DISS. ETH NO. 27741

TRACKING PASSENGERS TO ANALYSE TRAVEL BEHAVIOUR DURING PUBLIC TRANSPORT DISTURBANCES

A dissertation submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

ALESSIO DANIELE MARRA
M.Sc. Università degli Studi Roma Tre

born on 20 October 1993
citizen of Italy

accepted on the recommendation of

Prof. Dr. F. Corman, examiner
Prof. Dr. M. Bierlaire, co-examiner
Prof. Dr. O. A. Nielsen, co-examiner

2021

ABSTRACT

Public transport networks are characterized by daily disturbances, as delays or cancelled runs, which may negatively affect the passengers. In fact, in case of disturbances, the travel time may increase, or some routes may become unavailable, resulting in a discomfort for passengers. In these cases, passengers' behaviour can vary significantly, from sticking to the original plan to choosing a different route, mode or destination. Therefore, understanding passengers' behaviour in case of disturbances is a complex and relevant task, given the heterogeneous nature of both the passengers and the disturbances. In fact, the same passenger may react differently from other passengers or in case of different disturbances. To analyse this aspect, a large number of data is needed, including long-term observations of several passengers. In this regard, GPS tracking is a promising technology, allowing a long-term data collection of precise information on the travel behaviour of several users.

This dissertation aims to exploit GPS tracking to understand public transport passengers' behaviour in case of disturbances. In particular, a major focus is the development of algorithms for automatic mode detection and collection of long-term travel diaries from GPS data. Therefore, the travel diaries of a large amount of users are exploited to analyse route choices in public transport, with a focus on network disturbances.

To understand passengers' behaviour in case of disturbances, this research pursues four different objectives, which are addressed in four different chapters of the thesis. A fifth chapter shows the application of the GPS tracking and the proposed methodology to study passengers' behaviour during the COVID-19 pandemic.

Chapter 2 focuses on automatic collection of travel diaries from tracking data. A case study is carried out to test a low-battery consumption smartphone application, allowing long-term passive tracking with a low burden on respondents. Several algorithms are developed for trip, activity and mode detection, to automatically collect travel diaries of different users. In particular, the proposed mode detection algorithm exploits past travel information from the users to improve its detection accuracy, and is able to detect the specific public transport vehicle used.

Chapter 3 focuses on identifying the available alternatives for a given public transport trip. A novel choice set generation algorithm is proposed,

identifying all and only the relevant alternatives, which more realistically are taken into consideration during the choice process of the passenger. The algorithm is computationally efficient and able to work with different information provisions. It is evaluated on a large-scale tracking survey, obtaining high coverage (more than 94%) and the estimation of meaningful models. In this regard, multiple route choice models are estimated to understand the most likely information that passengers consider when choosing their route.

Chapter 4 focuses on evaluating public transport disturbances. A novel definition of disturbance impact is proposed, able to describe different degrees of disturbances based on the potential impact on passengers. Several disturbances are identified from realized public transport data, and their impact is analysed on simulated passengers' trips. Therefore, the relationships between the disturbance characteristics and the impact are analysed to identify the main characteristics of a disturbance.

In Chapter 5, the methods and the results of the previous chapters are combined, to analyse passengers' route choices in case of different network disturbances. Namely, information on passengers' route choices, available alternatives, network conditions and disturbances is collected in a large-scale integrated dataset. The passengers' chosen routes are compared with the available alternatives with and without disturbances, and the effects of disturbances on travel cost are evaluated. In particular, the analysis identifies that small disturbances and delays have a significant effect on travel cost, although they have marginal effects on route choice. Moreover, new available connections, unavailable in the timetable, are often not exploited by passengers.

Finally, Chapter 6 shows the application of GPS tracking and the proposed methodology, to understand travel behaviour in exceptional conditions, such as the COVID-19 pandemic. In this regard, a tracking survey is carried out during the first pandemic wave in 2020 in Zürich. Since the respondents participated also in a former survey in 2019 (the one described in Chapter 3), travel behaviours in a pre-pandemic period and in a pandemic period are compared. Among the main outcomes, a sharp reduction of travel distance is observed during the pandemic, in conjunction with the implemented restrictions. In this regard, public transport was more affected than private modes. Moreover, the estimation of a route choice model for public transport shows important differences between the two periods in the perception of costs related to transfers and travelling by train.

In summary, this dissertation shows a full process to understand public transport passenger behaviour, with a focus on route choice and public transport disturbances. Each step of the process, including the data collection, the estimation of a behavioural model, and the evaluation of disturbances, is addressed proposing a novel methodology, filling the related gaps identified in literature. In this regard, the leitmotif of the thesis is given by the GPS tracking. In fact, the availability of highly detailed long-term travel diaries, thanks to this technology, allows the observation of passenger behaviour in specific conditions, as in case of disturbances or during the COVID-19 pandemic.

The results of this dissertation can be beneficial for operating companies, infrastructure managers and the public transport industry in general. In fact, understanding passengers' behaviour in case of public transport disturbances is a crucial element to design a response to them. The GPS tracking, moreover, proved to be a powerful means to observe passengers' choices, which can replace stated preference surveys in certain applications. Ultimately, the analyses in this work are also useful for passengers, since operators may exploit them to provide a more passenger-oriented service.

ZUSAMMENFASSUNG

Öffentliche Verkehrsnetzwerke charakterisieren sich durch tägliche Störungen, wie Verspätungen oder Ausfälle, welche die Passagiere negativ beeinflussen. Bei Störungen kann sich die Reisezeit erhöhen und in manchen Fällen können gewisse Routen von Passagieren nicht mehr gewählt werden, was eine Unannehmlichkeit für die Passagiere darstellt. In solchen Fällen kann das Verhalten der Passagiere signifikant variieren: Passagiere können an ihrem ursprünglichen Plan festhalten, eine andere Route, ein anderes Verkehrsmittel, oder auch ein anderes Ziel wählen. Das Verständnis des Verhaltens von Passagieren bei Störungen ist daher eine komplexe und relevante Aufgabe, weil sowohl die Passagiere als auch die Störungen heterogen sind. Verschiedene Passagiere reagieren möglicherweise unterschiedlich auf eine Störung und die Reaktion eines spezifischen Passagiers hängt von der konkreten Störung ab. Um diesen Aspekt zu analysieren, wird eine grosse Datenmenge benötigt, einschliesslich Langzeitbeobachtungen von mehreren Passagieren. In dieser Hinsicht stellt das GPS-Tracking eine vielversprechende Technologie dar, die eine langfristige Datenerfassung präziser Informationen zum Reiseverhalten verschiedener Benutzer ermöglicht.

Diese Dissertation hat das Ziel, mittels GPS-Tracking das Verhalten von Fahrgästen im öffentlichen Verkehr bei Störungen zu verstehen. Der Hauptfokus liegt speziell auf der Entwicklung von Algorithmen zur automatisierten Erkennung von Verkehrsmitteln und Erfassung von Langzeitreisetagebüchern aus GPS-Daten. Die Reisetagebücher von einer grossen Anzahl von Passagieren werden genutzt, um die Routenauswahl im öffentlichen Verkehr zu analysieren, wobei der Schwerpunkt auf das Verhalten der Passagiere bei Störungen im Netzwerk liegt.

Um das Verhalten der Passagiere bei Störungen zu verstehen, verfolgt diese Studie vier verschiedene Ziele, die in den vier Kapiteln der Arbeit behandelt werden. In einem fünften und letzten Kapitel wird eine Anwendung des GPS-Trackings aufgezeigt und eine Methodik zur Untersuchung des Verhaltens der Passagiere während der COVID-19-Pandemie vorgeschlagen.

Das Kapitel 2 konzentriert sich auf die automatische Erstellung von Reisetagebüchern aus Tracking-Daten. Eine Fallstudie wird durchgeführt, um

eine Smartphone-Anwendung mit geringem Batterieverbrauch zu testen, die eine langfristige passive Verfolgung der Teilnehmer der Studie ermöglicht, welche für die Teilnehmer einen geringen Aufwand darstellt. Für die Erkennung von Reisen, Aktivitäten und Verkehrsmittel wurden verschiedene Algorithmen entwickelt, um automatisch Reisetagebücher verschiedener Benutzer zu erfassen. Insbesondere nutzt der vorgeschlagene Algorithmus Informationen bezüglich früherer Reisen der Benutzer, um seine Lokalisierungsgenauigkeit zu verbessern und kann das gewählte öffentliche Fahrzeug erkennen.
Das Kapitel 3 konzentriert sich auf die Ermittlung der möglichen Alternativen für eine gewünschte Fahrt mit öffentlichen Verkehrsmitteln. Es wird ein neuartiger Entscheidungsalgorithmus vorgeschlagen, der alle relevanten Alternativen identifiziert, die während des Auswahlprozesses des Passagiers realistischer Weise berücksichtigt werden. Der Algorithmus ist effizient und kann mit verschiedenen Informationsformaten verwendet werden. Er wird anhand einer gross angelegten Tracking-Umfrage ausgewertet, wobei eine hohe Klassifikationsgenauigkeit (mehr als 94%) und die Schätzung aussagekräftiger Modelle erreicht werden. In dieser Hinsicht werden mehrere Routenauswahlmodelle geschätzt, um dieInformationslage der Passagiere, basierend auf deren Routenwahl, zu repräsentieren.
Das Kapitel 4 beschäftigt sich mit der Auswertung von Störungen im öffentlichen Verkehr. Eine neuartige Definition für die Auswirkung einer Störung wird vorgeschlagen, mit der unterschiedliche Störungsgrade basierend auf potentiellen Auswirkungen auf die Fahrgäste beschrieben werden können. Aus den beobachteten Daten des öffentlichen Verkehrs werden mehrere Störungen identifiziert und ihre Auswirkungen auf simulierte Reisen von Fahrgästen analysiert. Dabei werden die Beziehungen zwischen den Störungseigenschaften und den Auswirkungen analysiert, um die Hauptmerkmale einer Störung zu identifizieren.
Im Kapitel 5, werden die Methoden und Ergebnisse der vorherigen Kapitel kombiniert, um die Routenwahl der Passagiere bei verschiedenen Netzwerkstörungen zu analysieren. Informationen in Bezug auf die Routenwahl der Passagiere, verfügbare Alternativen, Netzwerkzustand und Störungen werden zu einem grossen Datensatz zusammengefasst. Die von den Passagieren gewählten Routen werden mit den verfügbaren Alternativen mit und ohne Störungen verglichen und die Auswirkungen von Störungen auf die Reisekosten werden bewertet. Die Analyse zeigt insbesondere, dass kleine Störungen und Verspätungen einen erheblichen Einfluss auf die Reisekosten haben, obwohl sie nur geringfügige Auswirkungen auf

die Routenwahl haben. Darüber hinaus werden neu verfügbare Verbindungen, die im Fahrplan nicht verfügbar sind, von Passagieren nicht häufig gewählt. Schliesslich zeigt Kapitel 6 die Anwendung des GPS-Trackings und der vorgeschlagenen Methodik, um das Reiseverhalten unter aussergewöhnlichen Bedingungen wie der COVID-19-Pandemie zu verstehen. In diesem Zusammenhang wird während der ersten Pandemiewelle im Jahr 2020 in Zürich eine Tracking-Umfrage durchgeführt. Da die Befragten auch an einer früheren Umfrage im Jahr 2019 teilgenommen haben (die in Kapitel 3 beschrieben ist), kann Ihre Verhaltensweise von zuvor zu jener während der Pandemie verglichen werden. Insbesondere konnte eine starke Verringerung der Reisedistanz während der Pandemie in Verbindung mit den umgesetzten Beschränkungen festgestellt werden. In dieser Hinsicht war der öffentliche Verkehr stärker betroffen als private Verkehrsmittel. Darüber hinaus zeigt die Schätzung eines Routenwahlmodells für den öffentlichen Verkehr wichtige Unterschiede zwischen den beiden Zeiträumen bei der Wahrnehmung der Kosten im Zusammenhang mit Umsteigevorgängen und bei Zugreisen.

Zusammenfassend zeigt diese Dissertation einen vollständigen Prozess zum Verständnis des Passagierverhaltens im öffentlichen Verkehr, wobei der Schwerpunkt auf der Routenwahl und den Störungen des öffentlichen Verkehrs liegt. Für jeden Schritt des Prozesses, einschliesslich der Datenerfassung, der Schätzung eines Verhaltensmodells und der Bewertung von Störungen, wurden neue Methodiken vorgeschlagen, die die in der Literatur identifizierten Lücken schliessen. Das Leitmotiv der Arbeit ist dabei das GPS-Tracking. Die Verfügbarkeit hochdetaillierter Langzeit-Reisetagebücher ermöglicht, dank dieser Technologie, die Beobachtung des Passagierverhaltens unter bestimmten Bedingungen, beispielsweise bei Störungen oder während der COVID-19-Pandemie.

Die Ergebnisse dieser Dissertation können für Betreiber von öffentlichen Verkehrssystemen, Infrastrukturbetreiber und die öffentliche Verkehrsbranche im Allgemeinen von Vorteil sein. Tatsächlich ist das Verständnis des Verhaltens der Fahrgäste bei Störungen des öffentlichen Verkehrs ein entscheidender Aspekt, um eine Reaktion auf die Störungen zu modellieren. Darüber hinaus erwies sich das GPS-Tracking als eine leistungsstarke Methode zur Beobachtung der Entscheidungen der Passagiere, die in bestimmten Anwendungen stated-preference Befragungen ersetzen könnte. Letztendlich sind die Analysen in dieser Arbeit auch für Passagiere nützlich, da Betreiber diese nutzen können, um einen passagierorientierteren Service anzubieten.

RIASSUNTO

Le reti di trasporto pubblico sono caratterizzate da disturbi quotidiani, come ritardi o corse cancellate, che hanno un impatto negativo sui passeggeri. Infatti, in caso di disturbi, il tempo di viaggio può aumentare, o alcuni percorsi possono non essere più disponibili. In questi casi, il comportamento dei passeggeri può variare significativamente, dall'attenersi al piano iniziale allo scegliere un diverso percorso, mezzo o destinazione. Pertanto, capire il comportamento dei passeggeri in caso di disturbi è particolarmente rilevante, ma anche complesso, data la natura eterogenea sia dei passeggeri che dei possibili disturbi. Infatti, uno stesso passeggero può comportarsi in maniera diversa da altri passeggeri o in caso di diversi disturbi. Per analizzare quest'aspetto del comportamento di viaggio, è necessaria una grande quantità di dati, contenente osservazioni a lungo termine di molti passeggeri. A questo proposito, il tracciamento GPS è una tecnologia promettente, che consente una raccolta a lungo termine di informazioni precise sul comportamento di viaggio di utenti diversi.

L'obiettivo di questa tesi è capire il comportamento dei passeggeri in caso di disturbi nella rete di trasporto pubblico, tramite l'uso del tracciamento GPS. In particolare, uno degli obiettivi principali è lo sviluppo di algoritmi per l'identificazione automatica del mezzo di trasporto e la raccolta di diari di viaggio a lungo termine a partire dai dati GPS. In questo modo, i diari di viaggio di un gran numero di utenti possono essere utilizzati per analizzare le scelte nel trasporto pubblico, e in particolare in caso di disturbi.

Per capire il comportamento dei passeggeri in caso di disturbi, questa ricerca persegue quattro diversi obiettivi, ognuno affrontato in un diverso capitolo della tesi. Un quinto ed ultimo capitolo mostra l'applicazione del tracciamento GPS e dei metodi proposti per studiare il comportamento dei passeggeri durante la pandemia di COVID-19.

Il Capitolo 2 si concentra sulla raccolta automatica di diari di viaggio da dati di tracciamento. In esso è descritto uno studio di caso condotto per testare un'applicazione per smartphone, che consente un tracciamento passivo a lungo termine, con un basso consumo della batteria e senza gravare sui partecipanti. Diversi algoritmi sono proposti per identificare dai dati GPS i viaggi, le attività e i mezzi di trasporto utilizzati, e per raccogliere automaticamente diari di viaggio di diversi utenti. In particolare, l'algoritmo

proposto per l'identificazione del mezzo di trasporto sfrutta informazioni dei precedenti viaggi degli utenti, per migliorare la propria precisione; inoltre, esso identifica anche l'esatto veicolo utilizzato nel trasporto pubblico.
Il Capitolo 3 si concentra sull'identificazione delle alternative disponibili nel trasporto pubblico per un dato viaggio. Un nuovo algoritmo per la generazione delle alternative è proposto, il quale identifica tutte e sole le alternative rilevanti, ossia che sono realisticamente considerate durante il processo di scelta del passeggero. L'algoritmo è computazionalmente efficiente e può considerare diversi tipi di informazione disponibili per i passeggeri. Viene valutato su uno studio di tracciamento su larga scala, dove ottiene un'alta precisione (più del 94% di percorsi identificati) e consente la stima di modelli comportamentali significativi. A questo riguardo, diversi modelli per la scelta del percorso sono stimati, per capire l'informazione più probabile che i passeggeri considerano durante la loro scelta.
Il Capitolo 4 si concentra sulla valutazione dei disturbi nel trasporto pubblico. Una nuova definizione di impatto di un disturbo è proposta, capace di descrivere disturbi di grado diverso sulla base del loro potenziale impatto sui passeggeri. Diversi disturbi sono identificati da dati empirici dei trasporti pubblici, e il loro impatto è analizzato su passeggeri simulati. Sulla base di ciò, vengono analizzate le relazioni tra le caratteristiche dei disturbi e il loro impatto, per identificare le caratteristiche principali di un disturbo.
Nel Capitolo 5, i metodi e i risultati dei precedenti capitolo sono combinati, per analizzare i percorsi scelti dai passeggeri in caso di diversi disturbi nella rete. In pratica, informazioni sui percorsi scelti, le alternative disponibili, le condizioni della rete e i disturbi sono raggruppate in un grande dataset integrato. I percorsi scelti dai passeggeri sono paragonati alle alternative disponibili con e senza disturbi, e gli effetti dei disturbi sul costo di viaggio sono valutati. In particolare, l'analisi identifica che piccoli disturbi e ritardi hanno effetti significativi sul costo di viaggio, mentre effetti marginali sulla scelta del percorso. In aggiunta, nuove connessioni disponibili, non esistenti secondo l'orario programmato, spesso non sono utilizzate dai passeggeri.
Infine, il Capitolo 6 mostra l'applicazione del tracciamento GPS e della metodologia proposta, per capire il comportamento di viaggio in condizioni eccezionali, come la pandemia di COVID-19. A questo riguardo, uno studio basato sul tracciamento è stato condotto durante la prima ondata pandemica nel 2020 a Zurigo. Siccome i partecipanti sono stati già tracciati

in uno studio precedente del 2019 (quello descritto nel Capitolo 3), è possibile confrontare i comportamenti di viaggio in un periodo pre-pandemia e durante la pandemia. Tra i risultati principali, una drastica riduzione della distanza di viaggio è osservata durante la pandemia, in concomitanza con le restrizioni implementate. A questo proposito, il trasporto pubblico è stato colpito maggiormente rispetto a quello privato. Inoltre, stimando un modello di scelta del percorso nel trasporto pubblico, si notano importanti differenze tra i due periodi nel costo percepito viaggiando in treno e cambiando mezzo di trasporto.
In sintesi, questa tesi mostra un processo per comprendere il comportamento dei passeggeri nel trasporto pubblico, in particolare riguardo la scelta del percorso in caso di disturbi. Ogni parte di questo processo, inclusa la raccolta dati, la stima di un modello comportamentale, e la valutazione dei disturbi, è affrontato proponendo una metodologia nuova e colmando i relativi gap identificati nella letteratura. A questo riguardo, il leitmotiv della tesi è dato dal tracciamento GPS. Infatti, la disponibilità di diari di viaggio molto dettagliati e a lungo termine, grazie a questa tecnologia, consente di osservare il comportamento dei passeggeri in condizioni speciali, come nel caso di disturbi o durante la pandemia di COVID-19.
I risultati di questa tesi possono essere utili per società di trasporti, gestori di infrastrutture, e l'industria del trasporto pubblico in generale. Infatti, capire il comportamento dei passeggeri in caso di disturbi nel trasporto pubblico è un elemento cruciale per progettare una risposta ai disturbi stessi. Il tracciamento GPS, in aggiunta, ha provato di essere un mezzo efficace per osservare le scelte dei passeggeri, che può rimpiazzare in diverse applicazioni i sondaggi basati su preferenze dichiarate. Infine, le analisi in questo lavoro sono utili indirettamente anche per i passeggeri, poichè gli operatori possono utilizzarle per fornire un servizio più orientato ai passeggeri stessi.

ACKNOWLEDGEMENTS

First of all, I would like to thank Prof. Francesco Corman, for giving me the opportunity to join his team and come to Switzerland for my PhD. I have really appreciated the great support received over the years and the nice environment created at the Institute for Transport Planning and Systems.
I would like to thank professor Michel Bierlaire and professor Otto Anker Nielsen, as co-examiners of my PhD thesis, for taking the time to read it and to attend my PhD defence.
I would like to thank all researchers and colleagues that contributed to my research. A first thank goes to Prof. Kay W. Axhausen, who helped me during my first year at ETH. The suggestions and advices received in that period were particularly useful to successfully start my PhD. I thank Dr. Henrik Becker, with whom I collaborated on my first project, for his great help, in particular with the difficult task of conducting travel surveys.
I thank all my colleagues at IVT, past and present, with whom I enjoyed the working collaboration, but also the beautiful moments outside of work. A special thanks goes to Beda, with whom I started my PhD and I shared not only an office, but all the good and bad moments of this experience. I thank Nuannuan for her patience and the fruitful discussions. A big thanks goes to Marcel, Kimia, Allister and the other IPA's assistants, who shared their teaching experience with me, but more importantly one of the funniest trip I have done so far.
Finally, I want to thank all my family and friends, in Zürich and not, who always supported me and encouraged me in my work.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1

INTRODUCTION

1.1 RESEARCH MOTIVATION AND OBJECTIVE

Public transport is a sustainable alternative to private transport, and has seen an increase in usage in many countries in the last years. To remain a valuable choice, compared to private modes, a high service quality must be guaranteed, which meets the needs of passengers. In this regard, for an effective planning and design of public transportation systems, high-quality data are very important. In particular, observing passengers' behaviour and travel choices is a key element to plan and provide a more passenger-oriented service.

Travel surveys are a powerful tool for this purpose, and are widely employed in literate (Cottrill et al., 2013; Stopher and Greaves, 2007). In this context, GPS tracking is an emerging technology, allowing a long-term data collection of precise information on travel behaviour of several users. In fact, with the help of a GPS device, or a smartphone application, it is possible to track the movements of the owner, and to infer a travel diary from them. Tracking data provide several advantages compared to traditional surveys, based on interviews or requesting to manually report a travel diary. The burden on users is strongly reduced, which allows a long-term data collection, spanning several days. Moreover, if the data are collected via smartphone application, the data collection is easily scalable to several users. GPS tracking allows the collection of highly detailed and precise information, such as travel distance, exact departure time and visited locations. Therefore, they can complement traditional surveys, which are not suited to collect these data. In addition, tracking data can be integrated with external data sources, such as data on land use or on weather conditions, to discover additional information. For example, the integration with Automatic Vehicle Location data (AVL) of public transport operators allows discovering which public transport lines and vehicles are used by a tracked user.

On the other side, GPS tracking has its own issues and challenges. If the tracking is based on a smartphone application, the battery consumption must be minimized. Therefore, a trade-off between data quality (including sampling frequency) and battery consumption must be found. Data processing is also an additional challenge. In fact, if the users do not provide travel diaries manually, they have to be inferred from the raw GPS data. The algorithms to derive travel diaries from GPS data, such as mode detection algorithms, are a recent research topic. Therefore, travel diaries derived in this way are rarely applied to date within transportation planning (Harrison et al., 2020), and the aforementioned advantages are not yet exploited in research and practice. In this regard, the collection of long-term travel diaries, with highly detailed travel information, appears particularly useful for understanding the behaviour of travellers in exceptional conditions.

In this thesis, we investigate passengers' behaviour during public transport disturbances. This aspect is of great importance, given the possible negative effects of disturbances on passengers and service providers. In fact, a reliable service is particularly important for passengers, since an eventual disturbance may cause them discomfort. Moreover, disturbances affect also service providers, which may face revenue losses or additional costs due to the management of the disturbance or required additional services.

Understanding passengers' behaviour during disturbances is a complex task, which requires a large amount of data and non-trivial analyses, as represented in Figure 1.1. In particular, realized observations of both disturbances and passengers are needed, to observe their behaviour during disturbances. Moreover, the available alternatives for the observed trips need to be identified, to understand why passengers choose certain routes instead of others.

This thesis proposes a methodology divided in several steps, which, starting from raw GPS data of passengers and locational data of public transport vehicles, understand passengers' behaviour during disturbances. In particular, in this thesis, we first focus on identifying travel diaries from tracking data, developing a mode detection algorithm. Secondly, we focus on route choice in public transport, identifying the available alternatives for the observed trips. Therefore, we analyse passengers' behaviour in case of normal network conditions, disturbances in the network, and during the COVID-19 pandemic.

Below we report the main motivations and objectives related to each step.

Typically, to derive travel diaries from GPS data, different algorithms are developed, such as activity detection or mode detection algorithms. To collect realistic travel diaries, a high accuracy is required for these algorithms,

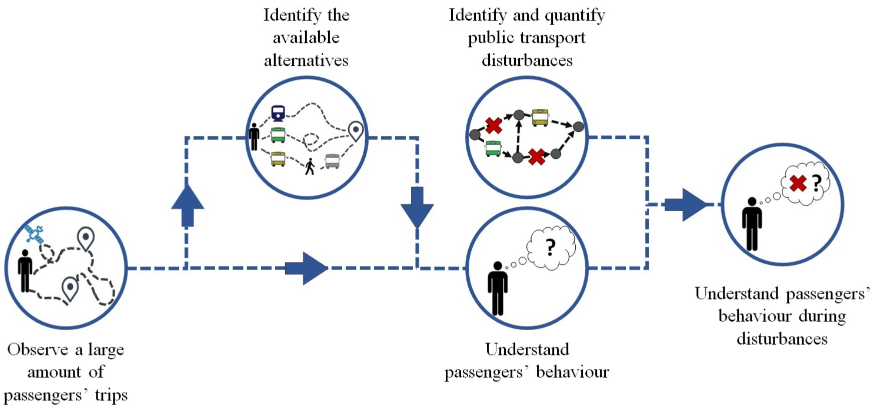


FIGURE 1.1: Required data processing to understand passengers' behaviour during disturbances.

which depends on: the quality of the GPS data and their frequency; the type of algorithms used; the data used for training (if needed); the test case (e.g. city, context and final users). In this context, this thesis proposes both a smartphone application, which passively collects GPS data without affecting significantly the battery consumption, and a set of algorithms, including a mode detection algorithm, to infer travel diaries from the collected data.

To understand the behaviour of public transport passengers, in particular why they choose certain routes, it is important to identify the alternative routes available to them. This problem is referred in literature as *choice set generation*, which consists in identifying the available alternatives, which more realistically are taken into consideration by passengers. Therefore, a choice set generation algorithm is a useful tool to understand travel behaviour from tracking data. Nevertheless, such an algorithm has to deal with several challenges: it must identify the route chosen by the user, only relevant and realistic alternatives, and have a low computation time. Moreover, since a passenger might be aware or not of network disturbances, such as delays or cancelled runs, a choice set generation algorithm should be able to model the possible information available to passengers. In this thesis, we propose a choice set generation algorithm, which solves the above-mentioned challenges, and it is tested on tracking data. In addition, the algorithm can identify which information provision should be assumed to best represent passengers' behaviour.

Knowing the chosen route of different users and their available alternatives, it is possible to understand the route choice criteria under general network conditions, estimating a route choice model. In contrast, if the interest is on route choice in case of disturbances, two major challenges arise, namely the definition of disturbance and the data collection.
Regarding the definition of disturbance, there is not a common measure in literature quantifying it. Most of the works analyse the effects on passengers of specific disturbances, in their specific test case (Leng, 2020; Van der Hurk, 2015; Yap et al., 2018). Typically, they consider disruptions, which are major events, such as a big service interruption in the transport network. In contrast, smaller disturbances, such as delays or cancelled runs, are rarely analysed. Moreover, a single disturbance can potentially have a different impact, depending on several factors, such as: the location in the network; the passengers involved; the time of the day. Without a clear quantification or categorization of the possible disturbances, at the current state, there is not a clear knowledge of the potential impact of a disturbance. In this regard, in this thesis, we propose a metric quantifying the impact of disturbances on passengers. Therefore, we analyse which characteristics of a disturbance are related with higher impact on passengers. This information can be particularly relevant for operating companies, to predict the effects of disturbances and plan an appropriate response to them.
Regarding the data collection during disturbances, passengers' behaviour is seldom observed, given the unpredictability of disturbances. Therefore, most of the work rely either on simulation or analysing a single disruption (Cats and Jenelius, 2014; Leng, 2020). In this context, tracking data can play a key role. In fact, passive tracking allows to collect realized data of several users for a long time. In this way, all the disturbances occurred during a study period can potentially be analysed. In literature, passengers' behaviour in case of disturbances has not yet been analysed from a large-scale survey based on tracking data. In this regard, this thesis fills this research gap, observing route choices of several users in case of different disturbances.

Finally, in this thesis we show how tracking data are a valuable resource to understand travel behaviour during the COVID-19 pandemic. In this period, restrictive policies were implemented in several cities in the world, to reduce contacts between people and their movements. Thus, the comparison of tracking data before and during the pandemic can shed some light on how the restrictions and the pandemic affected people's travel behaviour. Given that the pandemic outbreak is a recent event at the time of writing, very little has been explored in this research field.

As main contribution, this thesis shows the benefits of tracking data in the context of travel behaviour analysis and transportation planning. Moreover, it shows algorithms and automatic procedures to infer long-term travel diaries and understand travel behaviour in specific conditions, as in case of disturbances or during the COVID-19 pandemic.
This research can be beneficial for operating companies, infrastructure managers and the public transport industry in general. We show how GPS tracking is a powerful tool to observe passengers' choices in different contexts. In particular, understanding passengers' behaviour in case of disturbances is important to design a responsive and reliable public transport service. Moreover, a better understanding of passengers' behaviour is also beneficial for passengers, since it helps operators to provide a more passenger-oriented service.

1.2 RESEARCH QUESTIONS

1.2.1 *Research question*

Public transport disturbances have negative effects on both passengers and operators, which are hard to quantify. Knowing the potential impact of disturbances and how passengers react to them is important for operators to better counteract the disturbances. Nevertheless, observing passengers in case of disturbances is particularly challenging, given the unpredictability and rarity of the occurrence of a disturbance. In this regard, tracking technologies are a powerful resource to observe passengers' behaviour, given the long-term duration of a tracking study and the high level of detail of the acquired information. Moreover, GPS tracking allows to observe passengers in exceptional conditions, such as in case of disturbances or during the COVID-19 pandemic.
Based on that, the research question of this thesis can be formulated as follows:

How can tracking technology be exploited to understand passengers' behaviour in public transport, and in particular during different disturbances in the network?

1.2.2 *Research sub-questions*

The research question described above presents several challenges, which can be extended in different sub-questions. Each chapter of this thesis answers one of the following sub-questions.

1) How to derive travel diaries from long-term GPS tracking, based on a smartphone application?

This research question, explored in Chapter 2, aims to test a smartphone application for GPS tracking. The application must have a low battery consumption, to guarantee a low burden on the users and long-term data collection. Therefore, a following objective is to develop novel algorithms to automatically derive travel diaries from these GPS data. The algorithms need to deal with low-frequency GPS data, and must identify several information of a user, such as: activities, trips, stages, modes used, and, in case of public transport, also the exact lines and vehicles used.

2) How to identify the available alternatives for a public transport trip, which more realistically were considered by a passenger, given a certain knowledge of current network conditions?

This research question, explored in Chapter 3, aims to identify a methodology for generating the available alternatives for a given public transport trip. The proposed method (or algorithm) must be computationally efficient and accurate in identifying among the alternatives the one chosen by a real passenger. Moreover, it should be possible to identify the available alternatives based on different information provisions, representing different knowledge of network conditions for a passenger.

3) What is the impact of public transport disturbances on passengers? What are the main characteristics of disturbances affecting their impact?

This research question, explored in Chapter 4, highlights the lack of a common definition and quantification of disturbances in literature. In particular, a common metric, defining the potential impact of a disturbance on a passenger, is missing. Such a metric would allow comparing the effects of different disturbances, in different transport networks, ranging from small delays to big service interruptions. Moreover, it is not known which characteristics of a disturbance (such as its duration or the location in the network) contribute most to its severity.

4) How different network disturbances affect route choice of public transport passengers?

This research question, explored in Chapter 5, connects directly to the previous one, but referring to the effects of disturbances on passengers. In fact, in literature there is not a clear knowledge of passengers' behaviour in case of disturbances. Moreover, a great heterogeneity exists both between passengers and between disturbances, which needs to be investigated. In fact, the same passenger may react differently to other passengers or in case of different disturbances.

5) Is it possible to understand travel behaviour during the COVID-19 pandemic from tracking data? How the route choice criteria differ from those in a pre-pandemic period?

This last research question, explored in Chapter 6, asks if tracking data, and the derived travel diaries, are useful to analyse the effects on travel behaviour of exceptional events, such as a pandemic. In fact, the comparison of two long-term travel diaries, collected during a pre-pandemic and a pandemic period, may be useful to this end. An aspect of travel behaviour that is worth analysing is the route choice of passengers, since their criteria may be different between the two periods, due to possible travel restrictions and a different perception of the safety of public transport systems.

1.3 RESEARCH SCOPE AND BOUNDARIES

To answer the aforementioned research questions, this thesis is based on realized observations of both passengers and operations. Given the large heterogeneity of existing public transport networks, and all possible types of disturbances occurring, not all possible contexts are analysed. Therefore, in this section, we present the research scope in terms of study area, types of public transport disturbances analysed, and further research assumptions.

1.3.1 *Study area*

Regarding the public transport networks, we perform our analyses in Switzerland. In particular, the main city we consider is Zürich, where our tracking studies were collected. Other two Swiss cities, Basel and Bern, have also been considered in Chapter 2 and Chapter 4, to strengthen the proposed methods and results. Zürich is the largest city in Switzerland, with more than 400.000 inhabitants. The city is characterized by a multi-modal public transport network, integrating buses, trams and trains, with a single payment scheme. Public transport service is perceived as efficient and reliable, and it is widely used among the population (mode share of 41%, Stadt Zürich, 2021).
Therefore, our research focuses on a multi-modal public transport network in urban environment. In fact, the analyses on both passengers' behaviour and disturbances are based on urban trips. Among the observed trips of the tracked passengers, only the ones inside Zürich were analysed in detail. Inter-urban trips, connecting two different cities, represent a very different case study, given the limited number of modes and routes available. In contrast, the public transport network of Zürich is considered particularly efficient, and several alternative routes are available for most of the trips.
In any case, our research methods are not dependent on the city of Zürich or a specific public transport network, but rather they are based only on realized data of passengers and operations. Therefore, to apply the research methods, long-term GPS data of passengers and AVL data of the relevant public transport network are needed.

1.3.2 *Public transport disturbances*

Different types of disturbances can occur in a public transport network. An important distinction is between planned and unplanned disturbances. Planned disturbances, such as street maintenances, are well known in advance by service operators, which may prepare actions to counteract them, such as re-routing or providing additional services. Passengers may also be aware of them in advance. In contrast, unplanned disturbances, such as delays or accidents, are not known in advance by both operators and passengers. In this case, operators may react to a disturbance, informing the passengers or adapting the service provided; meanwhile, passengers may react changing the route, mode or destination.
Another characteristic of disturbances is their impact on passengers, which

is generally related to their size. Although there is not a common definition in literature of impact or size (see the third research sub-question), a small disturbance refers, typically, to a little inconvenience to both passengers and operators, such as a delayed bus; while, a large disturbance refers to a large inconvenience, as the closure of an entire train station for several hours. Moreover, small disturbances are generally frequent (daily occurrence), while large disturbances are rarer (few occurrences per year). A clear boundary between small and large disturbances does not exist, but rather the impact of a disturbance can span a relatively wide range (Yap, 2020).

In this work, we focus on small unplanned disturbances. We consider as a disturbance any deviation of the operations from the timetable. Practically, the disturbances considered in this work are in the form of delays, cancelled (or additional) runs, faults of vehicles, and re-routing. Combinations of disturbances are also considered, e.g. concurrent delays of different vehicles are considered as a single disturbance (more information in Chapter 4 and 5). Moreover, we assume the passengers may react to a disturbance only changing their route. Essentially, we assume a small disturbance does not justify changing to a private mode or changing the planned destination.

1.3.3 *Further assumptions*

This thesis relies on the quality of the AVL data of public transport operations. We assume the AVL data provided by the Swiss public transport companies are reliable and without errors. Moreover, a public transport network can be modelled entirely from the AVL data, which can describe both the planned public transport service (based on the timetable) and the actual service, with the real network conditions occurred.

We use random utility models to analyse route choice in public transport of the tracked users. These models assume passengers are rational and they choose the route, which maximizes their utility. We describe the utility as a function of several travel time components (e.g. walking time and time on a bus) and transfer-related parameters. Therefore, models based on different theories, as prospect theory, are not considered.

We assume passengers' route choices are not affected by capacity constraints or crowding. This is not a particularly strong assumption, given the high quality of service of public transport in Zürich, and the rarity of denied boarding.

1.4 STATE OF THE ART AND RESEARCH GAPS

In this Section, we briefly introduce the state of the art and the main research gaps related to the research questions of this thesis. In the following chapters, instead, a more detailed literature on each specific topic is shown. GPS tracking is a promising technology, which makes it possible to collect long-term travel diaries, with low burden on respondents (Stopher et al., 2008). In particular, while GPS loggers require a substantial effort for their distribution (Bohte and Maat, 2009), smartphones are often carried by participants, which made their use for travel surveys more popular (Cottrill et al., 2013; Montini, 2016). Nevertheless, the major challenge of smartphone-based travel surveys is the battery consumption, due to the continuous collection of GPS data (Cottrill et al., 2013; Pendão et al., 2014; Prelipcean et al., 2017). It follows, to collect long-term travel diaries without a significant burden on respondents, a smartphone application consuming very little battery power is needed.

To derive travel diaries from GPS data, dedicated algorithms are used, such as for activity detection or trip segmentation. In this regard, the identification of the travel mode is the most challenging part, which is often addressed with a dedicated mode detection algorithm, based on machine learning (Dabiri and Heaslip, 2018; Nikolic and Bierlaire, 2017; Wu et al., 2016). The main drawback of such algorithms is the need of a manual labelling of users' movements, to build a training set and train the model. Moreover, we are not aware of any existing approach able to detect, in addition to the mode, the exact public transport line and vehicle used. This information is particularly relevant to study route choices, as shown in Chapter 3. Finally, the quality of these algorithms is dependent on the quality of the GPS data and their sampling frequency. Therefore, a trade-off needs to be found between battery consumption and quality of travel diaries.

Understanding route choices of passengers requires (in most of the studies) the knowledge of the set of available alternatives for each observed trip, known as choice set in literature (Gentile and Noekel, 2016; Prato, 2009). In fact, based on the routes chosen by passengers, and their respective choice sets, it is possible to estimate a route choice model, to understand the choice criteria of passengers (Anderson et al., 2017). In this regard, we are not aware of any study evaluating a choice set generation algorithm based on realized data of both operations (e.g. AVL data) and passengers (e.g. GPS data). In particular, the usage of AVL data would allow comparing choice sets based on different information provisions, and therefore understanding the information available to (or used by) passengers and their choices during network disturbances.

With the absence of realized data, several studies analysed passengers' behaviour during disturbances based on simulations (Cats and Jenelius, 2014; Leng, 2020; Van der Hurk, 2015). In these cases the behaviour is based on assumptions or defined under normal network conditions, which might not be realistic. In fact, passengers' behaviour during disturbances can largely deviate from undisrupted scenarios (Yap et al., 2018). On the other hand, stated preference surveys are based on hypothetical scenarios, and cannot observe the passengers' real choices. For these reasons, GPS tracking is a promising method to collect revealed preference data and objectively observe travel decisions (Harrison et al., 2020; Zhu and Levinson, 2012). Moreover, Sun et al. (2016) affirms that without real passenger travel data, the behaviour in case of disturbances of many passengers is not discovered. We found only one study in literature (Yap and Cats, 2020), analysing passengers' behaviour during different disturbances from real passenger data. Nevertheless, they focus on the impact of disturbances in different public transport stops. Moreover, in most of the studies the research scope is often limited to the specific test case. Most of the work focuses on one or a few large disturbances (Leng, 2020; Van der Hurk, 2015; Yap et al., 2018), while low attention is given to small disturbances. In addition, there is not a common measure to quantify the size of disturbances, which makes difficult a comparison among them.

In fact, a large variety of disturbances is possible, ranging from small delays to large service interruptions. As highlighted by Yap (2020), there is no explicit demarcation between disturbances, but rather a range of disturbances with different impact. In this regard, little is known about which characteristics of disturbances most influence their impact.

In the field of passengers' behaviour during disturbances, we identify the following main research gaps: the lack of studies based on realized data of multiple passengers during different public transport disturbances; the lack of a metric quantifying and comparing different disturbances; the focus of the literature only on a few large disturbances in specific test cases. We believe these limitations make the analyses of passengers' behaviour during disturbances difficult to generalize in practice and do not allow a full understanding of it. Therefore, this work aims to fill these research gaps, to make general conclusions on passengers' behaviour during disturbances.

Finally, an additional aspect observed in this thesis is the passengers' behaviour during the COVID-19 pandemic. Given the novelty of the topic, very few research works are available to date, mainly focusing on observing general mobility trends (Aloi et al., 2020; Jenelius and Cebecauer, 2020; Molloy et al., 2021). In this regard, this thesis focuses on route choice in public transport, which has not yet been observed in literature.

1.5 RESEARCH CONTRIBUTIONS

1.5.1 *Scientific contributions*

This thesis makes several contributions to scientific research. In fact, while each chapter contributes to a specific research area, the whole thesis represents a novel process to understand travel behaviour during disturbances from tracking data.
The main scientific contributions are the following:

Automatic collection of travel diaries from GPS tracking (Chapter 2)
We use a smartphone application with low battery consumption and minimal burden on the users, to collect GPS data over long periods (the app was developed under the supervision of the thesis' author at the Institute for transport Planning and Systems, ETH Zürich). We propose algorithms to derive travel diaries from low-frequency GPS data, identifying activities, trips, stages and transport modes used. The proposed mode detection algorithm stands out among the ones available in literature for the following properties: it is not based on machine learning, therefore it is unsupervised and it does not require a manual collection of a training set; it can detect also the exact public transport vehicle used, using AVL data of public transport operators; it uses past travel information of users to identify missing transfers and improve the detection accuracy.

Identification of choice sets for public transport trips based on different information provisions on network conditions (Chapter 3)
We propose a novel choice set generation algorithm for public transport, able to generate all available alternatives given specific constraints on duration and transfers. The algorithm is characterized by a very low computation time and high performance. In fact, testing the algorithm on a large-scale tracking dataset, it identifies most of passengers' trips (coverage of 94%). Moreover, estimating a route choice model (Path Size Logit)

from the generated choice sets, we obtained realistic values for the parameters.
A major contribution of the algorithm is the ability to work with different information provisions on network conditions. This allows comparing choice sets based on different information provisions, to identify the one that best represents the passengers' choices. In fact, different information provisions may lead to significantly different choice sets.

Evaluation of the impact of public transport disturbances on passengers (Chapter 4)
This study contributes to the understanding of the effects of public transport disturbances, proposing a novel metric quantifying the potential impact of disturbances on passengers. This metric can be applied to any type of disturbance, including short delays and long service interruption, allowing a comparison between different disturbances. In this work, we identify real disturbances from AVL data of Swiss public transport operations, and we evaluate their impact on simulated passengers' trips. Therefore, based on random forest regression and feature importance analysis, we identify the main characteristics of a disturbance influencing its impact. Such analysis contributes to vulnerability analyses, identifying specific conditions or locations that are particularly vulnerable to disturbances.

Observing passengers' behaviour during public transport disturbances (Chapter 5)
This study is the first one in literature, observing passengers' route choices during different public transport disturbances from tracking data. In particular, we compare each route chosen by the users, with the planned alternatives available (according to the timetable), and the actual alternatives available (in case of disturbances). In this way, we evaluate the effects of disturbances on users, in terms of excess journey cost, i.e. the additional cost experienced by a user respect to the expected one. Therefore, the main contribution of this work is given by the observation of how the route choice and the excess travel cost vary in case of different disturbances.
Regarding the disturbances, we evaluate the network conditions for each observed trip, based on a metric of service degradation, which provides a numerical scale of the potential impact of the current disturbances in the network. In this regard, we focus on small disturbances (e.g. delays and cancelled runs) and "good disturbances", which we define as deviations of the operations from the timetable, leading to less costly alternatives.

Observing travel behaviour and route choice during the COVID-19 pandemic (Chapter 6)
This study contributes to the very little research available to date on travel behaviour during the COVID-19 pandemic. We compare two tracking studies involving the same participants, one conducted in 2019 and the other in 2020 during the first pandemic wave in Switzerland. This comparison allows identifying the main differences in travel behaviour due to the pandemic. In this regard, we analyse travel distance, mode share, visited locations and regular trips. The main aspect that makes our work stand out from the literature is the analysis of route choice in public transport. In fact, we estimate two route choice models, based on the two different periods, to identify how the route choice criteria changed during the pandemic.

Contributions of the thesis (All chapters)
While each chapter focuses on a specific research topic, their union, forming this thesis, has also important contributions to scientific research. In particular, we show how tracking data are a valuable resource to understand different aspects of travel behaviour. Overcoming the two main problems of data collection, namely the consumption of the smartphone battery and a required active participation of the users, we are able to conduct several studies, collecting highly detailed information, necessary for our research objectives.
A further contribution of this thesis is the proposed methodological process, described by the different methods in each chapter, that, starting from the collection of tracking data, allows observing and analysing the actual choices of passengers during public transport disturbances. This process can be used and adapted for future research, also with different research goals, as done for instance in Chapter 6, to observe travel behaviour during a pandemic.

1.5.2 *Contributions to society*

This work helps public transport authorities and service providers to understand passengers' behaviour, both in the general case and in case of public transport disturbances.
A main contribution of this thesis is a series of methods to automatically collect travel diaries from tracking data. In this respect, the collection and

analysis of travel diaries, based on traditional surveys, is complicated and expensive. In fact, a significant burden is given to respondents in compiling travel diaries, which causes most surveys asking respondents to report travel information during only one randomly chosen day (Greene et al., 2016), although research clearly shows day-to-day variations of travel behaviour (Pas and Sundar, 1995). Therefore, the collection of longitudinal data, based for example on a smartphone application that continuously tracks a user, can significantly improve the quality of studies based on travel diaries. In particular, a study based on tracking places a lower burden on respondents, offering high spatio-temporal precision of the information collected (Vij and Shankari, 2015).

This type of data collection can be implemented by service providers to enhance their understanding of passengers' behaviour. In fact, tracking data provide different and complementary information to traditional travel surveys, given their longitudinal nature and the high level of detail on passengers' movements. Service providers can enhance their already existing smartphone applications, with passive tracking and automatic collection of travel diaries, based on the methods described in Chapter 2. Furthermore, the limited effects on battery consumption show the burden on the users is reduced to a minimum, consisting only of installing the smartphone application.

The collection of a personal travel diary for each user is particularly important to provide a more user-oriented service, which is a benefit for both service providers and passengers. In fact, travel diaries have several applications, such as providing personalized route recommendations, or personalized information on network conditions. For instance, in case of a disturbance in the public transport network, it is possible to inform only the users that are potentially affected by the disturbance. Finally, collecting tracking data requires a single fixed cost, given by the development of a smartphone application, and limited costs over time, since the data collection is fully automatic on the users' smartphones. Management costs of the technical infrastructure for data storage and processing needs to be considered. In contrast, traditional surveys, are repeated with a certain frequency, with an additional cost for each survey.

A second contribution of this work is the proposal of an efficient and precise choice set generation algorithm for public transport (see Chapter 3). Such an algorithm finds a natural application in route recommender systems, which, given an origin and a destination, suggest the available routes for a user. Proposing only relevant routes and having a short computation time are two of the main characteristics required by route recommender systems. In this regard, the proposed algorithm is designed to run in a few seconds on a standard computer and to identify most of passengers' trips (coverage of 94%).

We analyse what is the right size of a choice set, both to guarantee a high coverage and to include only relevant paths. Such analysis helps improving route recommendation, since suggesting a low number of routes may not include the right option for a user, while a high number of routes may represent an excess of information, which may negatively affect the user. Finally, the proposed algorithm can generate choice sets based on different information provisions on network conditions, therefore assuming that the user may or may not be aware of network disturbances. This is particularly useful for service providers, to identify on which information their users rely to make their choices. In that way, providers can improve specifically for each user how information is provided, for instance improving the delivery time or better informing in case of disturbances.

