This means that less focus is put on making connections for trips that require a change from one train or bus to another, and less focus on the fact that trains and busses may have large differences in the number of passengers. That is, the fact that a delay of a train is less severe than a delay of a full train, at least in the perception of the general traveller.

The common critique directed towards the traditional measures of punctuality (and variability) is fourfold. First, the measures are focused on the delay of the vehicles, and not on the passengers. Weighting the delays with the number of passengers affected may produce a more useful measure. Secondly, cancelled trips are not included. Third, the punctuality and variability focus on individual journey (or, legs) and not on the travel chain that the traveller makes. Fourth, the punctuality threshold (seeing trains with a delay of less than five minutes as “on-time”), mean that the difference in a delay of 6 minutes and a delay of 30 minutes is not visible in the statistics (percentages of trains “on-time” at a specific station).

For Sweden, statistics about *train punctuality* is currently not a part of the official statistics. In Trafikanalys (2013), possible measures and techniques for such statistics are discussed, and the suggested measure for the delay at a station i as

$$\sum_{i=1}^{I} \acute{d}_i \frac{\acute{z}_i n_i}{\acute{n}_i}$$

where $\acute{n}_i$ is the average number of arrivals at the station i per day, n_i is the number of arrivals on a specific day, $\acute{z}_i$ is the average number of trains arriving at station i, per day and $\acute{d}_i$ is average delay for all trains at this station during a day.

From the traveller’s perspective, the arrival time variability and the punctuality may not be the most appropriate measure, and the concept of "schedule delay early" and "schedule delay late", may be more appropriate. The schedule delay is the difference between the actual arrival time at the destination compared to the preferred time of arrival, independent of a timetable or service frequency. This concept is complex, since the preferred arrival time normally is unknown for the operator. In this project, for the analysis of regional level services, this concept of schedule delay will not be used. This limitation implies that the results may be less relevant for measuring social costs as a consequence of travel time variability and punctuality.

Many of the above presented measurements are well defined for single legs of transport. When using the passenger-centric measures, the measures needs to be defined also for the case where the traveller change from one vehicle to another, or changes mode. In Lee et al. (2013) it is stated that around 28% of the rail passengers continue their journey from a station with public transport.

For journeys that include at least one transfer, the variability and punctuality measures need to be extended to capture the effects of delays in terms of missing connections. Computing the delay for the passengers is complex; if the connections for the passengers are maintained, the delay will be the delay of the last leg of the trip, but if a connection is missed, the delay is depending on the new travel plans.

The ”punctuality 2.0” measure, as described in Wolters (2016), uses the following expression for the *passenger punctuality*:

$$PP = \frac{T_{promised}}{T_{total}}$$

Here, T_{total}, is the total number of trips, determined by analysis of the fare cards data, and $T_{promised}$ is the number of trips where the service have arrived on schedule, that is, within a delay threshold of, for example, 5 minutes. In order to get an exact value, the fare card data need to include information on both boarding and alighting stations and the time stamp of those. By necessity, error in the time stamps from discrepancies between the actual boarding and the tap-in and the alighting and the tap-out, may affect the accuracy of the PP value.

The “punctuality 2.0” measure is an aggregate measure. When analyzing individual stations, individual services, and sequences of services, the measure gets more complicated. Normally (see e.g. Lee, 2013) the measures used are tailored for services with short headways and for long headways.

For direct passengers, with only one public transport service leg in their trip, the additional waiting time can be defined as in Section 2.2.3. The additional travel time experienced in the case of holding (which can be extended to be used also for a delay) is, as defined in Lee (2013),

$$T_{lij}^{hold} = \begin{cases} 0 \wedge if j \neq h_n \\ D_{lij}^{sched} - D_{lij}^{actual} \wedge if D_{lij}^{sched} \geq D_{lij}^{actual} \\ 0 \wedge if D_{lij}^{sched} \leq D_{lij}^{actual} \\ h^{max} \wedge if D_{lij}^{sched} - h^{max} \geq D_{lij}^{actual} \end{cases}$$

The *average in vehicle time per passenger* travelling through stop *j* is

$$T_{lj}^{hold} = \sum_i T_{lij}^{hold} \times \alpha_{lij}^{through}$$

and the *average additional in-vehicle time per passenger* on line *l* is

$$T_l^{-vehicle} = \frac{\sum_j T_{lj}^{hold} \times N_{lj}^{through}}{N_l^{board}}$$

where T_{lj}^{hold} is the in-vehicle time due to holding, per passenger in vehicle *i* at stop *j* on line *l*, h_n is the stop where holding is applied, h^{max} is the maximal holding time, $\alpha_{lij}^{through}$ is the proportion of travellers passing through stop *j* in vehicle *i*, $N_{lj}^{through}$ is the total number of travellers passing through stop *j*, and N_l^{board} is the total number of passengers on line *l*.

The *additional time for transferring passengers*, transferring from line *l* to line *m*, at stop *j*, can be computed based on the number of travelers making and missing the connection. These numbers are computed based on the actual arrival time of vehicle *i* at stop *j* on line *l* and the departure time of the vehicle *i* at stop *j* on line *m*. For the passengers missing the connection, additional transfer time is added based on the schedule for line *m*.

4. Applications

4.1. Hub Level

For the hub level, the data used for this variability analysis is the same as the one used in “D1.2: Form and flow dynamics at interchange, urban and regional network layers”. A network of cameras was installed in the pedestrian underpasses of Lausanne station which allowed the collection of detailed tracking data. This tracking data consists of a set of x and y coordinates and timestamps for each pedestrian. For further explanation, please refer to D1.2. The pedestrian tracking data available in Lausanne does not contain information about which train a pedestrian comes from or goes to, hence evaluating the Transfer reliability is not possible. Nevertheless, when simulations will be performed, assumptions can be made and therefore this indicator could be evaluated.

This data was collected as major construction work is going to take place in Lausanne station and further understanding of the pedestrian flows in the underpasses was required to know whether a third underpass needed to be built. For further information about the case study and the context, please see D5.1.

Observing, quantifying and understanding the variability which takes place inside the station of Lausanne could not be done without the previously defined variability indicators, namely travel time, travel distance and velocity. These three indicators are calculated for different locations and time intervals in the station and the results are presented in the present section. Two different spatial aggregations were performed; the first is considering all pedestrians crossing the western underpass (PIW) from north to south and vice-versa; the second concerns a small scale and is only pedestrians entering PIW from the "Tekoé ramp" and accessing platform 3/4, and vice-versa naturally. To observe the variability in time, two different strategies are used. The first considers the identification of high and low demand periods, while the second is day-to-day variations.

Travel time variability is the key indicator in this analysis as it will define whether a passenger can catch his connection or not. When considering the variability under different loads of the infrastructure (the number of pedestrians in the underpass is used to identify high and low demand periods, called accumulation), no noticeable difference is visible, see Figure 2. The distributions of travels times of pedestrians crossing the western underpass from north to south or vice versa are plotted based on three temporal aggregations, and no difference is visible.

When focusing on a shorter trip, namely the "Tekoé ramp" to platforms 3/4, a significant difference does appear. Unlike the previous case, the distributions of travel between the high accumulation and low accumulation are shifted. Figure 3 shows these distributions; while the variances and shape are similar, the means are clearly shifted. Therefore, travel times present a systematic difference depending on the accumulation inside the underpass, but not for all situations. This variability is important when building schedules as the time required to change platforms evolves based on the level-of-service within the station.

