

Measuring Passenger Travel Time Reliability Using Smart Card Data

Bagherian, Mehdi; Cats, Oded; van Oort, Niels; Hickman, M

Publication date

2016

Document Version

Accepted author manuscript

Citation (APA)

Bagherian, M., Cats, O., van Oort, N., & Hickman, M. (2016). *Measuring Passenger Travel Time Reliability Using Smart Card Data*. Abstract from TRISTAN 2016: The 9th Triennial Symposium on Transportation Analysis, Oranjestad, Aruba.

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

Measuring Passenger Travel Time Reliability Using Smart Card Data

Mehdi Bagherian ^{a*}, Oded Cats ^b, Niels van Oort ^b and Mark Hickman ^a

^a *School of Civil Engineering, The University of Queensland, Australia*

^b *Department of Transport and Planning, Delft University of Technology, The Netherlands*

Abstract

Service reliability is a key performance measure for transit agencies in increasing their service quality and thus ridership. Conventional reliability metrics are established based on vehicle movements and thus do not adequately reflect passenger's experience. In the past few years, the growing availability of smart card data allows shifting the reliability measures from vehicle's to passenger's point of view.

This research introduces two new passenger-oriented measures of transit travel time reliability and a method to measure them using on-board smart card transactions data. These measures reflect both punctuality (deviation from the schedule) and predictability (day to day variation) of the service over the selected spatial-temporal scale. The analysis approach and the mathematical formulation are presented and then applied to the public transport network of The Hague, the Netherlands. The developed method can evaluate passenger-oriented reliability at various spatial-temporal levels, from a single origin-destination pair to a network-wide evaluation.

1. INTRODUCTION

Travel time reliability is widely accepted as a key determinant of transit service performance for both operators and travellers. Walker (1) has defined reliability as “Our ability to trust a transit operation”. Similarly, Abkowitz (2) described it as “the invariability of service attributes which influence the decisions of travellers and transportation providers”. In this regard, transit agencies are increasingly seeking to improve the reliability of transit services so as to positively influence passengers’ experience and achieve a higher transit ridership. Depending on the available data and transit operator’s policy about reliability, various reliability metrics are used in different cities.

Over the past few years, as the corollary of using smart card Automated Fare Collection (AFC) to collect revenues, the availability of massive and disaggregated passenger’s transactions data for transit planning and operational purposes has significantly increased. Depending on the fare collection scheme, invaluable source of travellers’ data such as departure and arrival times, passenger’s origins and destinations, selected transit routes, and transfers information are available or can be inferred by applying a series of assumptions. As a result, AFC data has attracted a lot of attention by transit agencies and researchers in the last several years. Pelletier, et al. (3) performed a comprehensive literature review on smart card data and showed their range of application, from strategic to tactical and operational levels. Amongst the studies that relied on smart card data, origin-destination demand estimation (e.g.4), travel patterns identification (5), ridership prediction (6), activity detection (7) and reliability measures (8-14) can be mentioned. In the case of the latter, AFC data can not only provide a better and more accurate measure of the existing indicators, but new reliability metrics can also be defined for a better representation of service performance for its users. Implementation of passenger-oriented reliability metrics is amongst the new possibilities AFC data has provided for transit agencies.

An increasing trend to shift from service-based reliability metrics to passenger-oriented ones can be seen in the literature (e.g. 12, 13, 15, 16). This is mainly due to the supply-oriented nature of traditional measures of variability and their potential discrepancy from the reliability when accounting for passengers’ travel patterns. Consequently, policies and strategies designed to improve service reliability may not necessarily result with the envisaged effects when considered from passenger’s perspective. Such shift toward passenger-oriented metrics thus seems crucial to reach a higher ridership and increasing agencies revenues.

Numerous measures of travel time reliability are developed and implemented in practice. Diab, et al. (14) performed a comprehensive literature review on bus service reliability and classified these measures into passengers’ and transit agencies’ perspective. In another study, Gittens and Shalaby (17) performed a review on the existing user-based reliability indicators and presented them through five different classes (travel time, schedule adherence, headway regularity, wait time, and composite indicators). They identified travel time and waiting time variabilities as the chief indicators of a transit service reliability. In another study, van Oort (11) conducted a survey to see how transit agencies measure their service reliability in practice and reported that the existing measures are mainly based on vehicle data. He observed the lack of a passenger-oriented measure of transit reliability in practice and accredited its necessity for a transit agency, suggesting AFC implementation as a viable dataset to develop and utilize reliability metrics in practice. To consider passenger impacts, Van Oort (18) suggested two metrics focusing on the extension of passenger travel time and its distribution. Similarly, Lee, et al. (19) deal with one of the shortcomings of supply oriented metrics, being transfers. They perform analyses of service reliability of multi-leg journeys.

A couple of studies proposed reliability metrics based on passenger’s travel time experience. These metrics reflect the predictability of the service for the passengers based on their experienced travel times. Uniman, et al. (9) defined the Reliability Buffer Time (RBT) measure as the time a passenger should allocate as buffer time to secure on time arrival to their destination with a pre-specified degree of uncertainty. They suggested the difference between 95th and 50th percentile travel time values as the representative of typical and acceptable levels, respectively. They extended this work by classifying the performance into recurrent and incident-affected ones. Then they introduced Excess Reliability Buffer Time (ERBT) metric as the difference between 95th percentile travel time

values in overall (i.e. including incident affected records) and typical dataset. These measures are appropriate in the context of high frequency services where passenger arrival pattern can be considered to be uniform.

In conjunction with predictability metrics, Transit service punctuality (i.e. adherence to the timetables) ones are the second category of the transit travel time variability, reflecting how close the service is to the promised (i.e. scheduled) one. London Underground defined Excess Journey Time (EJT) as the difference between the scheduled and experienced arrival time to the destination. It can be presented as the additional time that may be added to the scheduled travel time due to incidents, queuing and crowding delays (12). Zhao, et al. (10) developed this measure and performed a reliability measurement on the London rail network (Overground) using Oyster smart card data. Oyster users tap their card upon entries and exits of the system thus an assignment module was used to find passenger route choices, including transfer locations. In addition, as the tap incidents were happening before entering the platform, the data reflect not only the in-vehicle time, but a combination of waiting time, queuing delays, in vehicle delays and egress times.

One common feature of the regularity (e.g. RBT) and punctuality (e.g. EJT) metric is that both have been established for the case where smart card data validation is performed at stations. In other words, transaction records provide information about the location and time of passenger entrance to the system and their exit at destination. Consequently, chosen routes cannot be directly observed and models such as assignment tools are used to estimate them. Besides, the available travel time data include walking and waiting times. Therefore, as acknowledged by Hendren, et al. (12), such reliability metrics do not provide actionable information for the transit agencies. The present work thus seeks to address this issue by providing a framework for delivering a practical reliability measure for transit agencies and operators. Furthermore, as the developed method is relying on in-vehicle tapping system, detailed data about transfers helps forming a more realistic measure of passenger-oriented reliability.

This research introduces and implements two passenger-oriented measures of transit travel time reliability using smart card data. These measures correspond to user experience of punctuality (deviation from the schedule) and predictability (day to day variation) of the service for any spatial-temporal level of interest. The main contributions of this paper are:

- Developing a passenger-oriented transit reliability metric in terms of both regularity and punctuality
- A method for measuring transit reliability indicators from raw smart card transactions dataset
- A tool to measure service variability for any targeted time period, mode, line, and area of interest
- Demonstrating the analysis approach by measuring the performance of transit service in The Hague, the Netherlands

The rest of the paper is organized as follows. In Section 2, firstly the proposed reliability measures and mathematical formulations to calculate and scale them are presented. Then, the analysis framework and data preparations are described. Section 3 presents the application of the developed tool and reliability metrics to measure the reliability of the public transport network in The Hague, the Netherlands, followed by a discussion on the results. The paper is concluded by summarizing the findings and suggesting directions for future research.

2. METHODOLOGY

This Section elaborates the proposed passenger-oriented reliability metric and the steps required to derive them from AFC data. Firstly, two reliability measures, reflecting regularity and punctuality of the service as a passenger point of view are introduced. Mathematical formulation of these metrics as well as their scalability to have aggregated measures (e.g. reliability of a specific line or mode) are then provided. Then, the developed tool to measure the reliability metrics is introduced and key data processing and inference challenges and remedies to tackle them are explained.

2.1. Passenger Oriented Measure of Reliability

The approach in this study is to directly deduce service reliability from observed passenger trajectories. The two proposed reliability indicators of this paper target passenger-oriented predictability (day to day variation) and punctuality (deviation from the schedule). The former considers day-to-day variations in travel experience, whilst the latter refers to the discrepancy from scheduled travel times. The measures are calculated at origin-destination stop level and can then be scaled to measure the travel time reliability of a stop or terminal, a set of selected stops (e.g. bus stops along a route), or the whole network. The latter is particularly relevant for monitoring and assessing service performance by comparing how network performance evolves and benchmark the performance by comparing different networks.