Regarding network disturbances, in this thesis we propose a novel metric quantifying the impact of any disturbance on passengers, from short delays to long service interruptions. The proposed metric can be used by service providers and network managers in case a disturbance occurs, to evaluate the severity of that disturbance. Moreover, they can identify which passengers are more or less affected by the disturbance, and then plan accordingly the interventions to be adopted for the management of the disturbance. Passengers can also benefit from such a metric. In fact, in case of a disturbance, service providers can inform just the passengers who may be most affected, without providing unnecessary information to the others. In this work, we identify which characteristics of a disturbance (such as the duration, the vehicles involved or the location) influence its impact on passengers. This may help operators in strategic analyses, identifying the conditions leading to larger disturbances and the vulnerable parts of the network. This allows operators to better plan measures for disturbance management, such as keeping vehicle reserves or providing alternative routes to the disturbed ones.

How passengers react to disturbances is a complementary aspect to the one just mentioned, which is also explored in this thesis. In particular, we focus on route choices in case of small disturbances in a public transport network. Observing which routes passengers choose in case of a disturbance helps service operators understanding its effects and which action to take to counter the disturbance. For instance, in case of a cancelled bus, if passengers wait for the next one of the same line, while other faster alternatives are available, it is likely passengers are not sufficiently informed. In this case, service providers can improve their information systems, to better communicate to the affected passengers both the current disturbances and the available alternatives to avoid them. Furthermore, a better understanding of passengers' behaviour helps infrastructure managers and service providers to estimate the usage of the network during disturbances, and therefore to adapt their supply in a passenger-oriented way. This is beneficial also for passengers, as they find a public transport service more suited to their behaviour and needs.

In this work, we observe how different passengers react to disturbances with different impact. In this regard, analysing the heterogeneity of both passengers and disturbances is a crucial aspect for disturbance management. In fact, the response to different disturbances needs to be tailored to each of them, according to its characteristics and the expected impact on passengers.

As last contribution, this thesis exploits tracking data to understand passengers' behaviour during the COVID-19 pandemic. The exceptional nature of this event and its rarity make this work one of the few study that is available on this topic. Therefore, such a study is particularly useful for decision makers and service operators, to understand how travel behaviour changes during a pandemic. Moreover, they can observe how an increase in contagion and/or in the restrictions implemented corresponds to a reduction in travel among the population. In this work, we also analyse route choices in public transport. In particular, we observe the changes in the route choice criteria during the pandemic, compared to a pre-pandemic period. Such information can be useful for transport providers to adapt their service during a pandemic. In fact, the public transport system and the passenger volume are recognized as important factors that increase the contagion (Cartenì et al., 2021). Therefore, crowding must be reduced to provide a safer transport service. In this regard, understanding route choices of public transport passengers helps to predict the traffic flow in the network, and therefore to adapt the service to the new situation.

1.6 OUTLINE

In this Section, we present the structure of this thesis and the dependencies between the chapters. Figure 1.2 shows the relationships between the chapters and the datasets used. Each chapter answers one of the research ques-

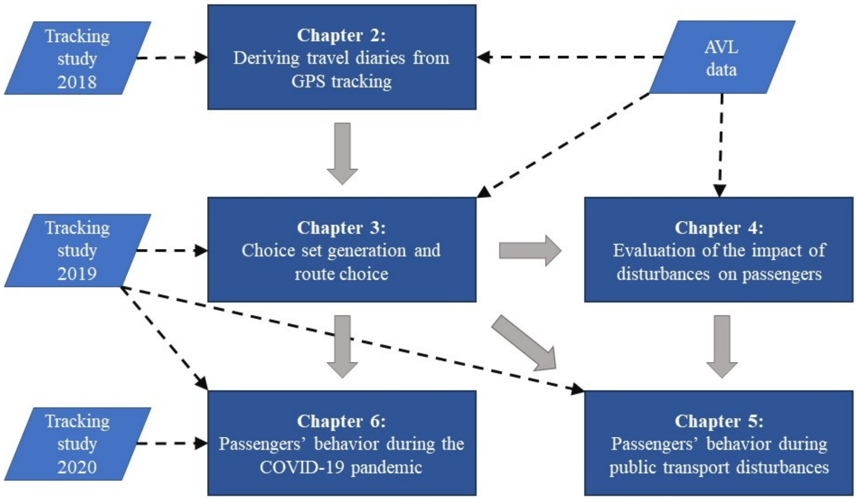


Figure 1.2: Thesis structure and dependencies between chapters.

tions identified in Section 1.2.2. Despite each chapter addresses a different problem, they are highly dependent in terms of data and methods used. The grey arrows indicate methodological dependencies between chapters, while the black arrows indicate the datasets used in each chapter.

Regarding the datasets, realized observations of both operations and passengers are used in this thesis. In particular, AVL data of public transport operations of the city of Zürich are used in all Chapters (in Chapter 2 also of Basel and in Chapter 4 also of Bern, two other Swiss cities). Instead, the observations of passengers are based on three different tracking studies, conducted in 2018, 2019 and 2020, respectively.

Chapter 2 focuses on the identification of travel diaries from GPS data. This chapter serves as the basis for the whole thesis, since it focuses on the data processing of the raw GPS data, collected with a smartphone application. A first tracking study is conducted for this chapter, to validate the proposed methodology. Moreover, AVL data are also used, as a key component of the mode detection algorithm. The raw data of the other two tracking studies, have all been processed as described in Chapter 2, to have travel diaries directly as input for the following Chapters, instead of raw GPS data.

Chapter 3 focuses on the choice sets generation for public transport trips, and the understanding of the route choice criteria under general network

conditions. A second tracking study is conducted for this chapter, which has been processed as described in Chapter 2. AVL data are used here, to represent the current network conditions, which the tracked passengers encountered during their trips. The methodology proposed in this chapter, consisting of a choice set generation algorithm and a route choice model, is used also in the following chapters.
Chapter 4 focuses on the definition of disturbances, the identification of their impact on passengers, and the identification of the main characteristics affecting their impact. Since this chapter focuses on disturbances, no tracking studies are used, but rather simulated passengers' trips. Here, the AVL data are used to identify real disturbances occurred in the public transport network.
Chapter 5 focuses on understanding passengers' behaviour during public transport disturbances. For this purpose, the methodology proposed in Chapter 3 on understanding passengers' route choices, and the one proposed in Chapter 4 on the evaluation of disturbances, are combined in this chapter. The same tracking study of Chapter 3 is used, as a set of real passengers' observations. AVL data are used also here to represent different network conditions and evaluate the impact of disturbances.
Chapter 6 focuses on understanding passengers' behaviour during the COVID-19 pandemic. An additional tracking study is conducted during this period. The participants of this study were selected among the ones of the previous study, to compare the behaviour of the same users between a pre-pandemic and a pandemic period. The choice set generation algorithm and the route choice model, described in Chapter 3, are also used in this chapter.
Finally, Chapter 7 reports the conclusions, summarizing the main findings and the implications for practice. Limitations of the research and future works are also discussed. An appendix is available after this chapter, containing clarifications to the previous chapters already published (the appendix is referred in the text with the superscript [Thesis Appendix]).
Since each chapter is based on a different journal article, some concepts described in previous chapters are summarized or repeated in the following ones. These repetitions are highlighted at the begin of each chapter. Moreover, due to the different publication date of the articles, on which the chapters are based, there are few discrepancies between the chapters, which we highlight here, to improve the readability. What is referred as *disruption* in Chapter 4, is referred as *disturbance* in Chapter 5. The metric of *disruption impact* defined in Chapter 4 is adapted in Chapter 5 and referred

as *service degradation*. Finally, Chapter 4 summarizes a previous version (almost identical) of the choice set generation algorithm described in Chapter 3.

REFERENCES

Aloi, A., B. Alonso, J. Benavente, R. Cordera, E. Echaniz, F. Gonzalez, C. Ladisa, R. Lezama-Romanelli, A. Lopez-Parra, V. Mazzei, L. Perrucci, D. Prieto-Quintana, A. Rodriguez, and R. Sanudo (2020). Effects of the COVID-19 Lockdown on Urban Mobility: Empirical Evidence from the City of Santander (Spain). *Sustainability* 12.9, 3870.

Anderson, M. K., O. A. Nielsen, and C. G. Prato (2017). Multimodal route choice models of public transport passengers in the Greater Copenhagen Area. *EURO Journal on Transportation and Logistics* 6.3, 221.

Bohte, W. and K. Maat (2009). Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. *Transportation Research Part C: Emerging Technologies* 17.3, 285.

Cartenì, A., L. Di Francesco, and M. Martino (2021). The role of transport accessibility within the spread of the Coronavirus pandemic in Italy. *Safety Science* 133, 104999.

Cats, O. and E. Jenelius (2014). Dynamic Vulnerability Analysis of Public Transport Networks: Mitigation Effects of Real-Time Information. *Networks and Spatial Economics* 14.3, 435.

Cottrill, C. D., F. C. Pereira, F. Zhao, I. F. Dias, H. B. Lim, M. E. Ben-Akiva, and P. C. Zegras (2013). Future Mobility Survey: Experience in Developing a Smartphone-Based Travel Survey in Singapore. *Transportation Research Record* 2354.1.

Dabiri, S. and K. Heaslip (2018). Inferring transportation modes from GPS trajectories using a convolutional neural network. *Transportation Research Part C: Emerging Technologies* 86, 360.

Gentile, G. and K. Noekel, eds. (2016). Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems. Springer Tracts on Transportation and Traffic. Springer International Publishing.

Greene, E., L. Flake, K. Hathaway, and M. Geilich (2016). A Seven-Day Smartphone-Based GPS Household Travel Survey in Indiana. *Transportation Research Board 95th Annual Meeting*.

Harrison, G., S. M. Grant-Muller, and F. C. Hodgson (2020). New and emerging data forms in transportation planning and policy: Opportunities and challenges for "Track and Trace" data. *Transportation Research Part C: Emerging Technologies* 117, 102672.

Jenelius, E. and M. Cebecauer (2020). Impacts of COVID-19 on public transport ridership in Sweden: Analysis of ticket validations, sales and passenger counts. *Transportation Research Interdisciplinary Perspectives* 8, 100242.

Leng, N. (2020). The role of information to passengers in public transport disruptions. Doctoral Thesis. ETH Zurich.

Molloy, J., T. Schatzmann, B. Schoeman, C. Tchervenkov, B. Hintermann, and K. W. Axhausen (2021). Observed impacts of the Covid-19 first wave on travel behaviour in Switzerland based on a large GPS panel. *Transport Policy* 104, 43.

Montini, L. (2016). Extraction of transportation information from combined position and accelerometer tracks. Doctoral Thesis. ETH Zurich.

Nikolic, M. and M. Bierlaire (2017). Review of transportation mode detection approaches based on smartphone data. *17th Swiss Transport Research Conference*.

Pas, E. I. and S. Sundar (1995). Intrapersonal variability in daily urban travel behavior: Some additional evidence. *Transportation* 22.2, 135.

Pendão, C. G., A. C. Moreira, and H. Rodrigues (2014). Energy consumption in personal mobile devices sensing applications. *2014 7th IFIP Wireless and Mobile Networking Conference (WMNC)*.

Prato, C. G. (2009). Route choice modeling: past, present and future research directions. *Journal of Choice Modelling* 2.1, 65.

Prelipcean, A. C., G. Gidofalvi, and Y. O. Susilo (2017). A Series of Three Case Studies on the Semi-Automation of Activity Travel Diary Generation Using Smartphones. *Transportation Research Board 96th Annual Meeting Transportation Research Board*.

Stadt Zürich (2021). Kennzahlen der Verkehrsentwicklung. `www.stadt-zuerich.ch/ted/de/index/taz/verkehr/webartikel/webartikel_kennzahlen_verkehrsentwicklung.html`.

Stopher, P. and S. P. Greaves (2007). Household travel surveys: Where are we going? *Transportation Research Part A: Policy and Practice* 41.5, 367.

Stopher, P., K. Kockelman, S. Greaves, and E. Clifford (2008). Reducing Burden and Sample Sizes in Multiday Household Travel Surveys. *Trans-*

portation Research Record: Journal of the Transportation Research Board 2064, 12.

Sun, H., J. Wu, L. Wu, X. Yan, and Z. Gao (2016). Estimating the influence of common disruptions on urban rail transit networks. *Transportation Research Part A: Policy and Practice* 94, 62.

Van der Hurk, E. (2015). Passengers, Information, and Disruptions. PhD Thesis. ERIM Ph.D. Series Research in Management. Erasmus Research Institute of Management.

Vij, A. and K. Shankari (2015). When is big data big enough? Implications of using GPS-based surveys for travel demand analysis. *Transportation Research Part C: Emerging Technologies* 56, 446.

Wu, L., B. Yang, and P. Jing (2016). Travel Mode Detection Based on GPS Raw Data Collected by Smartphones: A Systematic Review of the Existing Methodologies. *Information* 7.4, 67.

Yap, M. (2020). Measuring, Predicting and Controlling Disruption Impacts for Urban Public Transport. PhD Thesis. Delft University of Technology.

Yap, M. and O. Cats (2020). Predicting disruptions and their passenger delay impacts for public transport stops. *Transportation*.

Yap, M., S. Nijënstein, and N. van Oort (2018). Improving predictions of public transport usage during disturbances based on smart card data. *Transport Policy* 61, 84.

Zhu, S. and D. M. Levinson (2012). Disruptions to Transportation Networks: A Review. *Network Reliability in Practice*. Transportation Research, Economics and Policy, 5.

2

DERIVING TRAVEL DIARIES FROM GPS TRACKING

This chapter is based on the following article:

Developing a passive GPS tracking system to study long-term travel behavior

Alessio D. Marra[1] & Henrik Becker[1] & Kay W. Axhausen[1] & Francesco Corman[1]

[1]Institute for Transport Planning and Systems, ETH Zürich

Published in

Transportation Research part C: Emerging Technologies, 07.2019
DOI: https://doi.org/10.1016/j.trc.2019.05.006

Contributions

A. D. Marra: Conceptualization; Data curation; Methodology; Formal analysis; Writing - original draft; Writing - review & editing
H. Becker: Conceptualization; Data curation; Writing - review & editing
K. W. Axhausen: Conceptualization; Writing - review & editing
F. Corman: Conceptualization; Supervision; Writing - review & editing

Key findings

- Tracking system with low battery consumption
- User's travel behavior derived from low-frequency GPS data
- Unsupervised mode detection algorithm able to detect the specific public transport vehicle
- Past travel information from users used to improve the mode detection

Additional notes to this chapter

Some aspects of this chapter are clarified in the appendix of the thesis (see superscript [Thesis Appendix]).

ABSTRACT

This paper describes development and testing of a passive GPS tracking smartphone application and corresponding data analysis methodology designed to increase the quality of travel behavior information collected in long-term travel surveys. The new approach is intended to replace the pencil-and-paper travel diaries and prompted recall methods that require more user involvement due to requirements for manual data entry and/or high battery usage. Reducing the burden placed on users enables researchers to collect data over longer periods, thus improving the quality of travel behavior research. To reduce battery use the smartphone-based application collects GPS data less frequently than other methods. Therefore, new algorithms were developed to identify trips and activities, transport mode, and even the specific vehicle used by the traveler. An important finding was the significant advantage of using users' past data to improve mode detection results. The system was tested successfully in Zürich and Basel (Switzerland).

Keywords

Tracking; travel survey; public transport operations; mode detection; smartphone; GPS

2.1 INTRODUCTION

Until now travel diaries have been the primary source of travel behavior information on activity chains, trip patterns, mode choice and time use (Schlich and Axhausen, 2003). Unfortunately, conducting and analyzing travel diary studies is complicated and expensive. More importantly, since completing travel diaries places a significant burden on respondents, most surveys only ask respondents to report on travel during one randomly chosen day (Greene et al., 2016) although research clearly shows day-to-day variation in travel behavior (Axhausen et al., 2000; Pas and Sundar, 1995), which limits the political and operational value of one-day data (Susilo

and Axhausen, 2007).
The use of GPS data, for instance provided by smartphones, can significantly improve the efficiency of travel diary studies and the quality of information collected. Collecting GPS data from smartphones places a lower burden on respondents, offers greater spatio-temporal precision and has lower implementation costs (Vij and Shankari, 2015). The main drawback of using smartphones is their reliance on energy-intensive GPS services that quickly draw-down the smartphone battery, thus reducing the desire of travelers to use them.
The goals of this research were to develop a passive GPS tracking application that consumes very little battery power and a set of algorithms that can use this GPS data to provide detailed traveler behavior data. An application meeting the low power consumption objective was developed by reducing the GPS sampling frequency. The algorithms were designed to use this low(er) quality locational data to understand all user trips over a period of several weeks. These algorithms consisted of: activity and trip identification (dividing the users' records into activities and trips); trip segmentation (grouping trips into walking or using some means of transport); and, mode detection (identifying the transport means).
The key benefits of the application and algorithms are:

- The low battery consumption and limited (essentially zero) burden on the traveler means that it is possible to collect a great deal of positional data over long time periods.
- The user's travel behavior, in terms of activities, trips and transport modes used, is derived from low-frequency GPS data.
- The mode detection algorithm can also detect the particular public transport vehicle used using public transport operations data.
- Past travel information from users are used to identify missing transfers and improve the mode detection accuracy.

In particular, identifying the specific public transport vehicle used is not possible with existing mode detection algorithms and provides helpful information for understanding user travel behavior. The application passively collected location information from users approximately every 30 s. No interaction was required from respondents except to install the app and complete two short questionnaires at the beginning and at the end of the study. The smartphone application and algorithms were tested in Zürich. A further dataset, collected with a different smartphone application in the

city of Basel, was used to validate the proposed algorithms.
The paper is organized as follows: Section 2.2 describes the state of the art; Section 2.3 describes the smartphone application, the survey process and the datasets; Section 2.4 describes the data cleaning procedure; Section 2.5 describes the trip and activity identification algorithm; Section 2.6 describes the trip segmentation algorithm; Section 2.7 describes the mode detection algorithm; Section 2.8 presents the results of the data collection; Section 2.9 describes the validation procedure; and, finally, Section 2.10 presents conclusions.

2.2 STATE OF THE ART

GPS tracking makes it possible to collect long-term travel diaries while placing a very low burden on respondents (Stopher et al., 2008). Initial studies using GPS loggers for travel diary collection were promising, although they required a substantial effort to distribute the devices and to obtain additional information from respondents necessary to interpret the GPS records (Bohte and Maat, 2009; Montini et al., 2014; Oliveira et al., 2011; Schuessler and Axhausen, 2009).
Today, the focus has shifted away from GPS loggers towards smartphone applications (Cottrill et al., 2013), due to their easier administration and the development of automatic methods for detecting transport mode based on GPS data. Most of these methods are based on machine learning techniques and they often integrate GPS data with data from other smartphone sensors such as accelerometers (two related reviews are Wu et al. (2016) and Nikolic and Bierlaire (2017)). Kanarachos et al. (2018) reported the signals used by different mode detection algorithms, identifying GPS position, accelerometer, magnetometer, orientation, number of satellites, Horizontal Dilution of Precision and map information. Huang et al. (2019) compared the state of the art of mode detection based on GPS and mobile phone network data. They stated that the studies using GPS data tend to be more fine-grained, distinguishing among more transport modes.
Most of today's smartphone-based tracking systems use a prompted recall approach. This requires respondents to manually add details such as trip purpose, mode, group size, transit fare, parking fees etc. to each trip. Although some systems use statistical learning to make suitable suggestions to reduce the burden on respondents, a substantial amount of user interaction is still required to annotate or validate trip information. Cottrill et al. (2013) developed a prompted recall travel survey, in which par-

ticipants visit a website to validate activity and mode information. They concluded that battery life poses a major challenge and the participation process should be simple to ensure that the users do not feel overwhelmed. In this sense, to reduce battery consumption Lopez Aguirre (2018) used a mobile application, *Connect*, with a passive method that has a different sampling frequency according to the battery level. Nevertheless, this results in a inhomogeneous dataset. Cottrill et al. (2013), instead, proposed a "phased sampling", turning GPS off during sleeping periods (i.e. when user is doing an activity). Unfortunately, they stated the data quality is inevitably reduced during GPS sleep time. Pendão et al. (2014) reported information on the smartphone app *Moves*, which records the users' physical activity and identifies the type of motion. It minimises energy consumption, activating the GPS when detecting a known type of movement by the accelerometer. Despite this, it was observed that the app still has an impact on battery life. Regarding smartphone-based survey tools, there is an increasing number of them that collects both behavior and experience data. For instance, Raveau et al. (2016) extended the functionality of the system used by Cottrill et al. (2013) to collect both mobility information and measure happiness. To the authors knowledge, other (proprietary) apps were developed (e.g. rMove, 2019), which are not covered in detail.

Up until to now these smartphone-based tracking methods have mostly been tested on small datasets. For example, Tsui and Shalaby (2006) is based on 60 trips; and, Stenneth et al. (2011) recorded information for three weeks of travel by six people. Research based on large datasets have used dedicated GPS devices, making data collection complicated and expensive. Examples include Zheng et al. (2008) who collected information on 45 people over a period of six months; and Schuessler and Axhausen (2009) who used a dataset of 4882 people (requiring multiple waves to reduce the number of devices needed). An exception is Jariyasunant et al. (2015) that used a battery-saving application and collected complete data of 78 people for 3 weeks, without any person uninstalling the app for battery reasons. In this case, a questionnaire on battery usage to the participants could have given more insight on the perceived battery consumption.

Finally, it is important to mention Prelipcean et al. (2017), who developed MEILI, a battery-saving app to collect travel diaries from smartphones. They tested it only for one week and since the users had to manually annotate their trips, only about one third of them completed the study.

As this result shows, the problem of manually annotating trips places a significant hurdle on data collection. Therefore, a key tool for increasing the

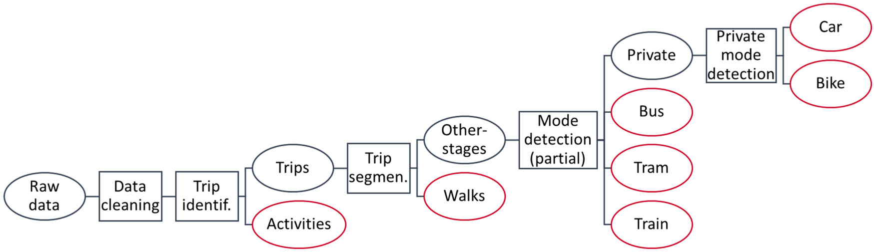


FIGURE 2.1: Sequence of algorithms used for mode detection. Rectangles indicate algorithms; ovals indicate data; red ovals represent the output.

efficiency of smartphone-based travel diary data collection and analysis is automatic mode detection. Several studies have used automatic mode detection from GPS data including Schuessler and Axhausen (2009), Stopher et al. (2005), Zhu et al. (2016), Zheng et al. (2008), Zhang et al. (2011). These studies divide the problem of mode detection into the following four tasks (with small variations):

- Data cleaning
- Trip and activity identification
- Trip segmentation
- Mode detection

Also this paper follows this structure, as shown in Figure 2.1. Since these tasks are often considered as separate problems, a short literature review is presented for each task.

2.2.1 *Data Cleaning*

Since raw GPS data may have systematic errors, it is necessary to identify and correct GPS errors before the data can be used in the next steps (Wu et al., 2016). The two main techniques used for data cleaning are filtering and smoothing.

Data filtering removes data that do not represent the user's real position. Several methods have been used to filter data. Ogle et al. (2002) used the position dilution of precision (PDOP, an accuracy measure based on the geometry of satellites) and the number of satellites for the filtering process. Gong et al. (2012) used the horizontal dilution of precision (HDOP) and the number of satellites to discard points. Schuessler and Axhausen (2009) used altitude values and sudden jumps in position to discard points. Ansari Lari and Golroo (2015) used speed to help identify GPS points for filtering.

The smoothing process is used to reduce the random noise present in the data. Jun et al. (2006) compared several smoothing techniques and found that a modified Kalman filter works best. Nitsche et al. (2014) also preprocessed the positional data using a Kalman filter. Schuessler and Axhausen (2009) used a Gauss kernel smoothing approach.

2.2.2 *Trips and activities detection*

The smartphone application records the user's position continuously throughout the day. Therefore, each user's data must be divided in trips and activities. An activity is given by a sequence of points near each other, indicating that the user is in the same place for a long period. In contrast, a trip is given by a sequence of points located apart from each other, representing the user's movement to a different place. A user's day is formed by activities alternating with trips.

Several techniques have been used to identify trips and activities. So far, to the best of our knowledge, all of them begin by first detecting activities and then defining trips. One of the most common techniques is to measure the time between two consecutive GPS points and compare it to a given threshold value. Different threshold values have been used ranging from 45 s (Pearson, 2001), 300 s (Wolf et al., 2004) and 900 s (Schuessler and Axhausen, 2009). A second technique is to define activities as periods when there are very low values of speed for a specified minimum amount of time (Tsui and Shalaby, 2006). A third technique is the density-of-points based method (Fan et al., 2015; Gong et al., 2012; Schuessler and Axhausen, 2009; Stopher et al., 2005). This method identifies activities where the density of points in a certain area is greater than a specific threshold. Schuessler and Axhausen (2009) defined a value of density for each GPS point by counting how many of the 30 preceding and succeeding points are within a 15 m radius. An activity occurs when there is a sequence of points with a density higher than 15 for at least 10 points or 300 s. Fan et al. (2015) mark a point t as part of an activity if the points within 2.5 min of t fall in a circle with diameter 200 m. Similarly, Gong et al. (2012) consider an activity as formed by points within 50 m of each other for more than 200 s.

2.2.3 *Trip segmentation*

Trip segmentation consists of dividing the users' trips into stages, which can be walk-stages or other-stages. To simplify terminology walk-stages are referred to as walks in the rest of the paper. A walk occurs when the user is walking or is waiting for transport in a single place. An other-stage occurs when a user travels using a mode of transport (car, bus, train or other vehicle).

Trip segmentation is the most challenging part of mode detection. There is no commonly accepted solution for this task described in the literature. Solutions include using additional sensors, or they rely on high sampling frequency ($\leq$ 10 s) for the GPS data.

All the trip segmentation algorithms found in literature were based in some way on speed. For instance, Biljecki et al. (2013) considered the user's stops as potential transition-points between modes. They identified a stop when consecutive points in an interval of 12 s did not have a speed higher than 2 km/h.

Shin et al. (2015) and Zheng et al. (2008) used acceleration to detect walks and stops. Since people usually walk or stop between two different transport modes, Zheng et al. (2008) used a threshold of speed and of acceleration to divide the points into walks and non-walks, then they merged segments of points of the same type according to rules depending on the segment length. Zhu et al. (2016) labeled points as walk or non-walk based on speed and acceleration threshold values, then adjusted the labels based on nearby points: if at least M (a value dependent on the number of points) of the previous and posterior points have a different label, then the point's label is changed. Zhang et al. (2011) used heading change to identify stops. Liao et al. (2006) used GIS information for trip segmentation, in particular the proximity to transition locations such as a bus stop. However, this approach is less reliable in cities with a high density of bus stops. Unfortunately, none of the algorithms outlined above has been tested on a dataset with a low sampling frequency similar to the dataset collected in this study.

2.2.4 *Mode detection*

Many methods have been studied to automatically collect information about users' travel behavior from raw GPS data. One of the major problems is identification of travel mode. A recent review by Wu et al. (2016) identified two categories of methods for mode detection: machine learning methods and hybrid methods, also relying partially on machine learning or on probabilistic models like Hidden Markov Models (Reddy et al., 2010). Nikolic and Bierlaire (2017) systematically reviewed the literature and found that the methodology adopted by all the studies was similar: first, some features are extracted from the sensors, then a training set is built to train a machine learning algorithm, and finally the algorithm is used to classify unseen data. For instance, Stenneth et al. (2011) compared different inference models including Bayesian Net, Decision Tree, Random Forest, Naive Bayes and Multilayer Perceptron. They were able to classify different transport means as car, bus, train, bike, walking and stationary. Fan et al. (2015) used a random forest to classify trip modes as bike, bus, car, train, walking and waiting, obtaining an overall accuracy of 86%. Nevertheless, their analyses are based on their travel survey method that was evaluated battery consuming. Dabiri and Heaslip (2018) used a deep learning technique, namely convolutional neural network, to predict the transport mode from raw GPS data. Reddy et al. (2010) built their classification system using accelerometer data in addition to GPS data and used a hybrid approach based on a decision tree and a Hidden Markov Model. Montoya et al. (2015) built a system based on a Bayesian network to infer the transport mode from smartphone data (GPS, wifi, accelerometer) and transport network information (e.g., public transport timetables). Patterson et al. (2003) presented a Bayesian model inferred in an unsupervised manner to distinguish between walk, drive or taking a bus. The research also showed that additional knowledge such as bus stop location can improve the algorithm results. Bantis and Haworth (2017) analyzed the relationship between personal and socio-demographic characteristics and travel mode choice using a Bayesian network. To reduce the complexity of a mode detection algorithm, Martin et al. (2017) combined dimensionality reduction techniques with machine learning algorithms, i.e. random forest. In that way, it is computationally simpler to run it on a smartphone.
Mode detection algorithms based on machine learning require manual labelling of user movements to train the model. This can be prohibitive when collecting data over a long period, since the labelling process requires sig-

nificant effort from users. Another shortcoming of existing approaches is that none of the mode detection algorithms can detect the exact public transport vehicle used by the traveler, but rather the generic mode (bus, train, etc...). In this regard, it is important to mention Carrel et al. (2015) who did not perform a mode detection analysis but used automatic vehicle location (AVL) data from the public transport network to identify public transport trips.

2.3 SMARTPHONE APPLICATION AND DATASETS

As part of this research, a smartphone app called *ETH-IVT Travel Diary* was developed to collect travel diaries over long periods of time while placing minimal burden on the users. The app was tested in a field trial with students at ETH Zürich. This section outlines app development and field testing. Analyses in the following sections are based on the records from this field trial.

2.3.1 *App design*

The travel diary app developed in this research was designed to be as easy to use as possible by not requiring regular interaction with the respondent and not substantially affecting battery life (therefore the GPS sampling frequency must not be too high). The travel diary app was developed for the Android operating system and made available on the Google Play store.
The app's user interface consisted only of a brief study description, a field to enter the respondent's identification code and a button to start data collection. Once launched, a process runs in background collecting GPS coordinates and timestamps until the end of the study period.
The app requests location data from the device's internal location services. To reach a balanced tradeoff between data quality and battery consumption, the app sends requests with different priorities at different intervals. In particular, a low priority request each 30 s and a high priority request each three minutes, that led to an average sampling frequency of 38 s in the collected dataset. These values were chosen during internal tests (technical details are beyond the scope of this paper). It is important to note that update of location information is at discretion of the device's internal location services. Typically, location is determined using GPS, Wi-Fi and cellular network. Kanarachos et al. (2018) compared different sensor fusion methods for battery energy saving and using a combination of these

three sensors reports the best results. Update frequency and sensors used depend on the frequency and priority of location requests from all the smartphone apps and varies by operating system version and device.
The *ETH-IVT Travel Diary* app can also use location information collected from other apps by sending a zero-priority data request, which provides the information without triggering an update. This means the app data is more accurate when respondents simultaneously use fitness trackers or navigation apps. Finally, in contrast to prompted-recall approaches employed in earlier studies, no interface was provided for respondents to (re)view their records.

2.3.2 *Data collection*

Students enrolled in a civil engineering program at ETH Zürich ($N = 1209$) were invited to participate in the study via e-mail in late March 2018. The study consisted of four weeks of tracking with the app. Students were paid CHF 20 to participate. No reminder e-mail was sent.
A total of 48 students using an Android smartphone signed up to participate in the study. However, the particular smartphone operating system for 9 respondents blocked the data collection and therefore their data could not be used. The resulting data set, referred to as the "Zürich dataset" in this paper, consisted of the travel diaries of 39 students and two of the co-authors. In total, 1 053 days of travel diaries were collected, which corresponds to an average of 25.7 days per respondent.
At the end of the study, 35 respondents completed the exit survey providing feedback on the app. They rated app user-friendliness as high and 80% stated that battery consumption was acceptable. This is a good result given that the app requires respondents to have their GPS turned on at all times, which increases battery consumption.

2.3.3 *Validation dataset*

A second dataset was used for validation and is referred to as the "validation dataset"[Thesis Appendix 1]. The validation dataset contains the ground truth of modes taken by users in addition to their GPS data. The validation dataset was collected with a different smartphone application in the city of Basel (Switzerland) during early 2018 (following the setup described in Becker et al. (2018) and Becker et al. (2017)). It contains GPS data from 625 users, with an average of 7.4 days of travel each. The ground truth was

obtained by using a proprietary automatic segmentation procedure that identifies stages when the user is not moving for a certain amount of time. Next, these stages are presented to the users through a web interface and the users are asked to manually specify the mode used for each stage. Asking users to label stages rather than individual GPS points reduces work for the users.
Interestingly, although users had the ability to correct the segmentation, they didn't do it very often. This means the ground truth may be less accurate than in reality. For instance, if a user walks and then takes a bus, the whole trip may have been labelled as bus, then the information about the walk is not present in the dataset. In the following sections, we always refer to the Zürich dataset except for the private mode detection (Section 2.7.3) and the validation (Section 2.9), where we refer to the validation dataset.

2.4 DATA CLEANING

The data cleaning process consisted of filtering and smoothing. Two main features were used to filter erroneous GPS points: the speed and the angle between points. Since the smartphone app collected only GPS coordinates and timestamps, other features such as those described in Section 2.2.1 were not available.
The speed for each point was derived from the previous point. Points with a speed equal to zero were removed because in these cases it is likely that the smartphone merely returned the previous recorded position – not the real one. Points with a speed of over 150 km/h were also removed since they were above the maximum accepted speed.
The second feature used in the filtering process was the angle between points (this feature has not been used in previous research to the best of our knowledge). Here, any point that forms a very small angle with the following and previous points, and which is far away from the previous point, is considered to be a false GPS point. More specifically: all points with an angle less than 15 degrees and a distance greater than 60 m from the previous point were removed. This rule was applied iteratively to all GPS points and the procedure was repeated until no more points are removed. The distance threshold was determined empirically; it was necessary because a small angle between near points can often occur if the user is walking or stopping. A limitation of this approach is that it assumes only one out of three consecutive points can be wrong, which means it cannot detect false points in case of two consecutive wrong points.

Source: map from openstreetmap.org.

FIGURE 2.2: Data Cleaning: Comparison of original data with cleaned data. Colour represents the time (from orange to red).

After completing the data filtering, the Kalman Filter was applied to smoothen the two space coordinates (longitude and latitude) of the GPS data. The Kalman Filter can deal with inaccurate observations and its efficacy increases with the frequency of the observations. For example, with a sampling frequency of 1 s, false points can be significantly corrected, while with a sampling frequency of 38 s (the average frequency of data collection in this study), only small adjustments are made to the user's trajectory. Figure 2.2 shows the application of data cleaning to part of a user's day, recorded with a high sampling frequency ($\approx$ 1 s). In this example, some erroneous points are removed, because of their small angle. Furthermore, the smoothing process adjusted the trajectory of the points in the bottom-left part of the figure.

2.5 TRIP AND ACTIVITY IDENTIFICATION

The next step was to identify trips and activities. In this research, an activity was defined as a user remaining within 250 m of the same point for at least 10 min. In turn, a trip is identified as a movement between two activities. The density-of-points method was used to identify trips and activities.

This method differs from those mentioned in Section 2.2.2 in that the algorithm does not rely on the number of GPS points (e.g. Schuessler and Axhausen, 2009) or on their frequency. In fact, the algorithm cannot rely on them, because the frequency of point data is too low. Therefore, in this

Algorithm 1 Trip and activity identification.

for each *point P* **do**
  if there are following points in a radius[1] long *activityRadius* for at least *activityTime* **then**
    P is the starting point of the activity
    The last point in the radius is the ending point
  end if
end for

[1]The center of the radius for each point Q following P is in the average point of all the points from P to Q.

research an activity was defined when there are at least 2 successive points within a 250-m radius (*activityRadius*) for at least 10 min (*activityTime*). At least 2 points are required because with 1 point it is not known if there is an activity or if no signal has been received. The iterative algorithm is presented in Algorithm 1. The 250-m radius used in this research is higher than used by others (e.g., 15 m in Schuessler and Axhausen (2009) and 30 m in Stopher et al. (2005)). It was made necessary due to the low precision of the GPS data. On the other hand, very short walks, starting and ending near an activity, are quite likely to really be part of the activity rather than an actual trip. When identifying activities and trips it is possible that the first and last points of an identified activity are in reality points from the previous or following trip. Fan et al. (2015) address this problem by adding an additional step. In this step, the starting and ending points of an activity are refined based on the time difference and the distance between points. Therefore, we added two rules to the algorithm to better assign these points. These rules rely on a variable called center of mass. Center of mass is computed as the average of the coordinates of all the activity's points.

- If the distance from the activity starting point to the center of mass is greater than two times the average distance of all points to the center of mass, an activity is not identified and the algorithm is repeated from the next new activity starting point.

- If the distance from the activity end point to the center of mass is greater than two times the average distance of each point to the center

FIGURE 2.3: Activity identification: The red points are in the activity, the green points are the last points of the previous trip, the blue points are the first points of the following trip. *Source:* map from openstreetmap.org.

of mass, then this point is removed from the activity and this rule is applied again for the new end point.

These additional rules improve the algorithm's ability to detect the break points between activities and trips, which helps improve mode detection. Figure 2.3 illustrates an example of activity identification. As shown in the figure, the first and last points, describing the user's arrival and departure, are not included in the activity, even though they are within the 250-m radius and they are not the farthest from the activity center of mass.
An additional rule was used to avoid the detection of false positive trips: a trip with an origin and destination very close to each other (less than 250 m) and with a short duration (less than 5 min), was merged with the previous and following activity to create one single activity. This addresses the problem of erroneous GPS position recording caused by proximity to cell sites or antennas.
Figure 2.4 summarizes the data collected in the Zürich dataset. The study period consisted of 40 days (21 March until 29 April 2018). The central part shows the number of trips detected for each user for all days in the study period. The top bar chart shows there are fewer trips made on Sundays than on other days of the week. The right bar chart shows that the number of trips is different among users: there are 96.7 trips per user on average, with two users making more than 200 trips during the study period.

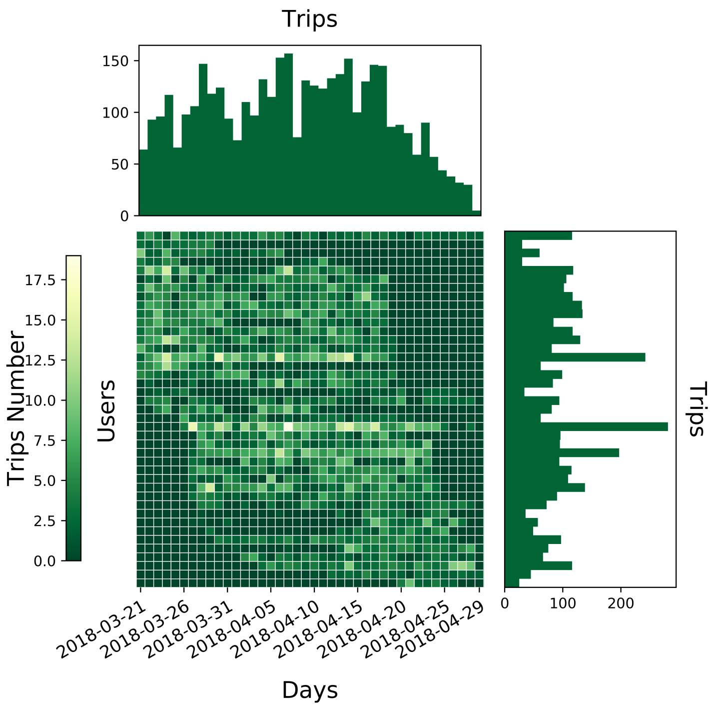


FIGURE 2.4: Number of trips for each user for each day. The users are ordered by the moment they installed the app. The number of trips is aggregated by user and by day in the two bar charts.

2.6 TRIP SEGMENTATION

Trip segmentation consists of dividing a user's trip into a sequence of walk and other-stages using a segmentation algorithm. In this research a walk was defined as a movement performed by walking or a transfer between different vehicles. An other-stage was defined as a movement performed using a vehicle. Therefore, a trip is formed by alternating walks and other-stages. After trip segmentation, all other-stages can be assigned a mode using a mode detection algorithm.
The GPS data recorded for this study have a low sampling frequency that varied considerably with different paths or users. Therefore, it was necessary to design a segmentation algorithm based on only GPS position and a derived speed that could work with irregular sampling frequency.

Parameter	Value	Parameter	Value
minSpeed	8.2 *Km/h*	minDuration	30 *s*
maxNearTime	30 *s*	stageMinDuration	50 *s*
Scale	0.8	walkMinDuration	70 *s*

TABLE 2.1: Values of the segmentation parameters.

Before applying the trip segmentation algorithm, the trip data was checked to identify any trips which had an absence of signal for more than 7 min. If so, the trip was divided into two different trips, to avoid errors during the segmentation.
The trip segmentation algorithm consists of four steps. Step 1 is label specification; here each GPS point is marked as either a walk or an other-stage according to the speed threshold *minSpeed*. In step 2 the label of each point is adjusted based on the label of the points adjacent in time. In other words, if a single point is labelled walk in the middle of a series of other-stage points, it is changed to an other-stage. Step 3 consists of grouping the consecutive points of the same type into sequences. Step 4 consists of merging the sequences according to rules depending on duration, distance and speed.
The segmentation algorithm is presented in detail in Algorithm 2. The parameters *minSpeed*, *maxNearTime* and *scale* are used for the label specification. The parameters *minDuration*, *stageMinDuration* and *walkMinDuration* are used to merge small stages into walks and other stages. Their values were automatically tuned during the algorithm validation (explained in Section 2.9). In particular: *minSpeed* was set to 8.2 km/h to reduce the number of wrongly detected other-stages. This is higher than the 6.48 km/h specified by Zhu et al. (2016), but this was appropriate since in this research it was possible to correct falsely identified walks later using the mode detection algorithm. The *stageMinDuration* was set to 50 s because it is unlikely that a user's other-stage would last less than 50 s. Similarly, the *walkMinDuration* was set to 70 s because it is unlikely that a user's walk is shorter than 70 s. The values used for all parameters are shown in Table 2.1.
The segmentation algorithm principally relies on the speed of each point (derived from the position and the time). Information on acceleration and heading change were not used, because they are not reliable with a low sampling frequency. Therefore, there are a few cases in which the trip seg-

Algorithm 2 Trip segmentation algorithm: Label specification (lines 2:20), merging rules (lines 22:25, 26:29, 30:33).

```
1: procedure SEGMENTATION(trip, minSpeed, maxNearTime, scale, minDuration,
   stageMinDuration, walkMinDuration)
2:     for each point ∈ trip do
3:         if point.speed < minSpeed then
4:             point.type ← walk
5:         else
6:             point.type ← other-stage
7:         end if
8:     end for
9:     repeat
10:        for each point ∈ trip do
11:            adjacentPoints ← previous and next points in maxNearTime
12:            M ← size(adjacentPoints) ∗ scale
13:            if more than M points in adiacentPoints are walk then
14:                point.type ← walk
15:            end if
16:            if more than M points in adiacentPoints are other-stage then
17:                point.type ← other-stage
18:            end if
19:        end for
20:    until no changes
21:    stages ← group all sequence of points of the same type
22:    repeat
23:        merge each stage with duration D < minDuration between
24:        previous and next stage if their duration > D
25:    until no changes
26:    repeat
27:        merge a other-stage between two walks if its duration <
   stageMinDuration
28:    until no changes
29:    repeat
30:        merge a walk between two other-stages if its average speed >
   minSpeed
31:        or its duration < walkMinDuration
32:    until no changes
33:    return stages
34: end procedure
```

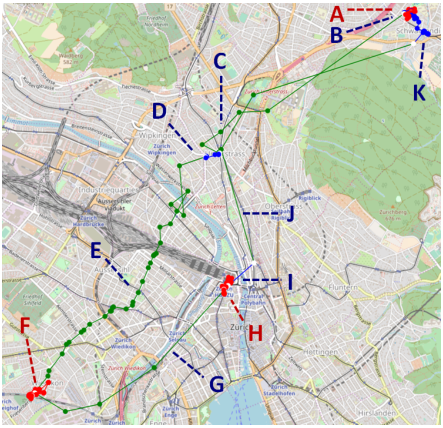
	Quantity	Per user day
Activities	3975	3.8
Trips	3965	3.8
Walks	8564	8.1
Other-stages	5548	5.3
Detected stages	1906 (34%)	1.8
Not assigned stages	863 (16%)	0.8
Ignored stages (outside of Zürich)	2779 (50%)	2.6
Past Data Detected stages	96	0.1

TABLE 2.2: Zürich dataset activities, trips and stage data obtained by the mode detection algorithm.

mentation will fail: a fast walk can be detected as an other-stage; a rapid change of buses, with the second departing shortly after the first arrives, can be detected as a single other-stage; a vehicle stuck in traffic for a long time can be detected as a walk. To overcome these problems, information obtained from the mode detection algorithm applied in the next step of the process was used to improve the trip segmentation. This is explained in Section 2.7.

Table 2.2 presents the results of the mode detection algorithm applied to the Zürich dataset. The top portion summarizes the number of activities, trips, walks and other-stages detected by the algorithm and the bottom portion the mode detection results. As shown in the top of Table 2.2 the algorithm found that users performed 3.8 activities per day and made 3.8 trips (based on a total of 1,053 days of valid data). These trips consisted of 8.1 walk stages and 5.3 other-stages. These results are reasonable for the study participants (university students).

Figure 2.5 presents an example of trip segmentation. The user's real path starts from the activity **A** where the user had a short walk (**B**) to take a tram (**C**). Then the user waited for a bus in **D**, took the bus (**E**) and arrived at **F** to stay there a while. Later the user took a train (**G**), stayed at the Zürich Main Station (**H**), walked to a stop (**I**), took a tram (**J**) and walked to home (**K**). Although the sampling frequency is different throughout the day, the segmentation algorithm is able to divide each trip correctly. More specifically, the figure shows that the frequency is lower when the user is



	Sampling Interval (s)
A	76
B	85
C	56
D	29
E	21
F	25
G	124
H	53
I	312
J	264
K	53

FIGURE 2.5: Trip segmentation of a user's day (from A to K): activities (red), other-stages (green), walks (blue). The table reports the average time between two points (sampling interval) for the stages and the activities. Source: map from openstreetmap.org.

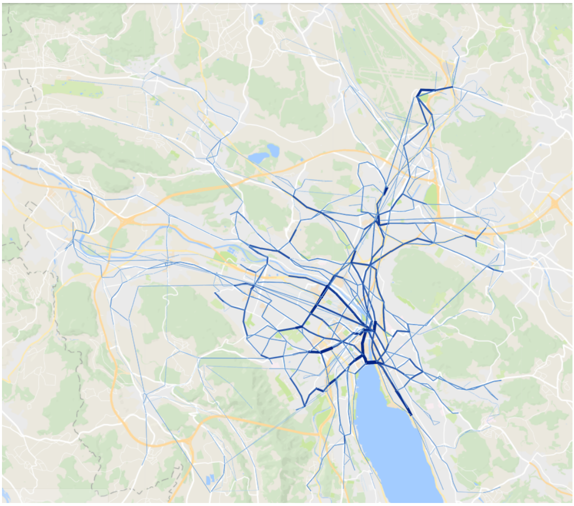
on a train (**G**) or in a tunnel (the upper-right part of **C** and **J**). It is also lower on **I**, because part of the main station is underground.

2.7 MODE DETECTION

Mode detection consists of assigning a specific mode of transport to the stages identified as other-stages in the trip segmentation. As outlined in Section 2.2.4, all the studies except Patterson et al. (2003), rely on inference models using a training data set to perform mode detection. Creating a training data set requires manually labelling of the transport modes, which limits the amount of data easily available.

In this study, the GPS data sampling frequency is a crucial parameter for the mode detection algorithm, because of the smartphone application's low sampling frequency (average of 38 s). Unfortunately, most published mode detection studies lack information about sampling frequency.

A specific mode detection algorithm was developed in this research to use this low sampling frequency GPS data (we verified that a moderate down-sampling of the data by 20% does not alter the conclusions and the results



Source: map from google.com.

FIGURE 2.6: Public transport traffic in Zürich: darker and larger lines represent more transport vehicles travelling on a working day (26-03-2018).

of the process to a major extent). The proposed mode detection algorithm is unsupervised, and it does not use any statistical inference model. Instead, it uses actual public transport operations data. Only Stenneth et al. (2011) used this type of data, and only to build features for their inference model.