Finally, day-to-day variations where investigated, but no significant differences are visible. In Figure 4 the distributions of travel time, travel distance and velocity are represented and in Figure 5 the travel time distributions for the problematic area between the "Tekoé ramp" and platforms 3/4 are plotted. In both cases, no day-to-day variations appear. When considering these figures, the variability at the hub level mainly comes from different levels of congestion inside the walking areas and not external effects which could impact the passengers walking experience. Finally, Figure 6 contains the distribution of scheduled transfer times for the Lausanne station. A peak is visible around 5 minutes.

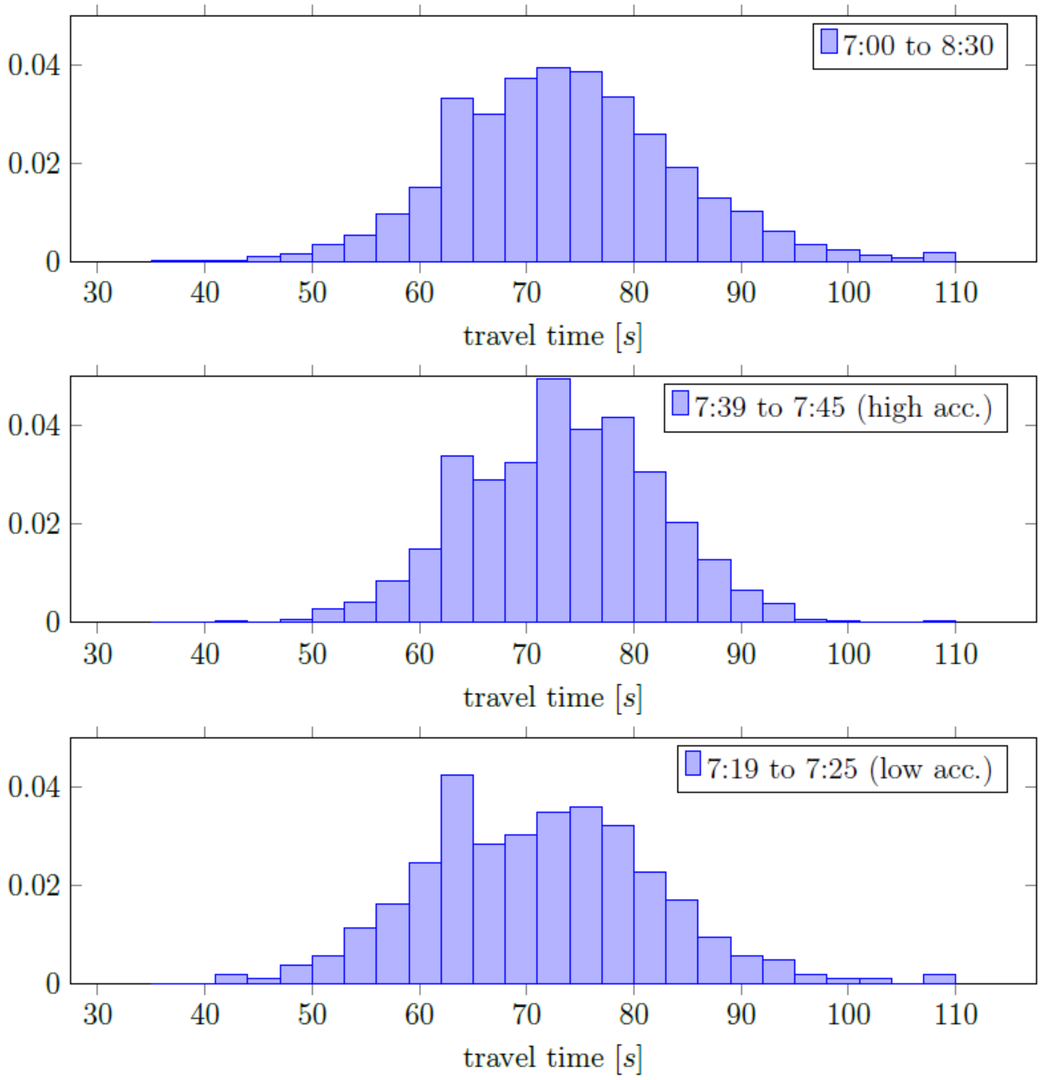


Figure 2 The travel time distributions of pedestrians crossing the western underpass do not present significant variations when subset based on pedestrian accumulation (proxy for density). One could expect longer travel times when the accumulation is higher as individuals cannot walk at their free flow velocity.