The measures are formulated and implemented for a fare collection scheme where passengers validate their card upon boarding and before alighting each vehicle. The metrics account thus for service arrival time variability as well as transfer coordination. The time interval between the first in-vehicle tap-in to the last in-vehicle tap-out is therefore considered, including transfer times but excluding the initial passenger waiting time as well as access and egress times from and to origin and destination stops. The approach taken in this study is to analyse passenger travel time reliability in terms of extreme events which were found to have a disproportional effect on traveller perception and future choices (20).

The following notations were used throughout this paper to refer to the aforementioned attributes:

$n \in N$	Recorded smart card transaction n in smart card set N
$s \in S$	The stop where either check in or check out is recorded there
$l \in L$	Set of defined transit routes in the network
$t_{n,i}^o$	tap out time passenger n at stop i
$t_{n,i}^c$	tap in time passenger n at stop i
$l_{n,i}$	Transit route that passenger n is boarded at stop i
$r_{t,i}^l$	Trip id of scheduled line l service to arrive at stop i at time t

Daily variability of passenger travel time

The first metric measures the daily variability of trip travel time for a given time of the day. From passenger point of view, it is important that trip travel times are predictable (i.e. past experiences are indicative of future ones) and exercise little day to day variation so that expectations concerning arrival times at the destination are met. Similar to the reliability buffer time (RBT measure) introduced by Uniman, et al. (9) to evaluate rail service regularity, a Daily Variation reliability (DV) metric was developed based on on-board tapping scheme to reflect the effect of vehicle's operation on passenger travel time reliability. In addition to the in-vehicle travel time variability, DV measure reflects the effect of coordinating services for possible transfers.

The DV metric reflects the regularity of a service based on the experience of the passenger over a period of time. It measures the ratio of an upper percentile (ζ) and the travel time in typical conditions as the indicator of trip travel time variability. For a given OD-Pair i and j , the daily variation measure during a targeted period of the day (τ) is formulated as follows:

$$R_{i,j}^{DV}(\tau) = \frac{k^{\zeta}(tt_{i,j}) - k^t(tt_{i,j})}{k^t(tt_{i,j})} \quad 1$$

$$tt_{i,j} = \{t_{n,j}^o - t_{n,i}^c \mid t_{n,i}^c \in \tau, n \in R_{i,j}(D)\} \quad 2$$

Where

$R_{i,j}^{DV}(\tau)$	Passenger oriented regularity index for OD pair i,j during time period τ
$k^{\zeta}(tt_{i,j})$	Upper percentile ζ of travel time sets between OD pairs i and j
$k^t(tt_{i,j})$	Representative percentile of the typical travel time between OD pairs i and j
$R_{i,j}(D)$	All the available journeys between OD pairs i and j in analysis days D

The 95th percentile and the median (i.e. 50th percentile) were suggested by Uniman, et al. (9) as the upper(ζ) and typical(t) percentile of travel times, respectively. Wood (13) also used the 95th percentile threshold, suggesting that passengers find a once a month chance of late arrival acceptable. This metric therefore measures the extent to which abnormal unusual travel time experiences deviate from the central value of the travel time distribution for which there is the same number of longer travel times as there are shorter travel times. In order to enable the scalability of the metric for different OD pairs, time periods, modes and systems, the DV measure is presented as a dimensionless indicator. Consequently, 1 minute variability in a 5 minute service and 10 minutes variability in trips that has a median value of 50 minutes are identical.

Deviation from scheduled travel time

The second reliability metric reflects the punctuality of the service within a defined time period. This measure, defined as passengers based Schedule Deviation (SD) measure, reflects the deviation of individuals' actual travel time from the scheduled one. Having a sufficient sample size for each OD pair, the SD measure is defined as the ratio of the excess delay to travel time at a targeted percentile (e.g. median or 95th). For a given origin-destination pair, travel time punctuality is measured using the following formulation:

$$R_{i,j}^{SD}(\tau) = \overline{f_{sd}}(t_{n,i}^c, t_{n,j}^o, t_{n,i}^p, t_{n,j}^p) \quad \forall t_{n,i}^a \in \tau, n \in R_{i,j} \quad 3$$

$$f_{sd}(t_i^c, t_j^o, t_i^p, t_j^p) = \frac{k^\zeta (tt_{i,j} - tt_{i,j}^p)}{k^\zeta (tt_{i,j})} \quad 4$$

$$tt_{i,j} = \{t_{n,j}^o - t_{n,i}^c \mid t_{n,i}^c \in \tau, n \in R_{i,j}\} \quad 5$$

$$tt_{i,j}^p = \{t_{n,j}^p - t_{n,i}^p \mid t_{n,i}^a \in \tau, n \in R_{i,j}\} \quad 6$$

Where:

$R_{i,j}^{SD}(\tau)$	Passenger oriented punctuality index of OD pair i,j during time period τ
$tt_{i,j}$	Observed travel times between OD pair i,j
$tt_{i,j}^p$	Scheduled travel times between OD pair i,j
$t_{n,j}^p$	scheduled arrival time of passenger n to stop j

The calculated SD measure needs more parameters and calculations to be calculated as a connection between passenger trajectories and service timetables has to be formed. For a transferred trip, in addition to origin and destination stops, considering transfer points, routes, and walking times are also imperative and calculations are more costly. In this study, SD measures were calculated for $\zeta = 50$ and $\zeta = 95$, representing the median ratio of variability and the one that passengers may experience 5% of their time, respectively. Section 0 shows the details of the procedure to extract scheduled travel time for each journey.

Scaling the measures

The measures that were introduced in previous section estimate the reliability of the service for each origin-destination pair. These values can be aggregated to represent the variability of any set of trips. In this study, three levels of aggregation, namely bus stop level, transit lines level and network level are studied using the following formulations:

$$R_s(\tau) = \frac{\sum_{j \in S} R_{s,j}(\tau) \cdot q_{s,j}(\tau)}{\sum_{j \in S} q_{s,j}(\tau)} \quad 7$$

$$R_l(\tau) = \frac{\sum_{i \in S_l} \sum_{j \in S_l} (R_{i,j}(\tau) \cdot q_{i,j}(\tau))}{\sum_{i \in S} \sum_{j \in S} q_{i,j}(\tau)} \tag{8}$$

$$R_M(\tau) = \frac{\sum_{i \in S} \sum_{j \in S} (R_{i,j}(\tau) \cdot q_{i,j}(\tau))}{\sum_{i \in S} \sum_{j \in S} q_{i,j}(\tau)} \tag{9}$$

Where:

- $R_s(\tau)$ Measures of reliability(either DV or SD) for stop s
- $q_{i,j}(\tau)$ Number of passengers travelling from stop i to j at time period τ
- $R_l(\tau)$ Measures of reliability(either DV or SD) for line l
- $R_M(\tau)$ Network level measure of reliability(either DV or SD)
- S_l The set of stops of line l

2.2. Tool for Analysing Passenger-Oriented Reliability

A set of input data and preliminary tasks, metrics, and post-calculations are required in order to obtain the proposed reliability indicators. Figure 1 illustrates the analysis framework, consisting of the required input data and data preparation and analysis steps in yielding passenger-oriented measures of transit service reliability. In this section we elaborate the components of this framework. An application of this framework is presented in section 3 to demonstrate how the tool can be used to evaluate network performance.