Figure 2.6 illustrates Zürich's extensive public transport traffic network of buses, trams and trains. The system is efficient and well used, in fact the city's modal split is 32% public transport, 33% walking, 21% motorized private transport, 12% bicycle and 2% other (Städtevergleich Mobilität 2015, 2018). Since the Zürich dataset was collected mainly from students, its modal split could be different from the citywide figures. The actual public transport operational data consisting of planned and actual arrival and departure times for all vehicles are available for all stops in Zürich (SBB Opendata, 2018)[Thesis Appendix 2].

The mode detection algorithm uses this operational data to label an other-stage as being carried out by bus, tram, train or otherwise a private vehicle. An addition to this algorithm described in Section 2.7.3 shows how to distinguish between cars and bikes for private vehicle stages. Moreover, an original contribution of the algorithm developed in this research is its abil-

ity to detect the exact public transport vehicle taken by the user.
The algorithm works as follows: given an other-stage, identify all the public transport stops in a radius *detectionRadius* near the starting point of the other-stage, next select all the vehicles stopping at one of these stops in ± *detectionTime* from the starting point. Repeat the same process for the end point of the other-stage. Next, find the intersection of the vehicles in the two groups (this is a list of all the vehicles passing near the user at the beginning of the other stage and at the end). Finally, apply a likelihood function to each element of the vehicle list to identify the most probable vehicle taken by the user. If the list is empty, that means a private vehicle was used for the trip. In this research the parameters were set to *detectionTime = 300 [sec]* and *detectionRadius = max(250,min(accuracy,400)) [m]* where accuracy is a value in meters provided by the smartphone for each GPS point.
To overcome the incorrect identification of walks during trip segmentation (e.g., the algorithm identifies a walk when, in fact, a vehicle is stuck in traffic), a further rule is applied: if a trip has the pattern "other-stage, walk, other-stage" and the two other-stages are not assigned to a means of transport, then the three stages are considered as a single other-stage, and the mode detection algorithm is computed again on the new other-stage. To overcome the problem of undetected walks, the user's past data are used (described in Section 2.7.2).

2.7.1 *Likelihood function*

This research used a likelihood function to determine which vehicle out of a set of possible vehicles best matches the user's other-stage, computing the degrees of similarity between the user's path and the paths of the vehicles. Other studies have also used probabilistic functions or a rule-based system for the mode detection although in a different manner. For instance, Schuessler and Axhausen (2009) and Tsui and Shalaby (2006) used fuzzy logic approaches with rules based on speed and acceleration of the GPS records.
The likelihood function to determine vehicle used, *L(v,s)*, compares the path of a user's other-stage *s* with the path of a vehicle *v*. It is the product of a function of space/time likelihood between the paths, *L′(v,s)*; and a scaling factor *T(v,s)*, which takes into account how much the same time/distance path from a given vehicle can explain the user's entire trip. The mathemat-

ical formulation of *L(v,s)* is shown below:

$$L'(v,s) = \lambda * TimeDifference(v,s) + (1-\lambda) * PathDistance(v,s) \quad (2.1)$$

$$T(v,s) = \textit{times v is a candidate vehicle for a other-stage in the same trip of s} \quad (2.2)$$

$$L(v,s) = L'(v,s) * T(v,s) \quad (2.3)$$

The function *L'(v,s)* is also a combination of two functions: *TimeDifference* and *PathDistance* which compare the user and vehicle paths in time and space.
TimeDifference is the sum of the difference between the vehicle and user departure times and the difference between the vehicle and user arrival times. *PathDistance* is the average Euclidean distance between the user's coordinates and the vehicle coordinates. The vehicle coordinates were calculated at the same timestamps as the user's coordinates (GPS points), by interpolating from the arrival time and the coordinates of each stop.
The *TimeDifference* and *PathDistance* values were scaled to be comparable as shown in Figure 2.7, considering a maximum value of *TimeDifference* = *detectionTime (300 s)* and a maximum value of *PathDistance* = *maxPathDistance (250 m)*. A value of 0.5 was chosen for λ because using only the *TimeDifference* can result in a false positive of matching a user with a vehicle that had a different path; meanwhile the *PathDistance* is not reliable if there are too few points from the user. Vehicles with *PathDistance* = 0 were discarded, except for trains since train path is hard to describe using only stop point data.
The scaling factor, T(v,s), is the number of times that the vehicle *v* appeared in one of the lists of candidate vehicles for other-stages of the user's in the same trip of *s*. This reflects the idea that if a single vehicle is a candidate match for different other-stages of the same trip, it is probable that the user took only one vehicle for all the other-stages.
To compute the likelihood that the user travelled on the given vehicle *L(v,s)*, the function *L'(v,s)* is multiplied by the scaling factor *T(v,s)*. It is important to note that the value of *L* is not comparable for different other-stages, because it is dependent on the quality of the user's path data. With low quality data, *L* tends to be low for the taken means; in contrast, with high quality data *L* approaches 1. Furthermore, vehicle path data is always of high quality, because they are based on the position of the stops (known) and on the actual arrival times (provided by the operator and which are subject to smaller errors than user smartphone GPS data).

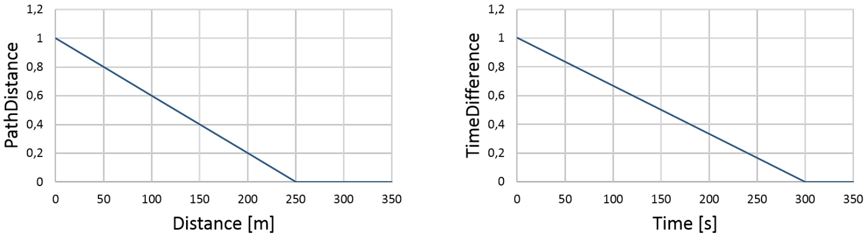


FIGURE 2.7: Mode detection likelihood: Scaling of the two sub-functions *PathDistance* and *TimeDifference*.

2.7.2 *Using past data to identify transfers*

The mode detection algorithm may miss transfers between two vehicles, especially when the transfer is performed quickly. Therefore, this research developed a new method for identifying missed transfers (i.e. points where the algorithm is unable to identify the transfer point and the two vehicles from/to) by applying the user's past data to improve trip segmentation. In other terms: when the segmentation does not detect a transfer (i.e. the user departs with a vehicle from a place A; arrives at place B with another vehicle; and the segmentation detects only one stage), the question is where to look for possible transfers points (a point C where the user might have changed the vehicle). In theory, the possibilities to connect two places by two public transport modes could be very large, especially in a dense network as the one considered in the test case. To this end, we prioritize the search for possible transfer points, where the user has been seen in the past. Note that those points do not describe a path, but only a possible point of transfer, and therefore would not be enough to identify a path, a priori.

This new method uses a personalized map of the places visited by each user from their travel history. The visited places consist of each activity (its center of mass) and the starting and ending points of each other-stage. For close together places (within 250 m), only the center of mass is considered, since they represent the same location.

After assigning modes to other-stages, the mode detection algorithm tries to detect missing transfers based on the user's visited places map. The places near the user's path (distance less than 400 m from any point on the path) are considered possible transfer points. Next, the algorithm tries to

detect if a transfer was feasible for each of these points. The mode detection is (re)computed from the starting point to the potential transfer point and from the potential transfer point to the ending point. If the mode detection algorithm can identify two public transport means that could support a transfer at this location, then it is assumed the transfer was made.
The process is based on the assumption that the user's travel behavior is recurring, especially during weekdays and for commuters. Therefore, places where the user has already been have a higher probability for the user to perform a transfer, especially if the user performed a transfer there on previous trips. This technique represents the first attempt, to our knowledge, of using the user's past data in a mode detection algorithm. This technique is particularly useful (and possible) for user datasets spanning multiple weeks with a large quantity of data.

2.7.3 *Private mode detection*

The main objectives of this study were to develop a smartphone application and mode detection algorithm to obtain travel behavior information with minimal impact on users. The methodology described in the previous sections is able to distinguish the used mode among walk, bus, tram, trains and private vehicles. Then, in this section, it is shown that is possible to integrate an additional module to distinguish between bicycles and cars for private stages. This module is not described in detail, since it is similar to classical mode detection algorithms (described in Wu et al. (2016) and Nikolic and Bierlaire (2017)), although the classification is performed only between two modes (bicycle and car).
The private mode detection used machine learning to identify modes. This required a ground truth and therefore the validation dataset was used to train and evaluate the private mode detection model. All the stages marked in the validation data as performed by bicycle or car were selected, then a set of features were extracted to represent each stage by a vector of features. The selected features were: number of points; length of the stage (*meters*); duration of the stage (*seconds*); average distance between two consecutive points; maximum speed; average speed; median speed; maximum acceleration; average acceleration; median acceleration; average angle formed by a point with the previous; median angle.
The validation dataset was divided in 70% for the training set and 30% for the test set. Then, several classification algorithms were tested: logistic regression, svm, decision tree and random forest. The one with the greatest

accuracy, defined as the percentage of correct detection, was selected for use in this study. The results of this procedure are described in Section 2.9.5.

2.8 ASSESSMENT OF ALGORITHM RESULTS USING ZÜRICH DATASET

Two methods were used to assess the algorithm and methods developed in this research. First the algorithm was applied to the Zürich dataset to test the overall ability of the smartphone application and algorithm to understand users movements. Second, the algorithm was applied to the Basel dataset (which included the ground truth of modes actually taken by the users) to determine how well the algorithm performed. This section describes testing with the Zürich dataset, Section 9 describes testing with the Basel dataset.

The lower part of Table 2.2 summarizes how the algorithm classified other-stages. As shown, the other-stages are divided into three groups: *detected stages* (34%), *not assigned stages* (16%), and *ignored stages* (50%). Stages are marked as *detected* if the mode detection algorithm was able to identify a mode. Stages are marked as *not assigned* if the algorithm could not identify a mode. Stages outside Zürich were ignored because this study relied on data from the city of Zürich. The *not assigned* stages principally consist of other-stages performed with a private vehicle but could also include public transport stages that were not detected due to low GPS quality or problems in the activity identification or trip segmentation steps. For instance, the *not assigned* stages could include false positives such as if the user is travelling in a car or on a bike directly behind a bus and the algorithm identifies bus as the transport mode. However, this case is considered rare because a car can overtake or has a different lane and a bicycle normally has a higher travel time.

The 96 stages detected using past data (representing 5% of the detected stages) would have been labelled as *not assigned* if the past data had not been used, with an increase of the *not assigned* group of the 11%. This indicates the importance of using information about the user's past travel behavior for mode detection. Its impact will be better analyzed in Section 2.9.3.

To measure the quality of the mode detection, the *TimeDifference* function (described in Section 2.7.1) was used. This function represents the sum of the difference of the departure times and the difference of the arrival times of the user and the detected vehicle. A low value indicates that the detec-

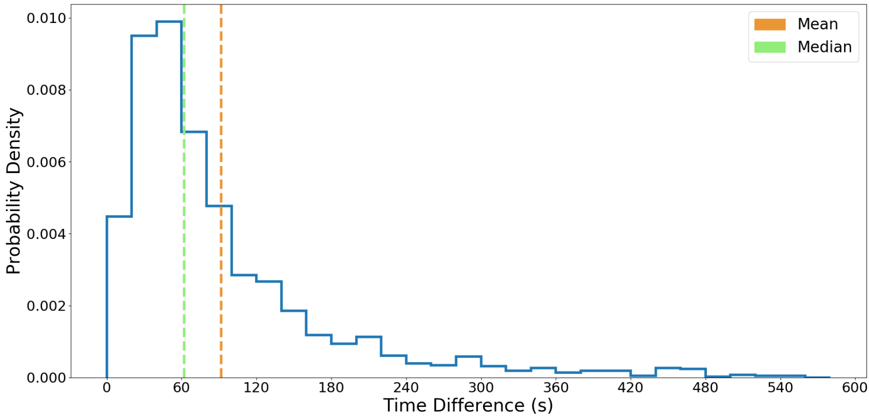


FIGURE 2.8: Distribution of *TimeDifference* for all the *detected* stages in the Zürich dataset (grouped each 20 *s*)

tion is correct, because the user and the vehicle were in the same places at the same times. The *PathDistance* function is not as good as an indicator because the GPS data describing a user's path can be quite noisy.

Figure 2.8 presents the distribution of the *TimeDifference* for all detected stages in the Zürich dataset. This value depends on two main factors: the trip segmentation and the sampling frequency. In particular, an erroneous trip segmentation can identify the beginning or the end of the other-stage at a point before or after the real beginning or end point. Importantly, due to the low sampling frequency there are often no points in the dataset representing the exact time the user boarded the transport vehicle. For this reason, the distribution's mean value of 91 *s* and a median value of 62 *s* for the *TimeDifference* can be considered good values and a strong indication of correct matching. Instead, with higher values, such as more than 300 *s*, the probability of a wrong detection increases.

2.9 ASSESSMENT OF ALGORITHM RESULTS USING VALIDATION DATASET

Showing that the average time difference between the users' paths and the detected public transport means is low demonstrates the validity of the proposed algorithm and shows good overall results but it does not show clearly the accuracy of mode detection. Therefore, the mode detection algorithm was also evaluated using the validation dataset collected in Basel, described in Section 2.3.3, which contains the ground truth data about user modes.

2.9.1 *Preparing the validation dataset*

The first step in using the Basel dataset was preparing the data. More specifically determining which labelled stages of the validation dataset can be considered valid for use in determining the accuracy of mode detection results. This step is needed because participants may have erroneously labeled a sequence of stages as a single stage, due to imprecise framing of the labelling question at the time of data collection or laziness in reporting. These cases are characterized by having fewer stages in the validation dataset compared to actual.

The process began by comparing each stage available in the validation dataset (labelled stage) with all the stages identified by the mode detection algorithm in the same time interval. If there was at least one stage with the same mode, the detection was considered correct. Figure 2.9 presents the distribution of number of stages detected by the algorithm for each labelled stage from the validation dataset. As shown, almost 50% of the stages perfectly match with one stage; 30% with two stages, which always include a walk stage; and very few with more than two stages, showing that the validation methodology provides a reasonable upper bound.

If there is only one stage detected for one labelled stage, then the labelled stage can be considered valid, since the two stages can be easily compared. If there are two detected stages for one labelled stage, then the labelled stage can be considered valid because one of the detected stages can be assumed to be a walk and every real trip includes a walk.

In cases where the algorithm identifies more than two stages there is some ambiguity as to whether the validation data has been correctly reported. In these cases if a labelled stage is detected by the algorithm as performed by both a bus/tram and a private vehicle, the stage is discarded as not valid, and not considered further for the validation, as it is impossible to associate a ground truth to it. This case is highlighted in orange on Figure 9; it occurs for only 6.6% of the stages. Further details on the combinations considered valid are available in the Appendix 2.11.1.

There were several shortcomings in the Basel dataset. First, actual public transport data was only available for half of the buses and trams in the network. For the other half, the algorithm used planned timetable data (two operators work in Basel, BVB and BLT, and the realized data are

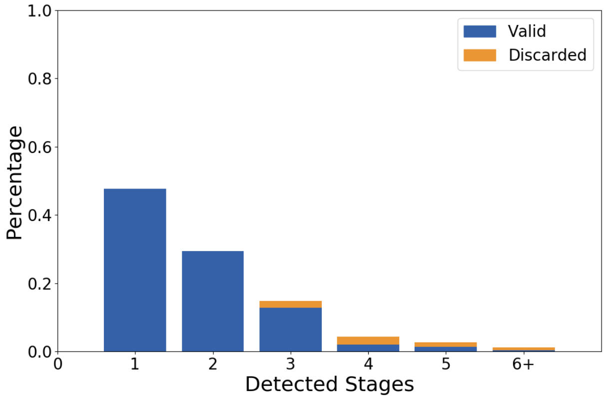


FIGURE 2.9: Distribution of the number of stages detected by the mode detection algorithm with an interval of time in common for each labelled stage from the validation dataset. In orange, stages not considered for the validation, since they are detected as performed by both bus/tram and a private vehicle.

provided only by BVB). Therefore, the algorithm cannot identify buses or trams when they are delayed by more than *detectionTime*. Second, there are on average only 7.4 days of data recorded per user, meaning that the user's travel history has only limited value in identifying possible transfers. Finally, to avoid errors in data collection, data from outside the city of Basel or that had stages with no signal for over 7 min were not used in the validation, to avoid errors due to the data collection.

2.9.2 *Tuning mode detection algorithm parameters*

A subset of the validation data consisting of approximately 400 days of tracking data (about 8%) was used to tune the parameters in the mode detection algorithm. Since the algorithm contains several parameters for each step (cleaning, activity identification, trip segmentation and mode detection), it was prohibitive to analyze all possible parameters. Therefore, only the five most relevant parameters (as assumed by the authors) were selected for tuning: three for segmentation (*maxNearTime*, *minSpeed* and *minDuration*) and two for mode detection (*detectionTime* and *maxPathDistance*). The other parameters were set by manual tuning as described in previous sections, except the *activityRadius*. This was set at 100 m, to better

Parameter	Values
maxNearTime (s)	{**30**, 60, 90}
minDuration (s)	{10, 20, **30**}
minSpeed (m/s)	{6.8, 7.5, **8.2**}
detectionTime (s)	{150, 200, 250, **300**}
maxPathDistance (m)	{150, **250**, 350, 450}

TABLE 2.3: Different values for the validation parameters. The selected best values are bold.

Mode	Total	Detected	Accuracy
All Stages	5659	4875	86.14 %
Walk	1520	1435	94.41 %
Bus/Tram	888	716	80.63 %
Train	123	84	68.30 %
Private	3128	2640	84.40 %

TABLE 2.4: Mode detection accuracy, express as percentage of correct detection.

align with data from a different smartphone application used in the Basel study.

Table 2.3 presents the different values considered for the five parameters. The set of evaluated combinations is the Cartesian product of those parameters. The final configuration was selected in order to have the highest average accuracy, in terms of percentage of correct detection, among the different modes. Further details are presented in Appendix (2.11.2).

2.9.3 *Accuracy of Mode Detection Algorithm*

After the mode detection algorithm was tuned it was used to detect modes from the rest of the validation dataset. Table 2.4 presents results of that analysis showing the accuracy of the mode detection algorithm in terms of percentage of correct detections. Since the validation dataset contains only one label for bus and tram, the two modes were considered as one.

The highest rate of correct detection was for walking. This shows the high accuracy of the segmentation algorithm. On the other hand, trains have the lowest rate of correct detection since the quality of user GPS data is lower inside stations and trains, and therefore it is harder to identify the exact GPS point when the train is departing (or arriving). The Bus/Tram stages were moderately well detected; the main problems for detecting them were the low quality of GPS data; segmentation errors; lack of user past data; and mainly the fact that actual public transport operations data were not available for all lines. Finally, the quality of private mode identification depended on the parameters *detectionTime* and *maxPathDistance*, chosen in order to have a balanced detection of public and private modes. In conclusion, the average accuracy of the mode detection algorithm was found to be comparable with the state of the art (the average accuracy of the works reviewed by Nikolic and Bierlaire (2017) is between 75% and 98%). Furthermore, these results can be significantly improved with more past data from users and actual public transport operations data for the entire network. Table 2.5 presents results from the mode detection algorithm with and without using the past user data. These results show the importance of users' past data, even though there are only 7.4 trips per user on average. In particular, the use of past data increases the accuracy of Bus/Tram stage detection by 4.1% and of train stages by 1.4%, by detecting missed transfer points. On the other hand, the slight decrease of accuracy for private mode stages is due to the detection of false transfer points. Table 2.6 compares the accuracy of the algorithm on the validation dataset using the realized public transport data (for the trains and the bus/trams with associated realized data) versus using only planned timetable data. As shown, the accuracy of public transport mode detection using planned schedule data is sharply lower. On the other hand, the accuracy for walks and private modes is almost the same (it is slightly different because their detection also depends on the public transport detection). These results confirm that the accuracy of the mode detection algorithm would have been greater if realized data had been available for all the public transport lines.

2.9.4 *Comparison with machine learning based mode detection*

The proposed mode detection algorithm was compared with a revised version of the algorithm proposed by Zheng et al. (2008). This algorithm was chosen because it is representative of common mode detection algorithm (it is based on machine learning techniques), it was tested on a relatively

Mode	Acc. with past	Acc. without past
All Stages	86.14 %	86.15 %
Walk	94.41 %	94.28 %
Bus/Tram	80.63 %	76.53 %
Train	68.30 %	66.93 %
Private	84.40 %	85.68 %

Table 2.5: Mode detection accuracy without using past data.

Mode	Accuracy Realized	Accuracy Timetable
All Stages	86.14 %	82.22 %
Walk	94.41 %	93.69 %
Bus/Tram	80.63 %	55.51 %
Train	68.30 %	60.33 %
Private	84.40 %	84.44 %

Table 2.6: Impact of realized data of public transport.

large dataset (45 people over a period of six months) and it has a large number of citations. Furthermore, it was evaluated on the accuracy of detected stages, making it easily comparable with the proposed algorithm. Both the algorithms rely on a previous segmentation and they were evaluated on the number of detected stages. After the segmentation (in common between the two algorithm), features were extracted from each segment and a random forest was trained on them. The dataset was divided in 60% training-set and 40% test-set, using cross-validation to estimate the internal parameters. The features used are the same described in Zheng et al. (2008). Since walks are detected during the segmentation, the model was trained only to detect Bus, Train or Private vehicles. For this reason, the post-processing proposed in Zheng et al. (2008) was not used, since its purpose was to connect other-stages with walk. The comparison of the two algorithms is shown in Table 2.7 and it is based on the same validation procedure explained in Section 2.9 (for this reason the walk accuracy is not exactly the same). The accuracy of the two methods is comparable for each mode, therefore we can assert that the proposed algorithm is comparable

Mode	Accuracy Proposed	Accuracy ML Based
All Stages	86.00 %	84.44 %
Walk	95.83 %	97.28 %
Bus/Tram	74.74 %	65.43 %
Train	71.64 %	73.91 %
Private	84.86 %	80.14 %

TABLE 2.7: Comparison between the proposed mode detection algorithm and a machine learning based algorithm.

if not better than a machine learning based algorithm, but in addition, it identifies also the exact public transport means used.

2.9.5 *Results of the private mode detection*

A private mode detection algorithm, described in Section 2.7.3, was used to distinguish private trips between bicycles and cars. The research tested several classification algorithms and selected the random forest algorithm with a maximum depth = 5 and 50 trees (since it performed best as defined by the highest percentage of correct detection). The training dataset for private mode detection needed to take into account that the Basel dataset contained many more bicycle trips than car trips. Therefore, each sample has a weight inversely proportional to the class frequencies, to equally train the classifier.
The confusion matrix is shown in Table 2.8. As shown the private mode detection algorithm had an overall accuracy of 86.75%. This good result is in spite of the fact that the segmentation procedure for the validation dataset was not perfect, as described in Section 2.3.3, and that no data were available from an accelerometer or other sources that could help distinguish between cycling and automobile travel.

		Detected			
		Bike	Car	Total	Accuracy
Real	Bike	915	92	1007	91 %
	Car	119	466	585	80 %
	Correct stages	915	466	1592	87%

Table 2.8: Confusion matrix and results of the private mode detection.

2.10 CONCLUSIONS AND FUTURE WORK

Results of this research confirm that GPS data collected from smartphones are a powerful means for understanding travel behavior and have several advantages over traditional survey methods. The research also clearly demonstrated that it is possible to develop a smartphone GPS tracking application that overcomes two of the main problems with earlier tracking technology by placing very low demand on the smartphone battery and requiring almost no work by the user. This ease of use makes it possible to easily track a large number of travelers for long periods of time, thereby significantly increasing the amount of data available for analyzing travel behavior.

Furthermore, the research shows that it is possible to understand the users' travel behavior based on only low-frequency and low-precision GPS data by designing and testing specific algorithms for activity detection, trip segmentation and mode detection for use with this type of data. The mode detection algorithm used in this research is an improvement over other methods because it only needs users' GPS traces and public transport network data. Most existing work described in the literature is based on supervised learning, requiring significant efforts to manually label data.

In addition to distinguishing between public transport and private modes, the algorithm also is able to detect the exact public transport vehicle used by the traveler. The method represents an original attempt to use actual operations data for a travel survey purpose. The algorithm also includes a method for exploiting the user's past data to detect transfers and thereby improve the quality of mode detection. The research results show clearly that this extra information helps improve the quality of results.

However, the results of this research have to be interpreted as a proofof-concept. The app was tested with a small[1] and very specific, urban, technology-savvy group. Yet, it can be expected that the app would work even better in other contexts: it does not require any respondent interaction and mode detection is hardest within cities, because of the small speed differences between modes. Anyway, a larger field test would be required to validate the results. Further limitations of this research are discussed in Appendix (2.11.3).

The research points out several paths for future work. First, the ability to unobtrusively collect data from users over long periods of time means it is possible to obtain much more interesting travel behavior data. For example, Figure 2.10 illustrates all the movements of one study participant over a 25-day period. Assuming that the participant's home and workplace are easily recognizable, there is a clear pattern in the morning trips from home to work. Although it is less regular, there is a similar pattern in the opposite direction in the evening. This illustrates a first step in discovering travel patterns and traveler choices that cannot be derived from a small dataset. Another path for future work is detection of the exact public transport vehicle. This is powerful information which cannot be obtained with a traditional mode detection algorithm. This precise data will make it possible to better analyze why people make specific travel choices (e.g., choosing a certain line), and to understand their main criteria for these choices. Remaining within the public transport mode, the long term tracking data made available by the proposed app could be complementary and/or alternative to smart card data tracking of users, which are able to measure only entrance/exit point of the user in the system in most common implementations (Pelletier et al., 2011).

The authors plan to use the methodology to collect a large-scale dataset, in terms of both users and days of tracking, which will support a wide variety of travel behavior research, such as identification of trip purpose (Montini et al., 2014). In particular, it could be possible to evaluate alternative plans of operations, within a specific mode of public transport, or including different modes, in passenger-aware models, similar to the bus bridging in case of disruption presented in Zhang and Lo (2018) or the train disruption resolution approaches proposed in Binder et al. (2017).

Having more precise recorded data and/or behavioral models of passenger route choices in public transport networks under delays is of crucial importance in evaluating the value of delays for specific vehicles. This is

1 Although the sample size is small, it is comparable with earlier research (Stenneth et al., 2011; Tsui and Shalaby, 2006)

FIGURE 2.10: Continuous tracking of a single user for different days. Activities in the same place have the same color, which goes from red to yellow according to the time spent doing the activity. A white space indicates absence of signal.

an open research gap, which needs to be tackled for instance for enabling non-discriminatory management of traffic, when multiple public transport modes and operators compete and complement each other's offer (see for instance Luan et al., 2017). Moreover, as identified in Carrel et al. (2013), providing correct and timely information to users would be a key factor in enhancing perception of public transport systems. The approach proposed could be very useful in segmenting users based on their diary, their risk-aversion, their attitude to partial, limited and dynamic information (Corman and Kecman, 2018) and their reaction to disruptions.

ACKNOWLEDGEMENTS

The authors wish to thank all study participants and the Swiss Federal Railways (SBB) for the open data.
Funding: This work was partly supported by projects funded by the Swiss Federal Road Office (project ASTRA 2017/006) and the Swiss Innovation Agency (Innosuisse, SCCER Mobility).
Declarations of interest: none.

2.11 APPENDIX

2.11.1 *Valid combinations of detected stages*

Table 2.9 shows the valid combinations of detected stages for each type of labelled stage, i.e. the combinations that can be used in evaluating the mode detection algorithm's accuracy.
The first four rows of Table 2.9 are cases when one stage is detected. In other words, the mode detection algorithm identifies one mode for a case where the validation dataset shows one mode. In the case of a one to one correspondence, the detected mode must be the same to be classified as correct (this is the case of the first column in Figure 2.9).
The next three rows are cases where the mode detection algorithm detects two stages, although the validation dataset shows only one stage. In these cases, the algorithm has detected a walk. When the algorithm detects two stages one of them is always a walk, since the trip segmentation algorithm divides a trip into alternating walks and other-stages. In other terms, a combination {Bus,Private} would be algorithmically infeasible, since there would be a small walk in between the two. Furthermore, in the real world it is common to walk before or after taking a vehicle. Since it is possible the user labelled the stage with only one mode (the one considered most relevant by the user), the detection is considered valid and correct either for a walk or for the other detected mode.
The next two rows are cases where the mode detection algorithm has detected three different stages for a single stage in the validation dataset. When the algorithm detects a labelled stage as performed by a train; either a bus or a private transport; and a walk (representing the transfer), the combination is considered valid, since it is common to take a bus or car to reach a train station.
The bottom two rows describe cases where the mode detection algorithm

Labelled stage / Detected stages	Bus	Walk	Private	Train
{Bus}	X			
{Walk}		X		
{Private}			X	
{Train}				X
{Bus, Walk}	X	X		
{Private, Walk}		X	X	
{Train, Walk}		X		X
{Bus, Walk, Train}	X	X		X
{Private, Walk, Train}		X	X	X
{Bus, Walk, Private}	discarded			
{Bus, Walk, Private, Train}	discarded			

TABLE 2.9: Valid combination of detected stages for each type of labelled stage. The correct are marked. The last two combination are discarded from the validation. The order of the modes is not relevant. Repetition of the same mode are not considered in the table. Combination with at least 2 stages and without a walk are not algorithmically feasible.

has identified combinations that include a bus and a private mode. A combination of this type is considered implausible and these data are discarded (this case is highlighted in orange in Figure 2.9).
The reason for discarding this data is that it is not possible to determine whether the algorithm made a mistake, or the user made a mistake when labelling the data. For instance, assume a user has performed a trip with 5 stages consisting of: a bus stage, a walk, a private stage, a walk, and again a bus stage (that would fit the scheme {Bus, Walk, Private} in Table 2.9). The user labels this as a single stage, associated to private mode. The algorithm correctly identifies the 5 stages but is unable to determine if this match is correct or incorrect.
Another possibility is that the algorithm is incorrect. For example, assume the user performs a trip by car, and the car drives so close to a bus that the algorithm assigns the first part of the trip to a bus stage, and something

similar happens on the last part of the trip. The algorithm would detect the same 5 stages listed above, although the user (correctly) would label this as a private trip.
Previous analysis has shown that user labelling in the original dataset was prone to misidentification of successive stages. For instance, users systematically identified a sequence of stages "walk, private, walk" as a single private stage. Therefore, it is not possible to exclude either of the two previously described possibilities. It is not clear whether it is the fault of the user, or the fault of the algorithm, and therefore we discard this labelled stage when determining the correctness and accuracy of the algorithm.
Finally, a labelled stage detected by the algorithm as an activity, like a slow walk near the same place, has been discarded for the validation, since it represents an activity and not a stage, according to our definition.

2.11.2 *Choice of Parameters*

Given the Cartesian product of the different parameters described in Table 2.3, the chosen configuration is the one with the greatest average accuracy among Walk, Bus/Tram, Train and Private vehicles. Since the number of Train samples is low (14 in the validation dataset) and the train accuracy is strongly dependent on *detectionTime*, both the best combination with *detectionTime* = 250 and 300 were tested and the second one is reported in the paper, since it has the highest accuracy. The importance of each parameter is reported in Figure 2.11 and the main consideration are the following:

1. Highest *minSpeed* increases the number of correct walk detection, since it relaxes the threshold to label a point as a walk point.
2. *minDuration* has a marginal impact compared to the other parameters.
3. A larger *maxNearTime* leads to wrong public transport detection, since it wrongly detects the correct starting point of a stage.
4. Low values of *detectionTime* and *maxPathDistance* lead to fewer trips being assigned to public transport mode and more trips being assigned to private mode. When lower values are used the algorithm tries to match the user's path only with the public transport means closer to the user during the trip.

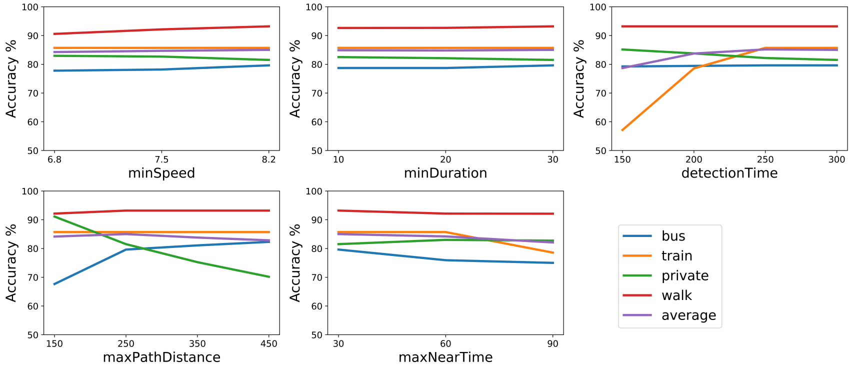


FIGURE 2.11: Parameters' importance: accuracy of different values of each parameter, fixing the other parameters to the values in the final configuration (*minSpeed* = 8.2, *minDuration* = 30, *detectionTime* = 300, *maxPathDistance* = 250, *maxNearTime* = 30).

Theoretically, with a correct segmentation and good quality of the data, the mode detection algorithm should work correctly with low values of *detectionTime* and *maxPathDistance*. However, when the quality of GPS data is poor, and the segmentation is incorrect, higher values are needed to guarantee the matching of user trips to public transport data. Therefore, improving the quality of GPS data and segmentation makes it possible to decrease *detectionTime* and *maxPathDistance*, and improve the quality of mode detection algorithm by reducing the number of false positives.

2.11.3 *Limitation of the research*

Given the nature of the data collection, tracking studies can be affected by different biases. For instance, despite smartphones have become widely spread across all population groups, the recruiting process is biased towards younger people, because of their familiarity with this technology. Also socio-economic dimensions, like educational level, can influence the decision of participation. For this reason, the recruitment process and the smartphone application need to be as simple as possible, to build a more representative dataset. Different people have different travel patterns, therefore a not representative dataset leads to biases in the observed travel behavior. The Zürich dataset consisted only of 41 persons and was collected mainly from students. Therefore, the usage of public transport is above the average population. In addition, all students have similar travel patterns as they share a common location (the university). These factors can affect the tuning of the algorithms and undermine their effectiveness with a different dataset. Another source of bias is the specific smartphone used, in fact only owners of Android devices could participate in the study. Therefore, a substantial part of the population were ignored, making the dataset less representative. In addition, the Android version and the other applications running on the smartphone can affect the data quality and the sampling frequency. In this sense, a systematic evaluation of the data quality on different devices was not performed. Despite these limitations, the use of two different datasets in this paper can contrast partially the aforementioned biases.

References

Ansari Lari, Z. and A. Golroo (2015). Automated Transportation Mode Detection Using Smart Phone Applications via Machine Learning: Case Study Mega City of Tehran. *Transportation Research Board 94th Annual Meeting*.

Axhausen, K. W., A. König, and R. Schlich (2000). Deskriptive Darstellung der Befragungsergebnisse des Projektes Mobidrive. Report. ETH Zurich.

Bantis, T. and J. Haworth (2017). Who you are is how you travel: A framework for transportation mode detection using individual and environmental characteristics. *Transportation Research Part C: Emerging Technologies* 80, 286.

Becker, H., F. Ciari, and K. W. Axhausen (2017). Modeling free-floating carsharing use in Switzerland: A spatial regression and conditional logit approach. *Transportation Research Part C: Emerging Technologies* 81, 286.

Becker, H., F. Ciari, and K. W. Axhausen (2018). Measuring the car ownership impact of free-floating car-sharing - A case study in Basel, Switzerland. *Transportation Research Part D: Transport and Environment* 65, 51.

Biljecki, F., H. Ledoux, and P. van Oosterom (2013). Transportation Mode-based Segmentation and Classification of Movement Trajectories. *International Journal of Geographical Information Science* 27.2, 385.

Binder, S., Y. Maknoon, and M. Bierlaire (2017). The multi-objective railway timetable rescheduling problem. *Transportation Research Part C: Emerging Technologies* 78, 78.

Bohte, W. and K. Maat (2009). Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. *Transportation Research Part C: Emerging Technologies* 17.3, 285.

Carrel, A., A. Halvorsen, and J. L. Walker (2013). Passengers' Perception of and Behavioral Adaptation to Unreliability in Public Transportation. *Transportation Research Record* 2351.1, 153.

Carrel, A., P. S. C. Lau, R. G. Mishalani, R. Sengupta, and J. L. Walker (2015). Quantifying transit travel experiences from the users? perspective with high-resolution smartphone and vehicle location data: Methodologies, validation, and example analyses. *Transportation Research Part C: Emerging Technologies* 58, 224.

Corman, F. and P. Kecman (2018). Stochastic prediction of train delays in real-time using Bayesian networks. *Transportation Research Part C: Emerging Technologies* 95, 599.

Cottrill, C., F. Pereira, F. Zhao, I. Dias, H. Lim, M. Ben-Akiva, and P. Zegras (2013). Future Mobility Survey: Experience in developing a smartphone-based travel survey in singapore. *Transportation Research Record: Journal of the Transportation Research Board* 2354, 59.

Dabiri, S. and K. Heaslip (2018). Inferring transportation modes from GPS trajectories using a convolutional neural network. *Transportation Research Part C: Emerging Technologies* 86, 360.

Fan, Y., J. Wolfson, G. Adomavicius, K. Vardhan Das, Y. Khandelwal, and J. Kang (2015). SmarTrAC: A Smartphone Solution for Context-Aware Travel and Activity Capturing. Report. Center for Transportation Studies University of Minnesota.

Gong, H., C. Chen, E. Bialostozky, and C. T. Lawson (2012). A GPS/GIS method for travel mode detection in New York City. *Computers, Environment and Urban Systems* 36, 131.

Greene, E., L. Flake, K. Hathaway, and M. Geilich (2016). A Seven-Day Smartphone-Based GPS Household Travel Survey in Indiana. *Transportation Research Board 95th Annual Meeting*.

Huang, H., Y. Cheng, and R. Weibel (2019). Transport mode detection based on mobile phone network data: A systematic review. *Transportation Research Part C: Emerging Technologies* 101, 297.

Jariyasunant, J., M. Abou-Zeid, A. Carrel, V. Ekambaram, D. Gaker, R. Sengupta, and J. L. Walker (2015). Quantified Traveler: Travel Feedback Meets the Cloud to Change Behavior. *Journal of Intelligent Transportation Systems* 19.2, 109.

Jun, J., R. Guensler, and J. Ogle (2006). Smoothing Methods to Minimize Impact of Global Positioning System Random Error on Travel Distance, Speed, and Acceleration Profile Estimates. *Transportation Research Record: Journal of the Transportation Research Board* 1972, 141.

Kanarachos, S., S.-R. G. Christopoulos, and A. Chroneos (2018). Smartphones as an integrated platform for monitoring driver behaviour: The role of sensor fusion and connectivity. *Transportation Research Part C: Emerging Technologies* 95, 867.

Liao, L., D. J. Patterson, D. Fox, and H. Kautz (2006). Building personal maps from GPS data. eng. *Annals of the New York Academy of Sciences* 1093, 249.

Lopez Aguirre, A. (2018). Processing crowdsourced data for the analysis of mobility behaviour. dissertation. Ghent University.

Luan, X., F. Corman, and L. Meng (2017). Non-discriminatory train dispatching in a rail transport market with multiple competing and collaborative train operating companies. *Transportation Research Part C: Emerging Technologies* 80, 148.

Martin, B. D., V. Addona, J. Wolfson, G. Adomavicius, and Y. Fan (2017). Methods for Real-Time Prediction of the Mode of Travel Using Smartphone-Based GPS and Accelerometer Data. *Sensors (Basel, Switzerland)* 17.9.

Montini, L., N. Rieser-Schüssler, A. Horni, and K. W. Axhausen (2014). Trip Purpose Identification from GPS Tracks. *Transportation Research Record: Journal of the Transportation Research Board* 2405, 16.

Montoya, D., S. Abiteboul, and P. Senellart (2015). Hup-me: inferring and reconciling a timeline of user activity from rich smartphone data. *ACM SIGSPATIAL, International Conference on Advances in Geographic Information Systems*.

Nikolic, M. and M. Bierlaire (2017). Review of transportation mode detection approaches based on smartphone data. *17th Swiss Transport Research Conference, Ascona*.

Nitsche, P., P. Widhalm, S. Breuss, N. Brändle, and P. Maurer (2014). Supporting large-scale travel surveys with smartphones - a practical approach. *Transportation Research Part C: Emerging Technologies* 43, 212.

Ogle, J., R. Guensler, W. Bachman, M. Koutsak, and J. Wolf (2002). Accuracy of Global Positioning System for Determining Driver Performance Parameters. *Transportation Research Record: Journal of the Transportation Research Board* 1818, 12.

Oliveira, M., P. Vovsha, J. Wolf, Y. Birotker, D. Givon, and J. Paasche (2011). Global Positioning System-Assisted Prompted Recall Household Travel Survey to Support Development of Advanced Travel Model in Jerusalem, Israel. *Transportation Research Record: Journal of the Transportation Research Board* 2246, 16.

Pas, E. I. and S. Sundar (1995). Intrapersonal variability in daily urban travel behavior: Some additional evidence. *Transportation* 22.2, 135.

Patterson, D. J., L. Liao, D. Fox, and H. Kautz (2003). Inferring High-Level Behavior from Low-Level Sensors. *UbiComp 2003: Ubiquitous Computing*.

Pearson, D. (2001). Global Positioning System (GPS) and travel surveys: Results from the 1997 Austin household survey. *8th Conference on the Application of Transportation Planning Methods, Corpus Christi*.

Pelletier, M.-P., M. Trepanier, and C. Morency (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies* 19.4, 557.

Pendão, C. G., A. C. Moreira, and H. Rodrigues (2014). Energy consumption in personal mobile devices sensing applications. *2014 7th IFIP Wireless and Mobile Networking Conference (WMNC)*.

Prelipcean, A. C., G. Gidofalvi, and Y. O. Susilo (2017). A Series of Three Case Studies on the Semi-Automation of Activity Travel Diary Generation Using Smartphones. *Transportation Research Board 96th Annual Meeting Transportation Research Board*.

Raveau, S., A. Ghorpade, F. Zhao, M. Abou-Zeid, C. Zegras, and M. Ben-Akiva (2016). Smartphone-Based Survey for Real-Time and Retrospective Happiness Related to Travel and Activities. *Transportation Research Record: Journal of the Transportation Research Board* 2566, 102.

Reddy, S., M. Mun, J. Burke, D. Estrin, M. Hansen, and M. Srivastava (2010). Using Mobile Phones to Determine Transportation Modes. *ACM Transactions on Sensor Networks* 6.2, 13:1.

rMove (2019). `https://rmove.rsginc.com/`.

SBB Opendata (2018). `https://opentransportdata.swiss/en/dataset/istdaten`.

Schlich, R. and K. W. Axhausen (2003). Habitual travel behaviour: Evidence from a six-week travel diary. *Transportation* 30.1, 13.

Schuessler, N. and K. W. Axhausen (2009). Processing Raw Data from Global Positioning Systems Without Additional Information. *Transportation Research Record: Journal of the Transportation Research Board* 2105, 28.

Shin, D., D. Aliaga, B. Tunçer, S. M. Arisona, S. Kim, D. Zünd, and G. Schmitt (2015). Urban sensing: Using smartphones for transportation mode classification. *Computers, Environment and Urban Systems* 53, 76.

Städtevergleich Mobilität 2015 (2018). `https://www.stadt-zuerich.ch/ted/de/index/taz/publikationen_u_broschueren/staedtevergleich_mobilitaet_2015.html`.

Stenneth, L., O. Wolfson, P. S. Yu, and B. Xu (2011). Transportation Mode Detection Using Mobile Phones and GIS Information. *19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*.

Stopher, P., Q. Jiang, and C. FitzGerald (2005). Processing GPS data from travel surveys. *28th Australasian Transport Research Forum, ATRF 05, Sydney, New South Wales, Australia*.

Stopher, P., K. Kockelman, S. Greaves, and E. Clifford (2008). Reducing Burden and Sample Sizes in Multiday Household Travel Surveys. *Transportation Research Record: Journal of the Transportation Research Board* 2064, 12.

Susilo, Y. and K. W. Axhausen (2007). How firm are you? A study of the stability of individual activity-travel-location pattern using Herfindahl Index. *11th World Conference of Tranport Research*.

Tsui, S. and A. Shalaby (2006). Enhanced System for Link and Mode Identification for Personal Travel Surveys Based on Global Positioning Systems. *Transportation Research Record: Journal of the Transportation Research Board* 1972, 38.

Vij, A. and K. Shankari (2015). When is big data big enough? Implications of using GPS-based surveys for travel demand analysis. *Transportation Research Part C: Emerging Technologies* 56, 446.

Wolf, J., S. Schoenfelder, U. Samaga, M. Oliveira, and K. W. Axhausen (2004). Eighty Weeks of Global Positioning System Traces: Approaches to Enriching Trip Information. *Transportation Research Record: Journal of the Transportation Research Board* 1870, 46.

Wu, L., B. Yang, and P. Jing (2016). Travel Mode Detection Based on GPS Raw Data Collected by Smartphones: A Systematic Review of the Existing Methodologies. *Information* 7.4, 67.

Zhang, L., S. Dalyot, D. Eggert, and M. Sester (2011). Multi-Stage Approach to Travel-Mode Segmentation and Classification of GPS Traces. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 3825, 87.

Zhang, S. and H. K. Lo (2018). Metro disruption management: Optimal initiation time of substitute bus services under uncertain system recovery time. *Transportation Research Part C: Emerging Technologies* 97, 409.

Zheng, Y., L. Liu, L. Wang, and X. Xie (2008). Learning Transportation Mode from Raw Gps Data for Geographic Applications on the Web. *Proceedings of the 17th International Conference on World Wide Web*. ACM, 247.

Zhu, Q., M. Zhu, M. Li, M. Fu, Z. Huang, Q. Gan, and Z. Zhou (2016). Identifying Transportation Modes from Raw GPS Data. *Social Computing*. Communications in Computer and Information Science, 395.

3

CHOICE SET GENERATION FOR PUBLIC TRANSPORT

This chapter is based on the following article:

Determining an efficient and precise choice set for public transport based on tracking data

Alessio D. Marra[1] & Francesco Corman[1]
[1]Institute for Transport Planning and Systems, ETH Zürich

Published in

Transportation Research part A: Policy and Practice, 12.2020
DOI: https://doi.org/10.1016/j.tra.2020.10.013

Contributions

A. D. Marra: Conceptualization; Data curation; Methodology; Formal analysis; Software; Visualization; Writing - original draft; Writing - review & editing
F. Corman: Conceptualization; Methodology; Writing - review & editing

Key findings

- Computationally efficient choice set generation algorithm with high precision in terms of coverage
- Analysis of choice sets' size and relevance of paths
- Choice sets used to estimate a route choice model for public transport from GPS data
- Identification of information provision that best represents passengers' choices

Additional notes to this chapter

Some aspects of this chapter are clarified in the appendix of the thesis (see superscript [Thesis Appendix]).

ABSTRACT

To understand the route choices of public transport users, it is important to know the information available to them, and the context present at that moment. In fact, each choice situation in a transport network has different characteristics and possibilities, also depending on the current status of the transport network. In this regard, travel diaries based on tracking technologies can capture precise observations for a long term. In this work, we exploit a large-scale tracking dataset, collected through a mode detection algorithm, to understand route choices of public transport users. We propose a choice set generation algorithm, able to cover more than 94% of the collected trips without any computational constraint. We compare the users' paths in the public transport network with different choice sets, under multiple performance indicators, including coverage, size, and fit. This latter is computed by the estimation of a Path Size Logit model.
The use of Automatic Vehicle Location (AVL) data allows comparing the available paths in terms of public transport vehicles used. We also consider different information provisions of network conditions and disturbances (full knowledge, no knowledge and current knowledge), and study which information provision best represents the choice set inferred by the observed users' behaviour. Estimating a Mixed Path Size Logit model, we identified high heterogeneity among the users in only a few aspects. Overall, a condition of no knowledge results as the best fit, i.e. users seem to take into account in a minor way the realized delays in the alternatives considered when deciding their public transport route.