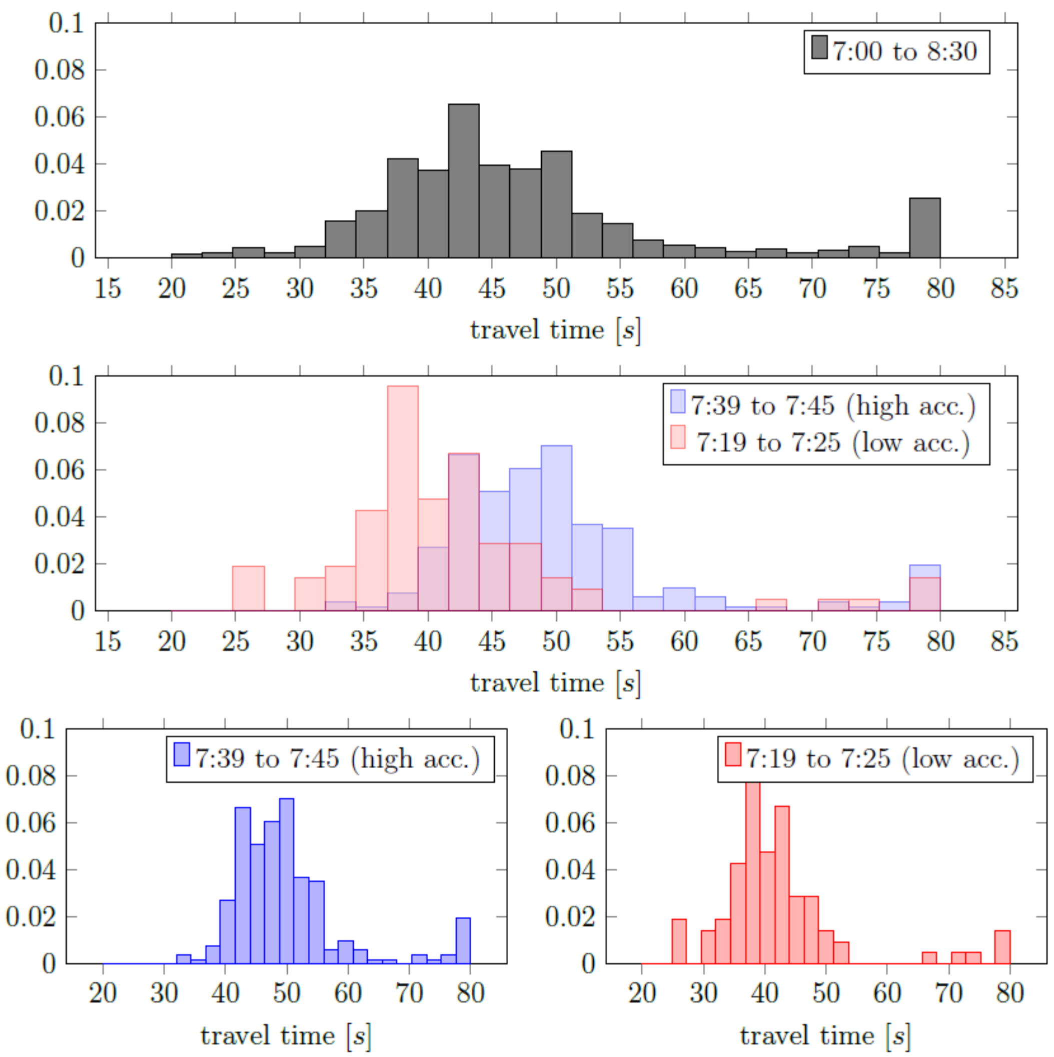


Figure 3: When analyzing the travel time distributions in the possibly most problematic location of the underpasses (Tekoé ramp to platform 3/4), one significant difference appears: the mean of the distributions is shifted. For the six-minute interval with high accumulation (7:39 – 7:45) the mean travel time is 50 seconds, while the mean travel time for the low accumulation interval is 43 seconds.

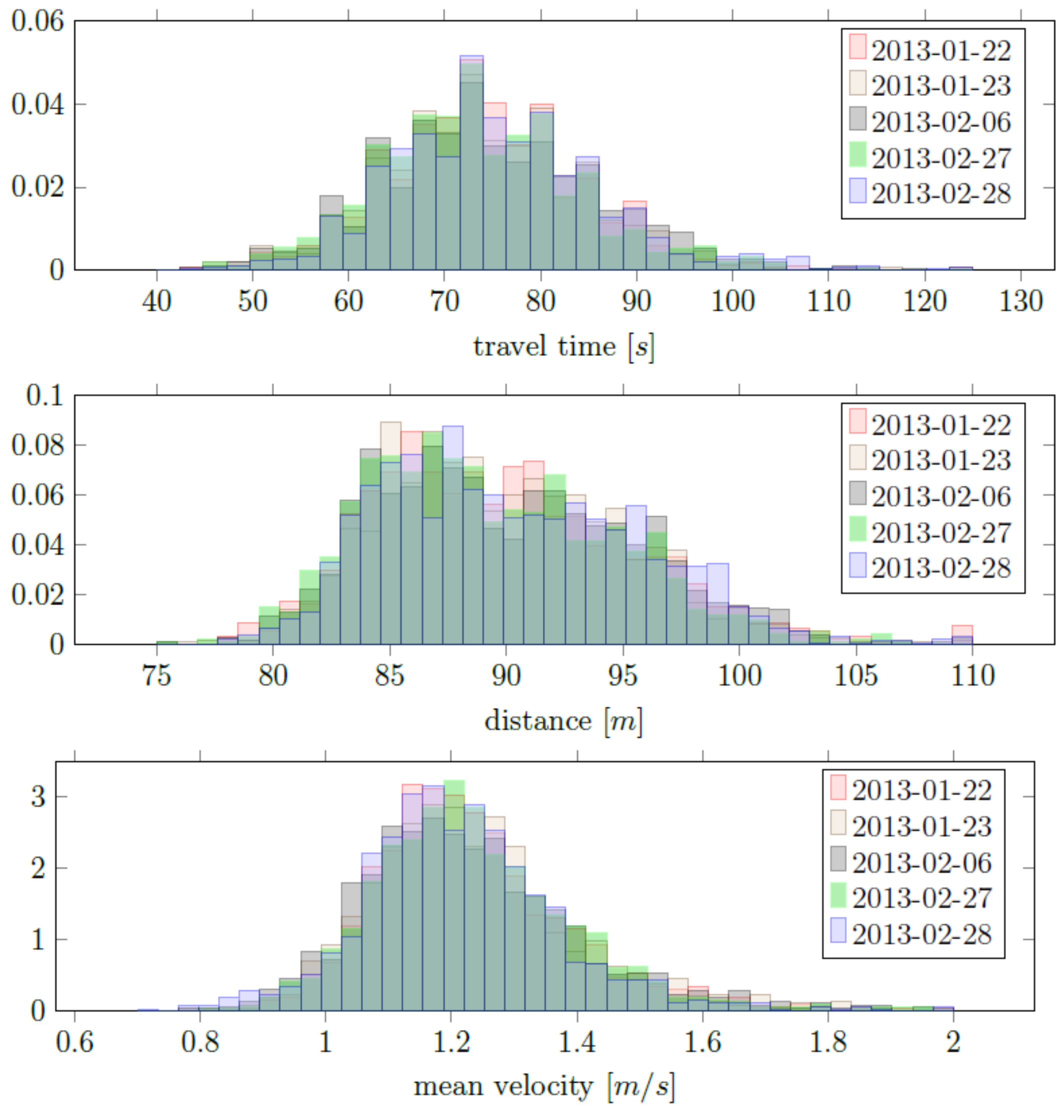
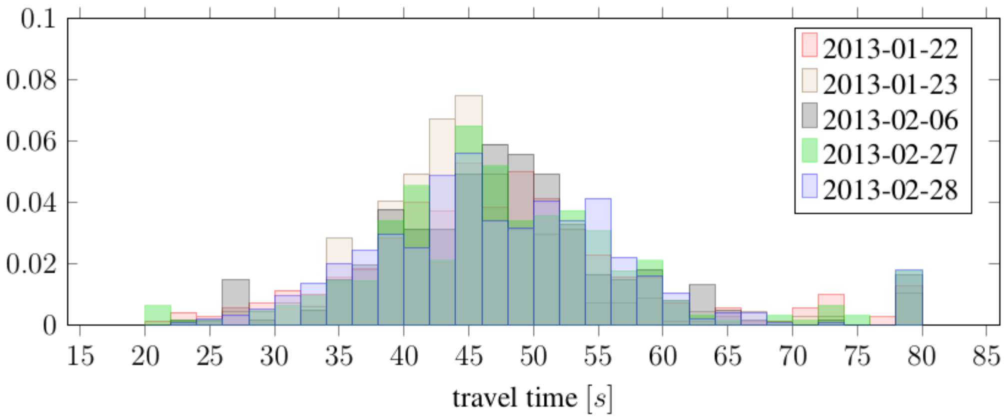
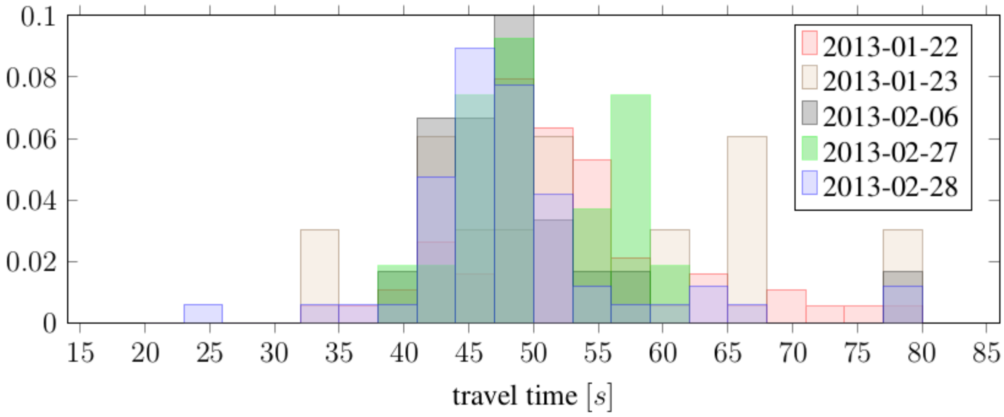


Figure 4: Variations between days should be kept to a minimum to ensure reliable experiences for the users. When considering the travel time, travel distance and mean velocity distributions this is the case. No significant differences appear between the days, the only remarkable element is the possible two-mode distribution of the travel distance.