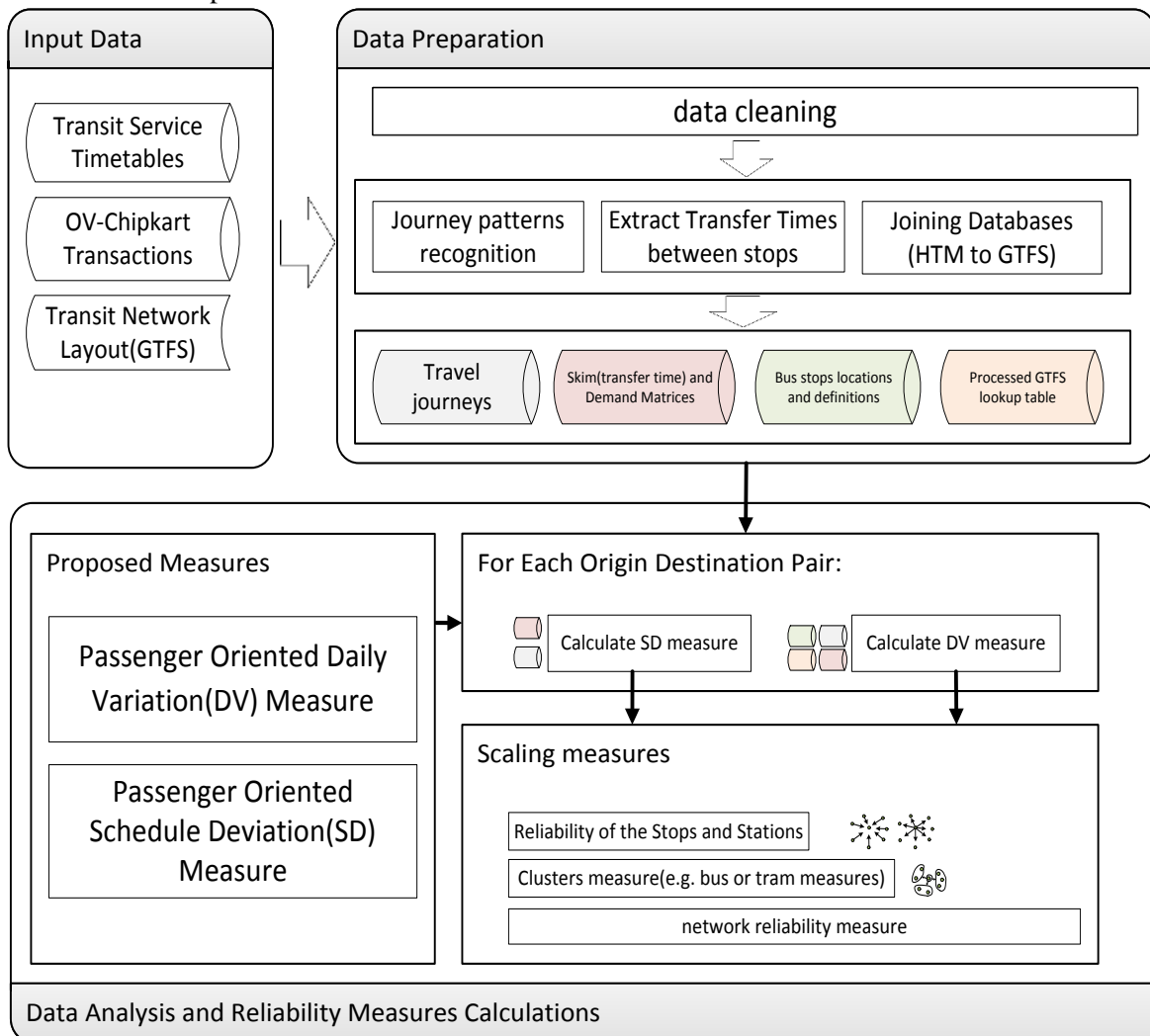


Figure 1 Proposed method to calculate passenger-oriented reliability

Input data

The presented methodology relies on three datasets: (1) recorded passenger transactions; (2) transit service timetables, and; (3) network layout including the stops locations and transit lines.

The main dataset is the smartcard transaction (SCT) data, recorded for each leg of the trip. Excluding cases with missing data, the raw data includes the time and location of passenger boarding and alighting, the transit line of the respective vehicle, and the passenger encrypted id. In order to deduce information about passenger travel time reliability, passengers' journeys need to be systematically inferred from single leg transactions.

General Transit Feed Specification (GTFS) (21) provides all the required data on network layout and service timetables as a set of standard CSV files. This data is processed to infer the scheduled service that is considered in SD (Eq. 3) when contrasting the planned and provisioned trip travel time. In contrast, the DV measure relies solely on transaction data.

Transaction datasets need to be pre-processed to extract the reliability metrics. Data preparation serves four main goals: removing outliers from missing values and incorrect records, forming journeys, linking transaction data with other datasets, and optimizing database structures to enable fast queries. These procedures are discussed in the following sub-sections.

From smartcard validations to passenger journeys

The very first step in data preparation process is to derive journeys from the tap-in-tap-out transactions. To do so, it is necessary to cope with data outliers. A series of filters was applied to handle outliers. After excluding data that contains missing values (e.g. missing tap-in or tap-out, no line identifier), Procedure 1 (Figure 2) is applied to construct the journey database. This procedure involves the specification of three parameters. Firstly, a minimum value of leg duration time (γ_{min}^{leg}) was defined to exclude transactions with short durations that are prone to fallacious records. The second parameter is the maximum duration of a single leg (γ_{max}^{leg}) to filter the trips that are abnormally long, mainly due to systematic error in recording the tapping timestamp. Finally, $\gamma^{transfer}$ is defined as the time interval between two successive legs with same card ID to classify the set of trips an individual performs on a given day into journeys. The output of this procedure is a database of the identified journeys, including an ID, date, number of transfers, and a list of details (line id, tap-in time and location, tap-out time and location) for each leg.

Procedure 1: Forming Journeys Database

Remove transactions with same location of tap-in and tap-out

Remove transactions between stops i, j with $t_{n,i}^o - t_{n,j}^c < \gamma_{min}^{leg}$

Remove transactions with abnormally long durations $t_{n,i}^o - t_{n,j}^c > \gamma_{max}^{leg}$

For each day in analysis period:

Group transactions using card ID

For each group:

Sort transactions using check-in timestamp

For each transaction r_n in R_n :

Form journey if $t_{r_n}^c - t_{r_{-1n}}^o < \gamma^{transfer}$

For each journey:

Calculate transfer times using walking distance

Calculate the number of transfers

Add journey to database

Figure 2 Developed Procedure to form journeys from transaction dataset

From passenger journey to reliability metrics

After the journey database has been generated, it is possible to obtain the DV metric for any given time window-location. However, to calculate the SD metric, journeys need to be matched to the

timetable dataset. This section first presents the approach adopted for linking the two datasets. Then, the algorithm for deriving the scheduled arrival time for each journey is described. Finally, a heuristic approach for restructuring the dataset in order to speed the processing time is presented.

Matching passenger journeys and service timetables

One of the necessary tasks of this study was to form a link between definitions of stops and lines and similar attributes in smart card transactions as these two databases was not consistently related to each other. To address this issue, a one-to-many line matching procedure was adopted so that a set of candidate GTFS lines was identified for each service line identified in the smartcard data. In each iteration the corresponding line is selected by considering the time discrepancy between the line recorded in the smartcard data and the scheduled arrival time of each of the candidate lines.

A similar approach was used in matching the stops in the two databases. The stops included in the SCT database refer often to both directions of the same line or a number of platforms. In this procedure, using stop coordinates, for each stop in the SCT database a set of GTFS stops was selected. The elements of this set were selected by applying two basic rules:

1. A GTFS stop is within an acceptable distance (e^d) from the SCT stop
2. There is at least one common line that serves both stops

For each transaction, the corresponding stop for each leg (tap-in or tap-out) was identified by checking if a corresponding trip can be made from origin stop to the destination one using the recorded transit line. In the unlikely event that more than one GTFS stop qualifies, the closest one (i.e. the stop that cause the minimum additional delay) is selected. The walking distances are then calculated between the stops introduced into the SCT database. Consequently, walking distances between the stops with same "parent-station" are uniformly assigned with a minimum value, γ_{min}^w , representing the minimum required time to perform a transfer.

Path Finding Algorithm

A path-finding algorithm is required for estimating the scheduled time of the service corresponding to each transaction. Using the boarding and alighting stops, service timetable, and the transit line of each leg of a journey, walking distance between the stops, and the timestamp of the first check-in, the corresponding scheduled journey can be inferred. A procedure was developed for obtaining the scheduled boarding and arrival times ($t_{n,i}^p, t_{n,j}^p$) for a direct (i.e. single leg) journey. For a given leg, firstly the corresponding candidate stops was determined. For each candidate stop pair, the connectivity via the recorded route name was checked. Then, the provided services that could potentially serve the passenger n were extracted. Finally, the service with minimum buffer from $t_{n,i}^c$ were identified and its timetable was used as $t_{n,i}^p$ and $t_{n,j}^p$ values. For a given tap-in time, the corresponding scheduled trip arrival time is determined by minimizing the discrepancy from passenger boarding time. This inference rule relaxes the assumption that the order in which buses arrive at stops corresponds to the order indicated in the timetable, reflecting the possibility of overtakings.

A method to calculate the scheduled travel time was also developed for non-direct journeys. In this regard, given a first tap-in time and a trajectory (lines and stops) as the arguments, a sequential procedure was designed for determining the scheduled arrival time at the destination. The transfer walking time is estimated based on the adjusted aerial distance between transfer points where a root of 2 was multiplied by the walking distances to convert aerial distances to distances on the network. Figure 3 illustrates a journey with one transfer and the procedure used for determining scheduled arrival times.