Keywords

public transport; choice set generation; tracking; route choice; passengers' information; AVL data

3.1 INTRODUCTION

The route choices of passengers in public transport networks can depend on different factors, which make complex the problem of their understanding. In order to understand the passengers' route choices and identify their main characteristics, it is necessary to identify the available alternatives. The literature identifies two problems: the choice model, determining the actual chosen route out of a small set (choice set); and the identification of a choice set from a large set of all possible alternatives called the universal set (Bovy, 2009). We focus on the latter problem, the choice set (CS) generation, which focuses on identifying all and only the relevant alternatives, which more realistically are taken into consideration during the choice process of the passengers. We solve this problem by means of a novel CS generation algorithm, able to identify all possible alternatives given some fixed constraints. Such an algorithm is fast and accurate, with a minimal computation time (in the order of seconds for a large network), with more than 94% of collected trips covered by the CS, and a high fit in the resulting choice models. In addition, the approach is flexible, as it can consider alternatives as sequences of public transport lines or vehicles, more detailed than the commonly used definition of alternative as a sequence of stops.
The different context and information available to passengers, when making each different choice, are difficult to model through a stated preference survey. Instead, revealed preference surveys relate to the real choices of passengers in their context and can potentially explain more details on the choices. Nevertheless, these surveys require more resources and are more difficult to analyse. In this context, passive tracking and mode detection algorithms can help to reduce the burden on the users and collect more easily a dataset, which can fairly well represent the passengers' choices.
We test our CS generation algorithm on automatically collected travel diaries from a large amount of users for long time (order of weeks). Therefore, we compare the chosen public transport route with the available alternatives, which are defined by considering different information provisions for the users. The use of realized operation data allows detecting the exact vehicle used and the available alternatives in that specific moment (i.e. with the realized conditions of the network, possibly delayed). For this study, we collected travel diaries of a sample of Zürich citizens travelling for 3 weeks in their city. Regarding the data collection methods, we refer to Marra et al. (2019). Therefore, the focus of this work is on the CS generation algorithm and its evaluation in terms of coverage, computation speed

and fit. About the latter, we estimated a Path Size Logit model, which validates the results of the CS generation algorithm.
We illustrate the capabilities of the algorithm on the analysis of the available alternatives based on different information provisions, to understand the observed route choices. Based on the available Automatic Vehicle Location (AVL) data, we studied different information provisions of realized network conditions and disturbances: no knowledge; perfect knowledge; and current knowledge, i.e. knowledge of the conditions at the begin of the trip. Those three conditions result in three different CSs.
We assume that if a CS, based on a specific information provision, performs on the observed choices better than another one, based on a different information provision, given the same choice model, the former CS describes better than the latter the alternatives considered by the users in their choice process. This allows understanding how information affected the CS, which we assume as a proxy for how the passengers used information that might have been available to them during their choice process. Therefore, the information is accounted in the CS generation, and the same Path Size Logit model is estimated with different CSs. In this sense, we obtained the best fit assuming no knowledge of network conditions, resulting in an R^2 value of 0.65.
Such a study has implications on the design of route recommender engines, and how to best inform users (frequency/availability of en-route updates) about the realized delays of the network.
The main contributions of this work can be summarized as follows:

- A computationally efficient (running in a few seconds on a standard computer) choice set generation algorithm is proposed, based on constrained enumeration, and able to work with dynamic operation description, and different information provisions. The algorithm is evaluated on a large-scale tracking survey, collected from GPS data and automatic mode and vehicle detection, obtaining high precision both in terms of coverage (more than 94%) and model estimation (high R^2 and reasonable parameters).

- Aspects of choice sets rarely considered in literature are analysed, such as: the CS size and relevance of a path; the assumed walking distance; trips with transfers; and two levels of details are considered, namely representing public transport stages as sequences of vehicles, or lines.

- The users' choices are evaluated according to different information provisions of network conditions: no knowledge, current knowledge and perfect knowledge. Multiple Path Size Logit models, also including panel effects, are estimated with different CSs to understand the information provision better representing the passengers' choices. The results show how information provision results in significantly different choices sets, which allowed understanding the most likely information provision that each user considers when determining the possible alternatives for their route choice. In addition, heterogeneity has been identified among users regarding information provision, perception of a few travel time components and assumed walking distance.

The paper is organized as follows: Section 3.2 presents the state of the art; Section 3.3 describes the datasets used and briefly introduce the mode detection; Section 3.4 presents the CS generation algorithm; Section 3.5 describes the way we evaluate the model, under different assumptions on information provisions; Section 3.6 shows the results; Section 3.7 discusses the results and their policy implications; Section 3.8 contains the conclusions.

3.2 STATE OF THE ART

3.2.1 *Choice set generation*

Transport networks tend to offer a large number of alternative paths between a given origin and a given destination. The enumeration of all possible alternatives is not feasible in practice and it is unlikely that passengers take into account all of them (Gentile and Noekel, 2016). Therefore, to model the possible choices considered, an often used paradigm splits the choice process into a selection of relevant alternative paths (CS generation), and a selection of a choice out of the selected alternatives (choice model); this latter can be finally observed in real life (Bovy, 2009). We focus on the former step. A CS generation algorithm must be able to identify all and only the relevant paths. In this sense, a CS must guarantee high coverage, in terms of the ability to observe the passenger's path. It must also guarantee high precision, in terms of including only relevant paths. However, what is a relevant path cannot be objectively defined, making challenging the assessment of the CS quality.
Most of the work on route CS generation are focused on car trips (Ras-

mussen et al., 2016) rather than public transport trips. In addition, most of the proposed techniques are used (in different works) both for road and transit networks. Nevertheless, the two networks have important differences, such as, the transit network has less possible connections in space, but instead both timetable and transfers must be taken into account.
Most of the CS generation algorithms proposed in literature are heuristics, which need to find a compromise between the quality of the CS and the computation time. They can be divided in deterministic and stochastic. Among the deterministic, a basic approach is a K-shortest path based algorithm (e.g. Yen, 1971), which is known to generate too similar routes (Prato, 2009). Labelling approaches consist on identifying the shortest path using different cost functions, assuming several objectives for the user (Ben-Akiva et al., 1984). Link elimination algorithms eliminate one or more links from the least cost path before identifying the next least cost path (Rieser-Schüssler et al., 2013). Nassir et al. (2015) proposed a time-dependent K-shortest path with link elimination and constraints on maximum transfers, walking distance and waiting time, to identify a set of possible access stops. Similarly, De la Barra et al. (1993) proposed a link penalty approach, which after identifying the least cost path, gives a penalty to the cost of links of the identified path.
A different group of deterministic algorithms is formed by the constrained enumeration methods, which rely on a different behavioural assumption. Instead of identifying minimum cost paths, they assume users choose routes according to several rules (Prato, 2009). In this context, Friedrich et al. (2001) proposed a branch and bound method that enumerates all possible paths given some constraints, as the maximum number of transfers and dominated connections. Hoogendoorn-Lanser et al. (2007) generated CSs in a multimodal transport corridor in the Netherlands, given constraints on several factors, such as time, space and money. Typically, rationality of the users can be assumed in order to filter those unreasonable paths which would not be rational choices; for instance a path visiting twice the same stop. Cats (2011) developed a recursive search method with exclusion of unreasonable alternatives using filtering rules based on walking distance, in-vehicle time and number of transfers. Tan et al. (2007) used a recursive search to find all possible paths, given constraints on the time and transfers. Since all these methods are heuristics, there is not a specific reason to prefer one approach to another. Nevertheless, constrained enumeration methods are the most used in the recent works and achieve the highest performances (Bovy, 2009), in terms of coverage level of observed routes

and estimation quality of estimated choice models.
Stochastic CS generation algorithms include a stochastic factor in the generation of each path. Therefore, repeated paths are computed and new paths are added to the current CS (Fiorenzo-Catalano, 2007; Frejinger et al., 2009). A stochastic factor can be considered both in network attributes (e.g. travel time, waiting time, etc.) and in their perception by the users. Fiorenzo-Catalano (2007) proposed a doubly stochastic approach for multi-modal networks, including private vehicles. In their work, both link attributes and behavioural parameters are randomized. They stated that, with a correct calibration of the underlying parameters, this method guarantees heterogeneous paths with reasonable computational costs.
As discussed by Rasmussen et al. (2016), in a schedule-based public transport network a path can be considered at different levels of detail. Some possibilities are: (1) vehicle level, in terms of vehicles used (i.e. specific run of a line, as in this work); (2) line level, in terms of lines used (Rasmussen et al., 2016); (3) stop level, in terms of stops passed, regardless of the lines used (Cats, 2011). Vehicle or line levels can be combined with stop level, considering also the stops where boarding, transferring and alighting occur. The chosen level of detail can affect significantly the computation time and the evaluation of the CS.
The evaluation of a CS generation algorithm in multi-modal public transport is rarely proposed, since a detailed revealed preference survey might be necessary. In fact, according to Meyer de Freitas et al. (2019), stated preference surveys have disadvantages in modelling route choices in a multi-modal network, given the high number of available alternatives and the response burden required. In addition, they can produce biased results due to framing effects (Beck et al., 2017; Meyer de Freitas et al., 2019). Bovy (2009) identified as common evaluation indicators the coverage of observed routes in the relative CSs and the estimation quality of an estimated choice model. The coverage is usually referred to as the percentage of observed routes that are fully (or partially) included in the relative generated CSs. Rasmussen et al. (2016) evaluated their algorithm using a questionnaire-based survey, resulting in a collection of travel diaries. The observed paths are therefore map-matched to the transport network. They evaluated the CS coverage (at stop level), obtaining a value above 90% (considering 100% of overlap) for some of the proposed configurations. Nevertheless, these configurations tend to build CSs with counterintuitive paths. In a previous work (Anderson et al., 2014), they showed a line-level coverage of 78% for CSs of average size of 40.4. Rieser-Schüssler et al. (2014) evaluated their

doubly stochastic CS generation algorithm using a GPS-based survey. Automatic mode detection and map-matching to the Zürich public transport network are performed, and validated by the users. Unfortunately, only 193 trips have been analysed, obtaining a coverage of 77%. Based on GPS data, there are also studies focused on car trips, such as Rieser-Schüssler et al. (2013). They applied a breadth first search with link elimination, obtaining a coverage of 63% for CSs of size 20 and 73% for CSs of size 100.
We believe some important aspects, which can provide an additional value to a CS generation algorithm, are missing in literature. To the best of our knowledge, small attention is given to the analysis of the CS size, in terms of which paths are relevant and should be included in a CS. A large CS size affects negatively the efficiency of the algorithm, while a small CS can affect its precision in terms of coverage. The CS size is an important aspect potentially affecting the correctness of the parameter estimation (see the recent discussion in Zimmermann and Frejinger, 2020). In addition, the definition of relevant path is subjective, dependent on assumptions of rationality (exploited in labelling approaches) and often hardly quantifiable or justifiable against real life behaviour, making difficult the selection of the CS size. A second missing aspect is that no work has ever used simultaneously realized data of both operations (AVL data) and passengers (tracking or automatic fare collection systems) to improve and adapt a CS generation algorithm. Finally, no work has ever evaluated different information provisions for passengers, comparing different CSs for the same trip. This comparison can be useful to understand which information a passenger has (and/or likely used) and if the passenger's choices adapt to network disturbances. In fact, a CS based on a given information provision represents the relevant alternatives, assuming that information available. Therefore, we assume, considering a small number of exemplary information provisions, the one associated to the CS matching best the passenger's choices represents best the information available. We based such an analysis on some of the assumption and results from the literature. In this sense, information acquisition is identified as a component for CS formation by Bovy (2009). In addition, Jiang et al. (2019) identified that some travellers do not make use of real-time traffic information and they assumed information is costly, in terms of acquiring and processing it. In this context, but for car trips, Ding-Mastera et al. (2019) investigated drivers' ability to plan ahead and utilize real-time information.

3.2.2 *Route choice behaviour*

Most of the mentioned work combined a CS generation algorithm with the estimation of a route choice behavioural model. The purpose is either to validate a CS generation algorithm estimating a route choice model (Rasmussen et al., 2016; Ton et al., 2018), or to use a CS to estimate a behavioural model and explain the users' behaviour in the specific test case (Anderson et al., 2014). In this context, Bovy (2009) argued the modelling steps of choice set generation and choice modelling should be explicitly separated since they are distinct mental processes of a traveller. One of the first works estimating transit route choice at individual level is Bovy and Hoogendoorn-Lanser (2005), using a Hierarchical Nested Logit Model and a Multi-Nested GEV Model. Nevertheless, their analysis is based on only 235 observations, and focuses on routes with train as primary mode. In recent works, for route choice estimation, the most used model is the Path Size Logit, a variant of the Multinomial Logit Model introducing a correction term for overlapping paths (Anderson et al., 2014; Montini et al., 2017; Tan et al., 2015). Anderson et al. (2014) estimated a Mixed Path Size Correction Logit in the multimodal public transport network of the Greater Copenhagen Area, obtaining an adjusted R^2 of 0.45. Rasmussen et al. (2016) tested different levels of stochasticity in their CS generation algorithm and analysed the stability of the parameters of the estimated route choice models. The highest R^2 obtained is equal to 0.66, even if the estimation of transfer-related parameters is reported to be affected by the level of stochasticity in their model. An interesting contribution is given by Montini et al. (2017), estimating a Path Size Logit on public transport trips, reported as travel diaries collected from GPS traces of people travelling in Zürich (only 273), obtaining an R^2 of 0.595. Nassir et al. (2015) adopted a Nested Logit model, including a correction term based on the Path Size Logit, to identify the access stop choice.

Finally, it is important to mention the existence of methods studying user behaviour without explicitly generating a CS, such as the recursive logit (RL, see Fosgerau et al., 2013). We see RL models as an alternative approach and applicable for a similar study (with the appropriate changes). To the best of our knowledge, and as reported in Zimmermann and Frejinger (2020), only one work applied the recursive logit to a multi-modal transport network (Meyer de Freitas et al., 2019)). The work reports on very heavy computational requirements for RL approaches, against which our methods compare favourably (see Section 3.6.2). Those computational requirements suggested the authors to simplify the transport network into a static network, i.e. without considering realized arrival/departure times of public transport. This would not allow understanding a key issue at the focus of the present paper, namely the impact of transit information in case of actually disturbed operations (See Section 3.6.4). Finally, without explicitly generating a CS, it is not easily possible to make considerations on the relevance of an alternative and the CS size (see Section 3.6.3).

3.3 TRACKING SURVEY AND DATASETS

A key aspect of the present work is the use of realized data for both passengers and operations. We based our analysis on urban trips of Zürich residents. Therefore, we used realized data of all the public transport vehicles travelling in the city, including trains, trams and buses (Swiss Federal Office of Transport, 2019). For each day, the arrival and departure times of each vehicle at each stop are provided, both planned and realized. This allows making more detailed analyses, compared to using only the planned timetable, since the different network conditions of each day are considered.

The passengers' data are collected through a tracking survey based on a smartphone application, the *ETH-IVT Travel Diary*, able to continuously collect GPS data for multiple weeks, without affecting the battery consumption (Marra et al., 2019). The user interaction is reduced to a minimum, consisting only in installing the application. Therefore, a mode detection algorithm is applied to identify the recorded movements. Zürich residents were invited to participate in the study, offering them a compensation of 20 CHF. The survey consisted of answering a starting and an ending questionnaire, installing the app and let it run for at least three weeks. Table 3.1 provides general statistics on the survey. Despite invited people were selected without any specific criteria, the participants are on average young and highly educated (most of the people with only mandatory school degree are still studying). This is probably due to the nature of the task, which requires familiarity with smartphones, or to the offered compensation, which can be more attractive for younger people. Nevertheless, the representativeness of the sample is not a key aspect for this work.

Period	03/04/2019 - 02/06/2019
Participants	172
average age	32.6
% female	43%
occupation	63% workers; 25% students; 7% both; 5% other
education	54% university or high professional; 36% high school; 10% mandatory school

TABLE 3.1: Tracking survey information.

3.3.1 *Mode detection*

The GPS raw data are processed to identify activities, trips and stages and automatically infer the transport mode used. First, the GPS points are classified as activities or trips, identifying areas with high-density of points as activities, and the connection between the activities as trips. Afterwards, a segmentation step divides trips in walk-stages and other-stages, based on the points' speed. Finally, a mode detection algorithm classifies the other-stages as public transport or private stages, comparing the users' GPS traces with the realized paths of the operations. In particular, if a user's stage begins and ends at the same time and at the same place of a public transport vehicle, the stage is marked as performed on that vehicle; otherwise, the stage is assumed as private, and it is classified as a car or bike stage afterwards. Therefore, a key aspect of the mode detection algorithm is the identification of the exact public transport vehicle used. In particular, the algorithm detects the following information for each public transport stage: the mode (Bus, Tram or Train), the line, the specific vehicle of that line, the user's departure stop and time, the user's arrival stop and time. More details on the mode detection algorithm are in Marra et al. (2019). We exploit this feature of the mode imputation, able to pin point the exact vehicle, in the analyses performed in the present paper regarding the choice set.
Table 3.2 shows the general results of the GPS data processing applied to the collected dataset. While activities, trips, walks and stages are always identified, the mode detection is limited to stages inside the city of Zürich. Therefore, considering only the trips inside the city, and assuming trips with transfers as performed by only one mode, the mode share is the following: walk 23%, public transport 38%, car 15%, bike 13%, mixed 10%

(public and private). Despite our dataset has not been selected to be representative of the population, the mode share is very close to official reports (Stadt Zürich, 2020, reporting 41% of modal split for public transport), especially for public transport trips, the only one studied in detail in this paper.

Tracked days	3785
Avg. days per person	22
Activities	14066
Trips	13913
Stages	22654
Walks/Transfers	29890
Stages outside Zürich	8472
Public transport stages	8849
Private mode stages	5333

TABLE 3.2: Results of the GPS data processing.

In this work, we consider only public transport trips, i.e. trips using only and at least one public transport vehicle. Therefore, we assume that people did not consider private or alternative transport modes for these trips. Table 3.3 shows more details on the public transport trips. Interesting is that 40% of the trips are performed with at least one transfer, making the route choice problem more challenging compared to direct connections.
With tracking and automatic mode detection, it is possible to build large-scale travel diaries with low efforts, since the burden on the users is reduced to a minimum. Knowing the exact public transport vehicle used for each trip would require a constant participation that can be hard to obtain for a long period. Despite this, some limitations exist, such as errors in the different steps of the mode detection, mainly due to the quality of the GPS data. This leads to missing or wrongly detected activities, trips or modes that add noise to the dataset. As shown in Marra et al. (2019), the average detection accuracy of the considered approach is 86.14% (68.3% for trains, 80.6% for bus and trams, 84.4% for private vehicles and 94.4% for walks). This accuracy refers to the percentage of stages correctly classified by the algorithm, and it is based on validation and testing on a similar dataset, collected in Basel (Switzerland), where the respondents manually labelled their stages. In any case, a wrong detection (misclassification) of a public

Urban public transport trips	2901
% # stages per trip	{1: 60%, 2: 29%, 3: 9%, 3+: 2% }
% modes used	{Tram: 52%, Bus: 38%, Train: 10% }
Avg. duration per trip	21.7 min
Avg. air distance per trip	2.88 km

TABLE 3.3: Information on public transport trips in Zürich.

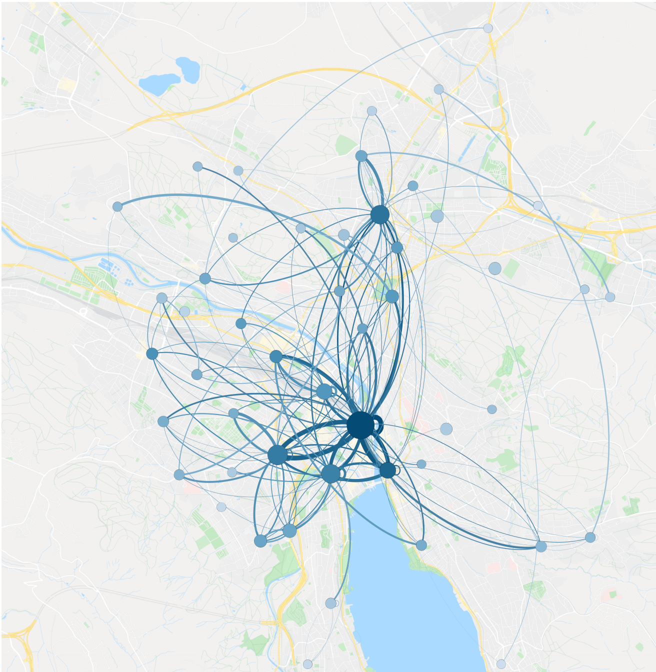


FIGURE 3.1: OD flows of the collected public transport trips in Zürich. The width of the arcs is proportional to the number of trips between an Origin and a Destination. Near locations are aggregated using the mean shift algorithm. Locations with 5 or less trips are not shown. Background from google.com/maps.

transport trip still represents a feasible path to reach the destination. For example, if a car trip is identified as performed by a certain bus, that bus actually did a path similar to the one of the car, at the same time. In fact,

the algorithm considers for a possible public transport trip only vehicles that did a path similar to the one of the user, according to a likelihood function. In addition, a misclassification of a car trip as a public transport trip with a transfer is rarer, since it involves three consecutive misclassifications (two vehicles and one walk for transfer). Therefore, in this work we assume correct the mode detection results. Despite these limitations, the following results were not significantly affected, since we obtained high coverage compared to the state of the art, and a R^2 value of a route choice model equal to 0.65.
Figure 3.1 shows the distribution in Zürich of the public transport trips considered. Most of the trips are located near the city center or the railway stations. Nevertheless, the dataset contains also peripheral trips and spreads through the whole city area.

3.4 CHOICE SET GENERATION ALGORITHM

In this Section, we describe a novel choice set generation algorithm, able to work with different information provisions. The proposed algorithm results efficient in terms of computation time, and with high precision, in terms of coverage and model estimation.
We propose a CS generation algorithm based on constrained enumeration, able to generate all possible alternatives given a maximum duration and a maximum number of transfers. As an alternative, we consider a path formed by walks and public transport vehicles. Same vehicles but different walks are assumed as the same alternative (e.g. same transfer between the same vehicles, performed at different stops). The algorithm is based on a schedule-based model, performing a Depth First Search (DFS) on a diachronic graph (time-expanded network). The graph is modelled from the AVL data of the public transport operations to model the timing of each trip of each vehicle.
The innovative features of the proposed algorithm are both methodological and related to its applicability. Regarding the methods, we used a different time representation from the literature, therefore a different graph. In addition, we propose a two-phases search to identify each path in the CS, reducing strongly the computation time. Regarding the applicability, the algorithm allows analysing choices both at line and vehicle level and is able to integrate different information provisions, to derive implications on information provisions to users. The information is provided as an input to the CS generation algorithm, in the form of transit events to consider. This

allows full flexibility on which information to consider, without affecting the correctness of the algorithm.

3.4.1 *Graph representation*

We use a transport network graph, where nodes represent specific events in a time-dependent fashion; and links connect them in an acyclic structure. Typically, in a schedule-based model, the transport network graph is represented using time-discretization, i.e. creating multiple nodes for each stop at different time intervals (Gentile and Noekel, 2016). In this case, a transit network and a walking network are connected at the stop location nodes. Instead, we decided to not discretize in time, but rather express all events in their realized time (event-based rather than time-based discretization). In particular, each node models the arrival or departure event of a vehicle at a stop.
Transfers are modelled with direct arcs between two nodes whenever a transfer is possible. In addition, not all possible transfers are allowed in our model, since we introduced constraints already at this step[Thesis Appendix 3], to further speed-up the paths search.
We consider as input a list of *stops* and a list of *vehicles* running in the network (retrieved from the AVL data). As a vehicle, we refer to a single service provided between the first and last stop of a planned run. The arrival and departure times of the vehicles at each stop represent the information provided (see Section 3.5.3).
The graph structure is described formally in Table 3.4. We considered a graph $G = (V, E)$, where each node in V is a triple $(vehicleId, stopId, A/D)$, representing the arrival or departure (A/D) of a public transport vehicle ($vehicleId$) at a stop ($stopId$). The nodes *Origin* and *Dest* are created for each OD, before computing the CS.
E models the trips of the vehicles, the possible transfers and the possible starting and ending walks. The first two groups of arcs (third and fourth rows) connect all the nodes of a single vehicle, modelling its trip. The third group of arcs models the transfers. The arrival of a vehicle is connected to the departure of another one if the following conditions hold: the walking distance is $\leq N$; the required walking time is sufficient; the waiting time is $\leq TD$. In addition, for each possible transport line, only the best transfer is considered. For instance, arriving with the bus A1 at 10:00 at a station S, where three buses of the same line (B1, B2, B3) are departing respectively at 10:10, 10:20 and 10:30, only the transfer to B1 is considered, since it is not

reasonable to wait more for the same bus line (unless special cases of denied boarding, that we do not consider). The Origin and the Destination of a trip are connected to the graph through the last two groups of arcs. The Origin is connected to all nodes departing in stops nearby, i.e. reachable from the starting point within a distance equal to N, as for the transfers. Even in this case, only the first feasible boarding for vehicles of the same line is considered. The Destination is connected to all the vehicles arriving at stops nearby (distance N).
Regarding the starting time, we include an extra adjustment factor to the starting time of the user's trip, to take into account possible errors due to the GPS quality and mode detection. In fact, wrong GPS position, missed walk or late trip starting time can make the user's trip infeasible in the graph. Therefore, the starting time is defined as $min(TS_t, FV_t - W_t)$, where TS_t is the trip starting time, FV_t is the departure time of the first vehicle used by the user and W_t is the walking time necessary to reach the used stop.
Modelling parameters are a maximum transfer distance of 700 meters (N), the average walking speed of 1.5 m/s and a maximum waiting time of 30 min (TD). The last one does not affect the CS quality, since in the mode detection a user not moving for more than 15 min (*ActivityTime*) is considered doing an activity. In addition, the effective frequency of public transport in our test case is generally much higher than 30 min.

3.4.2 *Selection of the alternatives*

For a given OD, the choice set generation algorithm works as follows:

- We restrict the transport network graph introduced above to the subgraph of only those nodes reachable by both the origin and the destination. This strongly reduces the search space, since only the vehicles able to connect the OD are considered. In addition, we consider only vehicles reaching the destination before 2 times (*maxTime*) the duration of the shortest path. Since the subgraph is time-expanded, it is directed and acyclic, and all the possible paths between two nodes can be efficiently found with a DFS. We perform a DFS to identify all possible combinations of vehicles that can be taken (i.e. sequences of vehicles without specifying the stops). A threshold of maximum 2 transfers (*maxTrans*) is also considered.

Node (v,s,A)	$\forall v \in vehicles, s \in stops : v\ arrives\ at\ s$
Node (v,s,D)	$\forall v \in vehicles, s \in stops : v\ departs\ from\ s$
$(v, s_i^v, A) \to (v, s_i^v, D)$	$\forall v \in vehicles, \forall i : 1 \leq i \leq v_{stops}$
$(v, s_i^v, D) \to (v, s_{i+1}^v, A)$	$\forall v \in vehicles, \forall i : 0 \leq i \leq v_{stops} - 1$
$(v,s,A) \to (w,z,D)$	$\forall v,w,s,z :\ distance(s,z) \leq N,$
	$walkTime(s,z) \leq z_{time}^{w,D} - s_{time}^{v,A} \leq TD,$
	$(w,z) = firstTransfer(v,s,w_{line})$
$Origin \to (v,s,D)$	$\forall (v,s,D) : distance(Origin, s) \leq N,$
	$walkTime(Origin, z) \leq s_{time}^{v} - Origin_{time} \leq TD,$
	$(v,s) = firstTransfer(Origin, v_{line})$
$(v,s,A) \to Dest$	$\forall (v,s,A) :\ distance(s, Dest) \leq N$
Notation:	
v_{stops}	number of stops passed by v
s_i^v	i^{th} stop of vehicle v
$s_{time}^{v,A/D}$	arrival/departure time of v at stop s
$firstTransfer(v,s,l)$	(w,z) : w is the first vehicle of line l and z is the nearest stop allowing a transfer from v and s to line l

Table 3.4: Public transport graph structure.

- For each combination of vehicles identified, there are multiple options to travel with the same vehicles, depending on the stops at which to get on, transfer and get off. These options have only marginal differences on the walks, and correspond to the same alternative according to our definition. Therefore, for each combination of vehicles, the path with shortest walks is selected. Practically, this can be obtained computing the shortest path on an even smaller subgraph, formed by only the nodes of the vehicles.

- The final CS is given sorting the list of paths by a simple cost function: the travel time with a transfer penalty of 5 min.

The problem of identifying a single path is decomposed in first identifying the combination of vehicles (i.e. stages), and then identifying the best option to travel with them (at which stops to get on, transfer and get off).

This results more efficient than performing only a DFS, identifying directly the final path.
Furthermore, the CS generation algorithm does not consider the following paths, which we assume unrealistic:

- paths using two times the same stop;
- paths using additional vehicles compared to others (therefore, with no improvement in travel time);
- paths using the same lines of less costly paths (according to the mentioned cost function).

These constraints can be easily considered during the DFS.
The only parameters affecting the CS size and inversely the computation time are *N*, *maxTime* and *maxTrans*. As we show in Section 3.6.1, the chosen values are large enough for our test case, Zürich, the city with the largest public transport service in Switzerland. We remark that the transfer penalty affects only the sorting of the CS and not its coverage. Douglas and Jones (2013) reviewed transfer penalty estimates in literature and they showed how there is no agreement on a common value, even if most of the estimates range between 5 and 9 minutes of travel time. Finally, since it is not possible to know without prior knowledge if a user is waiting for a transfer or is performing an activity, the minimum time for an activity (*ActivityTime*) is hard to determine. After a comparison of the results of this work, considering an *ActivityTime* of 10, 15 and 20 minutes, we evaluate that its value does not significantly affect the outcomes.

3.5 MODEL EVALUATION AND INFORMATION PROVISION

3.5.1 *Choice set evaluation*

We evaluated the CS generation algorithm according to four factors: the coverage, the size, the computation time and model estimation. With coverage we refer to the percentage of trips identified in their CS. High coverage shows the ability to capture the users' trips. We remark that we are considering paths at vehicle level, meaning that two paths are considered equal only if they used exactly the same vehicles. Coverage at line level is easier and more commonly found in literature. Therefore, we compared coverage both at vehicle and line levels. This leads to a different resolution from considering paths at stop level. In this last case, hyperpaths formed

by different vehicles connecting the same stops are considered; therefore the stops are assumed relevant, and not the vehicles.
The CS size should guarantee high coverage, even though this is seen often as detrimental to high precision, in terms of including only relevant paths. Since a good CS must contain all and only the choices relevant for the users, we analysed the increase in coverage and the performance of model estimation according to the CS size, to identify which paths can be considered relevant.
The computation time should be minimal, to guarantee the applicability of the algorithm in practical contexts, as route recommender systems. In this sense, we show the required time of the CS generation algorithm is significantly low (median of 7 s per CS).
Finally, a good CS should guarantee high estimation quality of an estimated choice model (Bovy, 2009), which we describe in Section 3.5.2.

3.5.2 *Estimation of route choice model*

In order to validate the CS generation algorithm and to identify the conditions resulting in a CS that better represents the users' behaviour, we estimated a behavioural model for the choice of public transport paths. Therefore, we estimated a Path Size Logit model (Equation 3.3) considering two possible levels of detail, vehicles and lines, and various information provisions. The software used for the estimation is Biogeme (Bierlaire, 2018). For each trip, we estimated a utility function $U_{trip,CS}(\overrightarrow{\beta})$ (Equation 3.1) based on the following parameters: tram (travel) time, bus time, train time, walk time, transfer time, # transfers, path size cost (PS). We denote the first and last walks as walk time, and the intermediate walks as transfer time. During a transfer, we could not discriminate between walk and waiting time, given the quality of the GPS data. In addition, we considered a penalty for each transfer. Monetary costs were not considered in this work, since inside the city of Zürich the price is fixed, and therefore it does not change among the alternatives.
The path size cost (Equation 3.2) is based on the formulation proposed by Bovy et al. (2008), considering the stages (vehicles or lines) forming each trip and the travel time as a measure of length, as in Tan et al. (2015). The PS is equal to 0 for a non-overlapping trip (i.e. its stages are only in this trip of the CS), and decreases negatively with the times each stage is considered in the CS. Considering other Path Size factors, as the classical one

proposed by Ben-Akiva and Bierlaire (1999), did not change significantly the results.

$$\begin{aligned} U_{i,CS}(\overrightarrow{\beta}) = \beta_{tram} * tram\ time + \beta_{bus} * bus\ time \\ + \beta_{train} * train\ time + \beta_{walk} * walk\ time \\ + \beta_{tt} * transfer\ time + \beta_{transfer} * \#transfers + \beta_{PS} * PS_{i,CS} \end{aligned} \tag{3.1}$$

$$PS_{trip,CS} = - \sum_{stage\ s \in trip} \frac{length(s)}{length(trip)} \ln(times\ s\ in\ CS) \tag{3.2}$$

$$P(trip|CS; \overrightarrow{\beta}) = \frac{e^{U_{trip,CS}(\overrightarrow{\beta})}}{\sum_{j \in CS} e^{U_{j,CS}(\overrightarrow{\beta})}} \tag{3.3}$$

3.5.3 *Information provision*

According to which data are used to generate the diachronic graph (planned or realized data), different information provisions of network conditions can be assumed, generating different CSs. We considered three different provisions: perfect information, assuming known all disturbances during the day; no information, assuming no disturbances; current information, assuming available the information of the next vehicles of each line departing near the user, at the begin of the trip. The perfect information provision is unrealistic, since it considers full knowledge of future network conditions, but it allows to identify the real available alternatives. The current information provision represents the case of a user checking the network conditions before departing. The no information provision represents a user relying only on the timetable. In this latter case, a connection missed due to a delay will still be present in the Timetable CS, but not in the Realized CS; the CSs will differ in the amount of possible alternatives.
This work aims to identify which information provision results in a CS, which represents best the users' behaviour. In this sense, Bovy (2009) explicitly modelled the information acquisition as a component for CS formation; and suggests that CS generation and choice modelling are distinct mental processes and they should be separated. Therefore, we evaluated the same Path Size Logit model proposed in Section 3.5.2, considering the three different CSs based on the different information provisions.
Again, we restate the underlying assumption of our study, which is: considering only a few possible cases of information provision, keeping the same model (Path Size Logit), but testing different CSs, the one performing best is more suitable to determine the choices considered by the users.

We highlight that the proposed analysis compares only the three proposed information provision. Possibly, additional information provisions could be taken into account. Further analysis in this direction are left for future work.

3.5.4 *Users' heterogeneity*

We analyse the heterogeneity among the users both in terms of information provision, and in terms of perceived costs of the different travel time components.
As stated by Bovy (2009), the CS formation process needs to consider heterogeneity among users. In fact, preferences, available information and other conditions are different among individuals making a similar trip, resulting in a different composition of the set of considered alternatives. In this sense, we report the same analysis described in Section 3.5.3, but for each single user separately.
Regarding the perceived costs of the different travel time components, to observe the heterogeneity among the users and possible panel effects, we estimated a Mixed Path Size Logit model (Anderson et al., 2014; Prato et al., 2014; Schmid et al., 2019). In this model, the βs are random parameters, distributed according to a probability density function $f(\beta|\theta)$. The probability to choose a path is therefore the following:

$$P(trip|CS) = \int \frac{e^{U_{trip,CS}(\overrightarrow{\beta})}}{\sum_{j \in CS} e^{U_{j,CS}(\overrightarrow{\beta})}} f(\beta|\theta) d\beta \tag{3.4}$$

The choice of the distribution f varies across the literature and types of parameters. Typically, the normal or log-normal distributions are chosen. If on one side the log-normal allows to restrict values to a certain sign, on the other it can provide a too broad distribution, given its long tail (Hess et al., 2005). For this reason, in this work we considered the normal distribution for all the parameters.
The model was estimated with *mixl* (Molloy et al., 2019), allowing computing for each user the individual-specific cost coefficients, conditional on the observed choices. In particular, the expected value of the parameters

for a certain user n ($\bar{\beta_n}$) is defined as follows (Sillano and Dios Ortúzar, 2005):

$$\bar{\beta_n} = \frac{\sum_{r=1}^{R} L_n(c_n|\beta_n^r)\beta_n^r}{\sum_{r=1}^{R} L_n(c_n|\beta_n^r)} \tag{3.5}$$

$$L_n(c_n|\beta_n^r) = \prod_{t=1}^{T_n} P(c_{n,t}|CS_{n,t}, \beta_n^r) \tag{3.6}$$

where R is the number of draws to compute the expected value, c_n is the set of choices of the user n, and T_n is its number of choices.

3.6 RESULTS

3.6.1 *Choice set coverage*

We evaluate the performance of the proposed algorithm already distinguishing the three different information provisions identified. For each public transport trip observed, three different CSs were computed, based on three different information provisions: perfect information, no information and current information, that we refer as realized, timetable and current CSs. Figure 3.2 shows the coverage of the CSs according to their size, considering both vehicles and lines. The figure shows also the adjusted R^2 of a route choice model, which is discussed later in Section 3.6.3. We can see that the timetable CS has always a higher coverage than the other two. Moreover, the 90% of coverage (line based) is already obtained among the first 13 paths for the realized CS and 9 paths for the timetable CS. A maximum size of 100 is set for the analysis (even if the algorithm identifies all possible paths, which can be more). Only 6.4% of trips are not covered in the first 100 paths of the realized CS, and 5.5% for the timetable CS.
6.4% of users' trips are not in the realized CS, and are not identified for the following reasons: 41% for a long walk; 28% for a too long duration; 14% are after the 100$^{\text{th}}$ position; 7% are dominated paths; 4% do more transfers; 4% do a fast transfer, which cannot match the walking speed considered in the model; 2% used two times a same stop. Therefore, the only reason that can slightly affect the coverage is the long walk. Unfortunately, increasing the walking distance by 250 meters (950 meters) increases the coverage only by 1.2%. Therefore, we consider valid our choice of the parameters. In addition, it should be considered that two sources of noise are present: the users, which can make mistakes or choose a particular path for external reasons; the mode detection, which can make mistakes, as identifying

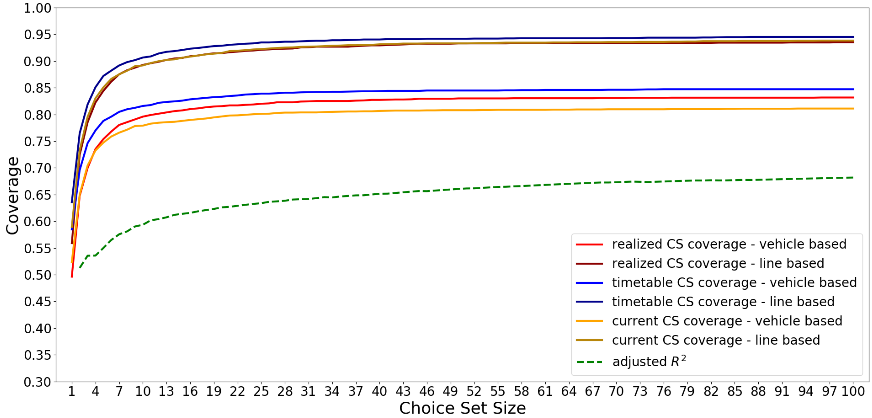


FIGURE 3.2: Choice set coverage based on different information provisions. The adjusted R^2 of the path size logit, line based, considering no information, is also shown.

a single trip between two activities, without identifying an existing activity in the middle.

3.6.2 *Computation time*

The computation time to compute a single CS is very short, even though no specific code optimization or parallelization was performed and a standard personal computer was used. On average, a graph representing a network during one day is formed by 500 thousand nodes and 10 million arcs. Nevertheless, the median of the computation time to generate the alternatives is 7 s and the average of the values below the 95 percentile (100 s) is 12 s. There is a small percentage of CSs requiring higher computation time, in the order of minutes. These are mostly long-distance paths, poorly connected, which require more transfers and travel time. In a practical context, further rules can be added to control and pre-process these specific cases. Since the estimation time of the Path Size Logit model is negligible (order of seconds), we can compare the order of magnitude of the required computation time by the proposed CS generation algorithm, and the recursive logit of Meyer de Freitas et al. (2019), both applied in Zürich. The recursive logit required around 1 min per observation and 80 GB of memory using

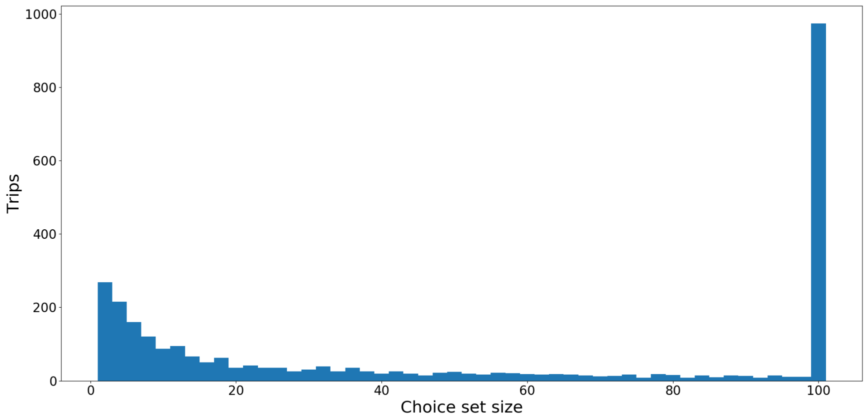


FIGURE 3.3: Distribution of size for the realized choice set. A limit of 100 paths is fixed.

a cluster; much higher than the 12 s and 7.5 GB required by the proposed algorithm, using a standard personal computer.

3.6.3 *Choice set size and relevance of a path*

Figure 3.3 shows the distribution of the size of the realized CS computed for each trip (the distributions of the other CSs are similar). Most of the CSs have a size of 100 (the fixed limit), even if there is a big portion (40%) with size less than 20. These are mostly trips with one or few options dominating the others in terms of travel time. Nevertheless, the CS generation algorithm obtains high coverage, even with relatively small CSs.
Despite the high coverage showed in Section 3.6.1, it is unreasonable that a passenger considers 100 alternatives to choose a path[Thesis Appendix 4]. In addition, an unnecessarily big CS increases the computational costs of following analyses, as route choice estimation. To identify a suitable amount of alternatives, we estimated the Path Size Logit model described in Section 3.5.2 with different CS sizes. Figure 3.2 shows both the variation of the coverage and the adjusted R^2 (of the Path Size Logit applied to the timetable CS, line based, reported dotted in green) according to the CS size. Both values grow more quickly for small CSs, and more slowly for bigger CSs. This shows that large size CSs may contain irrelevant paths that are not considered by the users. For our test case, we can identify 40

as a possible reasonable value of CS size. In fact, the coverage is 94% and the R^2 is 0.65, that do not differ significantly from the values obtained for size 100 (94.5% and 0.68) or even 500 (94.7% and 0.7).
Compared to the estimated parameters for a CS size of 40 (shown in Table 3.5), the ones estimated for a CS size of 100 are very similar (not shown for brevity). Moreover, a larger CS size brings marginally smaller changes, i.e. their values converge. In particular, with a CS size larger than 10, the estimated value of each parameter is already within 10% of the values at 100 (except the PS, which depends on the CS size). In conclusion, we can conclude that considering CSs with size larger than 40 does not change significantly the model interpretation and performances in our test case. Finally, it is important to notice that the CSs are sorted by travel time, including a transfer penalty of 5 min. Sorting the CSs by a utility function estimated by a route choice model determines a faster increase of the coverage (but not a higher value).
In this Section we analysed the trade-off between the CS coverage and model estimation on one side, and the CS size on the other. This represents a first step in analysing the CS size and irrelevant paths. Despite this, we acknowledge that the relevance of a path and the number of alternatives considered can vary across users and trip characteristics. We focus on some selected aspects of this heterogeneity in Section 3.6.8. We consider this analysis as a starting point and we leave further research for a future work.

3.6.4 *Performance of route choice model and information provision*

We estimated the Path Size Logit model both to validate the CS generation algorithm and to understand which information provision can better represent the users' choices, as explained in Section 3.5. Table 3.5 shows the results of the estimation with the three CSs, considering 40 as maximum size, and paths described at line level. The estimation considering paths at vehicle level is comparable and is discussed in Section 3.6.5. As a measure of goodness of fit, the adjusted R^2 obtained using the timetable CS is 0.65. Despite the results strongly depend on the test case considered, this value shows that our CS generation algorithm, based on modelling each alternative as a sequence of vehicles (or lines), is able to explain most of the users' choices. We consider the model reliable and usable for recommending policy actions (Garcia-Martinez et al., 2018, stated the same, given their model had an $R^2 = 0.51$). Other possible explanations for this value are:

Parameter	Timetable estimate	t-test	Realized estimate	t-test	Current estimate	t-test
In vehicle travel time (s)						
Tram	-1	-21.7	-1	-21.1	-1	-20.7
Bus	-1.14	-20.2	-1.11	-19.7	-1.15	-19.7
Train	-1.19	-12.4	-1.18	-11.2	-1.18	-11.7
Walking time	-2.56	-42.6	-2.6	-41.1	-2.63	-41.6
Transfer time	-1.06	-15.6	-1.17	-15.9	-1.16	-15.9
No. of transfers	-889	-27.4	-1050	-34.7	-940	-27.7
Path Size	55.4	2.49	64.5	2.68	96.2	4
Observations	2719		2687		2693	
Null log-likelihood	-7361		-7572		-7690	
Final log-likelihood	-2555		-2898		-2870	
$\bar{R}^2$	0.65		0.62		0.63	
Cross-val. $\bar{R}^2$	0.655		0.616		0.63	
St. dev.	0.01		0.014		0.017	
Scaling factor	0.0038		0.0033		0.0034	

Table 3.5: Path size logit estimates for CSs with different information provision, maximum size 40, line based. The parameters are scaled (multiplied by the scaling factor) to have the tram travel time coefficient equal to -1.

the nature and the size of the dataset, based on automatic mode and vehicle detection; the high coverage of the CS; users related factors, e.g. the average young age and the familiarity with smartphones (this can lead to a major use of route recommender systems).

Regarding the comparison of different information provisions, the estimation results of the three models are quite similar, but the timetable CS has the best performance. This confirms the better coverage obtained for this CS. To test the statistical significance of the values computed for the three models, we applied Monte Carlo cross-validation, with R^2 as performance measure. Specifically, we randomly divided the dataset into 75% training-set and 25% test-set. The model is trained with the former and evaluated on the latter. We repeated this operation 50 times and observed the distribution of the R^2. The mean of the R^2 for the three models (timetable, realized, current) are respectively 0.655, 0.616, 0.63 and the standard deviations are 0.01, 0.014 and 0.017. This shows a low variance of the model performances, which is smaller than the difference between the R^2 of the models. The same conclusion comes also from the hit-rate of the three models (i.e. percentage of times the most probable path is the one chosen by the user), which are respectively 68, 62 and 64%.

Looking at the estimated coefficients, all of them are significant. To better discuss the ratio of the parameters, we report the coefficients scaled, so that the tram travel time coefficient is equal to -1. This scaling does not affect the model or the interpretation of the results. The preferred transport mode in city is the tram, followed by the bus and the train[Thesis Appendix 5]. This is consistent with the literature, and in particular with studies in Zürich (Montini et al. (2017) estimated a preference for trams, and Meyer de Freitas et al. (2019) identified the same order of preference in one of their model). The walking time has higher cost than in-vehicle travel times, consistently with most of the literature (Anderson et al., 2014; Meyer de Freitas et al., 2019). The transfer penalty is estimated as around 15 min of travel time in tram. This is very close to the range estimated by Garcia-Martinez et al. (2018), between 15.2 and 17.7 min of in-vehicle travel time, in a multimodal urban network (Madrid). As shown by Douglas and Jones (2013), the range of transfer penalty values estimated in literature is broad, and the identified value is consistent with this range.

The estimated coefficients of the realized and current CSs are similar, showing that the disturbances in the starting area of the trip are the most relevant in the choices of the users. The main differences with the timetable CS are in the coefficients related to transfers, which are weighted more.

We highlight the results shown are dependent on our test case and can be different considering another city with a different transport system. In fact, we suppose the network reliability and the presence of disturbances or disruptions in the network can affect users' choices and their reliance on the timetable. In addition, a more realistic information provision would result in a gradual increase of information through the trip. Further investigations in these directions are left for future work.

We remark that the Null log-likelihood is different among the three models because the alternatives can be different among the models (and therefore, also the observations covered and used in the choice model). For instance, a disruption in the network could make a possible alternative available for choice in the timetable CS but not in the realized CS (as delays make the connection infeasible). Similarly, differences in the coverage among the CSs hint some minor variations in the amount of observations considered. This does not affect the comparison, since we aim to identify which data (i.e. information provision) performs better for model estimation, which we evaluated with both monte-carlo cross validation and comparing the hit-rates. Finally, we highlight that we considered for each model only trips covered in the respective CS, since the non-covered trips can include unrealistic paths, as deviations for intermediate stops, or long waiting for external reasons. This results in minor changes between the number of observations of the CSs within 1%; in other terms, including also these trips does not change significantly the results.

3.6.5 *Comparison vehicle and line based CS*

The coverage of a line based CS is obviously higher than of a vehicle based one, since same vehicle implies same line, but not vice versa. The difference in coverage is 9.8% for the timetable CS (see Figure 3.2). A trip in this 9.8% has a corresponding trip in the CS with the same lines but different vehicles. In addition, the same coverage of the vehicle based CS, considering a size of 100 paths (84.7%), is obtained with only 4 alternatives in the line based CS.