(a) Travel time histogram from 7:00 to 8:30 from the Tekoe ramp to platforms 3/4



(b) Travel time histogram from 7:39 to 7:45 (high accumulation) from the "Tekoé ramp" to platforms 3/4

Figure 5: Daily variations of travel time in the problematic Tekoé ramp to platform 3/4 area. During the 6-minute interval of high accumulation, the number of pedestrians passing in the area is not sufficient to produce a consistent histogram, hence the strong variations in (b).

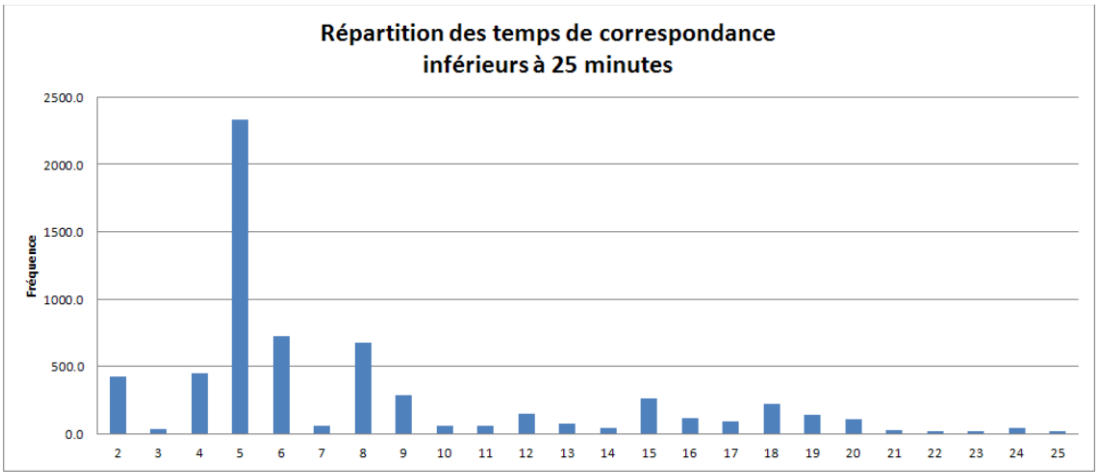


Figure 6: Scheduled transfer times according to the time table. Taken from an internal technical report produced by Nicolas Anken and Flurin Hänseler. The allowed transfer times are greater than 2 minutes "by construction", with a peak around 5 minutes.

4.2. Urban Level

This section shows the results of applying the passenger-oriented reliability indicators to the case study network. The case study network is the urban public transport network in The Hague, the Netherlands, operated by HTM. This network consists of 12 light rail and tram lines, together with 8 bus lines. From the 12 tram lines, two lines are light rail services operating on the agglomeration level as connection between the city of The Hague and satellite city Zoetermeer. The other 10 tram lines function as urban lines within The Hague. All 8 bus lines function on the urban level. Regional bus lines are not operated by HTM. On an average working day, about 290.000 passenger transactions occur over all tram and bus lines together.

To quantify reliability of the urban network level, we use smart card data from November 2015. All smart card data transactions from November 3rd to November 20th and from November 22nd to November 29th are used for this end. The data from November 21st showed to be inconsistent between AFC and AVL data, leading to the decision to exclude this day from the research. In The Hague, smart card penetration is about 93%, which means that a high percentage of all journeys is captured in this analysis. In the resulting data set, transactions from 19 working days, 4 Sundays and 3 Saturdays are used. In total, about 6.9 million smart card transactions are included in the analysis. Specifically for the journey level reliability indicator, we used a selection of this dataset, containing all data from November 3rd to November 16th, resulting in data from two full weeks (10 working days, 2 Saturdays and 2 Sundays).

From the passenger-oriented service reliability indicators as defined in chapter 3, we show the results of the application of the case study data to two of these indicators. First, we show the results of the daily variability indicator as defined by equation (11) in chapter 3. This indicator expresses the perceived journey travel time predictability using the ratio between the 95th percentile (as upper percentile) and the 50th percentile (as nominal percentile). We show results regarding journey reliability for several spatial aggregations as origin and as destination. This provides insight in the perceived journey reliability for passengers having their origin within a spatial area in the case study area, as well as for passengers having their destination within a certain spatial area of the case study area. This allows a comparison regarding perceived journey reliability between different spatial clusters. Second, results are shown of applying the hub reliability indicator as defined by equation (14) in chapter 3. This indicator expresses the number of transferring passengers at a specific hub missing their scheduled connection, relative to the total number of transferring passengers at that hub. Results are shown for one specific urban public transport hub from the case study network.

Table 1 below shows the quantification of equation (11) from chapter 3. It shows the difference between the 95' and 50' percentile perceived journey travel time as ratio of the 50' percentile value. All journeys are spatially aggregated based on the origin location of the journey. As can be seen, there are substantial differences in daily variability in perceived journey travel time for passengers starting their journey from different parts of the case study network. Especially passengers starting their journey in the satellite city of Zoetermeer, located most far away from all other areas, experience a relatively large journey travel time variability. Journeys starting from the city centre of The Hague also have a relatively high level of variability. On the other hand, journeys starting in the housing areas of The Hague (such as Segbroek, Scheveningen, Haagse Hout, Laak) and journeys starting in other small cities surrounding The Hague (Delft, Rijswijk) suffer less from perceived unpredictability.

Table 1: Perceived journey travel time variability per aggregated origin area of the case study area of The Hague

Aggregation origin location	Ratio (95' TT – 50' TT) / 50' TT
Den Haag Segbroek	1.32
Den Haag Scheveningen	1.34
Station Voorburg	1.38
Station Hollands Spoor	1.42
Delft	1.43
Rijswijk	1.44
Knooppunt Leyenburg	1.45
Centraal Station	1.46
Den Haag Leidschenveen-Ypenburg	1.48
Den Haag Haagse Hout	1.58
Den Haag Laak	1.66
Station Rijswijk	1.66
Station Den Haag Mariahoeve	1.69
Station Laan van NOI	1.70
Station Den Haag Moerwijk	1.76
Den Haag Loosduinen	1.88
Den Haag Escamp	1.94
Leidschendam-Voorburg	1.95
Den Haag Centrum	2.52
Zoetermeer	2.68

Table 2 shows the perceived journey travel time variability considered from the perspective of the destination location. All journey destinations within a certain area are spatially aggregated for this end. As can be seen, there are some similarities to Table 1. For example, Zoetermeer also suffers from relatively unpredictable journey travel times as destination as well. Cities as Rijswijk and Delft, where in general is quite some intra-zonal public transport usage, logically show quite predictable travel times in Table 2 as well. However, also some differences can be seen. For journeys starting in Segbroek and Scheveningen, Table 1 shows a relatively high travel time predictability. However, there is substantially more variability for journeys having Scheveningen and especially Segbroek as destination. This can be explained, since most public transport journeys are made between the city centre of The Hague and these areas. Given the high origin unreliability for journeys starting in the city centre, it is explainable that destination reliability of journeys ending in Scheveningen and Segbroek – but mostly originating in the city centre – is lower.