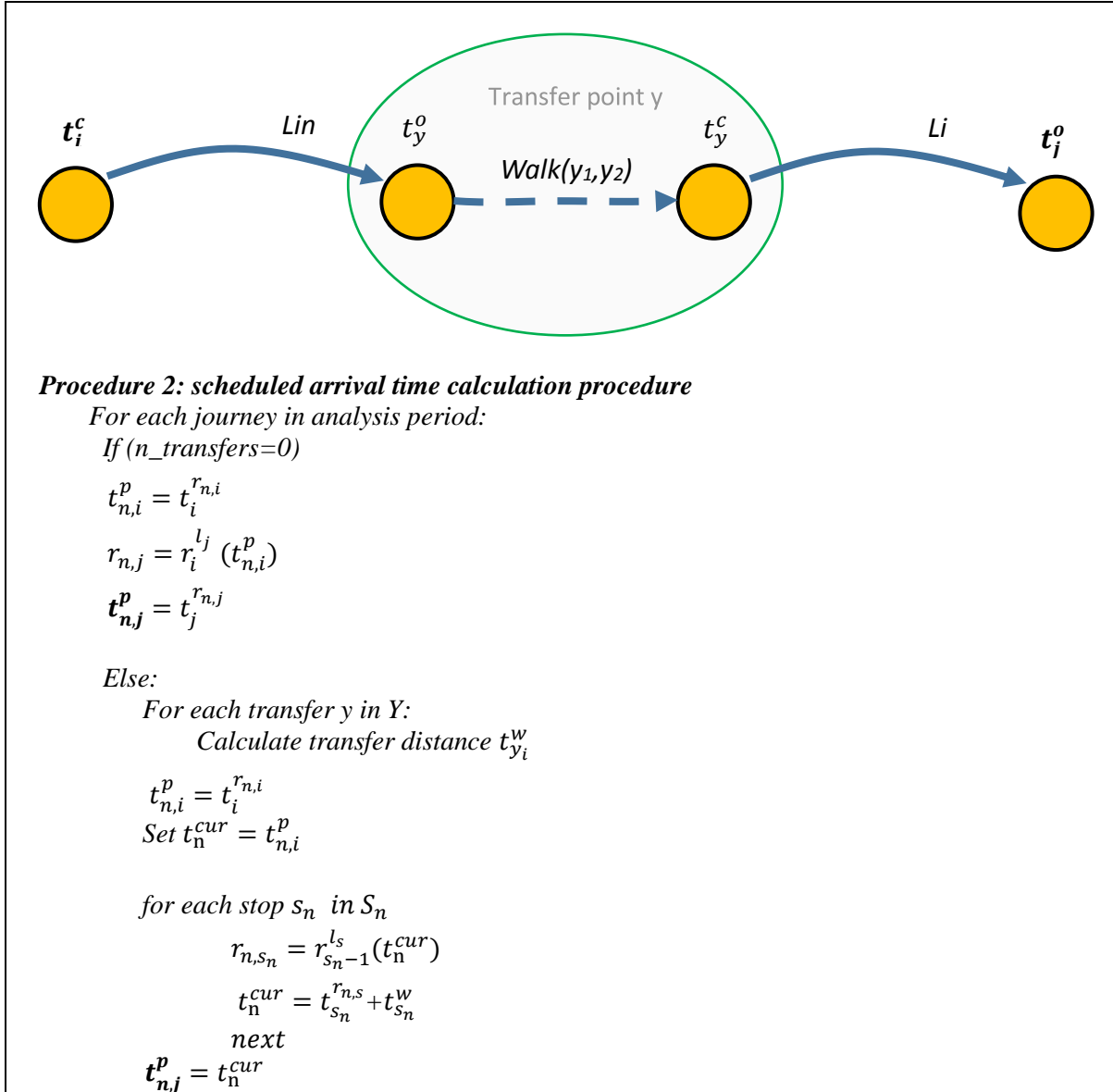


Figure 3 generic method to calculate the scheduled arrival time of the journeys

Once the scheduled arrival times of each journey are calculated, the measure of reliability for each OD pair for a given time period can be obtained. Unlike the DV measures that is derived from travel time distribution and thus requires a set of data for different days, the SD measure can be calculated for any individual journey since the timetable is used as the benchmark.

Restructuring datasets

The computational costs and feasibility of query and metrics calculation within a reasonable amount of time can become prohibitive for large datasets such as SCT. In the initial implementation of the proposed method, using the existing SCT database (transformed and optimized using SQL Server) and GTFS data files (loaded into memory once the software is initialized), the elapsed time for each query was noticeably high, refraining the evaluation of the simplest queries. For instance, running the program for two hours resulted with SD measures for around 2500 records, less than 0.05 percent of the monthly number of records. To improve the querying speed, two indexed look-up tables were formed to load journeys and timetables into memory and achieve a significant saving in computational cost of each query. Since the date and stop id are the two essential arguments for all queries, a two dimensional array with the size of D and S was formed and timetable and transactions

were loaded into this array in the initialization phase. Consequently, instead of searching over all days and all the stops for a corresponding line and time period, the search is conducted over the line and time period only. This improvement allows in the case study application to perform queries in a fraction of the time initially required (almost 30000 times faster). This made it possible to attain SD metrics for a typical AM peak period in a fraction of a second.

Passenger Reliability Tool Interface

The developed tool and metrics to evaluate passenger-oriented reliability of the transit service is available as a standalone executable application. The tool offers analysts, planners and decision makers an interface to obtain the results at the spatial-temporal level of interest. Figure 4 shows a screenshot of the interface. The application is using the language integrated querying (LINQ) component, allowing users to dynamically filter the target days, stops (origin and/or destination), mode sequence (e.g. bus, tram, first bus and then tram, etc.), analysis period (e.g. time of day, day of the week, specific dates), and transit line. Such flexibility along with the instant generation of passenger reliability metrics empowers transit operators and agencies by having a direct insight into the performance of the whole the network and individual network elements.

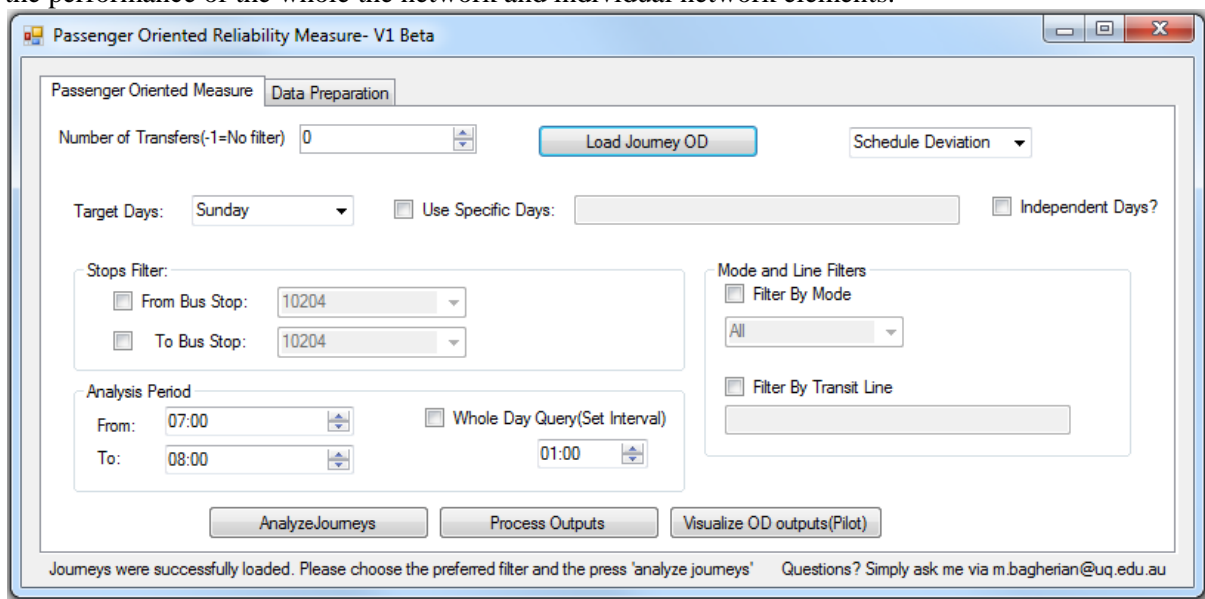


Figure 4 Main Interface of the developed Passenger-Oriented Reliability Measurement Tool

3. APPLICATION

This section shows the procedure and results of applying the suggested framework (including reliability metrics) on the The Hague transit agency's (HTM) services. This network consists of 12 tram and 8 bus lines and 923 stops (Figure 5). Transaction data from a single month (March 2015) period containing 8,177,434 records were analysed in this study. More details on the utilized smartcard data may be found in Van Oort et al. (2015). A representation of the network including the stops and defined routes can be found in <http://gtfs.ovapi.nl/htm/gtfs-kv1htm-20150227.kmz>. Using the developed tool, a study was performed using a top-down approach, to find out the sources of variability in the network during the analysis period.