We can find and quantify some sources for this difference. First, in case of delays, a user can take the same line as in the planned undelayed timetable, but probably different vehicles (i.e. they might get a delayed vehicle leaving within few minutes of another planned run of the same line). This applies to timetable and current CSs, which do not take into account the real network condition.

Moreover, walking speed has been assumed constant in our study, despite large evidence of consistent variations across users. Specifically, this has large impact at transfer points. In fact, it is difficult to establish if a user is able to transfer from a vehicle to another, especially when the time between the arrival of the feeder and the departure of the connected vehicle is very close to the assumed walking time. This is rather relevant as in our test case most public transport lines have a high frequency of operations.

Regarding estimation performances of a route choice model, we compared the estimation of a Path Size Logit model considering vehicle and line based CSs. Our hypothesis is that if the users reason in terms of vehicles, the line based model has lower performance, since it considers as a chosen

Parameter	Line Estimate	t-test	Vehicle Estimate	t-test
In vehicle travel time (s)				
Tram	-1	-21.7	-1	-19.7
Bus	-1.14	-20.2	-1.13	-18.2
Train	-1.19	-12.4	-1.16	-11
Walking time	-2.56	-42.6	-2.53	-39.8
Transfer time	-1.06	-15.6	-1.02	-13.7
No. of transfers	-889	-27.4	-943	-25
Path Size	55.4	2.49	20.5	0.82*
Observations	2719		2441	
Null log-likelihood	-7361		-6434	
Final log-likelihood	-2555		-2184	
$\bar{R}^2$	0.65		0.66	
Cross-val. $\bar{R}^2$	0.655		0.659	
St. dev.	0.01		0.018	
Scaling factor	0.0038		0.0039	

TABLE 3.6: Path Size Logit estimates for line and vehicle based timetable CSs with maximum size 40. * indicates a non-significant parameter ($|t| < 1.96$). The parameters are scaled (multiplied by the scaling factor) to have the tram travel time coefficient equal to -1.

alternative also an alternative equal to the real choice in the lines taken, but not in the precise vehicles taken. Instead, if the users reason in terms of lines, the two models should be comparable (since same vehicle implies same line). The results are shown in Table 3.6 and do not show remarkable differences. In other terms, there is not a significant improvement in considering vehicles, when studying users' choices. Despite vehicle based CSs are more realistic, line based CSs obtain higher coverage and similar model estimation. The results suggest that users reason at line level, therefore, in case of missed vehicle or connection, they generally wait the next vehicle (this is confirmed by different reasons in Section 3.6.7).

3.6.6 *Assessment of walking distance*

One modelling assumption of the CS generation algorithm is the maximum walking distance. Typically, in literature, this value is assumed as a constant, and its implications are seldom analysed. On one side, small values do not allow to capture certain paths (i.e. users might walk relatively long under certain conditions); on the other, large values assume higher willingness to walk (which also might not be a correct assumption, at times).
In this work, assuming for example that a public transport line ends within 500 meters of the destination, but no direct bus exists to reach the destination, and assuming a maximum walking distance of 700 meters, the algorithm will consider that users prefer to walk from the stop to the destination, rather than transferring to another public transport line, which will need to include penalties for transfer, and for waiting time.
This effect leads to a partial coverage, often analysed in literature as overlap percentage (see Anderson et al., 2014; Ding-Mastera et al., 2019; Prato and Bekhor, 2006). In other terms, a minor part of the trip, typically when leaving the public transport system, is not explicitly identified.
In our experiments, considering the timetable CS, 14.6% of the trips are partially covered, i.e. 94.5 - 14.6 = 79.9% have a perfect coverage. This latter concept refers to a path matching completely from the access to the public transport system to the egress. A way to overcome this problem is to enlarge the CS, including dominated paths that could be (in theory) considered irrelevant. To study this effect of partial coverage, we considered CSs including paths based on short walking distance, and paths based on long one. This will generate some partially dominated paths in the CS, as a short walking path will not consider long walking stages; and a long walking path will ignore using public transport services for very short distances.
In particular, for each trip we considered a CS formed by the union of two CSs, based on a walking distance of 300 and 700 meters, respectively. Figure 3.4 considers those larger CS including dominated paths related to short and long walking distance. The ratio of perfect coverage goes up to 85.5%, i.e. almost two fifths of the partially covered trips are perfectly covered ((85.5 - 79.9)/14.6). This comes at the cost of a slightly worse precision, i.e. the coverage grows slower in Figure 3.4 compared to Figure 3.2, proving that less relevant, actually dominated, paths are added. This analysis shows that in 14.6% of trips, people preferred to walk less than expected

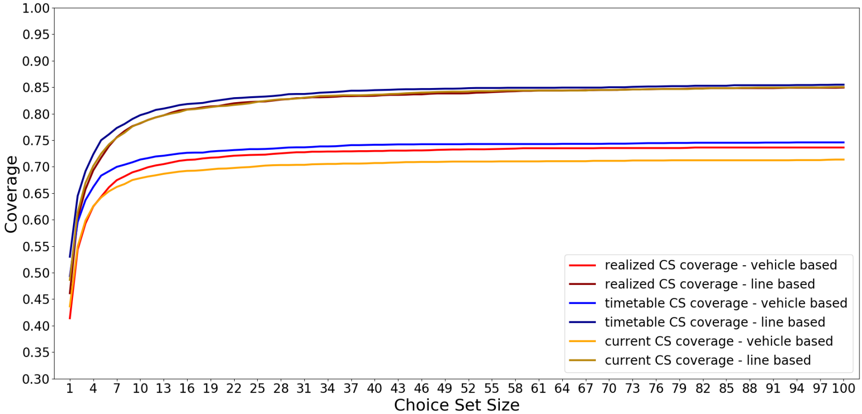


Figure 3.4: Perfect CS coverage considering two different walking distances in each CS (300 and 700 meters). Each CS is given by the union of two CSs based on the two walking distances.

(max 700 m) and took an additional vehicle instead. Two fifths of those 14.6% of trips are identified, considering walks long less than 300 m. For the remaining three fifths (9% of the trips), the additional vehicle replaced a walk which would have been less than just 300 m (maybe also related to crowded buses, or just mistakes).
In addition, denoting the mean walking distance of a user as *mwd* and its standard deviation as *swd*, we analysed the distribution of *mwd* and *swd* among the users. *mwd* has a mean of 240 meters and st. dev. of 118, while *swd* has a mean of 225 and st. dev. of 170. These values show a large inter- and intra-user variation of the walking distance, showing the considered walking distance is dependent by external factors, as probably the location or the weather.

3.6.7 *Trips with transfers*

We distinguish here the case of trips with transfers, to identify if there are differences in the coverage with different information provisions. Figure 3.5 shows the coverage only for trips with transfers. The coverage for line based CSs is comparable between the timetable and realized CSs. Regarding the vehicle based CSs, the realized performed better. In fact, the realized CS considers the real available alternatives and can better capture

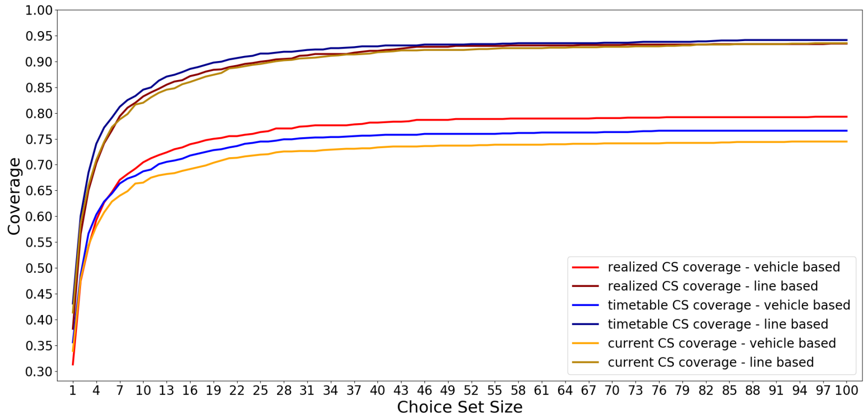


FIGURE 3.5: CS coverage for trips with transfers.

the available transfers. Nevertheless, the similar coverage for the line based CSs shows that, even if the timetable CS is not able to identify the correct vehicles, it can identify the correct lines. Finally, the current CS has less coverage because it considers disturbances only for the first vehicle, therefore considering unrealistic transfers.

3.6.8 *Users' heterogeneity*

In this Section, we analyse the heterogeneity of the results of the proposed analysis among the different users. In particular, we analyse the heterogeneity regarding the information provision and the heterogeneity on the parameters of the route choice model, in terms of how the different travel time components are valued. Further considerations on the heterogeneity of considered walking distance are described in Section 3.6.6.

3.6.8.1 *Heterogeneity regarding information provision*

Figure 3.6 shows the heterogeneity of the coverage of different CSs among the users. Each dot represents a user, located in the axis of the CS with highest coverage. For 35% of the users (the overlapping dots in the center), there is no difference in which information is provided, since the respective CSs are equally able to cover their trips. This probably shows that these users did not experience any significant disturbance. For the remaining users, the highest coverage is obtained with different information pro-

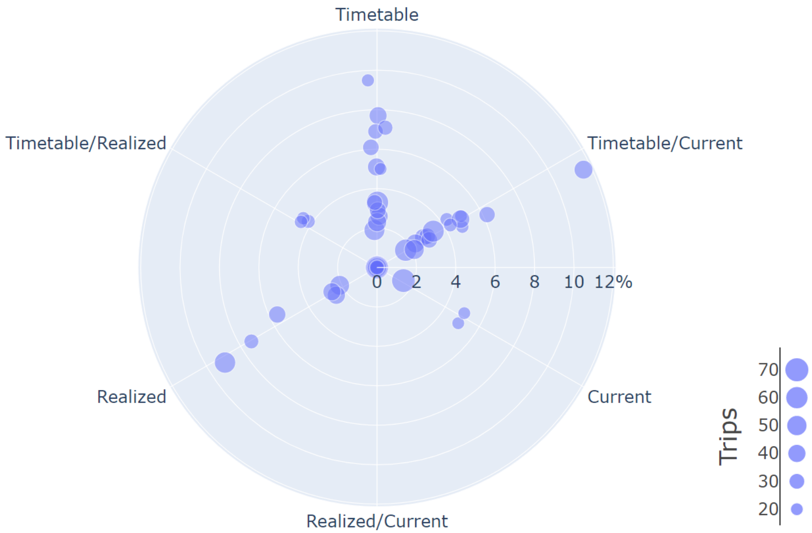


FIGURE 3.6: CS with the highest coverage, for each user. Each dot represents a user, and the axis indicates the CS with highest coverage (in case of Timetable/Realized, and Timetable/Current, the two CSs perform equally well). The distance from the center is the difference in coverage between the best CS (i.e. the axis) and the worst CS. The size of the dot is proportional to the number of trips (users with less than 20 trips are not shown). For example, the rightmost dot is a user with the timetable and current CS having equal coverage and 11% higher than the one of the realized CS. The dots at the origin of the axes are users for whom the three CSs perform equally well.

visions (22% by the timetable CS, 10% realized, 5% current, 28% timetable and another), even if the timetable CS performs better for more users. In other terms, generally, the timetable CS does not perform always best, and different users rely on different information provisions.

3.6.8.2 *Heterogeneity regarding route choice*

As described in Section 3.5.4, we estimated a Mixed Path Size Logit model to investigate possible panel effects. The computation time of this model is much larger than the one of the Path Size Logit (respectively 90 min, compared to 14 s on the same machine, an increase of more than 385-fold). Table 3.7 shows the estimated coefficients for the timetable CS (similar con-

Parameter	μ	t-test	std	t-test
In vehicle travel time (s)				
Tram	-1	-14.9	0.19	2.51
Bus	-1.21	-13.8	0.33	6.48
Train	-1.58	-7.57	0.96	8.05
Walking time	-2.64	-26.4	0.64	8.17
Transfer time	-1.02	-12.4	0.12*	1.44
No. of transfers	-899	-21.4	18.5*	0.56
Path Size	62.1	2.21	38.7*	1.54
Observations	2719			
Users	152			
Draws	500			
Null log-likelihood	-7361			
Final log-likelihood	-2420			
$\bar{R}^2$	0.67			
Scaling factor	0.004			

TABLE 3.7: Mixed Path Size Logit model estimates for line based timetable CS with maximum size 40. Parameters distributed according to a normal distribution. * indicates a non-significant parameter ($|t| < 1.96$). The parameters are scaled (multiplied by the scaling factor) to have the tram travel time coefficient equal to -1.

sideration can be made for the other CSs). The R^2 value is slightly higher than the one of the Path Size Logit (Table 3.5), proving a slight increase of performances. The mean value of all coefficients is significant, and in line with the value estimated by the Path Size Logit. Three of the coefficients have a moderate variation among the users (st. dev. between 19 and 27% of the mean value): the travel time in bus, in tram and the walking time. Remarkable is the variation of the in-train travel time among the users (st. dev. of 61%), showing that considering the train as a possible transport mode depends strongly on the user. Finally, the standard deviation of the transfer time and of the transfer penalty are not significant, showing that there is no reason to treat these parameters as random and that transfers are generally perceived homogeneously among the users. There-

Cluster	Users	Tram	Bus	Train	Walk	Transfer time	Transfers	Path Size
1	62	-1.10	-1.21	-1.55	-2.87	-1.07	-946	93.4
2	24	-0.98	-1.34	-1.82	-2.67	-1.08	-994	93.5
3	19	-1.02	-1.35	-1.45	-2.95	-1.06	-884	93.7
4	13	-1.10	-1.39	-1.15	-2.39	-1.06	-995	93.4

TABLE 3.8: Clustering of the users, based on the estimated coefficient of each cost component, using the k-means algorithm. The average values for each cluster are shown.

fore, from the moderate improvements of the R^2 value, and from the low standard deviation for most of the coefficients, we conclude that there is not a remarkable heterogeneity among the users on perceived costs, except regarding in-train travel time.

To understand if there are groups of users with similar perception of the costs, we estimated the most likely values of the parameters for each user (see Section 3.5.4) and we applied clustering on the users. We applied the K-means algorithm considering 4 clusters, which corresponds to the highest silhouette value (0.2). We considered 118 users, discarding the ones with an outlier in one of the parameters (below the 2nd percentile or above the 98th). Table 3.8 shows the results of the clustering. Among the clusters, two have more distinct characteristics: cluster 2 refers to people with a positive preference for trams and negative for trains and transfers; while cluster 4 refers to people with a preference for trains and for walking. Despite this, the low silhouette value indicates that within each cluster, there is not perfect homogeneity among the users, and that the global heterogeneity between users (which is large only concerning the train coefficient, see previous analysis) does not easily lead to groups with similar behaviour.

3.7 DISCUSSION AND POLICY IMPLICATIONS

In this Section, we discuss the results and related possible policy implications for transport planners and service providers.

The proposed work is based on a large-scale tracking survey, based on automatic data collection and mode detection. The high level of coverage of the generated CSs and significance of route choice model estimation show the applicability of this dataset in real context. Therefore, this work proves the feasibility for service providers to exploit location data provided by smartphone applications, to automatically collect large-scale surveys without burden from the users.

The novel CS generation algorithm obtained high precision in terms of coverage and minimal required size, high quality of model estimation, and high efficiency, in terms of computation time. This encourages the applicability of the algorithm both for model estimation and route recommender systems in practice. Several specific aspects regarding CSs, scarcely investigated in literature, also shed light on the choice process of users and their practical implications. First, the relationship between the CS size and the relevance of a path suggests that including more than 10 alternatives does not change substantially the coverage and the route choice model estimation significantly. This is relevant for transport planners, both for modelling purposes, as considering smaller CSs to speed-up the following analysis, and for policy making, as to show a sufficient number of alternatives on a route recommender system.

Also relevant for route recommender systems is the finding that, generally, users consider alternatives as sequences of lines, and not vehicles. In particular, in case of missed vehicle or connection, generally users have a low en-route replanning and they rather seem to wait for the next vehicle. This can be explained by a high service reliability; a lack of information for the passengers; or that the passengers tend to stick to their original plan. This is confirmed by the analysis of trips including transfers, showing the similar performance of all analysed CSs, when coverage in terms of line is considered.

In this work we analyse which information provision results in a CS, which represents best the users' behaviour. This has important policy implications, since it allows service providers to identify if their users rely on the provided information to make their choices. As a consequence, providers can adapt their information systems, for instance to improve route recommendation. In this sense, we assume that the information provision resulting in highest coverage and better model estimation, is the one better representing the general available information. The highest coverage and best fit is obtained with no information, showing that in general users rely on the timetable, rather than real conditions. This can be a signal suggesting service providers to better inform their users. In this regard, delivery time and correctness of the information might be improved; and/or the information system in case of delays or disruptions might be enhanced.

Moreover, information provision might be disregarded by users because its inclusion on the choice process is perceived as an effort (Jiang et al., 2019), and/or information is not deemed suitable enough, or sufficiently targeted, to the specific needs of the travellers.

On the modelling side, these results show that, for operations with these planned and operational characteristics, it is not necessary to use realized operation data to build a CS. In addition, they suggest to transport planners and policy makers what is the current value of reliability in practice. In fact, a highly reliable service can make users rely more on the usual choices, and having less need, or consideration, for updated information.

The estimation of the Path Size Logit assuming perfect and current information provided similar coefficients, showing that the disturbances in the starting area of the trip are the most relevant in the choices of the users. This suggests to policy makers to better exploit the possibility to counteract/notify disturbances at the begin of the users' trips. In addition, assuming the users have knowledge about disturbances, transfers are weighted more (compared to assuming no information), meaning that direct paths are preferred. This can be explained by a lower trust in the system in presence of disturbances, or by a greater preference for more comfortable paths without transfers.

We observed the users' heterogeneity, both in terms of perceived costs and information provision. A low heterogeneity is found in the perceived costs of the different travel time components, including transfers, but not for the in-train travel time. It follows that, even if the preferred mode is user-dependent, the perceived discomfort of a transfer is more uniform. Therefore, transport planners can focus on improving the quality of transfers, to decrease the general discomfort. Regarding the information provision, we compared different CSs for each user. Based on that, we identified high heterogeneity: different users rely on different information, despite the majority rely on the timetable. This result allows identifying which users are more (or less) informed and can be a basis for a personalized recommender system of a service provider. In fact, this suggests that certain specific users might appreciate (or even need) more information, which can be faced with personalized recommendations. Finally, users seem to have heterogeneity only in few aspects, namely the in-train travel time, and the considered walking distance. The latter variation is both inter- and intra-user, and probably it depends on trip purpose. In other terms, it might be worthwhile to consider alternatives based on both short and long walks. In this sense, route recommender systems should provide alternatives with different walking distances, or profile the users to understand the most likely situation.

3.8 CONCLUSIONS

This work presents an algorithm for efficient and precise computation of choice sets in public transport networks. As such, it represents a first attempt to study the users' choices in public transport with realized data of both passengers and operations. We exploit an automatic mode detection algorithm to collect travel data from a large amount of users for a long time, without requiring an active participation. The knowledge of what happened in reality allowed analysing precisely the users' choices in terms of public transport vehicles used. Our proposed computationally efficient CS generation algorithm is based on constrained enumeration, and performs well according to a series of performance factors, as follows. The running time of the proposed algorithm is a few seconds on a standard computer. The algorithm obtained a coverage above 94% within the first 40 alternatives. The estimation of a Path Size Logit model resulted in a high adjusted R^2 (0.65) and reasonable parameters' values, validating the applicability of the generated CSs. In this respect, the high value of the adjusted R^2 can be explained by different factors: the nature and the size of the dataset; the vehicle-based modelling of each alternative; the high coverage of the CSs; users' related factors.
We tried to fill the gap in literature, concerning the evaluation of the right CS size and the relevance of a path. Therefore, we identified a size of 40 as a reasonable maximum size to cover most of the observations and to understand the users' behaviour. In addition, we can consider paths defined as sequences of vehicles (or lines), with a different resolution compared to the commonly used sequences of stops. The latter, in fact, do not distinguish among vehicles passing the same stops. The comparison between line based CSs and vehicle based CSs hints at whether users reason in terms of line or vehicles. The results suggest generally users base their decisions on lines, and in case of a missed connection, they rather wait the next vehicle, than follow a different route.
Different information provisions of the network conditions were assumed for the users and the Path Size Logit model was estimated for each of them. We observed that, for most of the users, considering no information on the network conditions describes better their choices than assuming perfect information or current information, in terms of both coverage and model

estimation. This can be explained either assuming the passengers have no information of the network conditions, or that they stick to an original plan, based on the timetable. For trips with transfers, we observed the CS based on no information captures well the lines taken, but not the vehicles taken, where network information, such as delays, are needed.
We analysed the heterogeneity of users according to several aspects. Regarding perceived costs of the different travel time components, transfers are perceived homogeneously among the users, while there is a high variability on the perceived cost of in-train travel time. We identified that, despite the majority of the users rely on the timetable, different users rely on different information provisions. Finally, we observed a large inter- and intra-user variation of the considered walking distance.
For future work, we plan to integrate the CS generation algorithm with different information provisions into simulation tools (Leng and Corman, 2020) or optimization models (Corman et al., 2017), considering passengers' route choice in public transport. The potential of Recursive Logit approaches has been reported to come at high computational costs. Future research should investigate how improved graph modelling and estimation methods would allow for the identification of the respective advantages of the two approaches, as suggested in Zimmermann and Frejinger (2020), and their complementarity in determining different implications for public transport design and information provision.
Some assumptions related to the panel effects are worthwhile being investigated in future research: for instance, the possibility to exploit more in depth the longitudinal behaviour of each user, beyond the small panel effects found, is of potentially great interest, to determine the impact of trip purpose or external influences such as weather.
We also plan to explore other information provisions or adapted information during a trip. Additional information on the amount of alternatives considered by each user can improve the analysis of the CS size. In addition, we plan to compare the users' choices with the available alternatives in case of disruptions in the network, to identify how the users react to disturbances and in which cases a user sticks to a plan, or changes it.

ACKNOWLEDGEMENTS

The authors wish to thank all study participants and the Swiss Federal Railways (SBB) for the open data.
Declarations of interest: none.

REFERENCES

Anderson, M. K., O. A. Nielsen, and C. G. Prato (2014). Multimodal route choice models of public transport passengers in the Greater Copenhagen Area. *EURO Journal on Transportation and Logistics* 6.3, 221.

Beck, M. J., S. Hess, M. O. Cabral, and I. Dubernet (2017). Valuing travel time savings: A case of short-term or long term choices? *Transportation Research Part E: Logistics and Transportation Review* 100, 133.

Ben-Akiva, M. E., M. J. Bergman, A. Daly, and R. Ramaswamy (1984). Modelling inter urban route choice behaviour. *Ninth International Symposium on Transportation and Traffic Theory*.

Ben-Akiva, M. and M. Bierlaire (1999). Discrete Choice Methods and their Applications to Short Term Travel Decisions. *Handbook of Transportation Science*. Ed. by R. W. Hall. International Series in Operations Research & Management Science, 5.

Bierlaire, M. (2018). PandasBiogeme: a short introduction. Technical report TRANSP-OR 181219. *Transport and Mobility Laboratory, ENAC, EPFL.*

Bovy, P. H. L. (2009). On Modelling Route Choice Sets in Transportation Networks: A Synthesis. *Transport Reviews* 29.1, 43.

Bovy, P. H. L., S. Bekhor, and C. G. Prato (2008). The Factor of Revisited Path Size: Alternative Derivation. *Transportation Research Record* 2076.1, 132.

Bovy, P. H. L. and S. Hoogendoorn-Lanser (2005). Modelling route choice behaviour in multi-modal transport networks. *Transportation* 32.4, 341.

Cats, O. (2011). Dynamic Modelling of Transit Operations and Passenger Decisions. PhD thesis. KTH Royal Institute of Technology.

Corman, F., A. D'Ariano, A. D. Marra, D. Pacciarelli, and M. Samà (2017). Integrating train scheduling and delay management in real-time railway traffic control. *Transportation Research Part E: Logistics and Transportation Review* 105, 213.

De la Barra, T., B. Perez, and J. Anez (1993). Multidimensional path search and assignment. *PTRC Summer Annual Meeting, 21st, 1993, University of Manchester, United Kingdom*. Vol. P 366.

Ding-Mastera, J., S. Gao, E. Jenelius, M. Rahmani, and M. Ben-Akiva (2019). A latent-class adaptive routing choice model in stochastic time-dependent networks. *Transportation Research Part B: Methodological* 124, 1.

Douglas, N. J. and M. Jones (2013). Estimating transfer penalties and standardised income values of time by stated preference survey. *Australasian Transport Research Forum (ATRF), 36th, Brisbane, Queensland, Australia*.

Fiorenzo-Catalano, M. S. (2007). Choice set generation in multi-modal transportation networks. PhD thesis. TU Delft.

Fosgerau, M., E. Frejinger, and A. Karlstrom (2013). A link based network route choice model with unrestricted choice set. *Transportation Research Part B: Methodological* 56, 70.

Frejinger, E., M. Bierlaire, and M. Ben-Akiva (2009). Sampling of alternatives for route choice modeling. *Transportation Research Part B: Methodological* 43.10, 984.

Friedrich, M., I. Hofsaess, and S. Wekeck (2001). Timetable-Based Transit Assignment Using Branch and Bound Techniques. *Transportation Research Record* 1752, 100.

Garcia-Martinez, A., R. Cascajo, S. R. Jara-Diaz, S. Chowdhury, and A. Monzon (2018). Transfer penalties in multimodal public transport networks. *Transportation Research Part A: Policy and Practice* 114, 52.

Gentile, G. and K. Noekel, eds. (2016). Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems: COST Action TU1004 (TransITS). Springer Tracts on Transportation and Traffic. Springer International Publishing.

Hess, S., M. Bierlaire, and J. W. Polak (2005). Estimation of value of travel-time savings using mixed logit models. *Transportation Research Part A: Policy and Practice* 39.2, 221.

Hoogendoorn-Lanser, S., P. Bovy, and R. van Nes (2007). Application of Constrained Enumeration Approach to Multimodal Choice Set Generation. *Transportation Research Record* 2014.1, 50.

Jiang, G., M. Fosgerau, and H. K. Lo (2019). Route choice, travel time variability, and rational inattention. *Transportation Research Part B: Methodological*.

Leng, N. and F. Corman (2020). The role of information availability to passengers in public transport disruptions: An agent-based simulation approach. *Transportation Research Part A: Policy and Practice* 133, 214.

Marra, A. D., H. Becker, K. W. Axhausen, and F. Corman (2019). Developing a passive GPS tracking system to study long-term travel behavior. *Transportation Research Part C: Emerging Technologies* 104, 348.

Meyer de Freitas, L., H. Becker, M. Zimmermann, and K. W. Axhausen (2019). Modelling intermodal travel in Switzerland: A recursive logit approach. *Transportation Research Part A: Policy and Practice* 119, 200.

Molloy, J., B. Schmid, F. Becker, and K. W. Axhausen (2019). mixl: An open-source R package for estimating complex choice models on large datasets. Working Paper. IVT, ETH Zurich.

Montini, L., C. Antoniou, and K. W. Axhausen (2017). Route and mode choice models using GPS data. *TRB Annual Meeting Online*. Vol. 1204, 17.

Nassir, N., M. Hickman, A. Malekzadeh, and E. Irannezhad (2015). Modeling Transit Passenger Choices of Access Stop. *Transportation Research Record: Journal of the Transportation Research Board* 2493, 70.

Prato, C. G. (2009). Route choice modeling: past, present and future research directions. *Journal of Choice Modelling* 2.1, 65.

Prato, C. G. and S. Bekhor (2006). Applying Branch-and-Bound Technique to Route Choice Set Generation. *Transportation Research Record* 1985.1, 19.

Prato, C. G., T. K. Rasmussen, and O. A. Nielsen (2014). Estimating Value of Congestion and of Reliability from Observation of Route Choice Behavior of Car Drivers. *Transportation Research Record* 2412.1, 20.

Rasmussen, T. K., M. K. Anderson, O. A. Nielsen, and C. G. Prato (2016). Timetable-based simulation method for choice set generation in large-scale public transport networks. *European Journal of Transport and Infrastructure Research* 16.3, 467.

Rieser-Schüssler, N., M. Balmer, and K. W. Axhausen (2013). Route choice sets for very high-resolution data. *Transportmetrica A: Transport Science* 9.9, 825.

Rieser-Schüssler, N., M. Rieser, L. Montini, and K. W. Axhausen (2014). Exploring choice set generation approaches for public transport connection choice. *Swiss Transport Research Conference (STRC)*.

Schmid, B., S. Jokubauskaite, F. Aschauer, S. Peer, R. Hössinger, R. Gerike, S. R. Jara-Diaz, and K. W. Axhausen (2019). A pooled RP/SP mode, route and destination choice model to investigate mode and user-type effects in the value of travel time savings. *Transportation Research Part A: Policy and Practice* 124, 262.

Sillano, M. and J. de Dios Ortúzar (2005). Willingness-to-Pay Estimation with Mixed Logit Models: Some New Evidence. *Environment and Planning A: Economy and Space* 37.3, 525.

Stadt Zürich (2020). Mikrozensus Mobilität und Verkehr 2015 - `www.stadt-zuerich.ch/ted/de/index/taz/verkehr/webartikel/webartikel_kennzahlen_verkehrsentwicklung.html`.

Swiss Federal Office of Transport (2019). `https://opentransportdata.swiss/en/dataset/istdaten`.

Tan, M.-C., C. O. Tong, S. C. Wong, and J.-M. Xu (2007). An algorithm for finding reasonable paths in transit networks. *Journal of Advanced Transportation* 41.3, 285.

Tan, R., M. Adnan, D.-H. Lee, and M. E. Ben-Akiva (2015). New Path Size Formulation in Path Size Logit for Route Choice Modeling in Public Transport Networks. *Transportation Research Record* 2538.1, 11.

Ton, D., D. Duives, O. Cats, and S. Hoogendoorn (2018). Evaluating a data-driven approach for choice set identification using GPS bicycle route choice data from Amsterdam. *Travel Behaviour and Society* 13, 105.

Yen, J. Y. (1971). Finding the K Shortest Loopless Paths in a Network. *Management Science* 17.11, 712.

Zimmermann, M. and E. Frejinger (2020). A tutorial on recursive models for analyzing and predicting path choice behavior. *EURO Journal on Transportation and Logistics* 9.2, 100004.

4

IMPACT OF PUBLIC TRANSPORT DISTURBANCES

This chapter is based on the following article:

From Delay to Disruption: Impact of Service Degradation on Public Transport Networks

Alessio D. Marra[1] & Francesco Corman[1]
[1]Institute for Transport Planning and Systems, ETH Zürich

Published in

Transportation Research Record: Journal of the Transportation Research Board, 08.2020
DOI: https://doi.org/10.1177/0361198120940989

Contributions

A. D. Marra: Study conception and design; Data collection; Analysis and interpretation of results; Draft manuscript preparation
F. Corman: Study conception and design; Analysis and interpretation of results; Draft manuscript preparation

Key findings

- A disruption impact metric is defined to compare and analyse different public transport disruptions
- The impact of real disruptions occurred in Switzerland is evaluated on simulated trips
- The relationships between the disruption characteristics and the impact are analyzed to identify the main features affecting a disruption

Additional notes to this chapter

In this thesis, we do not make a rigorous distinction between the terms "disturbance" and "disruption", and we mainly use the first. Instead, in this chapter, we use the term "disruption", since it is used in the published article.
Section 4.4.2.1 summarizes a previous version of the choice set generation algorithm proposed in Chapter 3, using slightly different values for the parameters.
Some aspects of this chapter are clarified in the appendix of the thesis (see superscript [Thesis Appendix]).

4.1 ABSTRACT

Public transport networks (PTN) are affected daily by different types of disturbances. In fact, between a single delay and a long service interruption, there is a range of disruptions with different impacts, depending on their characteristics. Despite this, in literature, the common definition of disruption is a link closure for a certain amount of time. Low interest is given to different types of disruptions or to the connection between delays and disruptions. In addition, in multimodal PTN a physical link closure is not always observable, but rather people experience delays or cancelled stops on different lines. The aim of this work is to explore the relationship between delays and disruptions, analyzing different degrees of disruptions, in relation to duration, delay, size, and network characteristics. Real disturbances of the PTNs in Zürich and Bern, Switzerland, are analyzed to identify disruptions with different characteristics. Therefore, the disruption impact is computed on simulated origin–destinations (ODs), based on the sets of possible paths with and without the disruption. For this purpose, a choice set generation algorithm is used. Finally, relationships between the disruption characteristics and the impact are analyzed to identify the main features of a disruption.

4.2 INTRODUCTION

Public transport networks (PTN) are characterized by daily unexpected delays or cancelled trips. The impact of each of them depends on multiple factors, such as network characteristics and the entity of the disturbance. For instance, a cancelled trip of a bus travelling in a city center has a different impact than a bus travelling in a rural area. In addition, combinations of delays and failed trips can affect particular areas of the PTN more than single disturbances. Typically, a major exceptional event, in relation to duration and effects, is defined as a disruption. Nevertheless, there is not a clear distinction between small and big disruptions, but rather there is a continuous range of disruptions with different impact.

To discuss such a continuous spectrum of disruption it is proposed in this paper to relate a measurable quantity, that is, a disruption impact, with characteristics (features) of the disruption which can determine and explain most of the impact. In that way, knowing the characteristics of a disruption, public transport providers can proactively update operations to better deal with such events. An impact quantification would allow possible actions as: identification of critical locations and times and where to increase operational margins in running time or reserve vehicles; specific attention to some passenger groups which happen to be more vulnerable; and optimal response whenever a disruption has been just identified, including targeted alerts for passengers.

To identify different disruptions, real disturbances of the Zürich PTN and Bern PTN are extracted from one year of automatic vehicle location (AVL) data reporting planned and realized arrival and departure time at stops, and grouped through a clustering algorithm (1). Therefore, the impact of each disruption is estimated on different origin–destinations (ODs), considering a range of possible paths for the passengers. A disruption impact measure is defined, based on a discrete choice model and comparing the choice set of available alternatives in case of disruption with the one in case of no disturbance. Finally, the relationships between the disruption characteristics and the impact are analyzed through machine learning and feature importance metrics.

4.3 STATE OF THE ART

In previous works, the definition of disruption in PTN is generally simple and network based. Typically, a disruption is described as a node or a link failed for a certain amount of time, without traffic admitted through it (Cats and Jenelius, 2014; Rodríguez-Núñez and García-Palomares, 2014). This definition can be consistent with a railway/metro network and with long disruptions. Instead, for multimodal networks including buses, a disruption can be better defined from the operational perspective, taking into account delays or missed trips. In this regard, Sun and Guan (2016) analyze vulnerability from line operation perspective, but they consider only a metro network and a disruption as formed by cancelled trips.

In the literature of transport disruptions and vulnerability studies, few works examine public transport networks compared with road networks (Mattsson and Jenelius, 2015). In addition, most of them are focused only on metro or railways, instead of considering a multimodal PTN (Lu, 2018; Rodríguez-Núñez and García-Palomares, 2014; Van der Hurk, 2015). In that area, they analyzed the users' behavior in a multimodal network, but they considered only railway disruptions (Leng and Corman, 2020).

Most of the previous works are focused on identifying critical links or stations and less attention is given to analyzing the impact of disruptions with different characteristics. Burgholzer et al. (2013) described a disruption by its duration (2 h the smallest), its time of occurrence and the capacity reduction. Cats and Jenelius (2018) analyzed the relation between the extent of capacity reduction and its consequences on PTN performance, but they did not examine other disruption characteristics. One of the few works considering different characteristics of a disruption is Jenelius (2009), even if on a road network (e.g., road density, user travel time, traffic flow). He investigates the dependence of the effects of link closure on several indicators using a regression model.

Focusing on the methodology, Mattsson and Jenelius (2015) identified two distinct traditions in disruption analysis: topological vulnerability analysis, based on the topological properties of the transport network; and system-based vulnerability analysis, which represents also the demand of the transport systems. In the first group, Angeloudis and Fisk (2006) study the degree distribution of different subway networks and they simulate attacks on the stations to analyze the network robustness. In the second group, the interaction between demand and supply is simulated by means of transport system models. Typically, the passengers' behavior is modeled as the shortest travel time or using discrete choice models (Cats and Jenelius, 2014; Lu, 2018; Rodríguez-Núñez and García-Palomares, 2014; Van der Hurk, 2015). Therefore, the impact of a disruption is primarily measured based on the whole traffic in the network (Burgholzer et al., 2013; Cats and Jenelius, 2014; Cats and Jenelius, 2018). Computing the impact on passengers based on realized automated fare collection (AFC) and realized disruption (i.e., a complementary approach based on observed passenger choices) has been proposed concurrently to the present work, for larger and planned disruptions (Yap, 2020). To the best of the authors' knowledge, the impact is never analyzed on single ODs or considering the

entire choice set of a user.
A key missing aspect in literature is the analysis of short disruptions (in the order of minutes), although they are the most frequent disruptions people experience in daily trips. In addition, the relationship between disruptions and operational delays is seldom analyzed. Instead, it is reasonable to think that they are linked phenomena and there is not a strict boundary between them.

4.4 METHODS

The methods used to understand the impact of different types of disruptions on a PTN are schematically represented in Figure 4.1, and they can be divided into three parts. First, the concept of disruption is defined and several disruptions are identified from the AVL data; second, the impact of the disruptions is computed, based on a choice set generation algorithm and a behavioral model, applied to simulated ODs; third, disruptions' features are extracted and the relationships between them and the disruption impact are analyzed, through machine learning and features importance metrics.

4.4.1 *Disruption Identification*

The definition of a disruption as a (physical) link closure is not realistic in the case of a multimodal PTN. In fact, the network traffic is characterized by delays or missed stops (i.e., a bus that did not stop at a stop), that cannot all be described by a physical link closure (intended as no traffic allowed between two or several stops at a certain moment). Therefore, a new definition of disruption is necessary, able to include delays and failed trips, and to represent both short and long-term disruptions. An event is defined as an arrival of a public transport vehicle at a stop (considering departures does not change the analysis significantly). Therefore, a disruption is defined as a set of delayed or failed events near to each other in time and space. This definition is not strict, but it allows both to connect delays to disruptions and to determine many characteristics for disruptions, of which impact can be analyzed afterwards. For ease of understanding, a disruption is referred to as the cooccurrence of several disturbances in a certain area in a certain time.
To identify real cases of disruptions, AVL data are used, seeking clusters of delayed or failed events. To find the clusters, the ST-DBSCAN algorithm

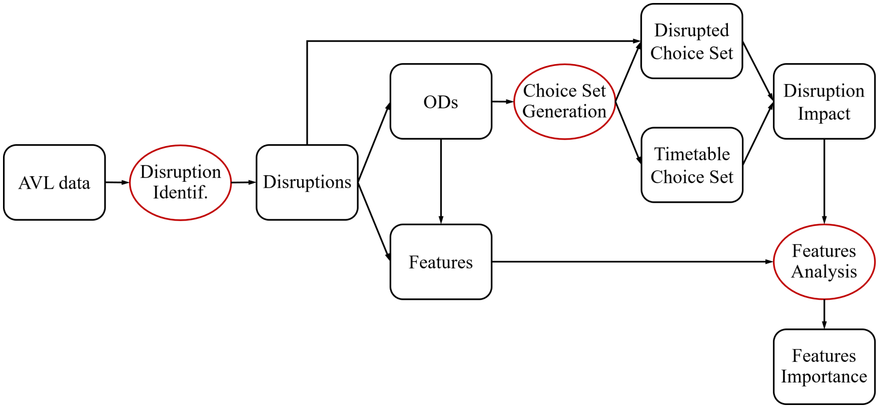


FIGURE 4.1: Graphical scheme of the proposed methods. AVL = automatic vehicle location; ODs = origin–destinations.

is used (Birant and Kut, 2007). This algorithm is a variant of the clustering algorithm DBSCAN, used to cluster spatio-temporal data. DBSCAN is a density-based algorithm that groups together points close to each other, based on a distance metric. In ST-DBSCAN both a spatial distance and a temporal distance are used, to form clusters of points (representing events) close in time and space. Given a set of delayed or failed events, ST-DBSCAN is able to detect groups of related events that satisfy the definition of disruption.
The algorithm takes in input the following parameters:

- *P*: set of points to cluster, with two spatial and one temporal coordinates (i.e., latitude and longitude of the stop, and time of the event)
- *epsSpace*: maximum spatial distance to consider two points as near
- *epsTime*: maximum temporal distance to consider two points as near
- *minPoints*: minimum number of points within *epsSpace* and *epsTime* to form a cluster

Therefore, the output is a label assigned to each point (event), representing its cluster. A point can also be marked as noise if it is not part of any cluster. In that way, an isolated event is not considered as part of any disruption. The parameter *epsSpace* regulates the spatial aggregation of the events in a disruption. Since each event is located in a stop of the network, *epsSpace*

determines if near stops can be part of the same disruption. *epsTime* behaves similarly in time.
Referring to one day of service, P is defined as the set of all the events with a delay $\geq$ *minDelay*. Failed events are considered as delayed events with a delay equal to the time difference with the next same event (same line at the same stop). The parameter *minDelay* acts as a threshold for too-small delays, since it is assumed they have a marginal impact compared to others. Furthermore, the $\Delta\varepsilon$ parameter of the ST-DBSCAN is not used ($\Delta\varepsilon = \infty$). For a detailed description of the algorithm, refer to Birant and Kut (2007).

4.4.2 *Disruption Impact Evaluation*

Since the aim is to consider also short disruptions (in the experiments *minDelay* is set = 6 min), it was decided to evaluate the impact of a disruption only on ODs directly affected by it, without considering capacity constraints or a full OD matrix of the network. Therefore, ODs were considered with Origin in the center of mass of the disruption, departure time at its beginning (the planned time of the first event of the disruption), and Destination chosen randomly among the stops of the network. The beginning of the disruption is chosen as departure time, since it is believed it is the time most affected by the disruption. In fact, an earlier departure allows considering alternatives starting before the disruption, while a later departure discards the first disrupted events.
For each OD, two choice sets are generated to model the possible paths with and without the disruption. Using the whole choice set, instead of a single optimal path, can better describe the disruption impact, since more possibilities for the user are taken into account.

4.4.2.1 *Choice sets generation*

Since the used choice set generation algorithm is not the focus of this paper, only the most important details of it are reported.
Two choice sets are generated: one based on the timetable (Timetable Choice Set), without considering disturbances in the network; the other considering as the only disturbances in the network the events of the disruption (Disrupted Choice Set). In this way, the impact of a disruption can be evaluated comparing the two sets of alternatives for an OD.
The PTN operations were modeled from the AVL data as a schedule-based model, using a time-expanded (diachronic) graph to represent each single trip of each vehicle. Nodes correspond to the arrival or departure of a pub-

lic transport vehicle at a stop, while arcs model the trips and the available transfers. The procedure to build the two choice sets is the same, since the difference is only in considering the disturbances to build the graph. In particular, a missed stop is modeled removing the related service nodes and arcs from the diachronic graph based on the timetable. A delay is modeled increasing the travel time (the weights of the arcs) and adjusting the possible transfers. The Origin and the Destination are connected to the network through walk arcs. The Origin is connected to all the nodes departing in near stops after the starting time of the trip. The Destination is connected to all the vehicles arriving at near stops.

To model the users' choices, a path-based assignment model was used, explicitly enumerating the relevant paths through a choice set generation algorithm. In particular, the algorithm is part of a well-known family of algorithms for choice set generation, called constrained enumeration (Bovy, 2009; Cats, 2011; Tan et al., 2007).

The proposed algorithm enumerates all the possible combinations of vehicles available for each OD and generates a feasible path for each of them. Given the amount of ODs that are analyzed in this work, and that two choice sets need to be generated for each OD, some constraints must be set to reduce the computation time. In particular, three maximum transfers are allowed; the maximum waiting time is 30 min; the maximum transfer distance is 350 m (used also to identify stops near the Origin and the Destination); and only the paths with a duration less than twice the one of the best path are considered. These constraints are also used in previous works and their values are assumed realistic for the transport systems of the two cities considered in this work (Cats, 2011; Nassir et al., 2015).

In addition, standard consolidation of paths is performed, removing unrealistic dominated paths, such as those having loops, using the same stop twice, or using additional vehicles compared with other paths, resulting in a longer travel time. Out of the set of paths using the same lines but different vehicles, only the shortest one was considered (i.e., if it is possible to take the first arriving vehicle of a line, it is assumed there is no reason to wait for the next). In relation to overlapping lines, that is, with some or all stops in common, they were considered as different alternatives. In fact, it is not obvious when to consider two lines overlapping. Since it is possible to get on (or transfer to) a certain vehicle from different stops, lines doing a common path can still be taken doing two different paths.

As the last parameter, the walking speed for transfers is set to 1.4 m/s. Instead, the walking times from the Origin to the first stop, and from the

last stop to the Destination are set to 0. This ensures that all the stops near the Origin are reachable at the departure time (the beginning of the disruption). In addition, since the Destination is a stop in the network, considering a walking time would have penalized near stops and lines not stopping exactly at the Destination.
In conclusion, with the presented constraints, the proposed algorithm is able to generate two choice sets for each OD, containing a path for each possible combination of vehicles to use. These two sets represent the available alternatives with two different conditions of the network: without any disturbance and in case of disruption.

4.4.2.2 *Disruption Impact*

The impact of a disruption on a certain OD was defined based on random utility theory and the multinomial logit model. The impact is defined as the difference of the weighted average travel cost of the two choice sets (Equation 4.1). The equation represents the difference of travel cost in case of disruption (first term) with the case of no disruption (second term). The cost of each path is weighted by the probability to use it, computed using the multinomial logit model. In particular, the unnormalized probability of choosing a path j in a choice set is equal to $e^{-\beta C_j}$. Therefore, the lower the cost C_j, the higher the weight for this path. The underlying assumption is that people choose different paths to reach a destination, and the percentage of people choosing a certain path decreases with the cost of it. Full information on the disruption is assumed for the user (i.e., the user knows the available alternatives). The cost function used C_j is the travel time with a transfer penalty of 5 min and the calibrated parameters are based on Montini et al. (2017). Douglas and Jones (2013) reviewed transfer penalty estimates in literature and they showed there is not a common used value, even if most of the estimates range between 5 and 9 min of travel time.

$$impact\,(od, dis) = \frac{\sum_{j \in P(od,dis)} e^{-\beta C_j} C_j}{\sum_{j \in P(od,dis)} e^{-\beta C_j}} - \frac{\sum_{j \in P(od)} e^{-\beta C_j} C_j}{\sum_{j \in P(od)} e^{-\beta C_j}} \tag{4.1}$$

$$P(od, dis) = \text{choice set for the given od and disruption.} \tag{4.2}$$

$$P(od) = \text{choice set for the given od without any disruption.} \tag{4.3}$$

Following, a numerical example evaluating the disruption impact is reported

Two alternatives are available for a certain OD, A and B, with respective costs $C_A = 5$ and $C_b = 10$ min. It is assumed $\beta = 0.1$ for simplicity. . Therefore, the second term of the impact (i.e., average travel cost without disturbances) is equal to 6:89.
A disruption causing a delay of 5 min is assumed to the alternative A (e.g. $C_A = 10$). The first term of the impact (i.e., average travel cost in case of disruption) would be equal to 10, and therefore, $impact = 3.11$ min. In conclusion, in this toy example a delay of 5 min in one of the alternatives caused an impact of 3.11 min to the OD. Nevertheless, if the delay would have affected the alternative B, the impact would have been only 0.8 min. This shows that a disruption affecting a less costly path has a higher impact.
In relation to passengers' information on the disruption, it was decided to assume full information instead of no information, since the former can describe better the resilience of a system, in relation to providing a good quality service and alternatives in case of disruptions. In fact, assuming no information, in case of a line failure or a very long delay, the passengers would still wait for the failed service, even if a much better alternative were available. Nevertheless, the full information case can assume a specific connection is known, available only thanks to a delay. Finally, analyzing some sort of partial information is left for a future work, given the additional challenges and assumptions to take into account.