Table 2: Perceived journey travel time variability per aggregated origin area of the case study area of The Hague

Aggregation destination location	Ratio (95' TT – 50' TT) / 50' TT
Station Rijswijk	1,27
Station Den Haag Moerwijk	1,31
Delft	1,37
Station Hollands Spoor	1,40
Station Den Haag Mariahoeve	1,43
Station Voorburg	1,43
Rijswijk	1,44
Leidschendam-Voorburg	1,50
Den Haag Scheveningen	1,51
Den Haag Haagse Hout	1,54
Den Haag Laak	1,66
Den Haag Loosduinen	1,69
Knooppunt Leyenburg	1,71
Den Haag Leidschenveen-Ypenburg	1,77
Centraal Station	1,81
Den Haag Segbroek	1,81
Overig	1,92
Den Haag Centrum	1,98
Station Laan van NOI	1,98
Zoetermeer	2,52
Den Haag Escamp	2,63

Below the first hub reliability criterion, in which the percentage of missed connections is determined for all transfer combinations, is quantified. For this end, the bus station hub 'Leyweg' is selected. This is an urban hub, served by three different bus lines and transfers occur frequently. The resulting figure illustrates that substantial differences in transfer reliability are derived at this hub. It also illustrates the importance of different transfers, given the transferring flow derived from the smart card data. The combination of these values gives insight in the reliability of different transfer connections, weighted by the passenger flow for each transfer connection.

from line	destination		to line	destination		transfering flow	transfering flow with scheduled connection	pax volume missed connections	% pax volume missed connections
21	Vrederust	terug	23	Scheveningen	terug	57	37	20	35%
21	Vrederust	terug	23	Kijkduin	heen	12	2	10	83%
21	Vrederust	terug	25	Vrederust	terug	96	53	43	45%
21	Vrederust	terug	25	Grote Markt	heen	59	33	26	44%
21	Scheveningen	heen	23	Scheveningen	terug	132	77	55	42%
21	Scheveningen	heen	23	Kijkduin	heen	77	33	44	57%
21	Scheveningen	heen	25	Vrederust	terug	8	1	7	88%
21	Scheveningen	heen	25	Grote Markt	heen	120	98	22	18%
23	Scheveningen	terug	21	Vrederust	terug	128	73	55	43%
23	Scheveningen	terug	21	Scheveningen	heen	11	4	7	64%
23	Scheveningen	terug	25	Vrederust	terug	292	178	114	39%
23	Scheveningen	terug	25	Grote Markt	heen	155	86	69	45%
23	Kijkduin	heen	21	Vrederust	terug	142	97	45	32%
23	Kijkduin	heen	21	Scheveningen	heen	90	62	28	31%
23	Kijkduin	heen	25	Vrederust	terug	350	210	140	40%
23	Kijkduin	heen	25	Grote Markt	heen	208	125	83	40%
25	Vrederust	terug	21	Vrederust	terug	156	115	41	26%
25	Vrederust	terug	21	Scheveningen	heen	84	70	14	17%
25	Vrederust	terug	23	Scheveningen	terug	224	161	63	28%
25	Vrederust	terug	23	Kijkduin	heen	112	55	57	51%
25	Grote Markt	heen	21	Vrederust	terug	19	8	11	58%
25	Grote Markt	heen	21	Scheveningen	heen	118	81	37	31%
25	Grote Markt	heen	23	Scheveningen	terug	365	196	169	46%
25	Grote Markt	heen	23	Kijkduin	heen	282	156	126	45%

Figure 7: Overview of transfer reliability for Leyweg hub in the urban bus network of The Hague.

4.3. Regional Level

For the regional level data analysis, two data sources are used, presented in "D1.2: Form and flow dynamics at interchange, urban and regional network layers". The first source is the tap-ins from the Blekingetrafiken public transport system, covering regional busses, local busses and regional trains. Based on this data set, passenger volumes are estimated. The second source is train arrival and departure times for national and regional trains in the Blekinge and Skåne region.

The initial analysis of the service variability will be based on the tap-in data. The data is processed using the procedures described in the D1.2 This processing generates passenger journeys. A number of attributes defines each journey. For each journey, the departure time from the origin stop (train station of bus stop) according to schedule, the scheduled arrival time to the destination stop, the tap-in time on the vehicle, and the travel card ID.

The first type of analysis that will be made is the transfers between lines at important train stations or bus stops. By analysing journeys that arrive and depart from a specific station where the arrival time and departure time is short, a transfer is identified. The journeys, for one specific day arriving, at the "Ronneby station" continuing from this same station within 20 minutes are illustrated in 8. From the figure it can be noted that the most transfers are made from route 23, which is a local bus in Ronneby, to route 150 which is a regional bus and to route 802 which is the regional Öresundståg train passing Ronneby with departures both eastward (towards Kalrkrona) and westward (towards Lund). Route 802 is the only train route passing Ronneby station. Route 150 and 600, also shown in the figure, are regional busses. The other routes are local busses serving the Ronneby regional area. For the example of Ronneby station, only transfers at the specific stop (the stop site at the station) are analysed. Other transfers, to and from nearby bus stops, can also be identified but are not part of this analysis.