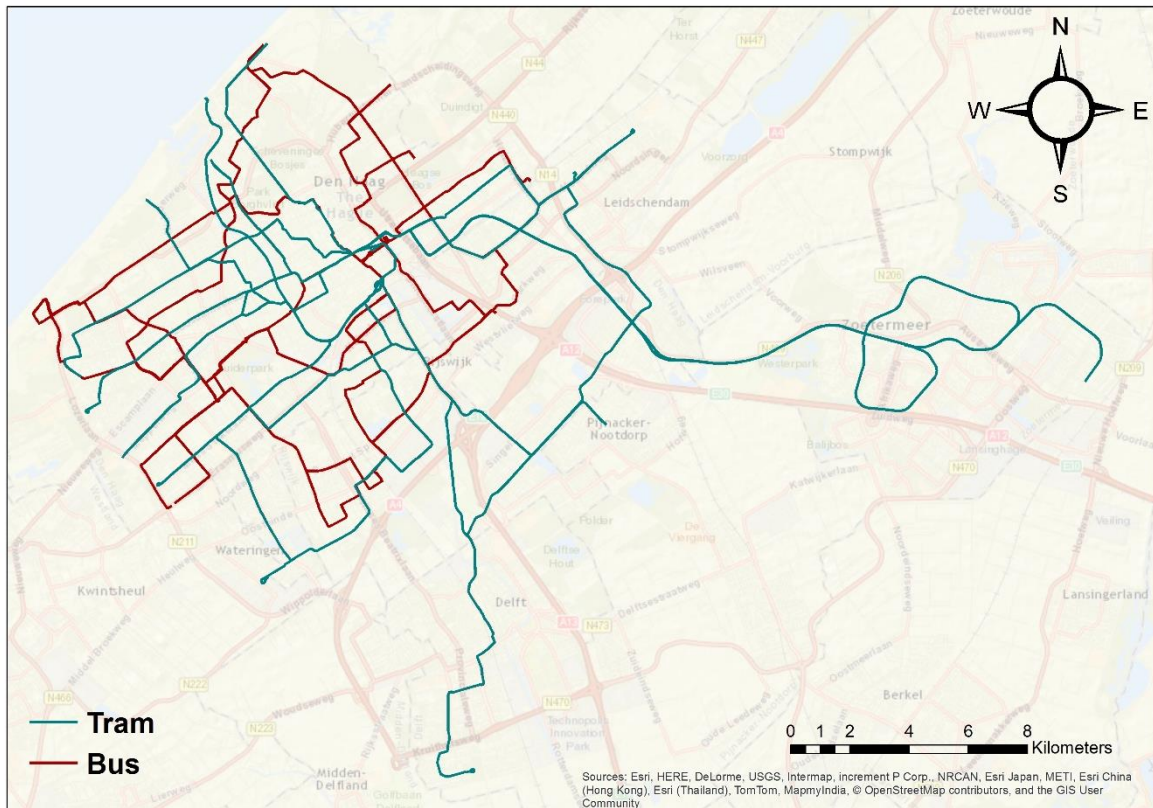


Figure 5 the public transport network of The Hague

In this section, firstly a review on the data preparation process is shown. Then, the descriptive statistic of the passenger trips, transferring behaviours, and spatial distribution of the demands are presented. The results of using daily variation and schedule deviation metrics to estimate passenger-oriented reliability of the service is then elaborated and discussed.

3.1. Descriptive Analysis of Transfer Patterns

The method to form passenger journeys database from transactions records (Figure 1) was applied to HTM database from March 2015. To remove the outliers, the incomplete transactions, the taps in same stop, and the ones outside the defined range of 60 seconds to 60 minutes (i.e. $\gamma_{min}^{leg} = 1min, \gamma_{max}^{leg} = 60min$, respectively) were excluded from the database. As a result, less than 1% of the records were discarded in the analysis. To compose the journeys, the definition established by the local public transport operator - identifying transfer for transactions within a time interval of up to 35 minutes (as $\gamma^{transfer}$) between two successive check-out and check-in - was adopted. Consequently, the procedure resulted with 6,265,185 journeys in the analysis period.

Figure 6 shows how the tap-in and tap-out times as well as average travel times varies over the day. The AM peak period is characterized by a short and well-defined peak in passenger load, whereas the PM peak period builds up over a long period and peaks at a level similar to the AM peak followed by a moderate decline. This pattern can be presumably attributed to the high share of part-time workers in the Netherlands. Journey duration (from first boarding time to last leg alighting time) is within the range of 15 to 20 minutes during most of the day.

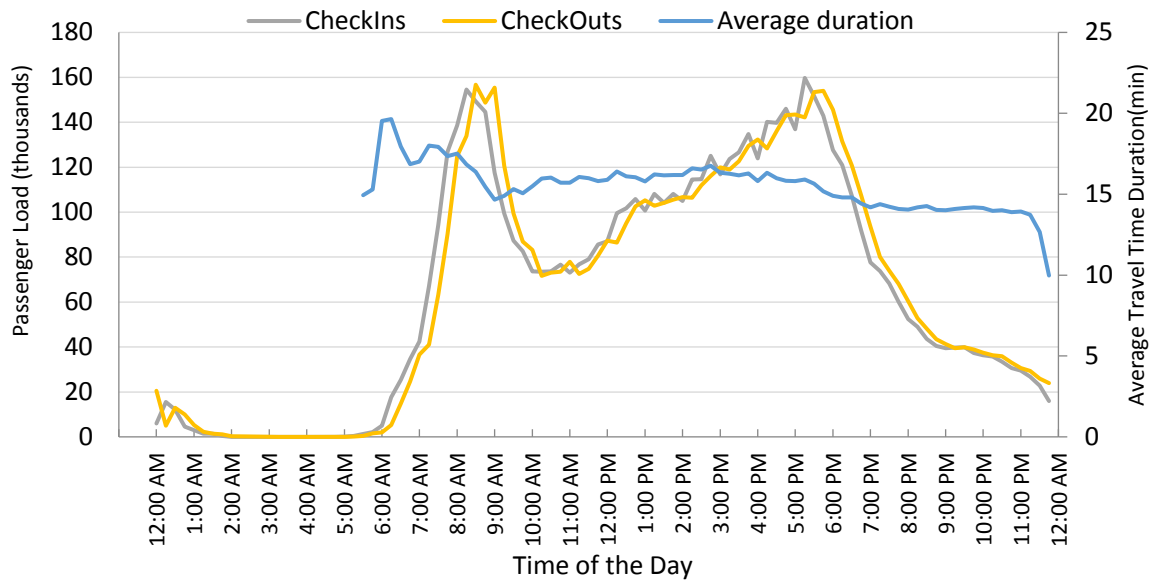


Figure 6 Daily Pattern of check-in, check-out and journey duration

Figure 7 shows the distribution of the observed transfer times. A pronounced peak can be observed at 3-4 minutes interval and 50% of the journey include transfer times that exceed 6:47 minutes. The average transfer time per journey is 9:35 minutes due to the long tail of the distribution which is sensitive to the transfer threshold parameter ($\gamma^{transfer}$). Note that the transfer times is calculated as the difference between two successive tap-out and tap-in in all the trips with transfer. Consequently, the transfer time includes the walking time between the stops as well as waiting for the downstream leg.

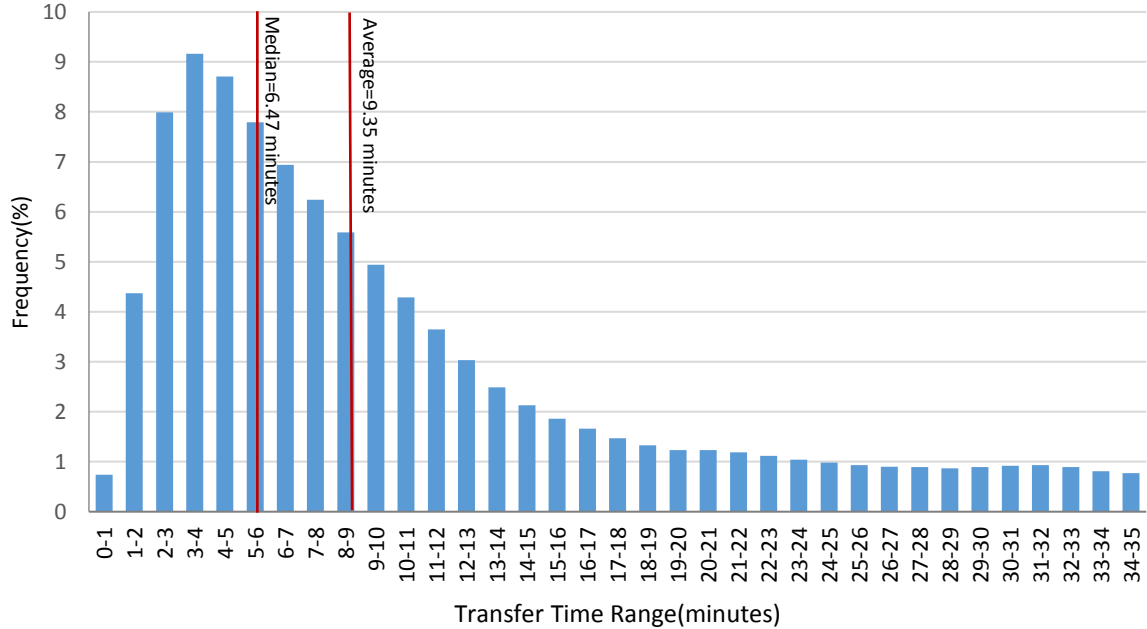


Figure 7 Distribution of Transfer Time Durations

Journeys were classified based on their number of transfer. Figure 8 shows the share and the travel time distributions of each class. It can be seen that around 80.5% of the trips are identified as direct journey that do not involve any transferring and journeys with one and two transfers have 17.1%

and 1.8% of total number of journeys, respectively. Only 0.6% of the journey were found to comprise of more than two transfers.