4.4.3 *Features Analysis*

Analyzing the relationship between an OD and the impact of a disruption, it is possible to determine how much the impact of the disruption on the OD depends on its characteristics and which of these are more important[Thesis Appendix 6]. First, 21 features were extracted for each OD, describing size of the disruption, duration, service frequency, network metrics, and other characteristics of the disruption and the OD. The list of features is shown in Table 4.1. The similarity between two trips is defined as the number of common stops divided by the average number of stops of the two trips. Therefore, the feature's importance in predicting the impact is analyzed computing the mutual information (MI) and using random forest regression[Thesis Appendix 7]. The mutual information is a measure of the amount of information one random variable contains about another (Cover and Thomas, 2006). Given two random variables X and Y, the mutual information is the following:

Feature	MDI	MI	Description
Frequency	0.078	0.203	Number of events per day (AVG)
Closeness	0.039	0.203	Closeness centrality (AVG)
OutDegree	0.047	0.181	# of reachable stops (AVG)
Betweenness	0.050	0.180	Betweenness centrality (AVG)
NumLines	0.079	0.156	# lines running at the stop (AVG)
ChoiceSetSize	0.119	0.146	Size of the timetable choice set
AvgTransfers	0.047	0.133	Avg. # of transfers in the timetable choice set
Events/Perimeter	0.034	0.133	# events / disruption perimeter
TripsSimilarity	0.048	0.131	Avg. similarity between disrupted trips
AvgTravelCost	0.079	0.122	Avg. travel cost in the timetable choice set
ClosenessDest	0.073	0.110	Destination closeness
Distance	0.080	0.098	OD distance
AvgDelay	0.044	0.087	Delay (AVG)
BetweennessDest	0.062	0.079	Destination betweenness
TotalDelay	0.042	0.062	Sum of delays of the disruption events
BusPercentage	0.001	0.048	% of buses involved respect to other means
TramPercentage	0.001	0.046	% of trams involved respect to other means
Duration	0.025	0.041	Disruption duration
Trips	0.016	0.031	# vehicles involved
Events	0.014	0.030	# events
TrainPercentage	0.001	0.021	% of trains involved respect to other means

TABLE 4.1: Feature Importance: Features Rankings Based on Mutual Information (MI) and Mean Decrease Impurity (MDI). AVG = the feature is computed as the average among the events of the disruption. Features are sorted by MI.

$$MI\left(X,Y\right) = \int_X \int_Y p\left(x,y\right) \log\left(\frac{p\left(x,y\right)}{p\left(x\right)p\left(y\right)}\right)\, dx\, dy$$

This metric determines the similarity between the joint distribution $p(x,y)$ and the product of the marginal distributions $p(x)$ and $p(y)$. In fact, if the two variables are independent, $MI(X,Y) = 0$.
Therefore, it is possible to rank the features by their mutual information with the impact.
Applications of this metric for a related task can be found in Chandrashekar and Sahin (2014). Indeed, present different feature selection methods, in-

cluding methods based on MI, to rank features by their importance. Nevertheless, this measure does not capture the relationships among features and it is possible that a feature has a high importance only if combined with others. In contrast, a random forest regression considers multiple features in one single model. To fit the regression model, 67% of the dataset was used as training set and cross-validation to estimate the parameters. The regression can show how much the features are able to describe the impact, and can rank them based on a metric called Mean Decrease Impurity (MDI) (Breiman, 2002; Louppe et al., 2013). Considering a feature x, its MDI value is computed as follows:

$$MDI(x) = \frac{1}{|T|} \sum_{T} \sum_{n \in T : v(s_n) = x} p(n) \Delta i(s_n, n)$$

where T is the set of trees in the forest; n is a tree node such that the split (s_n) is made on the feature x; $p(n)$ is the proportion of samples reaching n; $\Delta i(s_n, n)$ is the decrease (difference) of the considered impurity measure after the split s_n. In case of regression tree, i is the variance. For more details on feature importance in random forest, refer to Louppe et al. (2013). It is necessary to remark that particular attention must be given to correlated features, since this metric tends to distribute their importance.

These features have been preliminarily selected based on practical input and relevance, and review of the existing literature. Of course, a larger set of features could be considered, even though including additional features can reduce the interpretability of the results (in particular of the MDI) and cause overfitting. Therefore, a feature is included among the proposed ones, based on the following criteria: the feature describes a main disruption characteristic (as *avgDelay*, centrality measures, *events*); the feature has high importance (top part of the list); the feature has little correlation with the others (e.g., *closenessDest*, *busPercentage*, *choiceSetSize*). Among the excluded features none of them is considered worthy of mention. However, including different features does not impact the methodology of this work.

4.5 EXPERIMENTS AND RESULTS

4.5.1 *Disruption Identification*

For the experiments, 9 months (January 2018 to September 2018) of AVL data of the city of Zürich ($\approx$ 64 million of events) were used to analyze realized disruptions (Swiss Federal Office of Transport, 2020). These data are provided as open data from the Swiss Federal Office of Transport, and were used in other studies, confirming their validity (Büchel and Corman, 2019; Marra et al., 2019). In particular, Marra et al. (2019) showed how the same AVL data can be used to improve a mode detection algorithm, compared with using timetable data. Therefore, despite no report on the accuracy being provided, and it is not possible to exclude any error during the data collection, it is considered the data represent reasonably the real events in the network.

In Section 4.5.4, similar analysis is presented on the city of Bern. To identify disruptions with the ST-DBSCAN algorithm, for each day, all the events with a delay $\geq$ 6 minutes (*minDelay*) are selected for clustering. A threshold of maximum 3 h of delay is also considered to filter possible errors in the data.

The following values are assigned to the ST-DBSCAN parameters: *MinPoints* = 6, *epsSpace* = 250 m, *epsTime* = 4 min. How these values were chosen is explained in Section 4.5.3. In our experiments, 2955 disruptions were detected ($\approx$ 11 per day).

The disruptions were detected out of data spanning 9 months of service. Each day has a slightly different timetable, but, to allow comparisons, all those disruptions were projected to a "template" timetable, which is the same for all disruptions. In other terms, to avoid bias because of different timetables (e.g., during weekends or adjustments because of maintenance services), the disruptions considered in this analysis were only those with events (i.e., a public transport vehicle servicing a stop at a certain planned moment) that have been planned (and would have happened) also on a normal working day, which has been fixed w.l.o.g. to the October 1, 2018 (1,622 disruptions). If a disruption involves events not happening in that "template" day (e.g., a bus running only on Sunday), it is discarded from the analysis. Considering a different working day does not change significantly the results.

The spatial distribution is shown in Figure 4.2. It can be seen that most of the disruptions are located near the city center or railway stations.

Figure 4.3 presents the distribution of number of events and average delay per disruption. The number of events seldom becomes very high ($\geq$ 20), leading to clusters formed by events close to each other. This is also shown by the average number of stops involved in a disruption, which is 2.75. In relation to the amount of delay, the analyzed disruptions have a median of average delay of $\approx$ 10 min, showing that the analysis is focused mainly

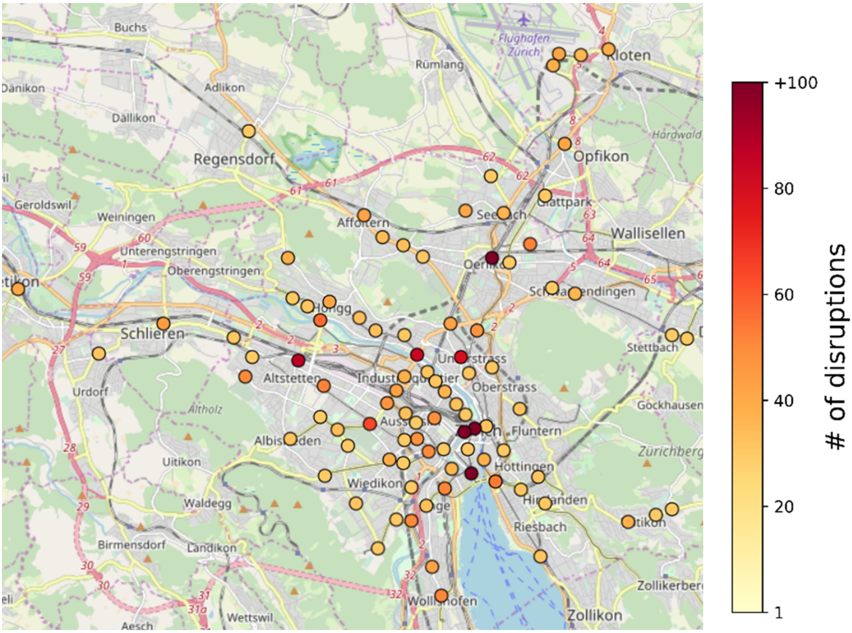


Figure 4.2: Distribution of disruptions in Zürich (from January 1, 2018 to September 30, 2018). Source: Map from openstreetmap.org.

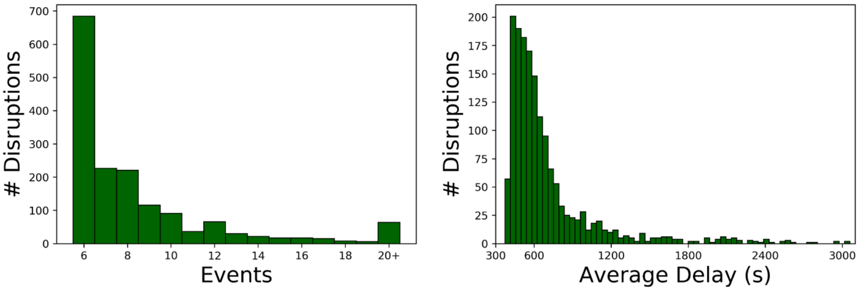


Figure 4.3: Distribution of number of events and average delay per disruption.

on small disruptions. In relation to the types of mode, on average, 97% of disruptions involve buses and trams, and 3% involve trains. The number of train disruptions is low because only 5% of the events (arrival of a vehicle at a stop) are made by trains in Zürich.

4.5.2 *Disruption Impact Analysis*

To evaluate the impact of the identified disruptions, for each disruption 15 different random ODs are created and the impact on each OD is computed, as explained in Section 4.4.2. Large choice sets were limited to the top 100 paths. The ODs not affected by the disruption (i.e., none of the vehicles involved in the disruption is ever used to reach the destination) are discarded (17%). In total, 19,731 OD pairs were analyzed.
It is necessary to remark that, since the impact function is based on the multinomial logit model, a disruption can also have a negative impact. For instance given an OD, if a disruption affects the worst path in its choice set, the probability to choose a better path increases, reducing the average travel cost.
The relationship between the features of each OD and the impact are analyzed as explained in Section 4.4.3. The random forest regressor gives an $R^2 = 0.41$. It is hard to assess the goodness of fit of the regression, since there are no studies with which to compare the results. Nevertheless, it is remarked that the main focus of this work is not the development of a prediction model, but to assess if it is possible to explain the impact of a disruption from its characteristics. Therefore, it can be appreciated how 41% of the variance in the disruption impact can be explained by the identified features. This proves that it is possible to (partially) predict the impact of a disruption (as defined by the authors) from its characteristics.
Given that two choice sets need to be computed for each OD, the number of possible ODs to analyze is limited in this paper by computational constraints. Therefore, the size of the dataset was selected when the corresponding model does not perform significantly better than a model based on a lower size. Namely, considering half of the dataset (9,865 ODs), the R^2 of the regressor decreases by 6%; while considering three-quarters of the dataset, it decreases by only 4%. Given that the full dataset, considering 15 ODs per disruption, does not perform significantly better than a subsample of it, it is assumed that considering even more ODs may still not significantly affect the results and the conclusions.
The results of the features importance analysis are shown in Table 4.1. Given the complexity of the task and the high correlation among the features, the values in Table 4.1 must be judged as useful to make general conclusions and not as strict rankings[Thesis Appendix 8]. One of the most relevant features in both the metrics is the *frequency* (of service). This proves that a high-frequency service can contrast delays or single failures. Slightly

less important are three network metrics (out-degree, closeness, and betweenness centrality), proving that the impact of a disruption depends on its location and connectivity in the network. These metrics are computed on a static network with a node for each stop and arcs weighted by the travel time. Two stops are connected if they are connected by a service or if they are distant by less than 350 m. Despite a high MI, they have a lower MDI for their high correlation among them and other features. The disruption density ($events/Perimeter$) has a moderate influence, showing that an increase of disturbances in the same area has a greater impact. Instead, features with a lower influence are the duration and the number of events of the disruption. It is interesting that network metrics of the destination do not have a high influence on the impact, proving that it is more important to go away from the disrupted zone. Finally, the type of mode involved in the disruption is not relevant ($trainPercentage$, $tramPercentage$, $busPercentage$).

The relationships between some of the features and the impact are shown in Figure 4.4. The impact decreases with higher $frequency$, $choiceSetSize$ and $outDegree$. This is realistic, since high values of these features correspond to a better quality of service in the disrupted area. With the increase of $betweenness$, the impact first slightly decreases, but then it increases again. This shows that a disruption has higher impact in a poorly connected area or in a hub, and less in intermediate zones. Considering the delays of the disruption events, the impact increases with $avgDelay$ until a certain value ($\approx$ 17 minutes), then an increase of delay is no longer important. Finally, the impact decreases with $numLines$. This shows that if several lines are available in the disrupted area, a disruption has lower impact.

To make the analysis as realistic as possible, it is based on a behavioral model, considering the set of available alternatives. In this sense, the validation of the proposed methods with realized passengers' data, such as AFC data or tracking data, can strengthen the findings.

Nevertheless, this validation leads to additional challenges, such as the identification of passengers affected by the disruption; the retrieval of information on the initial plan; and the delay estimation for each passenger. Therefore, this type of validation is left for a future work.

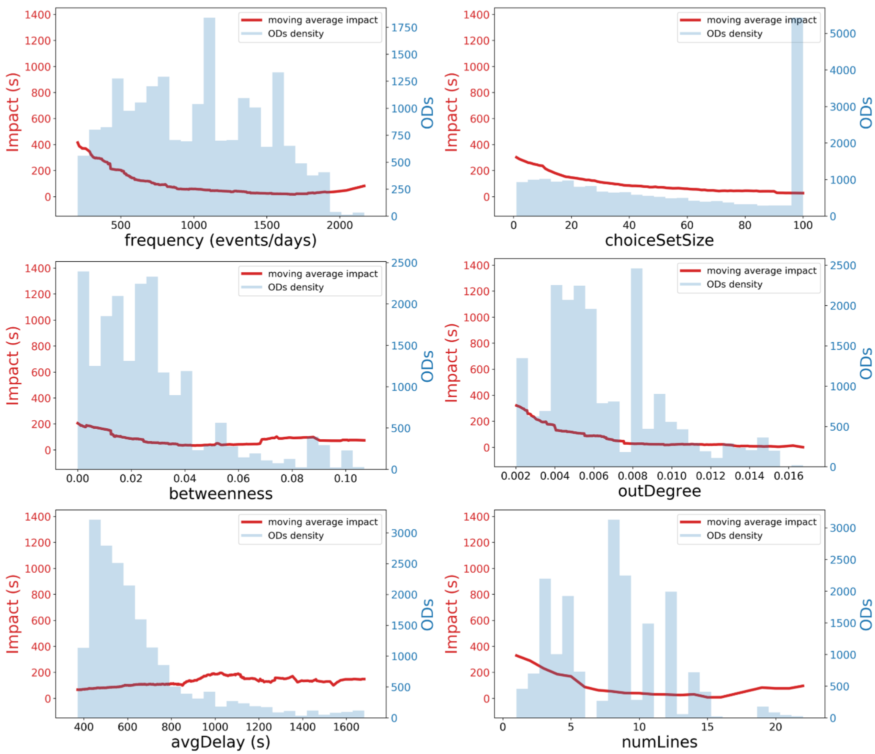


Figure 4.4: Relationships between features and impact.
ODs = origin-destinations.

4.5.3 *Tuning of the Disruption Identification Parameters*

Since the aim is to identify disruptions with different characteristics, there is not a specific target to tune the parameters of the disruption identification algorithm. Nevertheless, some boundaries must be respected to obtain feasible results. In particular, for each day, the algorithm must both identify disruptions and not report too-broad clusters. In fact, too-strict values lead to no disruption identified; in contrast, slack values lead to big disruptions expanding in the network. To tune the parameters, the disruptions which were analyzed were those identified with different combinations of values for the four parameters in a tuning period of 30 days: *minDelay* $\in \{4, \mathbf{6}, 8\}$; *epsSpace* $\in \{150, \mathbf{250}, 350\}$; *epsTime* $\in \{2, \mathbf{4}, 6\}$; *minPoints* $\in \{4, \mathbf{6}, 8\}$. The final values (in bold) were chosen as follows:

	Zürich	Bern
Area considered (km²)	330	64
Public transport stops	987	365
Bus/tram lines	126	25
Events per day (≈)	235,000	73,700
Avg # connected stops per stop	5.34	4.53
Std # connected stops per stop	3.7	3.8
Avg stop distance (meters)	1,156	735

TABLE 4.2: Zürich and Bern Public Transport Networks (PTNs) comparison. The Bus/tram lines referred to are those that provided automatic vehicle location (AVL) data in the analyzed period.

- $epsSpace = 250$. (Using 150 leads on average to less than two stops per disruption; and using 350 leads to less than two events per stops.))
- $minPoints = 6$. (Using 4 leads on average to 47 disruptions per day [that can be considered unrealistic]; and using 8 does not identify disruptions in most of the cases [59% of the analyzed days].)
- $minDelay = 6$. (Using 4 leads on average to 48 disruptions per day [unrealistic]; and using 8 does not identify disruptions in most of the cases [55% of the analyzed days].)
- $epsTime = 4$. (Given the chosen values of the other parameters, $epsTime = 2$ does not identify any disruption in 59% of the days; and 6 leads to events less close in time compared with 4 [disruptions 55% longer]; therefore 4 is preferred to analyze more compact disruptions.)

With the selected parameters, in the tuning period, disruptions were identified in 66% of the days, with an average of 8.35 disruptions per day and a maximum of 64 in one day. Each disruption has an average of 9.4 events and 1.4 missed stops, and the median of its events' delays of 10.2 min. Given the high number of events per day (Table 4.2), the number of disruptions per day is not considered to be too high.

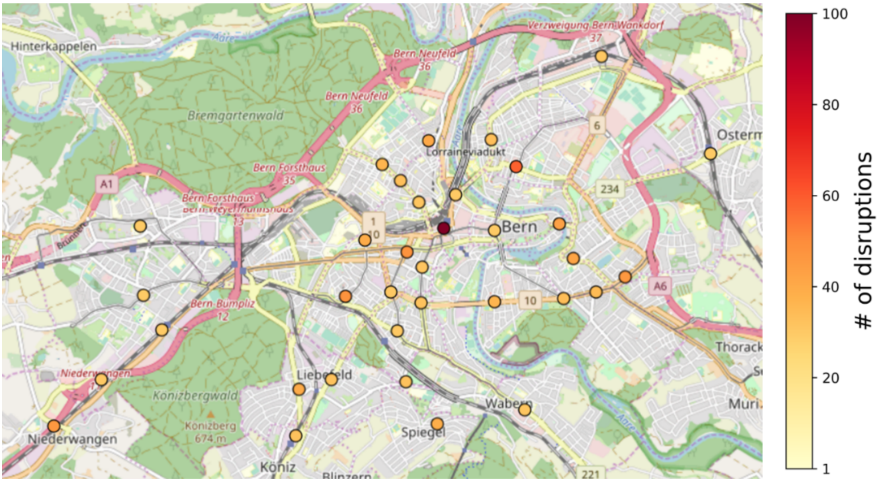


FIGURE 4.5: Distribution of disruptions in Bern (from 01-01-2018 to 31-12-2018). Source: Map from openstreetmap.org.

4.5.4 *Bern Network Analysis*

To strengthen the results with respect to a possible bias because of the specific PTN, the same analysis was repeated for a different city of Switzerland, Bern. The major differences between the PTN of the two cities are highlighted in Table 4.2. The network of Bern is smaller in relation to area and service, and the geographical characteristics of the cities are also different, given, for instance, the presence of the lake in Zürich. The same disruption identification algorithm was applied in Bern, considering data of all the days in 2018. Therefore, 376 disruptions were analyzed and their distribution is shown in Figure 4.5. As in Zürich, the highest number of disruptions is near the main station. For the impact analysis, 20 different ODs were selected per disruption. After discarding ODs not affected by the disruption, a total of 5,897 ODs were analyzed. The random forest regressor gives an $R^2 = 0.60$, that is higher than the one from Zürich. This can be explained by the smaller size of the PTN in Bern; therefore, it is easier to identify heterogeneity among different disruptions' locations.
The feature importance analysis is shown in Table 4.3. Comparing Table 4.1 and Table 4.3, the order of the features is similar for both the tables (with some exceptions). This demonstrates that the features of an OD can explain the disruption impact independently of the PTN. In this sense, it

Feature	MDI	MI	Feature	MDI	MI
Betweenness	0.067	0.474	Duration	0.054	0.209
Closeness	0.056	0.473	AvgTravelCost	0.069	0.167
Frequency	0.056	0.467	Trips	0.021	0.142
OutDegree	0.039	0.381	BusPercentage	0.009	0.131
TripsSimilarity	0.040	0.363	TramPercentage	0.007	0.118
Events/Perimeter	0.033	0.357	ClosenessDest	0.052	0.103
AvgDelay	0.046	0.308	Events	0.023	0.097
TotalDelay	0.054	0.265	Distance	0.072	0.096
ChoiceSetSize	0.115	0.258	BetweennessDest	0.049	0.087
NumLines	0.083	0.255	TrainPercentage	0.002	0.074
AvgTransfers	0.056	0.234			

TABLE 4.3: Bern feature importance: Features Rankings Based on Mutual Information (MI) and Mean Decrease Impurity (MDI). Features are sorted by MI.

can be confirmed that the frequency of service, the choice set size, and network metrics play a key role on the disruption impact. In conclusion, this section showed that the same analysis can be applied to different networks, obtaining similar results.

Despite testing the analyses with two different PTNs, it is acknowledged that the whole variety of cases was not covered. PTNs in different countries are subject to different regulations and developed in different manners. In addition, they can include different transport modes, such as a metro system. On the contrary, the Swiss public transport system is characterized by a short distance between the stops and a high reliability of service. Given that, it is remarked that the methodology is independent of the specific test case, while the tuning of the disruption identification algorithm and the feature importance values may be subject to variations. Given the objective effort in the collection and evaluation of long-term AVL data, the analysis of additional PTNs was left for future work.

4.6 CONCLUSION

The classical definition of PTN disruption as a physical link closure has been overcome in this study. A new definition is given, based on the combination of delays and missed stops, to represent disruptions with different characteristics and relate them with small disturbances in the PTN. This approach allows analyzing the impact of a disruption from small disturbances, and considering their co-occurrence as the actual disruption. In fact, it was identified that the amount of delay has a positive influence on the impact of a disruption, but only until a certain value. In addition, the heterogeneity of the generated disruptions allowed the analysis of both small groups of delays and big disturbances in the network. AVL data of the cities of Zürich and Bern was used to identify real cases of disruptions. This fills the gap in the literature on short (in the order of minutes) disruptions analysis in multimodal public transport. The focus was on road-based public transport services, as they occur more frequently in the dataset and the statistical strength of the analysis is larger. An extension to other modes, such as railways or metros, is of course possible. Possibly some ad-hoc adjustment of the disruption identification will be needed, to include also topological proximity, in case of track-based systems; but it is felt that the proposed methods are still valid to analyze these specific disturbances.
An additional aspect in this analysis that differs from the literature is the level of detail. Instead of computing the disruption impact on the whole network, it was modeled on single ODs affected by the disruption, based on two different choice sets, allowing consideration of the impact at a fine-grained level and analysis of it for different types of OD. Therefore, a choice set generation algorithm was used to generate the possible paths with and without disturbances in the network. It was shown for Zürich that there is a strong relationship between the impact of a disruption and its characteristics, and the disruption characteristics were ranked according to their influence. In particular, the service frequency, the choice set size of the considered OD, and network metrics of the disruption area play a key role in the disruption impact. In contrast, destinations' metrics are less important. An interesting finding is that the size and the number of events in a disruption are less relevant than the characteristics of its location. Finally, the analysis was repeated on a different city, Bern, showing that the main findings can be generalized and they are not limited to the particular case study.

From an industrial perspective, this paper helps operators in strategic analyses, by providing quantitative assessment of the impact of different disruptions in a PTN, thus prioritizing contingency measures, such as keeping suitable vehicle reserves, or providing sufficient buffer times in operations[Thesis Appendix 9]. From an operational perspective, the quantitative framework proposed can be triggered in real time to determine if measured delays are normal variations to be absorbed by existing buffer times; or if, instead, they need to be managed with specific processes (such as contingency plans or re-scheduling). For instance, a real-time impact computation of all events can be continuously computed, and only those events, that together will determine a large impact to passengers, prioritized for operator response. In this sense, operators can base their decision weighting the disruption impact by the relevance of the affected ODs. Finally, as a measure of analysis and evaluation, such a study allows authorities to design a passenger-based quality control, where not only operational aspects are considered, but also their joint impact toward passengers outcomes.
This paper represents a first step in the analysis of different types of disruptions and several future directions are possible, such as using a different disruption identification method or a different impact function, for instance including crowding factors in affected lines (Yap, 2020). Testing the same methodology on big disruptions, involving the whole network, can help to make a better distinction between small and big disruptions. Analyzing the disruption impact considering different information provisions for the users can help understand the importance of information systems for disruption management.
The test case refers to a public transport system working mostly on road-based buses, with a high quality. An analysis of cities with various degrees of exposure to disruptions would contribute to the state of the art of understanding how PTNs work. Moreover, combining this research with a tracking study and a mode detection algorithm, such as the one proposed by Marra et al. (2019), or with AFC systems, allows empirical evaluation of the findings and the users' behavior in case of disruptions.

DECLARATION OF CONFLICTING INTERESTS

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

FUNDING

The author(s) received no financial support for the research, autorship, and/or publication of this article.

REFERENCES

Angeloudis, P. and D. Fisk (2006). Large subway systems as complex networks. *Physica A: Statistical Mechanics and its Applications* 367, 553.

Birant, D. and A. Kut (2007). ST-DBSCAN: An algorithm for clustering spatial–temporal data. *Data & Knowledge Engineering* 60.1, 208.

Bovy, P. H. L. (2009). On Modelling Route Choice Sets in Transportation Networks: A Synthesis. *Transport Reviews* 29.1, 43.

Breiman, L. (2002). Manual on setting up, using, and understanding random forests v3. 1. *Statistics Department University of California Berkeley, CA, USA* 1, 58.

Büchel, B. and F. Corman (2019). The disruption at Rastatt and its effects on the Swiss railway system. *RailNorrköping 2019. 8th International Conference on Railway Operations Modelling and Analysis (ICROMA)*. Vol. 69, 221.

Burgholzer, W., G. Bauer, M. Posset, and W. Jammernegg (2013). Analysing the impact of disruptions in intermodal transport networks: A micro simulation-based model. *Decision Support Systems* 54.4, 1580.

Cats, O. (2011). Dynamic Modelling of Transit Operations and Passenger Decisions. PhD Thesis. KTH, Traffic and Logistics, Centre for Traffic Research, CTR.

Cats, O. and E. Jenelius (2014). Dynamic Vulnerability Analysis of Public Transport Networks: Mitigation Effects of Real-Time Information. *Networks and Spatial Economics* 14.3, 435.

Cats, O. and E. Jenelius (2018). Beyond a complete failure: the impact of partial capacity degradation on public transport network vulnerability. *Transportmetrica B: Transport Dynamics* 6.2, 77.

Chandrashekar, G. and F. Sahin (2014). A survey on feature selection methods. *Computers & Electrical Engineering* 40.1, 16.

Cover, T. M. and J. A. Thomas (2006). Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing). Wiley-Interscience.

Douglas, N. J. and M. Jones (2013). Estimating transfer penalties and standardised income values of time by stated preference survey. *Australasian Transport Research Forum (ATRF), 36th, 2013, Brisbane, Queensland, Australia*.

Jenelius, E. (2009). Network structure and travel patterns: explaining the geographical disparities of road network vulnerability. *Journal of Transport Geography* 17.3, 234.

Leng, N. and F. Corman (2020). The role of information availability to passengers in public transport disruptions: An agent-based simulation approach. *Transportation Research Part A: Policy and Practice* 133, 214.

Louppe, G., L. Wehenkel, A. Sutera, and P. Geurts (2013). Understanding Variable Importances in Forests of Randomized Trees. *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 1*. Curran Associates Inc., 431.

Lu, Q.-C. (2018). Modeling network resilience of rail transit under operational incidents. *Transportation Research Part A: Policy and Practice* 117, 227.

Marra, A. D., H. Becker, K. W. Axhausen, and F. Corman (2019). Developing a passive GPS tracking system to study long-term travel behavior. *Transportation Research Part C: Emerging Technologies* 104, 348.

Mattsson, L.-G. and E. Jenelius (2015). Vulnerability and resilience of transport systems – A discussion of recent research. *Transportation Research Part A: Policy and Practice* 81, 16.

Montini, L., C. Antoniou, and K. W. Axhausen (2017). Route and mode choice models using GPS data. *TRB Annual Meeting Online*. Vol. 1204, 17.

Nassir, N., M. Hickman, A. Malekzadeh, and E. Irannezhad (2015). Modeling Transit Passenger Choices of Access Stop. *Transportation Research Record: Journal of the Transportation Research Board* 2493, 70.

Rodríguez-Núñez, E. and J. C. García-Palomares (2014). Measuring the vulnerability of public transport networks. *Journal of Transport Geography* 35, 50.

Sun, D. (and S. Guan (2016). Measuring vulnerability of urban metro network from line operation perspective. *Transportation Research Part A: Policy and Practice* 94, 348.

Swiss Federal Office of Transport (2020). Actual Data - `https://opentransportdata.swiss/en/dataset/istdaten`.

Tan, M.-C., C. O. Tong, S. C. Wong, and J.-M. Xu (2007). An algorithm for finding reasonable paths in transit networks. *Journal of Advanced Transportation* 41.3, 285.

Van der Hurk, E. (2015). Passengers, Information, and Disruptions (No. EPS-2015-345-LIS). PhD Thesis. ERIM Ph.D. Series Research in Management. Erasmus Research Institute of Management.

Yap, M. D. (2020). Measuring, Predicting and Controlling Disruption Impacts for Urban Public Transport. PhD Thesis. Delft University of Technology.

5

PASSENGERS' BEHAVIOUR DURING PUBLIC TRANSPORT DISTURBANCES

This chapter is based on the following article:

How different network disturbances affect route choice of public transport passengers. A descriptive study based on tracking

Alessio D. Marra[1] & Francesco Corman[1]
[1]Institute for Transport Planning and Systems, ETH Zürich

Submitted to Transportation on 04.2021

Contributions

A. D. Marra: Study conception and design; Analysis and experiments; Writing of the manuscript
F. Corman: Study conception and design; Writing of the manuscript

Key findings

- The behaviour of passengers in case of disturbances is observed from a large-scale study based on GPS tracking
- The effects of different disturbances on passengers are analysed in terms of excess journey cost and route choice
- Disturbances are compared based on a metric of service degradation quantifying the disturbances affecting a specific demand
- The focus is on multiple small disturbances and "good disturbances"

Additional notes to this chapter

The metric of service degradation described in Section 5.5.2 is based on the definition of disruption impact in Section 4.4.2.2. The differences are in the definition of a disturbance and the cost function used.
The Appendix of this Chapter is an addition to the submitted paper.

5.1 ABSTRACT

Public transport networks are affected daily by disturbances with different entities. While big service disturbances (or disruptions) are rare, delays and cancelled runs are more frequent and they affect daily the passengers. Despite this, most of the work on passengers' behaviour in case of disturbances focuses on big disturbances and not on daily disturbances. In this work, we analyse how different network disturbances affect public transport passengers, regarding the chosen route and the travel cost. For this purpose, we exploit a large-scale travel survey, based on GPS tracking, and AVL data of public transport operations in Zürich (Switzerland). We propose a novel metric of service degradation to quantify the disturbances in the network, which are relevant for a specific passenger's trip. In that way, we can analyse passengers' route choices in case of different disturbances. In particular, our study evaluates the effects of disturbances on travel costs, comparing the passenger's chosen route with the available alternatives with and without disturbances. Our analysis identifies that small disturbances and delays have a significant effect on travel cost, although marginal effects on route choice. In contrast, "good disturbances", i.e. variations of the operations from the timetable, which generate less costly alternatives, have a significant effect on route choice. In particular, we identified that passengers do not exploit these new available alternatives, suggesting a need of better information in these cases. In addition, we observed that travel costs are generally higher in case of "good disturbances" than in case of no disturbance.
Keywords: Public Transport; Disturbances; Tracking; User Behaviour; Route Choice

5.2 INTRODUCTION

The presence of disturbances in a public transport network is an important aspect affecting the passengers' choices. Unexpected delays or, more in general, service degradation can significantly affect the users, forcing them to deal with unplanned events. In this regard, understanding the impact of a certain disturbance on a specific user is a complex task, which requires different sources of information. First, the disturbance needs to be identified and clearly defined. Second, the choice of an affected user must be observed, and the available alternatives identified. Third, there is the need of a reference choice (what would have been chosen if no disturbance occurred), with which compare the chosen alternative, in order to quantify the impact of the disturbance on the user.

In this work, we use the term disturbance to indicate any frequent and small unplanned service deviation, while the term disruption only when we report findings of external works using that word, which refers often to big service disturbances. Stated preference surveys are not particularly suited for this task (understanding the impact of disturbances on a specific user), since they are based on hypothetical scenarios, and the user's real choices under disturbances cannot be observed. Therefore, they can suffer of an inner bias (Lin et al., 2018). Several studies on disruption analysis based their work on simulations (Cats and Jenelius, 2014; Leng and Corman, 2020), in order to investigate the behaviour of multiple passengers under specific disruptions. Nevertheless, the behaviour of each simulated passenger is based on assumptions, or typically defined under normal conditions in the network. This might not be realistic, since the user behaviour under disruptions can largely deviate from undisrupted scenarios (Yap et al., 2018). In addition, the information available to each passenger needs to be specified, also based on assumptions. In this sense, despite in the last years the research attention on disruption impact is increasing (Marra and Corman, 2020b), there is not yet a clear understanding of specific aspects of passengers' behaviour under disturbances, compared to undisturbed conditions. In fact, Lin et al. (2016) identify low attention in literature on transit user behaviour in response to service disruptions.

For the above-mentioned reasons, revealed preference surveys are a promising data source to observe users' behaviour in case of disturbances. In particular, passive GPS tracking allows collecting a large amount of data for a long time, without an excessive burden on the users. In this regard, Zhu and Levinson (2012) recommend survey approaches based on GPS tracking, to objectively observe travel decisions and experiences, allowing more sophisticated behavioural analysis. Harrison et al. (2020) affirms tracking datasets provide a more finely grained picture of travel behaviour and accurate details on revealed trip choices. In addition, Sun et al. (2016) affirms

that without the support of real passenger travel data, the behaviour in case of disruptions of many passengers is not discovered. In this sense, with these data, no assumptions are needed on the passengers' behaviour or the available information, since the behaviour is directly observed.
This work aims to fill this gap, and understand, from realized data, "how different unplanned network disturbances affect public transport passengers, in terms of increased travel cost and deviation of the chosen route from the expected one in normal conditions?"
The scientific and practical challenges are in the determination of a coordinated set of procedures to perform: (1) the aggregation of data about multiple disturbances and disruptions into a single metric; (2) the identification of a most likely route, i.e. the alternative which, with higher probability, would be chosen in case of no disturbance; (3) the identification of the chosen route in case of disturbances. Comparing the last two, one can identify the changes in route and costs. If a relationship can be established between those changes and the disturbance metric, the role of the disturbances can be better understood.
We base this analysis on the hypotheses that the travel times and costs of disturbed trips may be different (typically, higher) from planned conditions; and that passengers may choose a different route to avoid a disturbance or its effects (passengers may also change their destination in case of big disruptions, which we do not consider in this study). We explore these hypotheses from a large-scale survey based on tracking. The resulting descriptive study is able to establish a relation between different disturbances and the observed route choices and travel costs of passengers.
Each disturbance may affect each passenger differently, depending on the disturbance characteristics, the passenger, and the trip of that passenger. Therefore, multiple observations of passengers' trips in case of different disturbances are needed, to understand their effects. In contrast, most of the work in literature focuses only on one or few disturbances, typically disruptions, which are also rare events (Yap and Cats, 2020). Little attention is given to small disturbances, such as delays or cancelled runs, which are much more common. No attention is given to quantify and compare different disturbances. This paper proposes a novel metric of service degradation, which quantifies the disturbances in the network, potentially affecting the passengers. Such a metric allows comparing, on the same numerical scale, conditions of no disturbance, small disturbances and big disturbances. Additionally, we observe the effects of "good disturbances", which we define as deviations of the operations from the timetable, lead-

ing to less costly alternatives. Those situations are for instance an earlier arrival/departure or new available transfers, which would typically be unavailable in planned conditions. This type of disturbance has not yet been analysed in the literature, despite our analyses show that it has a significant impact on passengers.
Regarding observations of passengers' trips, we exploit a large-scale GPS tracking survey, based on automatic mode detection. In addition, for each observed public transport trip, we identify the available alternatives, using a choice set generation algorithm based on Automatic Vehicle Location (AVL) data. This large dataset allows estimating a general route choice model of public transport passengers; and provides very detailed information in case a disturbance occurs (compared to aggregated passenger flows). To the best of our knowledge, this is the first work addressing the understanding of behaviour in case of different unplanned disturbances from a large-scale survey.
The main contributions of this paper can be summarized as follows:

- A metric of service degradation is defined, which quantifies the level of disturbances in the relevant sub-network for a specific demand, defined by an origin, a destination and a departure time. The metric is based on a discrete choice model, and it quantifies the expected increase of cost for a certain demand, due to the current network conditions. This metric allows evaluating the level of service degradation in the network, which directly affects a specific user. We use this metric to aggregate, over multiple cases of individual disturbances, their impact on passengers' route choices.

- We evaluate the impact of the disturbances on a specific user in terms of excess journey cost (EJC), i.e. the additional cost experienced by the user respect to the expected one. In particular, we compare the user's chosen route with the set of available alternatives in case of disturbances, and the set of alternatives in planned conditions.

- We observe the behaviour of public transport passengers, under different disturbances, through a large-scale travel survey, based on tracking. Information on route choice, including exact public transport vehicles used and current network conditions, are available for each observed trip. The focus is on multiple small disturbances, instead of the more studied big disturbances (or disruptions). In particular, we analyse 2901 trips, each affected by different network conditions and disturbances.

- We identify a general trend of the effects of disturbances. The main insights are as follows: even if disturbances negatively affect the users' trips (in terms of increased travel cost), they marginally affect the users' route choice. In contrast, in case of "good disturbances", such as an early departure of a public transport vehicle, the users do not exploit the new available alternatives, and their chosen route is often different from the most likely route in case of no disturbance.

The paper is organized as follows: in Section 5.3, we present the state of the art on users' behaviour under disturbances; Section 5.4 describes the datasets; Section 5.5 presents the methodology; Section 5.6 illustrates the results of the analyses; Section 5.7 shows the conclusions.

5.3 STATE OF THE ART

In the last decade, the attention on the management of public transport disturbances has shifted from an operation-oriented perspective to a passenger-oriented perspective (Krishnakumari et al., 2020; Van der Hurk, 2015; Zhu and Goverde, 2019). The impact of a disturbance is not only evaluated in terms of operations' delay, but the effects on passengers gained a central role. In particular, most recent works analyse the disturbance impact on passengers, focusing on their routes. This is also the focus of the current paper. In contrast, different studies, such as Rahimi et al. (2020) and Shires et al. (2019), aim to understand how passengers react to disturbances, in terms of change of mode or plan.
Table 5.1 reports some of the most recent works on public transport passengers and disturbances, summarizing the different datasets used for passengers' trips, and how the disturbance impact on passengers is evaluated. The list of works does not aim to be comprehensive of all available works, but wants to highlight the current state of the art, its missing aspects and where this paper positions itself (see last row). In addition, we remark that our focus is on the effects of disturbances on passengers, in terms of increased travel cost and deviation of the chosen route from the expected one. Therefore, travel time reliability (Carrion and Levinson, 2012), expected delays and past disturbances experienced by passengers are not considered in this paper, since we focus on the effects of a disturbance at that moment.

Work	Passenger Dataset	Type of Disturbance (amount)	Disrupted Trips
Yap and Cats (2020)	AFC	Disruption log (4263)	AFC estimation
Leng (2020)	Simulation	Unplanned disruption (1)	Simulation
Rahimi et al. (2020)	SP-RP survey	Hypothetical scenarios	Asked in survey
Shires et al. (2019)	SP-RP survey	Planned disruptions	Asked in survey for an old episode
Yap et al. (2018)	AFC (5 million trips)	Planned disturbances (4)	AFC estimation
Paulsen et al. (2018)	Simulation	Train delays (65 days)	Simulation
Ghaemi et al. (2018)	OD matrix	Railway disruptions (4)	Dynamic passenger assignment
Lu (2018)	OD matrix	Rail disruptions (167)	Shortest path
Antos and Eichler (2016)	AFC (8 days, 1 OD)	Various disruptions (6)	Data in disrupted days
Sun et al. (2016)	AFC	Rail disruption (1)	Shortest path
Van der Hurk (2015)	Simulation	Rail disruptions (3)	Shortest path
Cats and Jenelius (2014)	Simulation	Unplanned disruption (1)	Simulation
This paper	**GPS tracking (2901 trips)**	**Small network disturbances (2901)**	**All observed trips**

Table 5.1: Recent works on passengers' behaviour in case of public transport disturbances (the table continues).

Work	Planned Trips	Disturbance Impact Evaluation	Focus[a]
Yap and Cats (2020)	AFC estimation	EJT	Station
Leng (2020)	Simulation	EJT	Trip
Rahimi et al. (2020)	-	-	Change of mode or plan
Shires et al. (2019)	Asked in survey	-	Change of mode or plan
Yap et al. (2018)	AFC estimation	Prediction model	Network
Paulsen et al. (2018)	Simulation	EJT	Trip
Ghaemi et al. (2018)	Passenger assignment on timetable	EJC	Network
Lu (2018)	Shortest path	Resilience measure	OD
Antos and Eichler (2016)	Data in normal days	Travel time distribution comparison	OD
Sun et al. (2016)	Shortest path	Average delay, max delay, punctuality rate	Network
Van der Hurk (2015)	Shortest path	EJT	Network
Cats and Jenelius (2014)	Simulation	EJC	Trip
This paper	**Expected travel cost (timetable and realized)**	**EJC (timetable and realized)**	**Trip**

a: "Trip" if the focus of the analysis is each single trip, which is modelled; "Network" if the focus is on the overall traffic, i.e. the disturbance impact is evaluated at network level; similarly "Station" and "OD" evaluate the disturbance impact on stations or ODs.

5.3.1 *Data Sources*

Quantifying the impact of a specific public transport disturbance can help understanding its severity, which passengers are most affected, and how to respond better to the disturbance. Unfortunately, this task is not trivial, since it requires the knowledge of both the passenger's realized trip and the planned trip (Paulsen et al., 2018). If the first is retrievable from tracking systems or Automatic Fare Collection systems (AFC), the latter can be retrieved only directly asking to the passenger, which is often unfeasible for large, long-term longitudinal datasets. In addition, the initial plan is not always fixed and can mutate during the time and the trip, changing either the destination, the mode or the route.
Passengers' behaviour in case of disturbances can be observed from different data sources. Among survey-based studies, Currie and Muir (2017) collected information through a web survey to better understand users' priorities and perceptions in unplanned railway disruptions. Lin et al. (2018) collected a joint Revealed Preference (RP) and Stated Preference (SP) survey to investigate mode choice in case of subway service disruptions. They highlighted that if on one hand SP surveys have an inherent bias, RP surveys may contain an insufficient number of observations. In addition, many factors affect passengers' decision making, such as the duration of a delay, the weather, the trip purpose and the available information (Leng, 2020). Therefore, the complexity of decision making shows the need of empirical studies for a better understanding (Lin et al., 2016).
Rahimi et al. (2020) and Shires et al. (2019) observed passengers' behaviour through a combined SP-RP survey. Rahimi et al. (2020) presented hypothetical disrupted scenarios to respondents, to understand which mode/action they choose, such as using a taxi or changing the destination. Similarly, Shires et al. (2019) asked respondents to remember a disrupted episode. With the absence of passengers' data, a different stream of research simulated passengers' behaviour (Cats and Jenelius, 2014; Leng and Corman, 2020; Paulsen et al., 2018). In this context, Van der Hurk (2015) use a simulation model to calculate the change in passenger flows due to a disruption. Nevertheless, they need to assume a behavioural model of how passengers react to a disruption, and the information available to each passenger. An alternative data source is a network OD matrix, used to assign the traffic in the network, and analyse the effects of disturbances in the network or for specific ODs (Ghaemi et al., 2018; Lu, 2018). In this case, the traffic assignment in case of disturbances becomes a crucial assumption.

A more detailed data source is an AFC system, which automatically collects the entrance and/or the exit of each passenger in the network, allowing inferring the passengers' trip. Limitations of AFC data are the non-trivial inference of the real trip, and the lack of knowledge on the real origin and destination of the trip (since in the data the trip starts directly at the first public transport stop). One of the few paper investigating public transport route choice under disturbances from AFC data is Yap et al. (2018). Focusing on 4 planned disruptions, they identified that the in-vehicle time of rail-replacing services and its waiting time are perceived worse than the ones of normal services. Finally, we have no evidence of any study exploiting tracking data to observe passengers' behaviour in case of disturbances. In car traffic, in contrast, a recent work (De Moraes Ramos et al., 2020) analysed route choice in a congested network from GPS trackers. Instead, Paipuri et al. (2020) showed that phone data can be used to extract information on the user-equilibrium gap (a concept very close to the EJC of this work) for specific ODs and times of the day, which is otherwise possible only with simulations.

5.3.2 *Type of Disturbances*

Regarding the type of disturbances analysed, most of the works focuses on big disturbances, which are rare and unique events, instead of small network disturbances, which have a daily frequency (Marra and Corman, 2020b). Among the recent works, we identified only Paulsen et al. (2018) focusing on small disturbances. Nevertheless, they analysed only train delays, through a simulation model, while our work focuses on different public transport disturbances and real passengers' observations. Table 5.1 reports the number of disturbances of different studies (see third column). Most of the studies focuses on only one or few disturbances (Lin et al., 2016), making their findings difficult to generalize. In contrast, Yap and Cats (2020) built a prediction model from 4263 disruptions, to predict the disruptions frequency and their impact on passengers. None of the works in literature defined a metric quantifying the severity of different disturbances or describing them, despite its importance has been acknowledged (Yap et al., 2018). Such a metric allows comparing the passengers' behaviour under different types of disturbances, ranging from no disturbance to big ones, and including "good disturbances", as we do in this paper. In particular, we were able to compare 2901 different network conditions from realized observations, which is one of the largest number observed in literature (see Table 5.1).

5.3.3 *Disturbance Impact on Passengers*

In literature, the planned trip (i.e. a reference trip, to which the passenger's trip in case of disturbances is compared), is often modelled based on specific assumptions. Mainly, three alternative assumptions are used, all of them assuming no disturbance in the network: selecting the path with shortest travel time (Lu, 2018; Sun et al., 2016; Van der Hurk, 2015); simulating the passenger's behaviour (Cats and Jenelius, 2014; Leng and Corman, 2020; Paulsen et al., 2018); estimating a traffic assignment in the network or for specific ODs (Ghaemi et al., 2018; Yap et al., 2018; Yap and Cats, 2020). The shortest travel time and the simulation consider only one alternative as reference trip for a passenger, which can or cannot be based on some behavioural assumption. Instead, the traffic assignment loses the focus on each single passenger's trip, and analyses the effect of a disturbance at a broader level (for each OD or the whole network). In this sense, this work aims to keep the focus on each passenger's trip, but considering a more sophisticated reference trip, based on a choice set, comprising the available alternatives (see Section 5.5.3).
Regarding the evaluation of the disturbance impact to a specific user, there is not a common metric in literature, since it depends on the study purpose. The most common metric is the excess journey time (EJT), which is the difference between the travel times of the disrupted trip and the planned trip. Similarly, two works (Cats and Jenelius, 2014; Ghaemi et al., 2018) consider the excess journey cost (EJC, also called generalized travel cost), which is the difference of the travel cost of the disrupted trip and the one of the planned trip, with the cost obtained from a behavioural model. Considering the excess journey cost as evaluation criterion allows measuring the impact of a disturbance from the passengers' point of view (Ghaemi et al., 2018). Finally, other less common metrics are a resilience measure (Lu, 2018), maximum delay and punctuality rate (Sun et al., 2016), or using a prediction model to observe behavioural changes (Yap et al., 2018). In this work, we aim to compare the user's route choice with the expected one with or without disturbances. Therefore, we evaluate the EJC, as the difference between the cost of the user's trip and the expected cost among the available alternatives (instead of only one). Two different EJCs (timetable and realized) are considered, as explained in detail in Section 5.5.3. This

allows evaluating the disturbance impact on a user with respect to both the current condition and a condition of no disturbances.

5.3.4 *Missing Aspects*

We believe several aspects are missing in current works, which we consider in this paper. First, most of the work focuses only on big disturbances in the network, with little attention being paid to the more common small disturbances, or even to "good disturbances" (as defined in Section 5.2). In addition, previous studies mainly focus on one or few specific disturbances. Therefore, it is not trivial to generalize findings based on different contexts. We contribute to this aspect by the proposed metric in section 5.5.2.