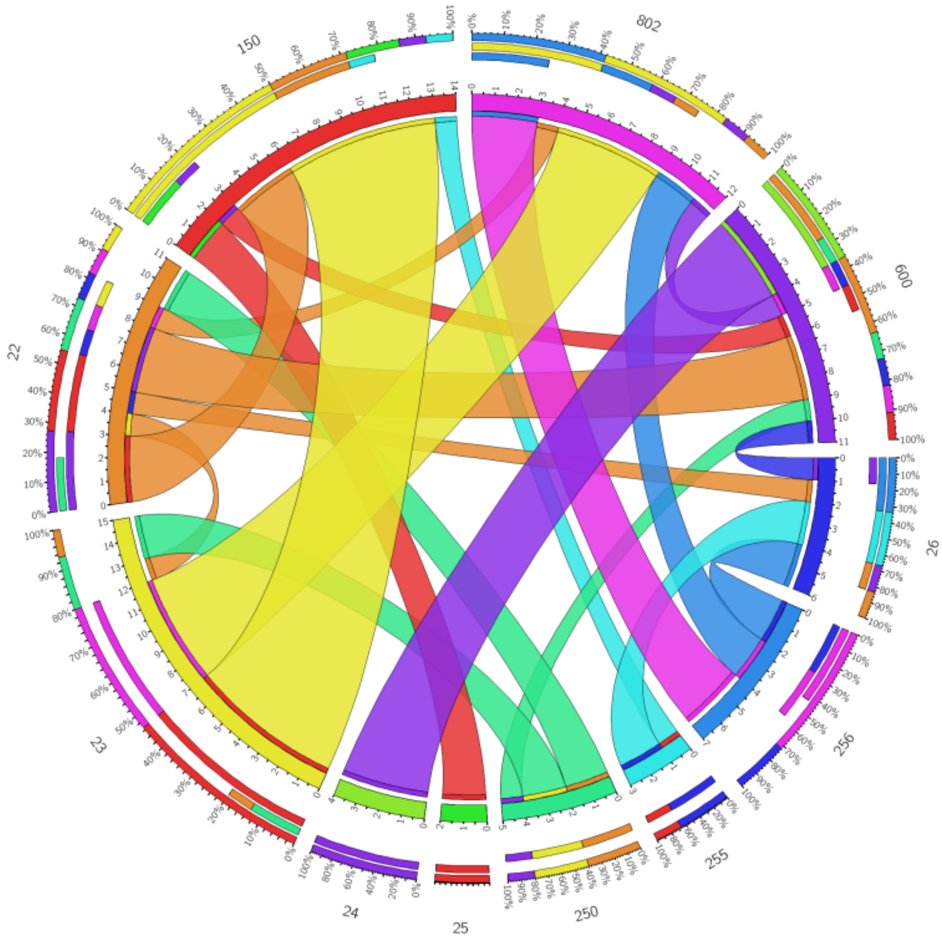


Figure 8: Route transfer relation at the Ronneby station one day in October 2016.

Each transfer shown in Figure 8 is associated with a transfer time. The transfer time distribution for all transfers at the Ronneby station, independent of the arrival and departure route, is shown in 9. This is the transfer time for the case where all used busses and trains arrive and depart according to their schedule.

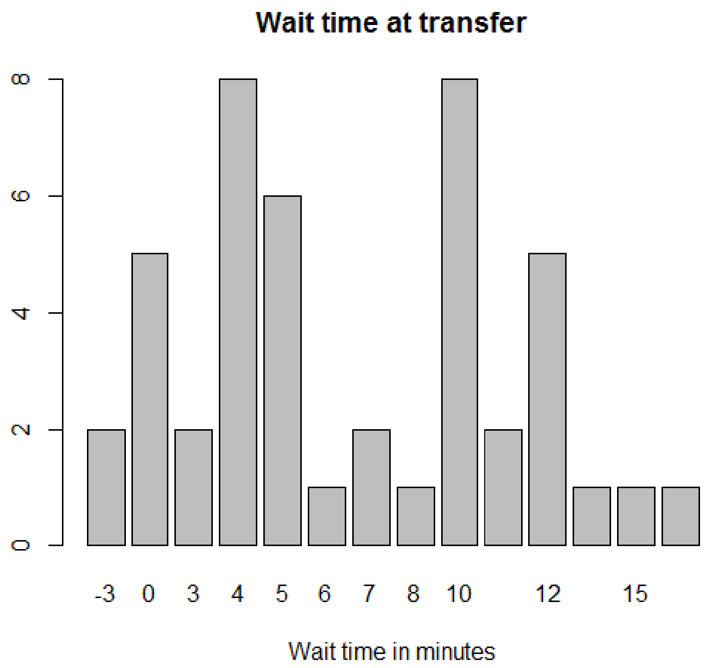


Figure 9: Wait time at Ronneby station, independent of arrival and departure route.

This data is not ideal for analysing the service variability. In order to apply the service variability analysis, the tap-in times need to be used. The tap-in can be made at any time after boarding the vehicles, although, in practice, the general behaviour is to tap-in when entering the vehicles. However, the relation between the departure time and the time when the passengers enter the vehicle may differ. Therefore, analysis of the tap-in distribution, with the aim of estimating the real departure time (which might differ from the scheduled time) needs to be done.

For the application performed in WP4, the real train arrival and departure times will be extracted from Trafikverket, and replace the scheduled times used in the figures above.

5. Conclusions and Prospects

5.1. Measuring across Levels

In this section a whole journey passenger reliability perspective will be explored, combining the measures in section 4 into integrated measures.

Although levels of analysis or measurement are not necessarily mutually exclusive, there are three general levels into which social science research, transportation in this case, may fall: micro-level, meso-level and macro-level. The term "level of analysis" is used to point to the location, size, or scale of a research target. Note that "level of analysis" is different from the term "unit of observation" in that the former refers to a more or less integrated set of relationship, while the latter refers to the distinct unit from which data have been gathered (Blalock, 1979). Thus, within TRANS-FORM project, the measurements of journey reliability from the passenger perspective, can be structured in these three levels as follows:

- **Micro-level** is the smallest unit of analysis where the research population is an individual in their social setting. This corresponds to the analysis of the movements of pedestrians in the Lausanne station (Switzerland case study).
- **Meso-level** indicates a population size or a coverage area that falls between the micro- and macro-levels. However, meso-level may also refer to analyses that are specifically designed to reveal connections between micro- and macro-levels. In the project, it corresponds to the reliability measurement of passengers' travel experience as in The Hague (The Netherlands case study).
- **Macro-level** analyses generally trace the outcomes of interactions over a large area, such as the regional level and aggregate analysis of flows which corresponds to the reliability analysis of the selected regions within the Sweden case study.

This three-layered (micro-meso-macro) framework for the reliability analysis provides knowledge at different levels of detail of the passenger journey experience. At the micro-level, the focus is on the individual, at the meso-level, it is on local hubs and passenger choices, and at the macro-level, the focus is on system-wide transportation .

As explained in "D1.2: Form and flow dynamics at interchange, urban and regional network layers", 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. So, the urban and regional public transport network levels are not directly connected with each other, but they are connected by means of the hub level. This means that the hub level constitutes the intermediate level. As a consequence, **measuring reliability of the whole journey from the passenger perspective could be made using the hub level as a common link for reliability measurements.**

5.2. Outlook

This deliverable reviewed, proposed, formulated, implemented and applied a range of public transport service variability indicators. Service variability was conceptualized from a passenger point of view and considered in terms of the experience when travelling through a hub, a multi-modal urban or regional public transport network and the interactions between these elements. While both reliability and robustness aspects were reviewed, the focus in the indicator development and application was on reliability indicators which relate to recurrent disturbances.

The following tasks in WP2 are concerned with representing the passenger flows effects at different public transport network levels - from hub through metropolitan services to large-scale train systems. Different tools for modelling pedestrian traffic flows, passenger route choice and public transport operations and infrastructure are coupled to allow accounting for interactions between system layers. Hence, it will allow to analyse service variability beyond the single element level and consider passenger transfer experience. This is especially valuable in the event of a service disturbance or traffic management intervention that will be examined in WP3. The control strategies and passenger predictions that will be developed will be integrated into a traffic management tool. The passenger-oriented indicators developed in this deliverable can be used for (1) evaluating system performance under alternative (demand, operations, disruptions, control) scenarios based on either empirical or simulated data, and; (2) steering the operations to better cater for passengers' experience for example by taking control decisions which minimize selected indicators.