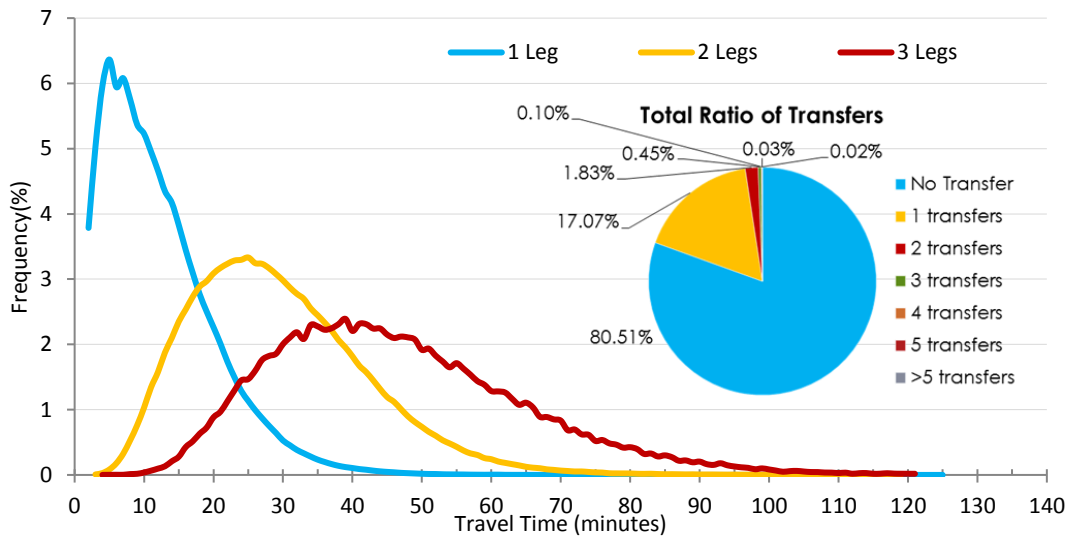


Figure 8 Travel time by number of transfers and a break-down of the journeys by number of transfers

Demand spatial distribution was also investigated using the origin and destination of the trips and their corresponding coordinates. Figure 9 shows the schematic map of the trip generation on an average weekday. As can be seen, a significant ratio of the demand is originates at few stops. In fact, 20% of the bus stops accounted for the generation of more than 70% of the trips. An analysis of the demand matrix indicates that all the journeys recorded during the analysis period were associated with merely 18% of the possible OD pairs in the case study area. This uneven demand distribution stresses the importance of using passenger-oriented measures in evaluating service performance.

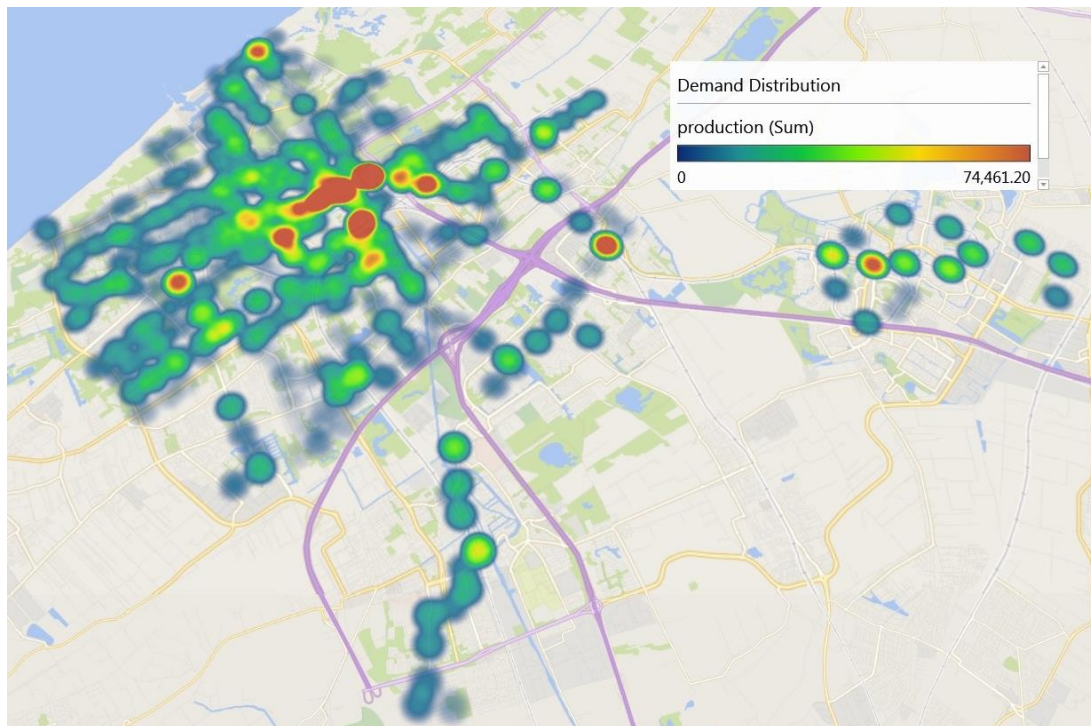


Figure 9 Schematic View of Demand Distribution across the case study network(Based on smart card data)

The main output of the data preparation module is the journey database that can be used to measure reliability for various spatial-temporal levels, as illustrated and discussed in the following section.

3.2. Results

The proposed framework along with the passenger-oriented reliability metrics was implemented for the HTM dataset. According to the operator's records, no major incident or failure occurred during the analysis period and services generally operated as planned". In the following, we present the results of our analysis at a network-wide level for the entire month of March 2015 as well as an illustration for a specific instance which is used to illuminate the potential insights that can be gained by further investigating the spatial and temporal variations in service performance.

The DV metric was first calculated for different days and time periods. For each OD pair, the DV measure was calculated if the sample size exceeds a pre-defined value (10 observations here) using Equations 1-2. The values were then scaled to the network level using Equations 7-9. Figure 10 shows the results of calculating the network level DV for different time periods. It can be seen that during the analysis period, AM peak travellers experienced more reliable services than during other periods if they rely on the total journey time experienced on former trips on the same day of the week. For example, extremely long travel times on weekdays are 25% longer than the median travel time during the afternoon peak period. It can also be observed that travellers have the highest level of variability on Monday morning and afternoon as well as Friday PM peak compared with the corresponding time periods on other days of the week, suggesting that the start and end of the working week are subject to more unusual travel conditions.

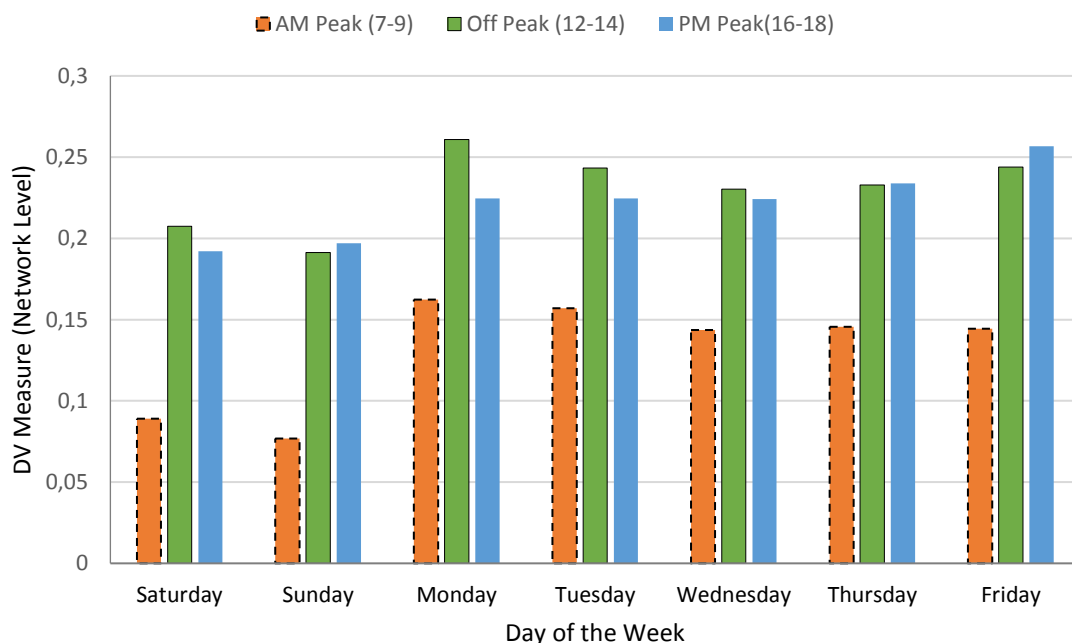


Figure 10 Comparison of Travel Time Predictability Measure (DV) for Different Days of the Week

To gain a better insight into the daily variation patterns, the SD of the AM peak period was calculated for each day of the month and is displayed in

Figure 11. In addition, the corresponding passenger load across the network is depicted. Most weekdays are characterized by a SD value in the range of 0.25-0.35. This implies that 5% of the passengers experience a travel time that is 25 to 35% longer than the scheduled journey travel time. In other words, an average commuter will experience once in two weeks a travel time that exceed by more than 25% the travel time expected based on the schedule. It can be observed that weekends are characterized by lower demand levels and higher service reliability. Notwithstanding, lower service reliability does not necessarily occur when higher passenger loads are observed. For example, March 25 has a highest level of SD, although it is not the day that experienced the highest demand level. This day, along with March 18 which can be considered a representative of a typical day for comparison purposes, were selected for further investigation.