As no work provided a rigorous quantification of the level of disturbances, the conditions of no disturbances, small and big disturbances could not be numerically compared with each other. The typical approach of defining one or very few disruption scenarios does not allow analysis of general trends. Therefore, it is not known at the current state how different disturbances may affect the same public transport passenger. Moreover, no work investigated how users react to different disturbances in a public transport network, which allows more general conclusions, not tied to one or few disrupted scenarios.

The only works analysing how users react to disturbances, focusing on each single trip, are based on simulations, therefore there are no empirical studies. In particular, no work analysed a large-scale revealed preference survey, such as a tracking-based survey, focusing on route choice in case of different disturbances. The only identified work focusing on route choice is Yap et al. (2018), but they focused only on few disruptions (4), which moreover were planned, and lasted multiple days.

5.4 DATASETS AND TRAVEL DIARIES

This work aims to empirically observe public transport passengers' route choice in case of disturbances. Therefore, observations of both passengers and operations are needed. For this purpose, we exploit a unique combination of different datasets and methods, allowing, for each passengers' public transport trip, the observation of: the route choice with the vehicles taken; the network conditions; the available alternatives both with actual network conditions and without any disturbance. Regarding the operations, we used Automatic Vehicle Location (AVL) data to identify the network conditions (i.e. the disturbances) and the available alternatives. In particular, we used AVL data of the public transport services in Zürich, containing the planned and realized arrival and departure times of each public transport vehicle at each stop. From these data, we identified the level of service degradation for each trip, as described in Section 5.5.2.

Regarding the passengers, we collected GPS tracking data from a smartphone application, the ETH-IVT Travel Diary, able to continuously collect GPS data for a long period (Marra et al., 2019). We collected data of 172 Zürich residents for an average of 22 days per person. Using a mode detection algorithm, we first divided the GPS traces into activities and trips, then the trips into stages, identifying the transport modes used. Since the mode detection algorithm uses AVL data of the public transport operations, it is also able to detect the exact public transport vehicles used. In Marra et al. (2019), the algorithm obtained an average detection accuracy of 86% and it has been proven to identify a realistic number of activities and trips. We refer to Marra et al. (2019) for more details on the mode detection algorithm.

Instead, we refer to Marra and Corman (2020a) for more details on the tracking survey, where it has been already described and used to test a choice set generation algorithm and to estimate a route choice model. Here, we summarize the main characteristics and limitations of the survey. Table 5.2 provides general information on it. Regarding the representativeness of the participants, they are generally younger and with higher education than official reports. The gender distribution is slightly skewed towards men, while income and household size follow quite regularly the real distribution, despite a lower share of participants in the lowest income range. The resulting mode share is realistic and follows the official reports, especially for public transport, which is the focus of this paper. In fact, in this work, we focus only on public transport urban trips, i.e. trips located inside the city of Zürich, and done by a combination of public transport and walking. Among the public transport trips, 40% of them have at least one transfer. Moreover, the tram and the bus are much more used than the train (10% of the trips), given the more frequent service offered in the urban area.

Despite the number of participants can be considered limited to be representative of Zürich population, we do not consider the size of the dataset as a major limitation. In fact, most of the socio-demographic characteristics reflects the official reports, as well as the mode share. In addition, Marra

Survey	This Paper	Zürich 2016
Participants	172	\\
Age	<18: 0%; 18-24: 33%; 25-34: 27%; 35-44: 21%; 45-54: 18%; 54>: 1%	<18: 15%; 18-24: 8%; 25-34: 22%; 35-44: 18%; 45-54: 13%; 54>: 24%
Gender	Female: 43%; Male 57%	Female: 50% Male 50%
Education	Mandatory: 10%; Secondary: 36%; Higher: 54%	Mandatory: 18%; Secondary: 34%; Higher: 48%
Household income (Monthly CHF)	<4000: 8%; 4000-8000: 24%; 8000-12000: 29%; 12000-16000: 14%; >16000: 9%; no answer: 16%	<5000: 24%; 5000-7500: 24%; 7500-12500: 31%; 12500-16666: 12%; >16666: 9% no answer: 0%
Household Size	1: 23%; 2: 25%; 3: 22%; 4: 22%; 5+: 8%	1: 22%; 2: 30%; 3: 18%; 4: 18%; 5+: 12%
Mode share	Public transport: 38%; Walk: 23%; Car: 15%; Bike: 13%; Mixed: 10%	Public transport: 41%; Walk: 26%; Car: 25%; Bike: 8%; Mixed: not reported

TABLE 5.2: Information on the tracking survey (the table continues). Official statistics of socio-demographic information in 2016 are shown (Stadt Zürich, 2021; Zürich Statistic Office, 2021). Income and mode share refers to 2015. This Table extends Table 3 in Marra and Corman (2020a).

Survey	This Paper	Zürich 2016
Study Period	03/04 – 02/06/2019	\\
Tracked days	3785	\\
Public transport trips in Zürich	2901	\\
Number of stages per p.t. trip (%)	1: 60%; 2: 29%; 3: 9%; >3: 2%	\\
Avg. duration per p.t. trip	21.7 min	\\
Avg. air distance per p.t. trip	2.88 km	\\

TABLE 5.2: Continue of previous table. Information on the tracking survey.

and Corman (2020a) show that the dataset leads to meaningful parameters of an estimated route choice model, in line with the values in literature. Small heterogeneity was also found among the users on perception of different travel costs. Moreover, regarding the scope of this paper (the effects of disturbances on public transport passengers), the dataset is one of the largest in the literature, in terms of number of trips and disturbances, given that only few works are based on realized observations (see Table 5.1).

5.5 METHODOLOGY

5.5.1 *Groundwork*

In this Section, we briefly summarize two preliminary methods, described in Marra and Corman (2020a), on which our work builds: a choice set generation algorithm and a route choice model. Therefore, we refer to that work for more details and for a review on route choice model estimation (see also Van der Hurk, 2015). In particular, the state-of-the-art model for route choice in public transport is the Path Size Logit Model (Nielsen et al., 2021; Yap and Cats, 2021), which we describe in this Section.
To identify the available alternatives for each trip of each passenger, we used a choice set generation algorithm based on constrained enumeration, a well-established technique in literature, which generates all available alternatives, given reasonable constraints, such as maximum travel time. The

algorithm showed high performance both in terms of coverage (identification of passengers' trips, 94%), selection of relevant alternatives and model estimation. Each alternative is described as a combination of walks and public transport vehicles, including possible transfers. In this paper, to compute our novel metric of service degradation for a certain demand, we generate two different choice sets: one in case of no disturbance (based on the timetable), and one with real conditions (based on AVL data). In that way, for each demand we know the alternatives available in planned conditions and the ones available in reality.
The second method is the Path Size Logit Model, a route choice model, which has been estimated in Marra and Corman (2020a) on the same tracking dataset of this work. This model is a variant of the standard Multinomial Logit Model, which includes a correction factor (PathSize), penalizing overlapping alternatives (i.e. alternatives with one or more vehicles in common). In particular, several coefficients were estimated, describing the following utility function:

$$\begin{aligned} U_j = \; -C_j = \; & \beta_{tram} * tram\ time + \; \beta_{bus} * bus\ time + \beta_{train} * train\ time \\ & + \beta_{walk} * walk\ time + \beta_{tt} * transfer\ time \\ & + \beta_{transfer} * transfers + \beta_{PS} * PathSize \end{aligned} \tag{5.1}$$

Walktime refers to the first and last walk, while *transfertime* to the intermediate ones. The waiting time is included in these two parameters, since we could not discriminate precisely between walk and waiting time, from the GPS data. Reliability parameters, as average delay of the public transport line or standard deviation of the travel time, were not included in the utility function, since they did not improve the model performance.
In this work, the cost of each alternative is determined from the cost function C_j, estimated from a choice set assuming no disturbances in the network. Table 5.3 reports the model estimation (see Table 5 in Marra and Corman, 2020a, for more details).
Here, we introduce some terminology used in the following Sections. We refer with *realized* (or using a superscript R) to each alternative, choice set or travel cost, based on the real (disturbed) network conditions, i.e. considering the disturbances occurring at that moment. In contrast, we refer with *planned* (or using a superscript T, from timetable) to each alternative, choice set or travel cost, based on planned (not disturbed) conditions, i.e. the timetable. *Chosenalternative* refers to the observed, actual trip of

Parameter	Estimate	t-test
Tram travel time (s)	-1	-21.7
Bus travel time	-1.14	-20.2
Train travel time	-1.19	-12.4
Walking time	-2.56	-42.6
Transfer time	-1.06	-15.6
Number of transfers	-889	-27.4
Path Size	55.4	2.49
Observations	2719	
Null log-likelihood	-7361	
Final log-likelihood	-2555	
$\overline{R^2}$	0.65	
Scaling factor	0.0038	

TABLE 5.3: Path Size Logit Estimation. The parameters are scaled (multiplied by the scaling factor) to have the coefficient of travel time in tram equal to – 1. See Table 5 in Marra and Corman (2020a) for more details).

a user, as determined by the tracking. $Leastcostlyalternative$ refers to the alternative in a choice set with lowest cost C_j, which is also the most likely alternative to be chosen, according to the Path Size Logit model. Despite the terms *planned* and *realized* are already mentioned and opposed in literature (e.g. in Van der Hurk, 2015), we acknowledge there is not a common terminology and different terms, as $perfectlyreliable$ and $actual$, are used to express similar concepts.

5.5.2 *Metric of Service Degradation for Various Disturbances*

To evaluate the passengers' behaviour in case of different disturbances, it is important to define clearly which disturbances are considered, and how to summarize insights over them. In this sense, this work focuses on small disturbances, like delays or single cancelled runs, rather than big disturbances, such as long service interruptions. Moreover, the same disturbance can have a large impact on certain passengers, while no impact on others, depending on several factors, as the origin, the destination and the starting time of their trips. Therefore, a metric that quantifies the disturbances affecting a certain user is needed, in order to compare the users' behaviour in case of different levels of disturbances.

We propose a novel metric of service degradation, which quantifies the level of disturbances for a specific demand, defined by an origin, a destination and a departure time. Namely, we measure how the current state of the network, with delays and cancelled runs distributed in the network, can affect a passenger travelling from O to D and departing at time W. This metric assesses the disturbance, and it is not based on what a user actually did, but on the set of alternatives available. The metric is based on the definition of *disruption impact* in Marra and Corman (2020b) (used to assess disruptions' effects on hypothetical demand), relating it to the Path Size Logit Model.

More formally, the service degradation, described in Equation 5.2, is the expected excess travel cost with the current network condition, computed as the difference between the *expected realized travel cost* ($\overline{C^R}$, expected travel cost considering the disturbances) and the *expected planned travel cost* ($\overline{C^T}$, expected travel cost without any disturbance). Each expected travel cost is based on a route choice model, and computed as a weighted average among the available alternatives at that time for that specific demand.

$$degradation\,(O, D, W) = \overline{C^R} - \overline{C^T} \tag{5.2}$$

$$\overline{C^R} = \sum_{j \in CS^R} P(j|CS^R) C_j \tag{5.3}$$

$$\overline{C^T} = \sum_{j \in CS^T} P(j|CS^T) C_j \tag{5.4}$$

$$P\,(i|CS) = \frac{e^{-C_i}}{\sum_{j \in CS} e^{-C_j}} \tag{5.5}$$

$$CS^R\,(O, D, W) = realized\ choice\ set\ for\ the\ given\ demand \tag{5.6}$$

$$CS^T\,(O, D, W) = planned\ choice\ set\ for\ the\ given\ demand \tag{5.7}$$

The cost of an alternative (C_j, Equation 5.1) is the negation of the utility function of a Path Size Logit Model, estimated as in Table 5.3. The weight of an alternative is its probability to be chosen ($P(i|CS)$). The difference of the two expected travel costs (i.e. the service degradation) gives a measure of the expected additional cost afforded in real conditions. Only the sub-network relevant to the OD is considered, since only the vehicles present in the choice sets are considered. In addition, disturbances on the most

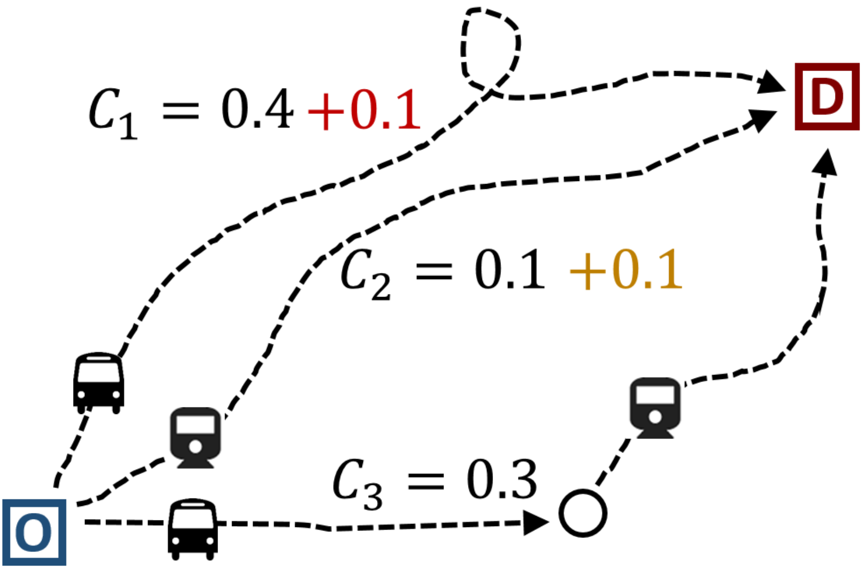


FIGURE 5.1: Example of a demand with origin, destination, available alternatives and relative costs. Possible increases of costs due to disturbances are highlighted in red and gold.

likely alternatives are weighted more.
In Figure 5.1, we report a small example, to explain better the service degradation metric. We assume the passenger has three available alternatives, with the costs shown in black (no unit measures in the example). Without any disturbance, the expected planned travel cost ($\overline{C^T}$) is 0.25. If due to a disturbance (e.g. a delay of a bus) the cost of the first alternative (C_1) increases by 0.1, the expected realized travel cost ($\overline{C^R}$) is 0.27, resulting in a service degradation of 0.02. Instead, if the same increase of cost (0.1) occurs in the second alternative, the expected realized travel cost becomes 0.29, and the service degradation 0.04, which is higher than the previous one. With this example, we want to show that the service degradation takes into account all the alternatives, but it focuses more on the less costly ones, since there is a higher probability to choose them. This is in accordance with the principle that a delay in an inconvenient alternative has a marginal impact, compared to a delay in the main alternative.
We want to highlight that the service degradation can also be negative, i.e. the current network conditions result in less costly available alternatives,

reducing the expected realized travel cost. This is the case of an early arrival/departure, or new available connections. Considering the same example in Figure 5.1, if due to an early departure, the cost of the first alternative decreases by 0.1, the expected realized travel cost becomes 0.22, resulting in a negative service degradation of -0.03 (i.e. a "good disturbance"). Summarizing, the defined service degradation metric results in a novel quantitative measure of the expected impact of any disturbance, and has the following advantages:

- It represents the expected increase of travel cost in the current network conditions for a given demand.
- It is based on the well-established random utility theory to determine the cost of each alternative. Therefore, a disturbance on a less costly alternative (with higher probability to be chosen) has a larger effect.
- It describes the level of disturbances from the point of view of the passengers, considering the effects on the available alternatives; and not from the operational point of view, such as in terms of vehicle delay or number of vehicles involved.
- It allows to measure both small and big disturbances, and to easily compare them. "Good disturbances", potentially reducing the passengers' travel cost, are also identified.

5.5.3 *Excess Journey Cost and Effects of Network Disturbances*

In this Section, we show the methodology to evaluate the effects of network disturbances on public transport passengers. In this regard, we quantify the level of the disturbances affecting a passenger, as the service degradation during that specific trip (O, D and W are respectively the origin, the destination and the departure time of the passenger). The hypothesis is that the service degradation influences both the chosen alternative and its travel cost (C_{trip}, as defined in Equation 5.1). Most of the work in literature (see Table 5.1) evaluates the impact of the disturbances on passengers in terms of excess journey cost (EJC) or excess journey time (EJT). Nevertheless, they consider only one alternative (e.g. the shortest path) as reference alternative to compare the one chosen by the user. Instead, in our study we evaluate the EJC, considering the set of available alternatives as set of reference alternatives. In particular, we define two different versions of EJC:

$$EJC^R\,(trip) = \; C_{trip} - \overline{C^R} \tag{5.8}$$

$$EJC^T\,(trip) = \; C_{trip} - \overline{C^T} \tag{5.9}$$

EJC^R is the difference between the cost of the chosen alternative by the user (C_{trip}) and the expected realized travel cost. While EJC^T is the difference between the cost of the chosen alternative by the user and the expected planned travel cost. In that way, we can observe how much the cost of the chosen alternative deviates from the expected one without disturbances ($\overline{C^T}$) or the expected one in case of disturbances and assuming full information on them ($\overline{C^R}$). The expected costs are calculated as the average cost of the available alternatives, weighted by the probability to choose them, as defined in the Equations 3 and 4. The available alternatives are respectively the ones in the choice set CS^R, if real conditions are considered, and the ones in CS^T, if planned conditions are considered. As a technical remark, we assume the cost of the chosen alternative (C_{trip}) being the same of the corresponding alternative in CS^R. This allows to remove the noise due to different walking patterns and speed between the user and the choice set generation algorithm, without changing the overall results (see discussion on walking distance in Marra and Corman, 2020a, for further details).

In the two EJC metrics, we consider as cost of the chosen alternative (C_{trip}) the actual cost (including possible delays), but a different reference cost ($\overline{C^T}$) or ($\overline{C^R}$). Therefore, the EJC^R can be also seen as the error of the route choice model (difference between the cost of the chosen trip and the expected cost assuming full information on the disturbances). In fact, we want to observe when the choice of a user deviates from the expected one. In contrast, the same interpretation is not true for the EJC^T, since it compares an actual cost (of the user) with planned costs. Therefore, it represents the increase in cost compared to that under planned conditions.[1]

Referring to the example in Figure 5.1, we can assume a large delay in the alternative 2, increasing its cost by 0.5 ($C_2 = 0.1 + 0.5$). Without showing the full calculations, it results $\overline{C^R} = 0.39$ and $\overline{C^T} = 0.25$. The service degradation is therefore 0.14. The most likely choice without disturbances would have been the alternative 2 ($C_2 = 0.1$). Instead, we can assume the user reacted to the delay by choosing the alternative 3 ($C_{trip} = C_3 = 0.3$). Therefore, $EJC^T = 0.05$, i.e. the user incurred in an increased cost, compared to a condition of no disturbance (where alternative 2 had lower

1 this and the previous paragraph have been improved compared to the submitted version.

cost). In contrast, $EJC^R = -0.09$, i.e. the user incurred in a journey cost lower than expected in disturbed conditions (in fact, the user chose the alternative with the lowest cost, given the large delay in the alternative 2). This represents the typical case of a user reacting to network disturbances, trying to reduce the journey cost, but still experiencing an increased cost compared to planned conditions ($EJC^T > 0$).
A similar example can be made for "good disturbances" (i.e. negative service degradation). We assume the cost of the alternative 2 reduced by 0.08 ($C_2 = 0.1 - 0.08$), due to for instance an early departure/transfer. It results, $\overline{C^R} = 0.21$, $\overline{C^T} = 0.25$ and the service $degradation = -0.04$. The most likely choice with or without disturbances is the alternative 2. We can assume if the user is aware of the early departure, the user takes alternative 2, with a lower cost than planned. Instead, if the user is not aware, we assume the user did not arrive in time at the stop to take the alternative 2 (even if it was possible). Therefore, the user chooses the alternative 3 ($C_{trip} = C_3 = 0.3$). In this case, $EJC^T = 0.05$, i.e. the user incurred in an increased cost, compared to a condition of no disturbance (where alternative 2 had lower cost). $EJC^R = 0.09$, i.e. the user incurred in a higher cost than expected in disturbed conditions, where alternative 2 is by far the best alternative. This represents the case of a good disturbance, which could potentially reduce the user's travel cost, but rather it increases it, since the user is unaware of it. We remark we proposed two simple examples to clarify the two EJC metrics; however, other examples are possible, describing different responses of passengers to disturbances.
In this work, we analyse how the two EJC vary, according to different values of service degradation, to evaluate how the travel cost varies in case of disturbances. We also compare how the hit rate of a route choice model varies according to different service degradations. In particular, we selected the most likely alternative among the ones available without any disturbance (CS^T), according to the estimated Path Size Logit model. In that way, we observe when the passengers' route choice adheres with the most likely one (in case of no disturbances) in terms of lines taken; and when it is less predictable, showing a reaction of the users to the current network conditions.
As shown in Section 5.3, the works on passengers' behaviour in case of disturbances are often based on behavioural assumptions, such as the choice of the shortest path or the usage of a simulation tool, given that the realized and planned trips are not always observable. In this work, the realized trip is observed through tracking; hence, there are no assumptions on it. The

only source of error is the accuracy of the mode detection algorithm (86%), which in any case currently cannot be overcome by alternative methods (AFC estimation can also lead to errors). In contrast, we assume the cost of the planned trip, as the expected one based on the Path Size Logit ($\overline{C^T}$ and $\overline{C^R}$). Given that the planned/intentioned trip is hardly identifiable, we believe our assumption (based on a choice set and a route choice model) is more realistic than the ones proposed in literature, based on a single shortest path or on a simulation. Finally, we remark that we choose the same cost function (i.e. values of the parameters) for all the users, to have a robust parameters' set, which can describe well the average behaviour (as done by Yap and Cats, 2020). In addition, in Marra and Corman (2020a) no significant heterogeneity was identified on cost perception among the users, except for the in-train travel time (the Mixed Path Size Logit did not perform particularly better than the Path Size Logit).

5.6 RESULTS

5.6.1 *Users' Choices in Case of Disturbances*

In this Section, we observe how the costs of the routes chosen by the users vary according to different network disturbances. We refer to network disturbances, as defined in Section 5.5.3, namely the level of service degradation in the sub-network of interest for a given OD.

Figure 5.2 shows the distribution of the service degradation among the collected public transport trips in Zurich (2901). The travel costs, and therefore the service degradation too, are scaled in terms of travel time in-tram ($C_J \rightarrow C_J/|\beta_{tram}|$ in Equations 3 and 4). Hence, the service degradation can be interpreted as the expected increase of travel time in tram (in seconds) for that trip with the current network conditions. We remark that a degradation of 600 seconds does not correspond directly to a delay of 10 minutes, but rather some of the available alternatives might be strongly delayed (or even cancelled), while others might be on time, and thus might partially compensate for the disturbances of the delayed alternatives. In this sense, even a small service degradation may strongly affect a passenger, if the chosen route is the only one delayed among all the alternatives. Most of the trips have no service degradation or low positive values, showing that the service runs on time most of the time. Nevertheless, 22% of the trips have a degradation greater than 240 (equivalent to a 4 minutes delay in tram if only one alternative is available), while 20% have a neg-

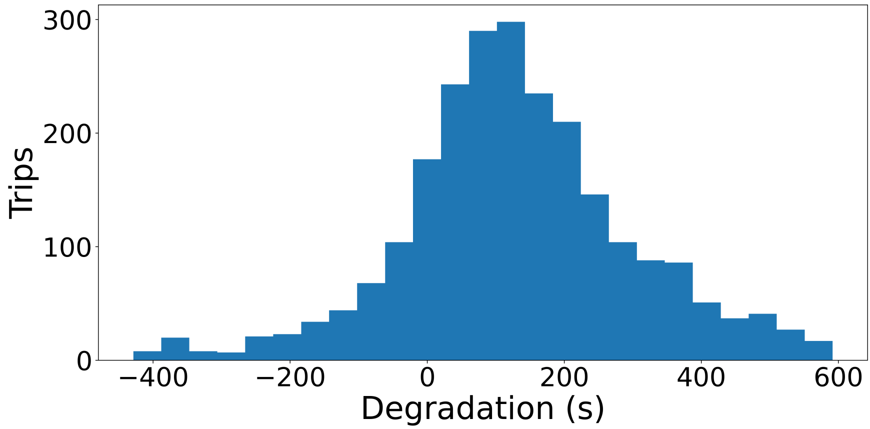


FIGURE 5.2: Distribution of service degradation among the public transport trips in the dataset. Outliers below the 4th percentile or above the 96th percentile are not shown.

ative degradation (i.e. "good disturbances", when the network conditions allow less costly solutions than without any disturbance). This distribution shows the novelty aspect of this work, which is the focus on many small disturbances and "good disturbances".

In Figure 5.3 , we compare each passengers' choice with the available realized alternatives (i.e. with disturbances), and the available planned alternatives. Namely, we compare each passengers' choice with two benchmarks, and we observe how the excess journey costs (EJC^R and EJC^T , as defined in Section 5.5.3) vary according to the service degradation (x-axis and grey line). Therefore, the y-axis represents the extra cost that the user incurred, compared to the expected one in either the planned conditions (EJC^T , blue line) or the realized conditions (EJC^R, orange line). The travel costs are scaled in terms of travel time in-tram, therefore the EJC can be interpreted as the increase of travel time in tram (in seconds) the passenger incurred. We highlight that the service degradation (grey line) represents the expected increase of travel cost with the current network conditions, i.e. how the blue line should look like, if users choose as expected with current conditions ($C_{trip} = \overline{C^T}$, namely if users follow the Path Size Logit). We report the moving median of the EJC, since we aim to observe the trend and the average behaviour. In fact, with the same service degradation, two different trips in the dataset can have different EJC, since they

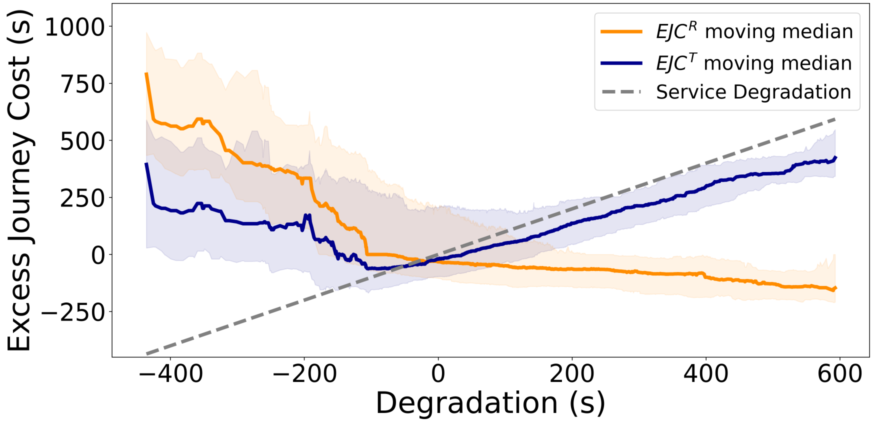


FIGURE 5.3: Relationship between service degradation and excess journey cost (timetable and realized). Moving median and range between 25 and 75 percentiles are shown (window of 120 seconds).

might have different origin, destination and network conditions (but same service degradation). The confidence intervals (25 and 75 percentiles) are also shown.

Looking at positive values of service degradation, $EJC^R < EJC^T$, since, by definition of service degradation, the realized travel cost ($\overline{C^R}$) is higher than the timetable one ($\overline{C^T}$), showing that disturbed operations result in higher travel costs. The EJC^T increases roughly linearly with the service degradation. Therefore, higher degradation implies more costs to passengers (compared to planned conditions). In contrast, the EJC^R decreases with the service degradation. Therefore, with higher disturbances in the network, passengers counteract the disturbances and choose a less costly available alternative, i.e. a path that compensates or avoids the degradation in fact.

Looking at negative values of service degradation, $EJC^R > EJC^T$, since, by definition of service degradation, the realized travel cost ($\overline{C^R}$) is lower than the timetable one ($\overline{C^T}$), showing that the current network conditions result in less costly alternatives. The confidence intervals of EJC^T and EJC^R (≈ 400 seconds) are larger than the ones for positive degradation (≈ 200 seconds), showing a higher variability. Looking at the median, the EJC^T is almost constant for a service degradation near 0, while it increases remarkably for values < -105, and up to a maximum where data are unavailable.

This shows that even with "good disturbances", the passengers' travel costs are higher than without disturbances. This means that the new network conditions are not exploited by the users, but instead they hinder them. In addition, the sharp increase of the EJC^R suggests that the new available alternatives (with lower cost) are rarely chosen by the users (see Section 5.6.2 for more details).

5.6.1.1 *Available information*

One remarkable advantage of realized data, compared to simulations, is that there is no need to make assumptions on the available information to passengers regarding the disturbances, since the chosen route is observed. Therefore, the effects of different disturbances can be evaluated without knowing if the user is aware of them or not. Nevertheless, the analysis of the EJC^R can give some insights on the information available to passengers. In particular, for negative values of service degradation, the EJC^R increases, showing that the users do not exploit the new available alternatives and/or have no information about them. In contrast, for positive values of degradation, the EJC^R decreases, showing that on average users choose less costly alternatives. This suggests a greater need of information in case of "good disturbances" compared to "bad disturbances", since the new less costly alternatives are less likely chosen. In this sense, we believe this analysis can be a starting point for future studies on information availability, in which the available information is not just assumed in a behavioural model, but instead it is inferred from realized observations.

5.6.1.2 *Summary of the findings*

We can summarize the key findings of this analysis, as follows: the effects of disturbances on travel costs are highly variable (especially those of "good disturbances"), and they may affect different passengers' trips in different ways. Nevertheless, a general trend of the excess journey cost with respect to the service degradation can be clearly observed. For positive values of degradation, the users experience an increased cost compared to normal conditions, nevertheless the gap between the user's travel cost and the least costly available alternatives reduces (lower EJC^R). Therefore, there is no particular need of information for users, since they already choose less costly alternatives. Instead, the operations' delay should be reduced, since users still experience a delay compared to normal conditions (higher EJC^T). In contrast, for negative values of degradation, the operations provide less costly alternatives than in planned conditions, but the users still experience an increased cost (higher EJC^T) and do not exploit the new available alternatives (higher EJC^R). Therefore, in this case there is need to inform better the users.

5.6.1.3 *Clarifications and limitation*

We remark that the proposed analysis lays on the assumption that the planned travel cost is based on a Path Size Logit model, describing the average behaviour for all the trips in the dataset. Therefore, it must not be seen as an analysis on when users perform better, but on when and how travel behaviour deviates from the expected one. Moreover, given that we are focusing on small disturbances (in the order of minutes), the assumption of similar user behaviour for different service degradations is more realistic, compared to the case of big service disturbances. In addition, our reference costs ($\overline{C^R}$ and $\overline{C^T}$) are more realistic and have more theoretical foundations than those used in literature (e.g. shortest travel time or simulation), since ours are based on random utility theory and a set of available alternatives.
We highlight that what we observed is the relationship between the excess journey cost and the service degradation. To better consolidate the dependencies identified between the two variables, and exclude possible external effects, we examined the correlation between the service degradation and several variables describing the trip. For brevity, we do not report in detail the relationships between all the analysed variables, since we did not identify any significant correlation between the service degradation and the following ones: average number of transfers, average walking distance, trip length, betweenness and closeness centrality of origin and destination (i.e. location in the network), service frequency and time of the day (despite a slightly higher degradation is found during the morning peak and in the evening). The service degradation is only slightly negatively correlated (-0.12) with the expected planned travel cost ($\overline{C^T}$). Namely, trips with higher cost tend to have a lower service degradation, while trips with lower cost a higher service degradation. This occurs because trips with higher cost (i.e. longer trips or with multiple transfers) tend to have more available alternatives, with similar costs, and therefore they are more resistant to disturbances.
As alternative approach to the one proposed, we included in the Path Size Logit two additional parameters, describing the delay of the first vehicle of each alternative (one if the delay is positive and the other if the delay is

negative, i.e. there is an early departure). The parameters were not found significant, showing the delay does not contribute linearly to the utility function (probably the user is not aware of it). This also shows that extending a Logit model with these parameters related to disturbances does not contribute to the understanding of route choice in case of disturbances. In contrast, from our approach, we observe that different service degradations affect differently the route choice of passengers, and with a certain variability, as shown in Figure 5.3.
Finally, we remark that we did not discriminate among the users in this analysis, to observe possible heterogeneity. Despite each user was tracked for a large number of days (22 on average), given that we are considering a continuous range of service degradation, it was not possible to collect enough trips for each user for the different values of service degradation (especially the negative ones, occurring in only 20% of cases). Therefore, the analysis of heterogeneity among users is left for a future work.

5.6.2 *Deviation from the Most Likely Route Choice*

In this Section, we analyse if and how the alternative chosen by each user in case of disturbances (i.e. observed in the tracking) is different from the most likely planned alternative, i.e. the alternative with higher probability to be chosen in case of no disturbances. In this way, we can see if the users choose differently in case of disturbances. A similar objective is analysed by Yap et al. (2018), which compared the prediction accuracy of two models, one trained in case of disruptions, and the other with regular service. Differently than them, in addition to showing that the prediction accuracy can be different in case of disturbances, we identify how the accuracy vary for different values of service degradation.
We define as hit rate, the percentage of times the Path Size Logit identifies as the most likely alternative the one chosen by the user. We consider the hit rate in terms of lines taken. Therefore, if the user does not take the first available vehicle of a certain line (e.g. because the run is cancelled), but waits for the next one, the prediction is considered correct. Figure 5.4 shows the hit rate obtained by the Path Size Logit, considering the planned choice set (CS^T, i.e. alternatives available in case of no disturbance), according to different values of service degradation. The hit rate (reported as a covered alternative in Figure, in blue) is above 70% for positive values of degradation (average 72%). This shows that in case of "bad disturbances", both with low and high degradation, the users choose more than 70% of

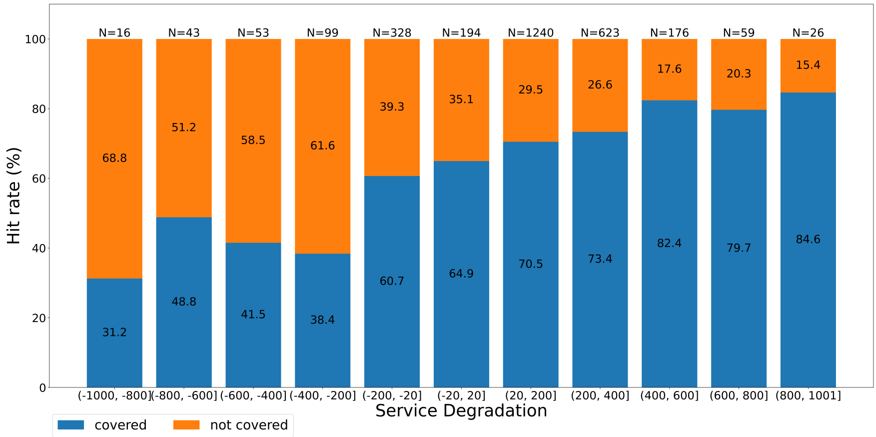


FIGURE 5.4: Hit-rate of Path Size Logit based on planned choice set (planned network conditions), according to different values of service degradation. N is the number of trips in that range of service degradation.

times the most likely alternative according to the timetable.
Instead, for negative values of service degradation the hit rate strongly decreases (average 54%). This shows that in case of "good disturbances" the users do not choose (or cannot choose) the least costly planned alternative, but rather they choose alternatives based on different lines. In other terms, they choose less likely the expected route.
Table 5.4 describes in more details the cases with negative degradation. In 84% of cases (first two rows), the degradation is negative due to the most likely realized alternative, which costs less than the most likely planned alternative. In contrast, in 16% of cases (last two rows) the degradation is negative due to other less likely alternatives. We focus on the mentioned 84%, given that there is a direct relationship between the negative degradation and the most likely alternative, which is easier to analyse. In particular, in 30% of cases the least costly alternatives in the planned and realized conditions are the same (i.e. in case of disturbances there is an earlier arrival). When the least costly alternative is the same with and without disturbances, considering all trips in the dataset (both delays and earlier arrivals), the hit rate is generally high (78%). Instead, considering only earlier arrival cases, the hit rate decreases to 70% (second row of Table 5.4). Therefore, earlier departures/arrivals affect passengers' chosen route. The

Case	Occurrence percentage	Hit rate
The least costly alternative with and without disturbances is different. The alternative in case of disturbances costs less.	54%	40% planned 25% realized
The least costly alternative with and without disturbances is the same. One of the vehicles arrives earlier than planned.	30%	70% (both)
The least costly alternative with and without disturbances is the same. No vehicle departs earlier, i.e. another available alternative costs less.	12%	67% (both)
The least costly alternative with and without disturbances is different. The alternative in case of disturbances costs more, i.e. other available alternatives cost less.	4%	65% planned 17% realized

Table 5.4: Occurrences of different cases leading to negative service degradation. The hit rate for each case is shown.

effect is equivalent to 10.3% ((78-70)/78) of passengers missing the planned vehicle, due to an earlier departure of that vehicle.
In 54% of cases with negative degradation (first row), there is a new available alternative, with a lower cost than the least costly planned alternative. In this case, 40% of users choose the planned alternative, while only 25% the realized one. Therefore, the users less likely choose and exploit the new available alternative, despite its lower cost (similar conclusions are derived in Section 5.6.1).

5.7 CONCLUSIONS

In this work, we observed passengers' behaviour in public transport, with a focus on route choice in case of disturbances. We identified as the main aspects missing in literature a focus on small disturbances and how different disturbances affect the passengers. In this sense, we defined a metric

of service degradation to quantify the network disturbances affecting each specific trip. This is the first metric defined in literature for quantification and comparison of various disturbances, with a focus on passengers, rather than the operations (as the vehicle delay). In particular, this metric allows the comparison of several network conditions, such as no disturbance, small disturbances, big disturbances, and "good disturbances".
An additional novel aspect of this paper is the use of realized observations of passengers and the focus on each individual trip. In this regard, we analysed a GPS-based tracking dataset, which allows observing the actual passengers' choices under different network disturbances.
We study the effects of disturbances on passengers based on the excess journey cost, derived from the chosen route and two different choice sets. In particular, we compared the passengers' chosen route with the available alternatives in case of disturbances and in case of planned conditions.
We identified that the effects of disturbances on travel costs are highly variable among different trips. Despite this, a general trend can be observed. In particular, "bad disturbances" negatively affect passengers, in terms of increased travel cost, despite they choose more likely the least costly available alternative in current network conditions. Instead, "good disturbances", leading to less costly available alternatives compared to no disturbance, also affect negatively the passengers, since in this case passengers do not exploit the less costly alternatives, and their travel cost is even higher than the expected one without the disturbances. This leads to the following suggestions to reduce the overall passengers' travel cost: in case of "bad disturbances", the operations' services must be improved; while, in case of "good disturbances", the passengers' must be informed about the new available alternatives. In particular, regarding "good disturbances", besides not being aware of the new available alternatives, passengers are also affected by vehicles departing earlier, which are chosen 10% of the time less, compared to on-time vehicles. Therefore, "good disturbances" affect users as much as bad disturbances, hence, service operators must increase the attention on them.
We remark that this is the first work observing directly the public transport route choice of different passengers in case of many small disturbances. Despite this, we acknowledge the following limitations of our study. The real intention of the users for each trip is unknown; therefore, we compared each users' choice with the expected choice defined by a behavioural model, which can best represent the average behaviour of the users in the dataset. For this reason, the results of this work should be interpreted as

observing how the users' route choice deviates from the expected one, in case of different public transport disturbances. In this regard, for future works, we plan to investigate other types of expected behaviour in addition to the two considered (without disturbances and with full information on disturbances), such as assuming current information on disturbances at the begin of the trip. Moreover, several external sources can affect the users' behaviour, such as the weather, which we did not take into account. Further investigations in this direction are left for future works. In addition to that, we plan to exploit longitudinal data to identify patterns and regularities from users' trips, and possible anomalies in case of disturbances[1].
Finally, we want to add some considerations that we learned from this study, about the analysis of disturbances from tracking data. Although tracking data are suitable to analyse the passengers' behaviour in daily life and in case of small disturbances, we believe the data collection still has some problems in case of big unplanned disturbances. In fact, since they are rare events, the study duration should be particularly long, and it is not guaranteed that a big disturbance will occur in that period. In addition, the sample must be large and heterogeneous enough to identify a large amount of users affected by the disturbance. In fact, even in case of a big disturbance, it is not guaranteed that the tracked users will be affected by it. The largest disturbance found in our dataset has a service degradation of 2093 seconds, experienced by a user traveling from and to a peripheral zone of the city. The only direct bus connection was delayed by 10 minutes and in addition, its run was cancelled in the middle of the trip, therefore the user had to wait 5 minutes more for the next vehicle of the same line (approximately 15 minutes delay and an additional transfer correspond to an increased cost of 2093 seconds, according to Equation 5.1). Despite this disturbance strongly affected a user, it did not affect the vast majority of travellers in Zürich, given its peripheral position.

DECLARATIONS

Funding: Not applicable.
Conflicts of interest/Competing interests: On behalf of all authors, the corresponding author states that there is no conflict of interest.
Ethics approval: approved by ETH Zurich Ethics Commission.
Consent to participate: Informed consent was obtained from all individual participants included in the study.

1 this paragraph has been improved compared to the submitted version.

ACKNOWLEDGEMENTS

The authors wish to thank all study participants.

5.8 APPENDIX: ALTERNATIVE METRIC OF SERVICE DEGRADATION

A possible metric of service degradation, alternative to the one proposed in Section 5.5.2, is the change in consumer surplus, named also welfare changes or logsum difference. This metric has been applied in literature as an evaluation measure for projects, comparing the conditions before and after a change (De Jong et al., 2007). The change in consumer surplus is defined as the difference between the expected maximum utility in a certain condition (e.g. during disturbances) and the expected maximum utility in another condition (e.g. without disturbances). In our test case, the metric can be calculated as follows (see De Jong et al., 2007; Zhao et al., 2012, for the formal derivation):

$$\Delta E = \frac{1}{\beta_{tram}}(ln(\sum_{j \epsilon CS^R} e^{U_j}) - ln(\sum_{j \epsilon CS^T} e^{U_j})) \qquad (5.10)$$

CS^R is the realized choice set and CS^T is the planned choice set, as defined in Section 5.5.2. The metric is typically divided by the marginal utility of income (in our case by β_{tram}, the coefficient of the travel time in tram).

The change in consumer surplus and our metric of service degradation have the following theoretical difference. The former is formally derived from the Multinomial Logit model, taking into account the error term in the utility. Our metric, instead, takes inspiration from practical traffic assignment and simulation models, which have to deterministically distribute the population among the available routes (according to the probability to choose them).

In practice, in our test case the two metrics are highly correlated ($\rho = 0.94$), and using the change in consumer surplus instead of the proposed metric of service degradation does not change significantly the results. Therefore, the same conclusions drawn in this work can be drawn using the change in consumer surplus to evaluate the service degradation.

References

Antos, J. and M. D. Eichler (2016). Tapping into Delay: Assessing Rail Transit Passenger Delay with Data from a Tap-In, Tap-Out Fare System. *Transportation Research Record* 2540.1, 76.

Carrion, C. and D. Levinson (2012). Value of travel time reliability: A review of current evidence. *Transportation Research Part A: Policy and Practice* 46.4, 720.

Cats, O. and E. Jenelius (2014). Dynamic Vulnerability Analysis of Public Transport Networks: Mitigation Effects of Real-Time Information. *Networks and Spatial Economics* 14.3, 435.

Currie, G. and C. Muir (2017). Understanding Passenger Perceptions and Behaviors During Unplanned Rail Disruptions. *Transportation Research Procedia*. World Conference on Transport Research - WCTR 2016 Shanghai. 10-15 July 2016 25, 4392.

De Jong, G., A. Daly, M. Pieters, and T. van der Hoorn (2007). The logsum as an evaluation measure: Review of the literature and new results. *Transportation Research Part A: Policy and Practice*. Selected Papers on Applications of Discrete Choice Models Presented at the European Regional Science Conference, Amsterdam, August 2005 41.9, 874.

De Moraes Ramos, G., T. Mai, W. Daamen, E. Frejinger, and S. P. Hoogendoorn (2020). Route choice behaviour and travel information in a congested network: Static and dynamic recursive models. *Transportation Research Part C: Emerging Technologies* 114, 681.

Ghaemi, N., A. A. Zilko, F. Yan, O. Cats, D. Kurowicka, and R. M. P. Goverde (2018). Impact of railway disruption predictions and rescheduling on passenger delays. *Journal of Rail Transport Planning & Management* 8.2, 103.

Harrison, G., S. M. Grant-Muller, and F. C. Hodgson (2020). New and emerging data forms in transportation planning and policy: Opportunities and challenges for "Track and Trace" data. *Transportation Research Part C: Emerging Technologies* 117, 102672.

Krishnakumari, P., O. Cats, and H. van Lint (2020). Estimation of metro network passenger delay from individual trajectories. *Transportation Research Part C: Emerging Technologies* 117, 102704.

Leng, N. (2020). The role of information to passengers in public transport disruptions. Doctoral Thesis. ETH Zurich.

Leng, N. and F. Corman (2020). The role of information availability to passengers in public transport disruptions: An agent-based simulation approach. *Transportation Research Part A: Policy and Practice* 133, 214.

Lin, T., A. Shalaby, and E. Miller (2016). Transit User Behaviour in Response to Service Disruption: State of Knowledge. *Canadian Transportation Research Forum 51st Annual Conference - North American Transport Challenges in an Era of Change*.

Lin, T., S. Srikukenthiran, E. Miller, and A. Shalaby (2018). Subway User Behaviour When Affected by Incidents in Toronto (SUBWAIT) Survey. *Canadian Journal of Civil Engineering* 45.

Lu, Q.-C. (2018). Modeling network resilience of rail transit under operational incidents. *Transportation Research Part A: Policy and Practice* 117, 227.

Marra, A. D., H. Becker, K. W. Axhausen, and F. Corman (2019). Developing a passive GPS tracking system to study long-term travel behavior. *Transportation Research Part C: Emerging Technologies* 104, 348.

Marra, A. D. and F. Corman (2020a). Determining an efficient and precise choice set for public transport based on tracking data. *Transportation Research Part A: Policy and Practice* 142, 168.

Marra, A. D. and F. Corman (2020b). From Delay to Disruption: Impact of Service Degradation on Public Transport Networks. *Transportation Research Record* 2674.10, 886.

Nielsen, O. A., M. Eltved, M. K. Anderson, and C. G. Prato (2021). Relevance of detailed transfer attributes in large-scale multimodal route choice models for metropolitan public transport passengers. *Transportation Research Part A: Policy and Practice* 147, 76.

Paipuri, M., Y. Xu, M. C. González, and L. Leclercq (2020). Estimating MFDs, trip lengths and path flow distributions in a multi-region setting using mobile phone data. *Transportation Research Part C: Emerging Technologies* 118, 102709.

Paulsen, M., T. K. Rasmussen, and O. A. Nielsen (2018). Modelling Railway-Induced Passenger Delays in Multi-Modal Public Transport Networks: An Agent-Based Copenhagen Case Study Using Empirical Train Delay

Data. *14th Conference on Advanced Systems in Public Transport and Transit-Data 2018*.

Rahimi, E., A. Shamshiripour, R. Shabanpour, A. (Mohammadian, and J. Auld (2020). Analysis of Transit Users' Response Behavior in Case of Unplanned Service Disruptions. *Transportation Research Record* 2674.3, 258.

Shires, J. D., M. Ojeda-Cabral, and M. Wardman (2019). The impact of planned disruptions on rail passenger demand. *Transportation* 46.5, 1807.

Stadt Zürich (2021). Mikrozensus Mobilität und Verkehr 2015 - `www.stadt-zuerich.ch/ted/de/index/taz/verkehr/webartikel/webartikel_kennzahlen_verkehrsentwicklung.html`.

Sun, H., J. Wu, L. Wu, X. Yan, and Z. Gao (2016). Estimating the influence of common disruptions on urban rail transit networks. *Transportation Research Part A: Policy and Practice* 94, 62.

Van der Hurk, E. (2015). Passengers, Information, and Disruptions (No. EPS-2015-345-LIS). PhD Thesis. ERIM Ph.D. Series Research in Management. Erasmus Research Institute of Management.

Yap, M., S. Nijënstein, and N. van Oort (2018). Improving predictions of public transport usage during disturbances based on smart card data. *Transport Policy* 61, 84.

Yap, M. and O. Cats (2020). Predicting disruptions and their passenger delay impacts for public transport stops. *Transportation*.

Yap, M. and O. Cats (2021). Taking the path less travelled: Valuation of denied boarding in crowded public transport systems. *Transportation Research Part A: Policy and Practice* 147, 1.

Zhao, Y., K. Kockelman, and A. Karlström (2012). Welfare calculations in discrete choice settings: An exploratory analysis of error term correlation with finite populations. *Transport Policy* 19.1, 76.

Zhu, S. and D. M. Levinson (2012). Disruptions to Transportation Networks: A Review. *Network Reliability in Practice*. Transportation Research, Economics and Policy, 5.

Zhu, Y. and R. M. P. Goverde (2019). Dynamic Passenger Assignment for Major Railway Disruptions Considering Information Interventions. *Networks and Spatial Economics* 19.4, 1249.