References

Al-Deek, H. & Emam, E.B., 2006. New methodology for estimating reliability in transportation networks with degraded link capacities, *Journal of Intelligent Transportation Systems*, 10(3) 117-129.

Angeloudis, P. & Fisk, D., 2006. Large subway systems as complex networks. *Physica A*, 3, 553–558.

Asakura, Y. & Kashiwadani, N., 1991. *Road network reliability caused by daily fluctuation of traffic flow*. In: Proceedings of the 19th PTRC Summer Annual Meeting, Brighton, 73–84.

Asakura, Y., Hato, E. & Kashiwadani, M., 2001. Stochastic network design problem: An optimal link improvement model for reliable network. *Proceedings of the 1st international Symposium on Transportation Network reliability*, Kyoto, Japan.

Bagherian, M., Cats, O., Van Oort, N. & Hickman, M., 2016. Measuring passenger travel time reliability using smartcard data. *Transport Policy*, under review.

Balcombe, R., Mackett, R., Paulley, Preston, J., Shires, J., Titheridge, H., Wardman, M. & White, P., 2004. *The Demand for Public Transport: A Practical Guide*, TRL Report, TRL 593.

Barker, K., Ramirez-Marquez, J.E. & Rocco, C.M., 2013. Resilience-based network component importance measures. *Reliability Engineering and System Safety*, 117, 89-97.

Bates, J., Polak, J., Jones, P. & Cook, A., 2001. The valuation of reliability for personal travel. T*ransportation Research Part E*, 37, 191-229.

Bell, M. & Iida, Y., 1997. *Network reliability: Transportation network analysis*. Chichester, England: John Wiley and Sons.

Berdica, K., 2002. An introduction to road vulnerability: what has been done, is done and should be done. *Transport Policy*, 9(2), 117-127.

Börjesson, M., Eliasson, J. & Franklin, J.P., 2012. Valuations of travel time variability in scheduling versus mean–variance models. *Transportation Research Part B*, 46, 855–873.

Bush, R., 2007. Does every trip need to be on time? Multimodal scheduling performance parameters with an application to Amtrak service in North Carolina. In *Proceedings of the 86th annual meeting of Transportation Research Board*. Washington, D.C.

Cats, O., Koutsopoulos, H.N., Burghout, W. & Toledo, T., 2011. Effect of real-time transit information on dynamic passenger path choice. *Transportation Research Record*, 2217, 46–54.

Cats, O., 2014. Regularity-driven bus operation: Principles, implementation and business models. *Transport Policy*, 36, 223-230.

Cats, O. & Jenelius, E., 2014. Dynamic vulnerability analysis of public transport networks: mitigation effects of real-time information. *Networks and Spatial Economics*, 14,435–463.

Cats, O. & Jenelius, E, 2015a. Beyond a Complete Failure: The Impact of Partial Capacity Degradation on Public Transport Network Vulnerability. *Transportmetrica B*, 2015, 1-23.

Cats, O. & Jenelius, E., 2015b. Planning for the unexpected: the value of reserve capacity for public transport network robustness. *Transportation Research Part A*, 81, 47-61.

Cats, O., West, J. & Eliasson, J., 2016a. A dynamic stochastic model for evaluating congestion and crowding effects in transit systems. *Transportation Research Part B*, 89, 43-57.

Cats, O., Yap, M.D. & Van Oort, N., 2016b. Exposing the role of exposure: Public transport network risk analysis. *Transportation Research Part A*, 88, 1-14.

Cats, O., 2016. The robustness value of public transport development plans. *Journal of Transport Geography* 51 (2016) 236–246

Cham, L.C. & Wilson, N.H.M., 2006. *Understanding bus service reliability: a practical framework using AVL/APC data*. Available at: https://dspace.mit.edu/handle/1721.1/34381

Chen, A., Yang, H., Lo, H.K. & Tang, W.H., 2002. Capacity reliability of a road network: an assessment methodology and numerical results, *Transportation Research Part B*, 36, 225-252.

Clark, S. & Watling, D., 2005. Modelling network travel time reliability under stochastic demand. *Transportation Research Part B*, 39, 119–140.

Criado, R., Hernández-Bermejo, B. & Romance, M., 2007. Efficiency, vulnerability and cost: an overview with applications to subway networks worldwide. *International Journal of Bifurcation and Chaos*, 17, 2289–2301.

Dehghanisanij, M., Flintsch, G.W. & McNeil, S., 2013. Vulnerability Analysis of Degrading Roadway Networks. *Proceedings of the 92th Transportation Research Board Annual Meeting*, Washington D.C.

Deng, Y., Li, Q., Lu, Y., Yuan, J., 2013. Topology vulnerability analysis and measure of urban metro network: The case of Nanjing. *Journal of Networks*, 8 (6), 1350-1356.

Derrible, S. & Kennedy, C., 2010. The complexity and robustness of metro networks. *Physica A*, 389(17), 3678–3691.

Dupuy, G., 2013. Network geometry and the urban railway system: the potential benefits to geographers of harnessing inputs from "naïve" outsiders. *Journal of Transport Geography*, 33, 85-94.

Department for Transport, UK, 2013. *Tendering Road Passenger Transport Contracts. Best Practice Guidance.*

Eliasson, J., 2007. The relationship between travel time variability and road congestion. In: *Proceedings from the 11th World Conference on Transport Research*, University of Berkeley, USA.

Engelson, L. & Fosgerau, M., 2011. Additive measures of travel time variability. *Transportation Research Part B*, 45, 1560–1571.

European Commission, 2011. WHITE PAPER. *Roadmap to a Single European Transport Area – Towards a competitive and resource efficient transport system.*

Edie, L. C. (1963). Discussion of traffic stream measurements and definitions, Port of New York Authority.

Garcia Pastor, A, López Lambas, M. E., 2005. *Quality issues in transport operation tenders. The difficult equilibrium between price and service level*. Association for European Transport.

Hakkesteegt P. & Muller, T.H.J., 1981. Research increasing regularity, *Verkeerskundige werkdagen*, 415-436 (in Dutch).

Hendren, P., Antos, J., Carney, Y. & Harcum, R., 2015. Transit Travel Time Reliability: 2 Shifting the Focus from Vehicles to Customers 3. *Transportation Research Board 94th Annual meeting*, 2015.

Homeland Security, U.S. Department of, 2010. *Transportation Systems Sector-Specific Plan, An Annex to the National Infrastructure Protection Plan*.

Hoogendoorn, S.P., Daamen, W. (2005). Pedestrian Behaviour at Bottlenecks. *Transportation Science*, 39(2), pp. 147-159.

Iida, Y. & Wakabayashi, H., 1989. An approximation method of terminal reliability of a road network using partial minimal path and cut set. *Proceedings of the 5th world conference on transport research*, Yokohama, Japan.

Jenelius, E. & Mattson, L., 2012. Road network vulnerability analysis of area-covering disruptions: a grid-based approach with case study. *Transportation Research Part A*, 46, 746-760.

Jenelius, E., Petersen, T. & Mattson, L., 2006. Importance and exposure in road network vulnerability analysis. *Transportation Research Part A*, 40, 537-560.

Knoop, V., Van Zuylen, H.J. & Hoogendorn, S.P., 2008. The influence of spillback modelling when assessing consequences of blockings in a road network. *European Journal of Transport and Infrastructure Research*, 8, 287–300.