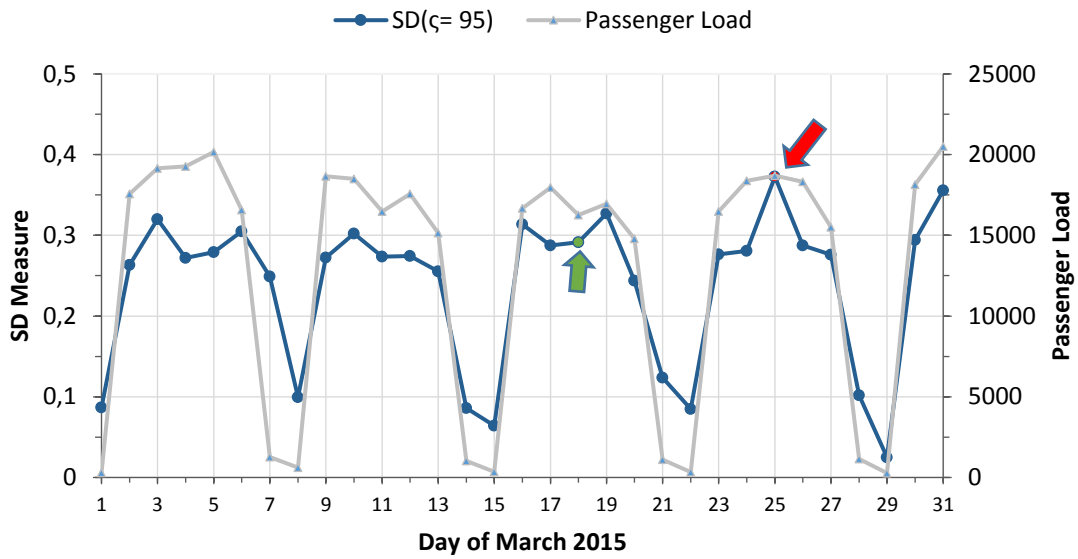


Figure 11 Passenger Load and Network Level Reliability Measure in Different Days (AM Peak Period)

Passenger reliability by mode

The HTM transit network consists of 12 tram and 8 bus lines. Reliability was analysed for each mode separately to compare their performance and the results are presented in Table 1. In addition to the higher passenger load, it can be seen that the reliability of buses and trams was worse on March 25 than on March 18. Comparing buses and trams, it can be seen that for passengers, trams were consistently less reliable than buses during the AM peak period of the selected days. To find out the reason of experiencing such variability in tram lines, further analysis was performed.

Table 1 Reliability Measures of Buses and Trams in AM Peak Period(7:30-8:30AM)

MEASURE		MARCH 18TH		MARCH 25TH	
		Tram	Bus	Tram	Bus
Passenger Load		13839	2030	15577	2665
Travel time [sec]		803	823	822	816
Absolute deviation [sec]	$\zeta = 50$	61	35	96	65
	$\zeta = 95$	206	135	305	176
Schedule deviation	$\zeta = 50$	0.09	0.04	0.16	0.08
	$\zeta = 95$	0.31	0.17	0.40	0.22

Passenger reliability per line

Schedule deviation was analysed for each specific tram line of the network using Eq. 8. Figure 12 shows the comparison of SD measure for different tram lines for March 18 and March 25. It can be seen that two lines (Line 3 and 4) experienced the greatest deterioration in reliability when comparing these two days. Tram Line 4 was selected for a more detailed analysis.

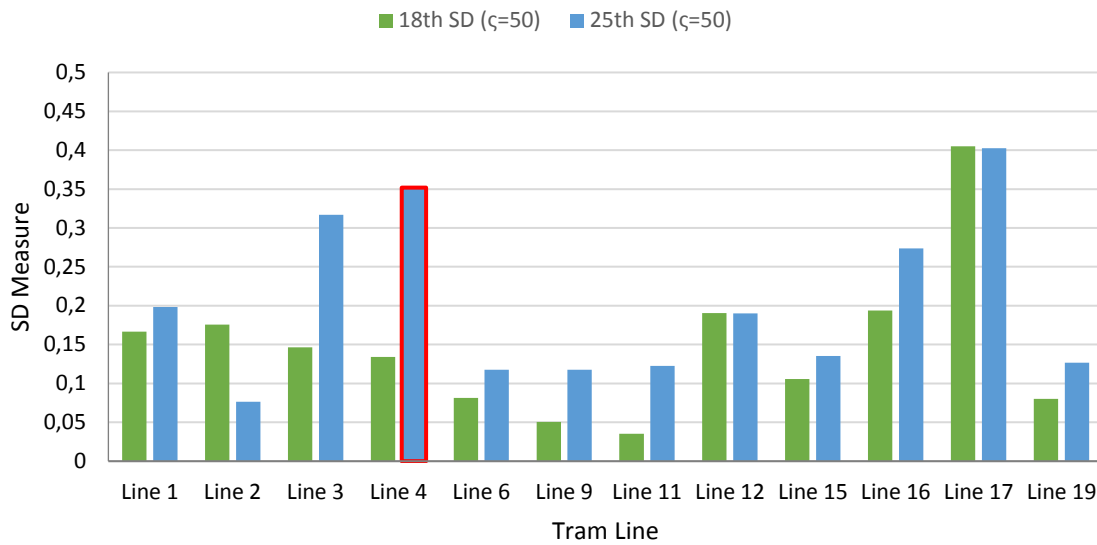


Figure 12 Comparison of Tram Lines Reliability Measure in two selected days (Line 4 is highlighted as the critical one)

Passenger reliability for a given line

The SD measure was calculated for the selected line for the whole analysis period of March 2015. As can be seen in Figure 13, line 4 had a SD measure below 0.15 (indicating 15% delay in arrival time) for the vast majority of days, with the exceptions of March 25 and March 28. This sudden change coincides with a noticeable increase in passenger load on this line.

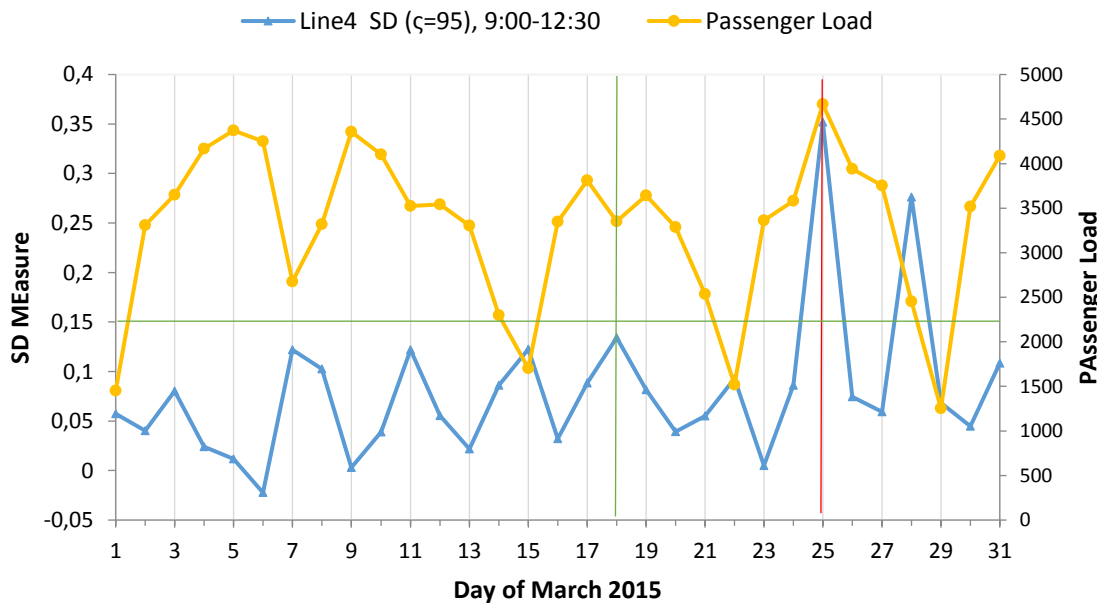


Figure 13 Passenger Load and Reliability Measure of Tram Line 4 in Different Days of the Analysis Period

The daily trend of Line 4 was compared in one hour intervals for both March 18 and March 25 (Figure 14). The SD measure does not follow the AM and PM demand peaks. On March 25, the most noticeable change in service reliability of line 4 occurred directly after the AM peak (9 AM- 1 PM) and PM peak (6-7 PM).

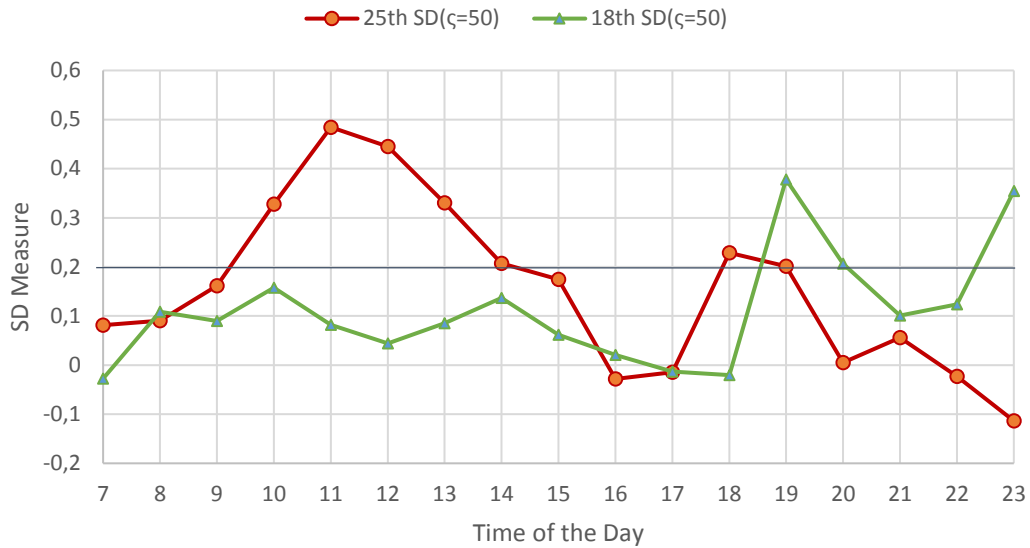


Figure 14 Comparison of Tram Line 4 Reliability Measure between 18th and 25th March 2015

Considering the variability changes in different times of the day in smaller intervals (15min), 9:45-10:15am was identified as the interval where a noticeable increase in service variability of tram line 4 occurs. For This period, the SD measure at stop levels was generated for each stop along the line. Figure 15 shows a schematic comparison of the person delay changes along the line on a typical day (March 18) and the abnormal day (March 25). While service reliability worsened for almost all stops during the time period selected for analysis, certain locations experienced a particularly extreme increase in passenger delay (denoted by red colour). Such a fine interval-scale insight can help transit agencies to precisely identify the variability sources and potentially design measures to mitigate them.

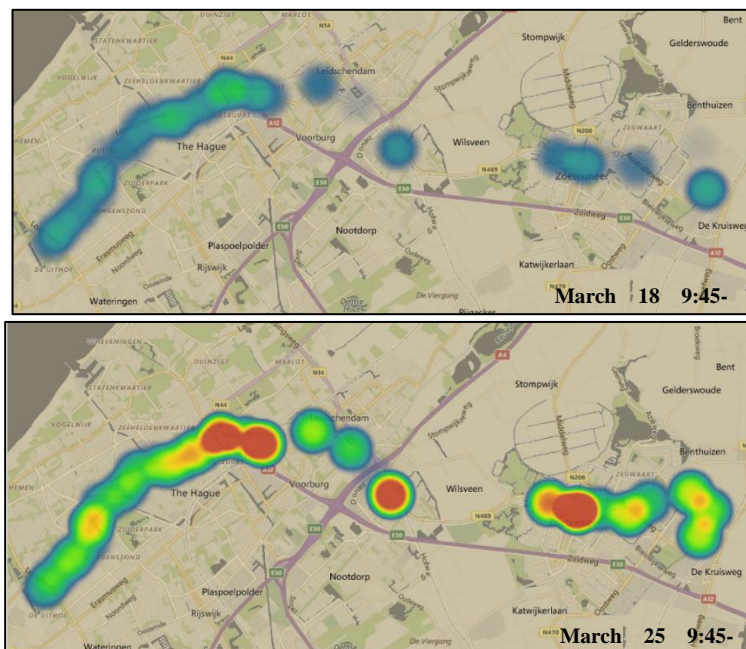


Figure 15 Person-Delay visualization of tram Line 4 for the targeted abnormal period

4. CONCLUSIONS

In this paper, a comprehensive tool to estimate the passenger-oriented measure of transit service reliability was presented. This method included two novel metrics to reflect both predictability and

punctuality of the network in the selected spatial-temporal window, measured at the passenger journey level. The methodology was presented along with the data analytics procedures. With the availability of reliability measure at origin-destination level, the method for measuring the reliability at stops, lines, transit modes, and for the network as a whole was developed as a tool with a user friendly interface. While the aggregated level of punctuality and regularity can be derived at the network level, disaggregate measures can aid agencies and operators in identifying the locations and times where maximum person-delay may occur, as demonstrated in the application. In the absence of tap in and tap out transaction data, methods to infer passenger routes and transfer locations need to be included in the data preparation and processing data to obtain passenger reliability metrics.

Passenger-oriented reliability metrics can lead to a paradigm shift in managing transit services from planning to operations. The introduction of incentive schemes based on passenger reliability metrics rather than vehicle-based performance measures will assist service providers in focusing on remedies such as preferential treatments, rescheduling the services, and transfer points coordination where and when most needed in terms of their consequences for passengers. Monitoring passenger reliability in real-time can facilitate steering operations towards passengers' experience. Research into passengers' perception of service reliability will enable differentiating between different journey components and account for their contribution to the overall passenger experience. Future research will also further improve the reports and visualizations produced by the tool. Different networks and transit services can be assessed to allow the analysis of performance evolution by comparing with past performance as well as benchmarking by comparing with peer networks. Transit agencies can utilize the tool to identifying the sources of unreliability by incorporating other sources of information (e.g. precipitations, incidents, etc.).

5. ACKNOWLEDGEMENTS

The study was supported by Haagsche Tramweg Maatschappij (HTM). Special thanks to March Drost, Rien van Leeuwen and Menno Yap for their support and input.

6. REFERENCES

- (1) J. Walker, *Human transit: How clearer thinking about public transit can enrich our communities and our lives*: Island Press, 2012.
- (2) M. Abkowitz, *Transit service reliability*, 1978.
- (3) M.-P. Pelletier, M. Trépanier, and C. Morency. Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, Vol. 19, No. 4, 2011, pp. 557-568.
- (4) J. Chan, "Rail transit OD matrix estimation and journey time reliability metrics using automated fare data," Massachusetts Institute of Technology, 2007.
- (5) X. Ma, Y.-J. Wu, Y. Wang, F. Chen, and J. Liu. Mining smart card data for transit riders' travel patterns. *Transportation Research Part C: Emerging Technologies*, Vol. 36, No. 2013, pp. 1-12.
- (6) N. Van Oort, T. Brands, and E. de Romph. Short term ridership prediction in public transport by processing smart card data. *Transportation Research Record*, No. 2015, pp.
- (7) N. Nassir, M. Hickman, and Z.-L. Ma. Activity detection and transfer identification for public transit fare card data. *Transportation*, Vol. 42, No. 4, 2015, pp. 683-705.
- (8) C. Morency, M. Trépanier, and B. Agard. Measuring transit use variability with smart-card data. *Transport Policy*, Vol. 14, No. 3, 2007, pp. 193-203.
- (9) D. Uniman, J. Attanucci, R. Mishalani, and N. Wilson. Service Reliability Measurement Using Automated Fare Card Data. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2143, No. 2010, pp. 92-99.
- (10) J. Zhao, M. Frumin, N. Wilson, and Z. Zhao. Unified estimator for excess journey time under heterogeneous passenger incidence behavior using smartcard data. *Transportation Research Part C: Emerging Technologies*, Vol. 34, No. 2013, pp. 70-88.

- (11) N. van Oort. Incorporating service reliability in public transport design and performance requirements: International survey results and recommendations. *Research in Transportation Economics*, Vol. 48, No. 2014, pp. 92-100.
- (12) P. Hendren, J. Antos, Y. Carney, and R. Harcum, Transit Travel Time Reliability: 2 Shifting the Focus from Vehicles to Customers 3, in *Transportation Research Board 94th Annual Meeting*, 2015.
- (13) D. A. Wood, "A framework for measuring passenger-experienced transit reliability using automated data," Massachusetts Institute of Technology, 2015.
- (14) E. I. Diab, M. G. Badami, and A. M. El-Genieidy. Bus Transit Service Reliability and Improvement Strategies: Integrating the Perspectives of Passengers and Transit Agencies in North America. *Transport Reviews*, Vol. 35, No. 3, 2015, pp. 292-328.
- (15) S. Chakrabarti and G. Giuliano. Does service reliability determine transit patronage? Insights from the Los Angeles Metro bus system. *Transport Policy*, Vol. 42, No. 2015, pp. 12-20.
- (16) O. Cats. Regularity-driven bus operation: Principles, implementation and business models. *Transport Policy*, Vol. 36, No. 2014, pp. 223-230.
- (17) A. Gittens and A. Shalaby. Evaluation of Bus Reliability Measures and Development of a New Composite Indicator. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2533, No. 2015, pp. 91-99.
- (18) N. Van Oort. Incorporating enhanced service reliability of public transport in cost-benefit analyses. *Public Transport*, No. 2016, pp. 1-18.
- (19) A. Lee, N. van Oort, and R. Van Nes. Service reliability in a network context: Impacts of synchronizing schedules in long headway services. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2417, 2014, pp. 18-26.
- (20) M. Friman, B. Edvardsson, and T. Gärling. Frequency of negative critical incidents and satisfaction with public transport services. I. *Journal of Retailing and Consumer Services*, Vol. 8, No. 2, 2001, pp. 95-104.
- (21) M. Catala, S. Dowling, and D. Hayward, *Expanding the google transit feed specification to support operations and planning*, 2011.