Kurauchi, F., Shimamoto, H., Ieda, Y. & Bell, M.G.H., 2004. Evaluation of public transport connectivity reliability using capacity constrained transit assignment model. *Proceedings of the 2nd international Symposium on Transportation Network Reliability*, Christchurch, New Zealand.
Lee, A., Van Oort, N. & Van Nes, R., 2014. Service reliability in a network context: Impacts of synchronizing schedules in long headway services. *Transportation Research Record*, 2417, 18-26.
Landex, A., & Kaas, A. H., 2009. Examination of operation quality for high-frequent railway operation. In *Proceedings of the 3rd international seminar on railway operations modelling and analysis*. Zürich.
Li, M., 2008. *Robustness for Road Networks – A Framework with Combined DTA Models.* Delft University of Technology, Faculty of Civil Engineering and Geosciences, Delft, the Netherlands.
Mackie, P., Worsley, T., Eliasson, J., 2014. Transport appraisal revisited. *Research in Transport Economics*, 47, 3–18.
Mattson, L. & Jenelius, E., 2015. Vulnerability and resilience of transport systems – a discussion of recent research. *Transportation Research Part A*, 81, 16-34.
Nikolic, M., Bierlaire, M., Farooq, B., and de Lapparent, M. (2016). *Procedia – Social Behavioral Science*, 162, 158-167.
Oliveira, E.L., Portugal, L.S. & Porto Junior, W., 2016. Indicators of reliability and vulnerability: Similarities and differences in ranking links of a complex road system. *Transportation Research Part A*, 88, 195-208.
OECD/International transport forum, 2010. *Improving reliability on surface transport networks*.
OECD/International transport forum, 2016. *Travel Behaviour Response to Major Transport System Disruptions Implications for Smarter Resilience Planning.* ITF Discussion Paper 2016-09
Pau, S.A., 2013. *Using Smart Card Technologies to Measure Public Transport Performance: Data Capture and Analysis.* Universitat Politecnica de Catalunya, Barcelona, Spain.
Pelletier, M.P., Trépanier, M. & Morency, C., 2011. Smart card data use in public transit: A literature review. *Transportation Research Part C*, 19(4), 557-568.
Reggiani, A., Nijkamp, P. & Lanzi, D., 2015. Transport resilience and vulnerability: the role of connectivity. *Transportation Research Part A*, 81, 4-15.
Rietveld, P., Bruinsma, F.R. & van Vuuren, D.J., 2001. Coping with unreliability in public transport chains: A case study for Netherlands. *Transportation Research A*, 35(6), 539-559.
Scott, D.M., Novak, D.C., Aultman-Hall, L. & Guo, F., 2006. Network robustness index: a new method for identifying critical links and evaluating the performance of transportation networks. *Journal of Transport Geography*, 14, 215-227.
Snelder, M., Van Zuylen, H.J. & Immers, L.H., 2012. A framework for robustness analysis of road networks for short term variations in supply. *Transportation Research part A*, 46, 828-842.
Sullivan, J.L., Novak, D.C., Aultman-Hall, L. & Scott, D.M., 2010. Identifying critical road segments and measuring system-wide robustness in transportation networks with isolating links: a link-based capacity reduction approach. *Transportation Research Part A*, 44, 323-336.
Sun, L., 2014. *Research on urban transit reliability using smart card data.* National University of Singapore, Singapore.
Schittenhelm, B., & Landex, A., 2009. Quantitative methods to evaluate timetable attractiveness. In *Proceedings of the 3rd international seminar on railway operations modelling and analysis*. Zürich.
Tampère, C.M.J., Stada, J. & Immers, B., 2007. Methodology for Identifying Vulnerable Sections in a National Road Network ‹www.mech.kuleuven.be/› (07.07.2014).
Taylor, M.A.P., Sekhar, S.V.C. & D'Este, G.M., 2006. Application of accessibility based methods for vulnerability analysis of strategic road networks. *Networks and Spatial Economics*, 6, 267–291.
Taylor, M.A.P. & Susilawati, 2012. Remoteness and accessibility in the vulnerability analysis of regional road networks. *Transportation Research Part A*, 46, 761-771.

Tseng, Y.Y., 2008. *Valuation of travel time reliability in passenger transport* (PhD Thesis). Vrije Universiteit, Amsterdam, the Netherlands.

Turnquist, M.A. & Bowman, L.A., 1980. The effects of network structure on reliability of transit service. *Transportation Research Part B*, 14, 79-86.

Uniman, D. (2009). *Service Reliability Measurement Framework using Smart Card Data: Application to the London Underground* (MSc Thesis). Massachusetts Institute of Technology.

Uniman, D., Attanucci, J., Mishalani, R. & Wilson, N., 2010. Service Reliability Measurement Using Automated Fare Card Data. *Transportation Research Record*, 2143, 92-99.

Van Oort, N. & Van Nes, R., 2004. Service regularity analysis for urban transit network design. *Proceedings of the 83rd Annual Meeting of Transportation Research Board*, Washington D.C.

Van Oort, N. & Van Nes, R., 2006. Reliability of urban public transport and strategic and tactical planning: a first analysis, In H.J. van Zuylen (Ed.), *Proceedings 9th TRAIL Congress*, TRAIL Research School, Delft, the Netherlands, 1-18.

Van Oort, N., 2011. *Service reliability and urban public transport design (PhD Thesis)*. Delft University of Technology, Delft, the Netherlands.

Van Oort, N., Brands, T. & De Romph, E., 2015. *Transportation Research Record, 2535*, 105-111.

Van Oort, N., 2014. Incorporating service reliability in public transport design and performance requirements: International survey results and recommendations, *Research in Transportation Economics*, 48, 92-100.

Van Oort, N., & Van Nes, R., 2009. Regularity analysis for optimizing urban transit network design. *Public Transport*, 1(2), 155e168.

Von Ferber, C., Holovatch, T., Holovatch, Y. & Palchykov, V., 2009. Public transport networks: empirical analysis and modeling. *European Physical Journal B*, 68, 261–275.

Von Ferber, C., Berche, B., Holovatch, T. and Holovatch, Y., 2012. A tale of two cities: vulnerabilities of the London and Paris transit networks. *Journal of Transportation Safety and Security*, 5(3), 199–216.

Vromans, M.J.C.M., 2005. *Reliability of Railway Systems* (PhD Thesis). Erasmus University, Rotterdam, the Netherlands.

Wood, D.A., 2015. *A framework for measuring passenger-experienced transit reliability using automated data.* Massachusetts Institute of Technology.

Yap, M.D., Van Oort, N., Van Nes, R. & Van Arem, B., 2015. Robustness of multi-level public transport networks: a methodology to qauntify (un)robustness from a passenger perspective. *6th International Symposium on Transportation Network Reliability (INSTR)*, Nara, Japan.

Zhang, J., Xu, X., Hong, L., Wang, S., Fei, Q., 2011. Networked analysis of the Shanghai subway network, in China. *Physica A*, 390, 4562-4570.

Zhao, J., Frumin, M., Wilson, N. & Zhao, Z., 2013. Unified estimator for excess journey time under heterogeneous passenger incidence behavior using smart card data. *Transportation Research Part C*, 34, 70-88.