Zürich Statistic Office (2021). Statistik Stadt Zürich - www.stadt-zuerich.ch/prd/de/index/statistik.html.

6

PASSENGERS' BEHAVIOUR DURING THE COVID-19 PANDEMIC

This chapter is based on the following article:

The impact of COVID-19 pandemic on public transport usage and route choice: Evidences from a long-term tracking study in urban area

Alessio D. Marra[1] & Linghang Sun[1] & Francesco Corman[1]
[1]Institute for Transport Planning and Systems, ETH Zürich

Submitted to Transport Policy on 05.2021 (under review)

Contributions

A. D. Marra: Conceptualization; Methodology; Data curation; Formal analysis; Visualization; Writing - original draft
L. Sun: Formal analysis; Visualization
F. Corman: Conceptualization; Writing - original draft

Key findings

- Two travel surveys based on GPS tracking are conducted in 2019 and 2020 to observe the effects of the COVID-19 pandemic on travel behaviour
- Analysis of travel distance, mode share, visited locations and recurrent trips
- Variations in route choice criteria are observed estimating two route choice models

Additional notes to this chapter

In this Chapter (Section 6.5.1), we estimate the same route choice model described in Section 3.5, but with data collected during the COVID-19 pandemic.

6.1 ABSTRACT

The COVID-19 pandemic strongly affected mobility around the world. Public transport was particularly hindered, since people may perceive it as unsafe and decide to avoid it. Moreover, in Switzerland, several restrictions were applied at the beginning of the first pandemic wave (16/03/2020), to reduce the contagion. This study observes how the pandemic affected travel behaviour of public transport users, focusing on route choice and recurrent trips. We conducted a travel survey based on GPS tracking during the first pandemic wave, following 48 users for more than 4 months. The very same users were also tracked in spring 2019, allowing a precise comparison of travel behaviour before and during the pandemic. We analyse how the pandemic affected users, in terms of travel distance, mode share and location during the day. We specifically focus on recurrent trips, commuting and non-commuting, observing how mode and route changed between the two different periods. Finally, we estimate a route choice model for public transport (Mixed Path Size Logit), based on trips during the two different years, to identify how the route choice criteria changed during the pandemic. Major differences were identified in the perception of transfers and of travel time in train.
Keywords: COVID-19; Public Transport; Tracking; Route Choice; User Behaviour

6.2 INTRODUCTION

The outbreak of the COVID-19 pandemic dramatically affected the world's population in early 2020. Mobility was particularly affected, since several governments imposed restrictions and disincentives to it, as lockdowns, remote working and closure of shops. Moreover, people tried to reduce their movements and social contacts, to reduce the risk of a contagion. Public transport suffered particularly from the pandemic, since passengers may perceive the system as unsafe and a possible source of infection (Aloi et al., 2020).

This work aims to understand how the COVID-19 pandemic affected travel behaviour of public transport passengers in Zürich, Switzerland. We focus on the effects of the pandemic in general, observing travellers in that period. Therefore, we do not isolate the effects of specific aspects, such as imposed restrictions, the progress of the infection, or passengers' perception of safety and risk. In this study, we exploit two long-term travel surveys based on GPS tracking, one collected in spring 2019 and the other during the first pandemic wave in 2020 (from 14 February to 13 July), including data of the very same 48 users. The surveys contain a travel diary for each user, including information on activities, trips, mode choices and route choices. The long duration of the surveys, the high level of detail of the collected information, and the possibility to track the same users during both an entire pandemic wave and the same period of the year before make this dataset a unique opportunity to observe the effects of a pandemic on travel behaviour.
We analyse several aspects of travel behaviour during the COVID-19 pandemic, which result in the following contributions:

- Two long-term travel surveys based on GPS tracking are conducted in 2019 and 2020 in Zürich, on the same users. The users were continuously tracked during the first wave of COVID-19 pandemic to observe changes in travel behaviour, compared to the previous year. General changes in travel distance, mode share and visited locations during the day are shown.

- From the travel surveys, we identify recurrent trips, and differentiate them in commuting and non-commuting, based on imputation of home and work locations. Variations in mode and route chosen are analysed for those trips.

- We estimate two route choice models for public transport, one on trips before the pandemic (2019) and the other during the pandemic (2020). This allows to identify the main criteria for route choice during the pandemic and compare them with the ones in 2019. The differences identified pertain to preferences towards transfers and trains. To the best of our knowledge, this is the first work in literature analysing route choice in public transport during the COVID-19 pandemic.

The paper is organized as follows: Section 6.3 presents the state of the art on travel behaviour during a pandemic, and on route choice models estimated in different periods; Section 6.4 describes the surveys; Section 6.5 presents the methodology; Section 6.6 presents the results; Section 6.7 discusses the results; Section 6.8 shows the conclusions.

6.2.1 *The first wave of the COVID-19 pandemic in Switzerland*

This Section summarizes briefly the evolution of the pandemic in Switzerland during the study period, with a focus on the main measures affecting mobility.
On February 25, 2020 the first case of COVID-19 was confirmed in Switzerland. Following, the number of reported infections increased quickly, and on March 16, more than 1000 daily cases were reported. The first pandemic wave continued until May, with a peak of 1464 reported cases on March 23. Afterwards, in May and June, the number of cases daily reported remained below 100 on most days. In July the cases started to increase towards the second wave in fall 2020, which is out of the analyses in this paper.
To mitigate the contagion, several restrictions were applied by the Swiss government. Starting from March 16, schools and most of the businesses were closed, with only essential stores and institutions remaining open. Working from home was implemented whenever possible.
Restrictions were applied also regarding movements. Traffic with neighbouring countries was strongly limited. People were not forced to stay at home (like in Italy, France and Austria) but a limit of 5 people for gatherings in public places was imposed. The Federal Office of Public Health recommended to avoid public transport at peak times, whenever possible, and especially for risk categories. Nevertheless, public transport services were maintained (sometimes with reduced service frequency, depending on cantons and the evolution of the pandemic).
The restricting measures were kept until April 26, and released in three successive steps, driven by the decrease of infections in the country. From April 27 only certain businesses were allowed to re-open. From May 11 mandatory schools, museums and most of the businesses, as shops and restaurants, were allowed to re-open. On June 8 universities and entertainment businesses re-opened to public. The limit of people for gatherings increased from 5 to 30. The general obligation to wear masks in public transport became mandatory only from the July 6.
Summarizing, the study period of this work can be divided in three phases: pre-lockdown (from 14 February to 15 March); lockdown (from 16 March to 27 April); post-lockdown (from 27 April to 13 July).

6.3 STATE OF THE ART

6.3.1 *Transport studies during the COVID-19 pandemic*

During a pandemic, the transport system and the passenger volume play a key role in the spread of the infection (Cartenì et al., 2021; Lau et al., 2020). On the other end, the pandemic itself affects the transport system and the passengers, which may drastically change their behaviour. This paper focuses on this second aspect. In this regard, restricting measures decided by authorities may forbid or discourage movements or specific transport modes. Moreover, public transport systems, especially if crowded, might be perceived as unsafe (Aloi et al., 2020), encouraging a shift to private modes, as cars or bikes. In fact, Badr et al. (2020) identified in the USA that behavioural changes were observable days to weeks before the restrictions, implying that individuals anticipated public health directives. This is also confirmed by this study and Molloy et al. (2021). Other individuals, in contrast, may be captive to public transport and need to use it in any case (Awad-Núñez et al., 2021).
Several studies analysed the effects of the COVID-19 pandemic on mobility in different countries. They are mostly based on online surveys (Abdullah et al., 2020; Bhaduri et al., 2020) or third-party data, such as ticketing data (Jenelius and Cebecauer, 2020), reports from google (Aloi et al., 2020; Tirachini and Cats, 2020), data from Baidu maps (Huang et al., 2020), or data from bike sharing systems (Chai et al., 2020).
Most of the work focuses on general mobility trends, such as the mode share or the traffic reduction. For instance, Aloi et al. (2020) identified a drop of 93% of public transport users in Santander (Spain) due to the imposed quarantine. Abdullah et al. (2020) collected data from various countries and identified during the pandemic a significant variation of trip purpose, mode choice, distance travelled, and frequency of trips for the primary travel. In particular, they observed a shift from public transport to private transport modes. Similarly, Bhaduri et al. (2020) identified in India a propensity to shift to private modes from shared ones, but also a significant inertia to continue using pre-COVID modes. They reported heterogeneity in the results based on age, income and working status. A different analysis is performed by Chai et al. (2020), which developed a behaviour pattern analysis framework based on usage of bike sharing, to measure the relevance between the change of behaviour and the progress of confirmed cases. Overall, they identify a decrease of less than 40% of

usage of bike sharing.
From online surveys and the above-mentioned third-party data, general changes in mobility and overall trends can be observed. In contrast, to observe long-term variations for specific users, longitudinal data are needed. In this regard, Jenelius and Cebecauer (2020) analysed mobility in Sweden from smart-card data, focusing on public transport ridership and reduction of the number of trips due to the pandemic. Molloy et al. (2021), instead, collected a GPS tracking panel of 1439 Swiss residents. They focused on behavioural shifts in terms of mode share, travel distance, travel speed, and socio-demographic variations.
In addition to Molloy et al. (2021), other studies and companies provided reports on mobility in Switzerland during the COVID-19 pandemic. Google (2021) and Apple (2021) provide reports regarding mobility in several countries of the world. Apple (2021) report the number of search requests by transport mode (public transport, walking and driving), while the reports of Google (2021) focus on the usage of different categories of places such as retail, supermarkets and public transport. Intervista AG (2021) carried out a mobility tracking study of the Swiss population, observing the daily distances covered and the purpose of mobility. In this regard, mobility tracking, besides providing highly detailed information, avoids some of the modelling problems of traditional surveys, such as the difficulty for respondents to describe their routes (Zhu et al., 2010).
To the best of our knowledge, at the time of writing, only two works analysed longitudinal data of several individuals (Jenelius and Cebecauer, 2020; Molloy et al., 2021), but they presented mainly aggregated results and general trends. Therefore, additional specific insights are extremely relevant, to answer several open questions on users' behaviour. This paper focuses on route choice in public transport and analyses how recurrent trips, both commuting and non-commuting, changed during the pandemic. In that sense, a major characteristic of the current study is the availability of tracking data collected both during spring 2019 and the entire first pandemic wave in spring 2020. Given that the pandemic was an unforeseen event, and there was no time to prepare a tracking study, which includes data of the previous year and just before the pandemic outbreak, our dataset represents a unique resource, which cannot be collected again.

6.3.2 *Route choice in public transport*

Route choice models in public transport are used to analyse or predict the chosen routes of one or multiple passengers in a public transport network. Most of the work divide the modelling in two steps (Anderson et al., 2017; Bovy, 2009; Marra and Corman, 2020): a choice set generation algorithm, enumerating the available alternatives; and a route choice model, estimating the passengers' behaviour. The choice set generation is a complex task, often solved using heuristics, since enumerating all possible alternatives is typically not feasible. In this work, we apply the choice set generation algorithm described in Marra and Corman (2020), based on constrained enumeration, a family of algorithm widely used in literature (Bovy, 2009; Cats, 2011; Prato, 2009). The algorithm has a coverage of 94% (identified trips), and is able to generate all the available alternatives given some constraints on the maximum duration and number of transfers of the trip (see the validation performed in Marra and Corman, 2020).

Regarding route choice, different models are used in literature (Prato, 2009). The most used one is the Path Size Logit, which is a variation of the standard Multinomial Logit, including a penalizing parameter for correlated/overlapping alternatives. Anderson et al. (2017) estimated a Mixed Path Size correction Logit in the public transport network of the greater Copenhagen area, from a revealed preference survey. Similarly, Montini et al. (2017) estimated the Path Size Logit from public transport trips in Zürich, collected from GPS data. Yap and Cats (2021) estimated a Path Size Logit to evaluate denied boarding in crowded public transport systems.

In this work, we estimate two Mixed Path Size Logit models: one with data of 2019 and the other with data during the pandemic in 2020. The Path Size Logit has been already estimated with success on the dataset of 2019 in Marra and Corman (2020). We refer to that work for further details on the model and the literature on both choice set generation and route choice models. We also believe that a comparison with alternative models (Nested Logit, see Nassir et al. (2015), or Recursive Logit, see Zimmermann and Frejinger (2020)) is out of the scope of this paper.

A main outcome of this study is the estimation of a route choice model during the COVID-19 pandemic, and its comparison with a model estimated before the pandemic. In literature, route choice has not yet been analysed during a pandemic. Moreover, a recent work (Weis et al., 2021) highlights that there are only few repeated studies in the field of transport planning.

Repeating a study allows observing changes in respondents' preferences, if the survey methodology and the sample characteristics stay consistent over time. To this end, Weis et al. (2021) analysed mode and route choice of Swiss population from a combined SP/RP (stated preference/revealed preference) survey in 2015, comparing it with one in 2010. They identified that willingness to pay indicators are rather stable in time, which is particularly relevant for their use in cost-benefit analyses.
In the current paper, the tracking of the very same respondents before and during the pandemic, and the use of the same methodology, guarantee a fair comparison between the two different periods.

6.4 DATASETS AND STUDY PERIOD

The travel diaries used in this study are collected during two different periods from the same group of users, which are all residents of Zürich. In spring 2019, 172 participants installed a smartphone app, the *ETH-IVT Travel Diary* (Marra et al., 2019), which continuously collect GPS data without affecting the battery consumption. On average each participant was tracked for 22 days. In February 2020 (few weeks before the outbreak of the pandemic in Switzerland), the same participants were contacted again, and 48 of them decided to participate again to the study. This time the participants were tracked until early July 2020, with an average of 112 days per user. Only data and trips collected inside the city of Zürich are considered in this analysis.
To derive travel diaries from the GPS data, we applied a mode detection algorithm, described in Marra et al. (2019). The algorithm automatically identifies activities, trips, stages and transport modes used. Each public transport stage is described with information on the line, the vehicle of that line, the departure stop and time, and the arrival stop and time. The mode detection algorithm has an average accuracy of 86.14% and has been already validated in previous studies. In particular, Marra and Corman (2020) used the same dataset of this study (the one of 2019), showing a realistic mode share and realistic estimations of route choice models.
Table 6.1 shows information on the users in our surveys and their representativeness (the information refers to 2020, despite it was not significantly different in 2019). We remark that the survey in 2020 contains only 48 of the 172 users of the survey in 2019, used in Marra and Corman (2020). Therefore, in this paper, we will consider only these users, for both years. Regarding the representativeness, our survey contains in general younger

		Survey 2020 (%)	Zürich 2016 (%)
	Users	48	-
Gender	Male	54	50
	Female	46	50
Age	<18	0	16
	18-24	23	8
	24-34	37	22
	34-44	21	18
	44-54	19	14
	>54	0	25
Education	Mandatory	6	18
	Secondary	29	34
	Higher	65	48
Household size	1	29	22
	2	33	30
	3	9	18
	4	21	19
	5+	8	12
Income	<4000	9	24 (<5000)
(monthly CHF)	4000-8000	29	24 (5000-7500)
	8000-12000	31	31 (7500-12500)
	12000-16000	13	11 (12500-16666)
	>16000	8	9 (>16666)
	No answer	10	0

TABLE 6.1: Comparison of socio-demographic characteristics between the survey and the official statistics in Zürich in 2016 (Zürich Statistic Office, 2021). The income information in Zürich Statistic Office (2021) is based on a survey in 2015 and the ranges are slightly different from the ones of our survey.

and highly educated participants. A possible explanation is the nature of the survey, requiring installing a smartphone application, which might not

	2019	2020
Period	03.04.2019 - 02.06.2019	14.02.2020 - 13.07.2020
Users	48	48
Avg. days per user	25	112
Activities	4617	12234
Trips	4597	12157
Trips inside Zürich	2266	6316
Car trips in Zürich	382 (17%)	1371 (22%)
Bike trips in Zürich	279 (12%)	1089 (17%)
Walk trips in Zürich	398 (18%)	1520 (24%)
Mixed trips in Zürich	244 (10%)	687 (11%)
Public transport trips in Zürich	963 (43%)	1649 (26%)
# transfers per p.t. trip (%)	{0: 58%, 1: 31%, 2: 8%, 3+: 3% }	{0: 68%, 1: 25%, 2: 6%, 3+: 1% }
p.t. modes used (%)	{Tram: 52%, Bus: 41%, Train: 7%}	{Tram: 52%, Bus: 40%, Train: 8%}
Avg. duration per p.t. trip	22 min	20 min
Avg. air distance per p.t. trip	2.99 km	2.35 km

Table 6.2: Comparison of travel diaries in 2019 and 2020. Mode share in Zürich in parentheses.

be attractive for older people. Gender, household size and income follow quite regularly the actual distribution, despite there are fewer participants in the lowest income range, and slightly more men than women.
Table 6.2 compares the travel diaries collected in 2019 and 2020. The duration of the data collection is much longer in 2020 (almost 5 months) compared to 2019 (2 months), and on average each person was tracked 4 times longer in 2020. This led to a larger number of trips and activities collected

in 2020. Despite this, due to the effects of the pandemic, the number of identified trips in 2020 is just 2.6 times that of 2019 (1.7 times for public transport).

In this work, we analyse only trips inside the city of Zürich, and we discard trips with an absence of signal longer than 7 minutes (as in Marra et al., 2019). In total, we analyse 2266 trips in 2019 and 6316 trips in 2020. The mode share in 2019 (reported in parentheses in Table 6.2) is close to the official one in 2016 (Stadt Zürich, 2021, 41% public transport, 26% walk, 25% motorized private mode, 8% bike), especially for public transport (43% instead of 41%), which is the focus of this paper. We remark that mixed trips (public and private transport) are not reported in the official statistics, but only in our survey. In contrast, the mode share in 2020 is remarkably different: the public transport share is strongly reduced (26%) in favour of walk, car and bike (more details in Section 6.6).

An additional difference between 2020 and 2019 is the number of transfers. Trips without transfers (i.e. only one public transport vehicle) increased from 58% to 68%, suggesting passengers prefer to reduce the number of transfers. This might come from a perception of each vehicle as additional source of contagion; we analyse this aspect in details in Section 6.6.3. A further difference concerns the length of the trips, which decreases from 2.99 km to 2.35 km. Finally, there is no particular difference in the mode share among public transport modes (tram, bus and train).

This dataset represents a unique data source, containing long-term travel diaries of several users before and during the pandemic. We believe the relatively small number of users (48) does not represent a limitation: most of the socio-demographic characteristics reflects the official reports, with few exceptions; the mode share in 2019 does not differ significantly from the real one; general characteristics of mobility, as mode share and travel distance during pandemic are in line with other studies in Switzerland (see Section 6.6.1). In addition, the data collection and the dataset of 2019 were already tested successfully in a previous work (Marra and Corman, 2020). In that work, no heterogeneity among participants was found regarding public transport route choice, which is the focus of this paper. This suggests that the results are significant and representative, without the need of a larger dataset (which would be in any case impossible to collect).

6.5 METHODOLOGY

6.5.1 *Route choice model formulation*

In this Section, we present the route choice model estimated for public transport trips. We estimate the same model both on data of 2019 and data of 2020, to observe the differences between before and during the pandemic. To understand route choices of public transport passengers, a route choice model requires two types of information: a set of observed choices, describing the routes chosen with different attributes (e.g. travel time, mode, transfers); a set of non-chosen routes for each observation (choice set), describing alternative choices discarded by the passengers. In our study, the observed choices correspond to the public transport trips in Zürich in Table 6.2, while the choice set for each trip is determined by the choice set generation algorithm described in Marra and Corman (2020). For each trip, up to 40 alternatives are identified, consisting of sequences of public transport vehicles alternated by walks. We consider an alternative matching the passenger's choice, when it has the same lines used by the passenger (taking the same line at a near stop is considered the same alternative). No information on network conditions and delays is assumed for the choice set generation, i.e. the alternatives are generated from the timetable.

We estimate a Mixed Path Size Logit model, which is an extension of the Path Size Logit, allowing for random taste variations across users. The Path Size Logit is a variant of the standard Logit, including a penalty for overlapping trips in the utility function. For each route, we consider the utility function in Equation 6.1, including travel time (in bus, tram and train), walking time, transfer time and the number of transfers. The walking time refers to the time between the start of the trip (at the origin) and the departure of the first vehicle (at the first stop), plus the time between the arrival of the last vehicle (at the last stop) and the arrival at the destination. The transfer time is the entire time during a transfer. Therefore, the waiting time is included both in the walking time and the transfer time, since the quality of the GPS data did not allow a precise discrimination between walking and waiting. Monetary costs were not considered in this work, since in Zürich there is a fixed price for public transport, which does not depend on the chosen route.

$$\begin{aligned} U_j = -C_j &= \beta_{tram} * tram\ time + \beta_{bus} * bus\ time \\ &+ \beta_{train} * train\ time + \beta_{walk} * walk\ time \\ &+ \beta_{tt} * transfer\ time + \beta_{transfer} * \#transfers + \beta_{PS} * PathSize_j \end{aligned} \tag{6.1}$$

$$PathSize_{trip} = - \sum_{stage\ s \in trip} \frac{duration(s)}{duration(trip)} \ln(times\ s\ in\ choiceset) \tag{6.2}$$

$$P(trip|choiceset; \overrightarrow{\beta}) = \frac{e^{U_{trip}(\overrightarrow{\beta})}}{\sum_{j \in choiceset} e^{U_j(\overrightarrow{\beta})}} \tag{6.3}$$

The *PathSize* is a penalty attribute, based on the formulation in Bovy et al. (2008), which penalizes alternatives using the same stage (public transport line). The penalty increases with the duration of an overlapping stage in the trip and the number of times the stage appears in the choice set.
To observe possible panel effects and heterogeneity among users in the perception of costs, we estimated a Mixed Path Size Logit model (Anderson et al., 2017; Prato et al., 2014; Yap and Cats, 2021). In this model, the coefficients ($\overrightarrow{\beta}$) are assumed random parameters following a probability density function $f(\beta|\theta)$. In literature, the normal and log-normal distributions are used for the parameters. Despite the log-normal distribution allows to restrict the values to only one sign, it may result in a wide distribution, given its long tail. Therefore, we assume normally distributed parameters, described by a mean and a standard deviation. A high standard deviation for a parameter indicates high heterogeneity in the perception of its cost among the users. The probability of choosing a trip is the following:

$$P(trip|choiceset) = \int \frac{e^{U_{trip}(\overrightarrow{\beta})}}{\sum_{j \in choiceset} e^{U_j(\overrightarrow{\beta})}} f(\beta|\theta) d\beta \tag{6.4}$$

The model was estimated with the software mixl (Molloy et al., 2019), using 500 draws to simulate the probabilities.
The Path Size Logit model (and the Mixed Path Size Logit) has already been successfully estimated on the dataset of 2019 in Marra and Corman (2020), which also discuss details on the performance and validity. In this work, instead, we estimate the same model on data collected during the

pandemic, from the same users, and we compare the two models and the estimated coefficients. While no remarkable heterogeneity estimating the Mixed model was identified in the dataset of 2019 in Marra and Corman (2020), we here analyse the heterogeneity in cost perception during the pandemic. The pandemic is an exceptional condition, and the passengers might consider also the risk of contagion during their choices, which may influence the perception of the travel cost components in different ways. We remark we estimate a single model, and not a model for each phase of the pandemic, since it would result in fewer observations per model and less reliable results.

6.5.2 *Identification of visited locations and recurrent trips*

To analyse how the pandemic affected recurrent trips, we need to understand from which location a trip is performed and to which destination. In the collected travel diaries, each activity of a user corresponds to a set of GPS points near to each other for a long time. Each activity is the end location of a previous trip and the start location of an upcoming trip. No semantic meaning is associated to the identified activities.

Therefore, in this work we apply an intuitive and effective method, to classify the activities in *home*, *work* (or secondary location) or *other* location. First, to identify activities representing the same location, we applied a clustering algorithm, the DBSCAN (Ester et al., 1996), which assigns a cluster to each activity. A simple rule-based approach can then classify the clusters (and the activities) in *home*, *work* and *other*.

The DBSCAN algorithm takes as input the GPS coordinates of all the activities of a user (mean point of each activity), and a maximum distance as parameter (100 meters). No minimum number of activities for a cluster is specified. The advantage of this algorithm is that it does not require to specify the number of clusters, since the algorithm just groups together activities near to each other. Each isolated activity will form a cluster of its own. As result of the clustering, the activities in the same cluster represent the same location (e.g. the home).

After identifying the clusters, we apply two simple rules to identify home and work locations. We name *home* location the one corresponding to the cluster with highest number of activities (weighted by their duration) during weekdays, between 23:00 and 06:00. We name *work* location the one corresponding to the cluster with highest number of activities (weighted by their duration) during weekdays between 09:00-12:00 and 13:00-17:00

(excluding the home cluster). The activities belonging to other clusters are marked as *other* locations.
The proposed method finds several correspondences in the literature. Bhadane and Shah (2020) compare different clustering algorithms to identify Region of Interest (e.g. home, work, post office), concluding DBSCAN suits well for spatial data clustering. Liu et al. (2019) identify individual activity clusters from geo-tagged tweets, applying an adapted version of DBSCAN. Xiong et al. (2020) clusters points of interest in regions using DBSCAN. Moreover, existing research identifies home and work locations as the most frequently visited stop during nighttime and daytime hours, respectively (Calabrese et al., 2013; Chen et al., 2014; Kung et al., 2014; Phithakkitnukoon et al., 2012).
In our test case, the average distance between two activities in the same cluster is 27 meters, which confirms they represent the same physical location. This method does not aim to be the state of the art in activity classification, but it is sufficient to show the general changes in users' location due to the pandemic. We remark that there are exceptions in users' behaviour which are not considered in this method, as night workers or people with multiple work locations.
After assigning a location (i.e. a cluster) to each activity of a user, it is possible to assign each trip to an origin-destination couple (OD). In other words, two trips starting and ending in the same locations can be assigned to the same OD. We call those *recurrent trips*. Moreover, we refer the recurrent trips between home and work (both directions), as *commuting trips*.
As a technical remark, we applied the clustering algorithm considering both the activities in 2019 and 2020, to have the same physical location (e.g. a supermarket) labelled as the same location/cluster in both years. In contrast, we identify the home and work location independently in 2019 and 2020, to identify people who potentially have changed home or workplace. Considering the two years independently or together for the clustering and/or the labelling does not change significantly the results.

6.6 RESULTS

6.6.1 *Mobility trends during pandemic*

In this Section, we show how the mobility changed during the different phases of the pandemic, studying mode share, travel distance and location during the day of the tracked users.

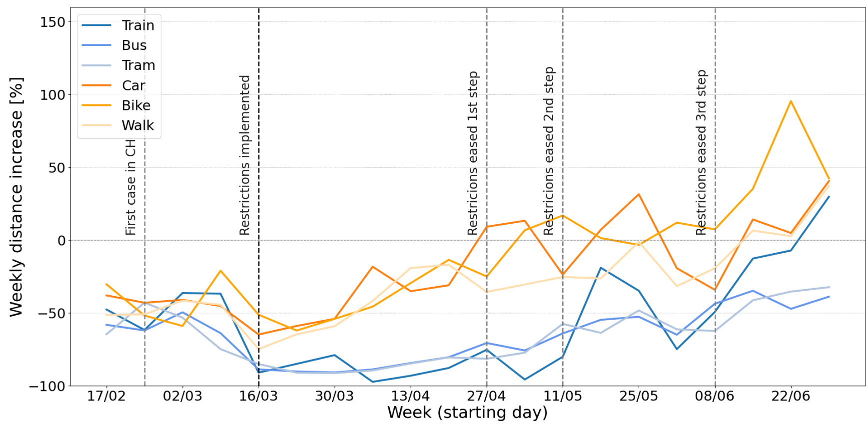


FIGURE 6.1: Travel distance increase during 2020. The baseline (0%) is the average travel distance in spring 2019.

Figure 6.1 shows the travel distance increase during 2020, compared to 2019. During the last weeks of February, at the begin of the outbreak, the travel distance for every mode is around 50% less than 2019. This can be explained both by a higher baseline (which is based on spring 2019, including Easter and other holidays), and by the first effects of the pandemic. In fact, Badr et al. (2020) show evidence of behavioural changes in US before the restrictions, indicating an anticipation of public health directives from the individuals. With the first restrictions implemented on 16 March, the travel distance drops significantly (more than 90% for public transport). With the first easing of the restrictions, the travel distance increases again. In June, with most of the restrictions removed, the travel distance of private modes reaches the values of 2019. In contrast, for public transport a decrease of around 40% remained (despite an increase for trains in the last days), probably due to ongoing policies, as the possibility to work from home, and a remaining perception of public transport as unsafe. We highlight that the same trend of travel distance was observed in two different surveys in Switzerland (Intervista AG, 2021; Molloy et al., 2021), based on larger samples of users, collected in the whole Switzerland. Despite our survey focuses on the city of Zürich, the trend of travel distance observed in the three studies is similar, which shows the validity and representativeness of our approach. An exception is the distance by bike, which was found increasing significantly during the pandemic by Molloy et al. (2021),

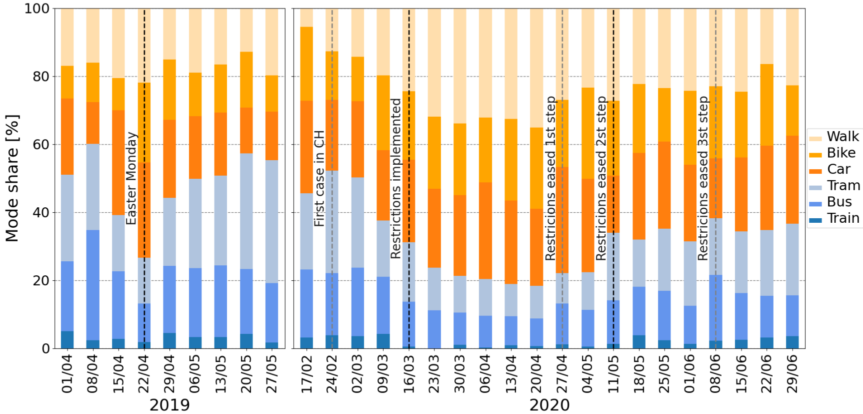


FIGURE 6.2: Mode share in 2019 and 2020 (number of trips).

while not by this work and Intervista AG (2021). A possible explanation is given by the different baseline, based on fall 2019 in Molloy et al. (2021), while on spring 2019 in this work. In any case, this paper does not focus on bike trips.

Figure 6.2 shows the mode share in 2019 and 2020. In February 2020, the mode share is similar to the one identified in 2019. With the restrictions in March, the share of public transport reduces in favour of walk and private modes. The share of public transport follows a similar pattern as for the travel distance, reaching a plateau (below 40%) at the second half of May, lower than the 2019 baseline (around 48%). Again, this can be explained by ongoing policies, as possibility to work from home, and by a perception of public transport as unsafe.

Figure 6.3 shows the location of the users in 2019 and 2020 during the day. In 2019, during weekdays, most of the users stays at home in the early morning and during the night, as expected. Around 8 and 18, there are the two travelling peaks, in conjunction with an increase and a decrease of people at work. In 2020, instead, the percentage of people staying at work decreases, from a peak of 50% to 14%. The trips also decreases, especially the ones in the morning. In contrast, the activities marked as *other*, i.e. everything else besides home and work, did not decrease substantially. In general, the location pattern during weekdays in 2020 is similar to the one in 2019 during weekend, with most of the trips occurring in the afternoon. Comparing the weekend in 2019 and 2020, the location of the users during

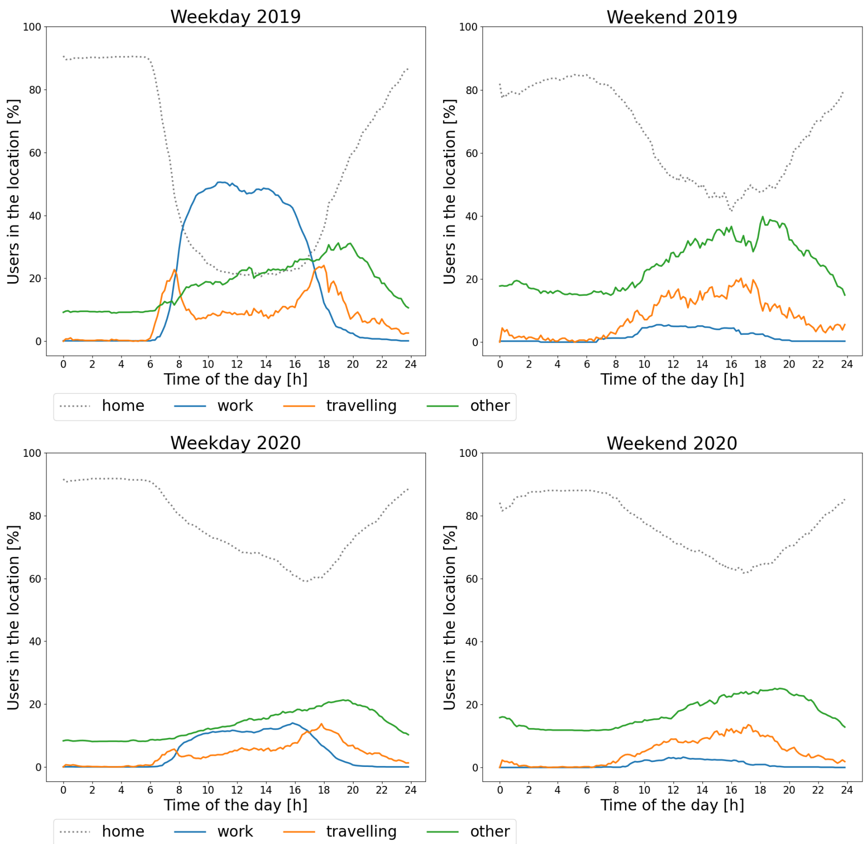


Figure 6.3: Location of the users during the day in 2019 and 2020. All days of all users are aggregated.

the day is similar, with the only exception of a higher percentage of people at home in 2020, unsurprisingly.

6.6.2 *Analysis of recurrent trips: commuting and non-commuting*

In this section, we analyse if in 2020 people choose a mode or route different from that of 2019 for recurrent trips. We selected for each user all ODs occurred at least 4 times in 2019, identifying 125 ODs. 51% of them occurred at least 4 times also in 2020, for a total of 64 ODs analysed.

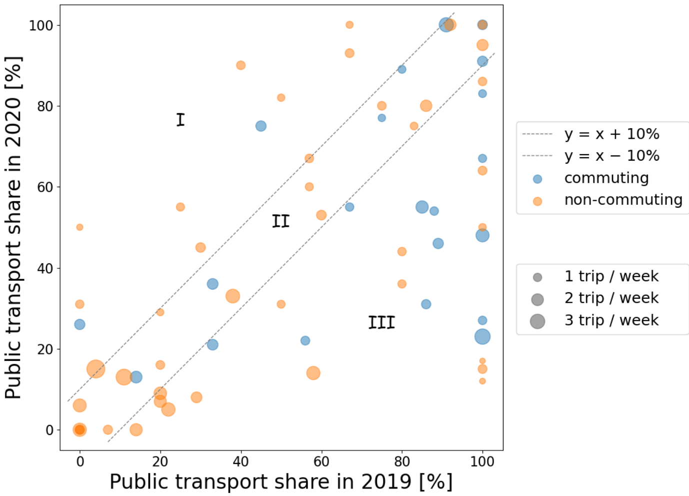


Figure 6.4: Mode choice of the ODs in 2019 and 2020. Each dot represents an OD, with its size representing the frequency. Axes represent the percentage of times the OD is performed by public transport in 2019 and 2020. Only ODs with at least 4 trips in each year are considered.

Figure 6.4 shows the public transport share for each OD in the two years. For example, a 80% in the x-axis means the user chose 80% of times public transport for that OD in 2019 (the remaining 20% includes walk, bike and car). We can divide the ODs in three groups: ODs with an increase in public transport share in 2020 by at least 10% (labelled I); ODs with a similar share between the two years (labelled II); ODs with a decrease in public transport share in 2020 by at least 10% (labelled III). The most of non-commuting ODs are in the second group (43%), compared to the third (36%) and the first (21%). Thus, for those trips the share of public transport decreased during the pandemic, but not significantly. In contrast, for commuting, the majority of users switched clearly from public transport to private modes (14% first group, 33% second group, 53% third group).
In general, no ODs are located in the top-left corner of the figure, representing a switch from private to public transport. The few ODs in the bottom-left corner, with a higher public transport share in 2020, can be imputed to the shorter study period in 2019. In fact, an OD that is rarely travelled by public transport may result in a 0% of public transport usage in 2019 and a small percentage (5-20%) in 2020.

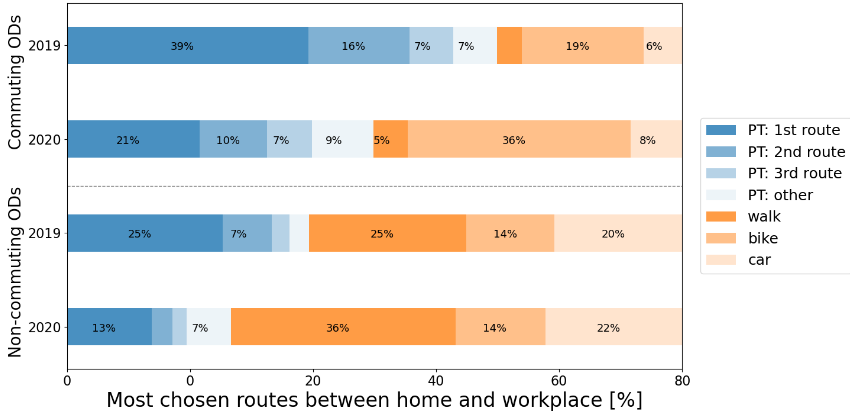


FIGURE 6.5: Frequency of most chosen routes for commuting and non-commuting ODs. Only ODs with at least 4 trips in each year are considered.

Figure 6.5 shows how the frequency of the most chosen routes in 2019 changed in 2020. For example, the first two bars show that on average the most frequent route in commuting ODs is chosen 39% of times in 2019, while the same route is chosen 21% of times in 2020. On average, the chosen routes for the same OD in the two years are different. For commuting ODs, the preferred public transport route is chosen less, since passengers try different routes and partially switched to private modes. The usage of bike greatly increased, matching the observed reduced public transport share. Also for non-commuting trips the preferred public transport route is less chosen (from 25% in 2019 to 13% in 2020), and walking trips significantly increased.

6.6.3 *Route choice*

Table 6.3 shows the estimation of the Mixed Logit in 2019 and 2020. The observations include all identified public transport trips, covered by the choice set generation algorithm. Very short trips, where using public transport is unrealistic (i.e. walking takes less than half the time of any public transport trip), are discarded. The estimated model in 2019 is the same as in Marra and Corman (2020) (Table 7), except that the number of observations is lower, since in this work we consider only the users, which are also

Parameter	2019	t-test	2020	t-test
In vehicle travel time (s)				
μ Tram	-1	-8.34	-1	-9.17
μ Bus	-1.17	-8.57	-1.16	-10.09
μ Train	-1.85	-4.97	-1.00	-4.54
μ Walking time	-2.39	-14.10	-2.44	-20.22
μ Transfer time	-1.06	-13.70	-0.91	-10.49
μ Number of transfers	-792	-13.04	-1008	-12.44
σ Tram	0.18	1.96	0.17	4.74
σ Bus	0.22	3.39	0.13*	0.51
σ Train	0.83	3.61	0.28*	1.09
σ Walking time	0.56	4.09	0.25	2.78
σ Transfer time	0.01*	0.12	0.01*	0.03
σ Number of transfers	88*	0.74	200	3.31
Observations	877		1427	
Null log-likelihood	-2352		-3544	
Final log-likelihood	-726		-1054	
Adjusted R^2	0.69		0.70	
Scaling factor	0.0040		0.0041	

TABLE 6.3: Mixed Logit estimated in 2019 and 2020. Parameters distributed according to a normal distribution. * indicates a non-significant parameter ($|t| < 1.96$). The parameters are scaled (multiplied by the scaling factor) to have the in-tram travel time coefficient equal to -1.

tracked in 2020. Nevertheless, the estimated coefficients and standard deviations are very close to the ones estimated in Marra and Corman (2020), showing a model robust to a reduction of users (48 instead of 172). The only difference is in the *PathSize* factor, which was not found significant in this work. We remark that the *PathSize* was not found significant also in one of the experiments in Marra and Corman (2020) and in Nielsen et al. (2021). They showed the *PathSize* may act both as a correction factor, penalizing overlapping alternatives, and as a positive factor, rewarding paths with more opportunities to reach the destination. Doubts on the validity of the *PathSize* were also raised in Duncan et al. (2020), which demonstrate issues with this model. Therefore, we also estimated the model without this parameter, as a Mixed Logit (testing different correction parameters is out of the scope of this work).

Here, we briefly discuss the model in 2019 to better understand the model in 2020. We report the coefficients scaled by the travel time in tram, to better discuss the rates of substitution among them. All mean values are statistically significant, and their sign and values are realistic and in line with the literature. The preferred mode is the tram, followed by the bus and the train, in accordance with previous works in Zürich (Meyer de Freitas et al., 2019; Montini et al., 2017). The walking time has a higher cost than the in vehicle travel time, as expected (Meyer de Freitas et al., 2019). The transfer penalty is around 15 minutes of travel time in tram, which falls near the range identified by Garcia-Martinez et al. (2018), between 15.2 and 17.7 min of in-vehicle travel time, in the multi-modal urban network in Madrid. Looking at the standard deviations, the ones of travel time in bus, tram and walking time are significant but low (between 18% and 23.4% of the respective mean value), showing low heterogeneity among the users. The standard deviations of the transfer penalty and the transfer time are not significant, showing that there is no heterogeneity in the perception of transfers among the users. Finally, the only parameter with a large standard deviation is the travel time in train (44.9%). This implies that 1.3% of the population is associated with a positive coefficient for the travel time in train. This is a limitation of using the normal distribution[1].

The model of 2020 has comparable coefficients of tram, bus and walk, as well as low standard deviations, which is non-significant for the bus. Remarkable differences are in the other coefficients. The in-train travel time is perceived with a lower cost, and it is comparable with the in-tram travel time. No heterogeneity was found related to this parameter. Looking at the transfer-related coefficients, compared to the 2019, the transfer penalty is much higher and the coefficient of the transfer time is lower. In addition, in 2020, there is heterogeneity in the perception of the transfer penalty.

6.7 Discussion

Mobility has been strongly affected by the COVID-19 pandemic, even before any restriction, observing a strong decrease and a slow recovery. The situation in June 2020 is not at the level of 2019, probably due to increased working from home, and perception of unsafe public transport. The ob-

1 clarification added in this thesis

served shift to private modes can have negative consequences if it remains after the pandemic, such as increased traffic congestion and pollution. Therefore, keeping an attractive public transport system, especially with policy measures to improve its safety, or its perception, would be crucial. The obligation for wearing masks (as of 06 July 2020) is indeed moving in this direction.

Looking at the locations visited during the pandemic, people spent much more time at home compared to the previous year, especially during weekdays. The workplace, instead, is visited roughly 3.5 times less. These changes show the impact of working-from-home, and temporary closures of most of working activities. Accordingly, the time spent travelling also decreased, and the daily routine during weekdays switched from a morning and an afternoon peak, to a single peak in the afternoon. If the working-from-home share is higher in the future, the demand will decrease, and especially the demand peaks, with a potential impact towards supply and infrastructure needs.

Focusing on recurrent trips, we identified that the chosen modes and routes strongly changed due to the pandemic. Public transport usage generally decreased in favor of private modes. For commuting trips, the increase was greater for bike; while, for non-commuting trips, the share of walk increased. This shows there is a large percentage of people willing to switch to private modes to avoid public transport. In this regard, the increase in bike usage is a positive trend, which can be maintained after the pandemic, given sufficient investments in bike infrastructure and bike sharing systems.

Regarding the public transport route choice, in 2020 the preferred route is chosen less, in proportion to other routes or private ones. In other terms, the routes are chosen more equally and there is a smaller gap between the most chosen route and the second chosen one. Among the possible factors (different daily routines, route choice criteria or safety perception), we tested the crowding, but without identifying a clear relationship between the chosen route and the expected crowding (considering the average occupancy rate of public transport lines in 2019). Further investigations on crowding and safety perception are left for future works.

The estimation of a route choice model in 2019 and 2020 showed the route choice in public transport was affected by the pandemic. In fact, we identified important differences and similarities between the two periods (see Table 6.3). While the cost perception of tram, bus and walking is similar, the one for train and transfers is different. We hypothesize for the lower cost perception of the travel time in train that trains, being large vehicles, guarantee (or are perceived to guarantee) more easily an adequate social distance. Moreover, trains are used for short urban trips (rarely more than three stops). Passengers would therefore be for a very short period with other people, which could be potentially perceived as a threat (wearing a mask was not mandatory during the study period).

The transfer penalty in 2020 is much higher than in 2019 (by 27%). We link this again to perceived contagion risks, as an additional transfer means more and different people encountered during the trip. In contrast, the perception of transfer time is lower in 2020 and it is lower than the travel time in any vehicle. As most of the transfers in Zürich takes place outdoors, the waiting time at a transfer stop could be perceived safer than being in a vehicle. No heterogeneity among users was found in 2020 related to trains, showing a general lower cost perceived by all the users. Instead, the transfer penalty is perceived differently among the users in 2020, while not in 2019.

Overall, the identified changes in route choice criteria help predicting the traffic flow and prepare a response for current or future waves or pandemics. The new estimated coefficients and the resulting utility function can be used for a more accurate traffic assignment model, which can predict the use of the public transport service during a pandemic. Such a model helps public transport planning, for instance, identifying potentially crowded vehicles, which may expose people to a high risk of contagion. Similarly, the model can also be used to recommend routes with a lower risk of contagion for the user. In this regard, the collected data hint towards redesigning the network to increase direct connections and reduce transfers, or increasing frequency and capacity of the lines for which crowding is expected.[1]

While the differences in mode share and travel distance can be explained both by personal factors (as safety perception) and by the implemented restrictions, the differences in route choice can be imputed mainly to personal factors. In fact, the settings for route choice (public transport offer, available alternatives and travel times) remained substantially the same as before the pandemic. The only differences in the public transport service are a general reduction of crowding and a more reliable service. A marginal change in the public transport offer occurred from March 30 to May 4, 2020, consisting in a reduction of the frequency of some tram and bus lines, mainly during peak-hours, from a run every 7.5 minutes to ev-

1 this paragraph has been extended compared to the submitted version.

ery 10 minutes. Anyhow, this period corresponds to the highest restrictions and therefore a small fraction of observations in our dataset. We thus believe changes in personal preferences are the main factor determining the observed changes in route choice.

A limitation of the route choice analysis is the limited inclusion of trip purpose (as 2020 had much less work trips than 2019). We are aware that trips with different purposes might result in different choice criteria, but we preferred estimating a single model for all trips. Dividing by trip purpose would result in fewer observations per model and less reliable results.

As a thought exercise, we could think of the trips in 2019 as predominantly work-oriented and the ones in 2020 as predominantly leisure-oriented. In such a case, we can look if the identified differences match the literature comparing working and leisure trips. In Switzerland, Weis et al. (2021) identify for both the transfer penalty and the transfer time a higher cost (respect to the travel time) in non-working trips. Nielsen et al. (2021) identify in leisure trips a higher rate of substitution with in-bus time (i.e. a higher cost), for in-train time, transfer penalty and transfer time. In our case, instead, we have identified a different situation, with higher transfer penalty (in 2020), but lower costs for transfer time and in-train time. Therefore, we link the identified differences between the two periods to the pandemic, and not to the different trip purposes.

6.8 CONCLUSIONS

In this work, we observed travel behaviour during the COVID-19 pandemic from GPS tracking. This technology proved to be an efficient method, to collect long-term travel diaries without a significant burden on the users. The resulting dataset used here is an unrepeatable opportunity to observe travel behaviour during a pandemic. Moreover, the observation of the very same users already in 2019 allows a precise comparison of travel behaviour before and during the pandemic.

We observed how the mode share and the travel distance changed during the different phases of the pandemic. Public transport modes resulted as the most affected ones, with a reduced traffic persisting even after the first wave. We exploited the long-term nature of the dataset to observe how recurrent trips changed in 2020, in terms of mode and route choice. The share of public transport decreased, in favour of private modes, with a significant increase of bike usage for commuting trips. Moreover, public transport users have not anymore a precise preferred route, and they often

choose different routes for the same OD.
We estimated two route choice models, based on trips before and during the pandemic, identifying important differences in perception of travel time in train and transfers. Given an already exiguous literature on comparing route choices of the same population in different periods, our work represents an important contribution on understanding how travel behaviour evolves in time, especially during a pandemic.
Therefore, for future work, we encourage the repetition of long-term surveys in different years. An interesting possibility is comparing a pre-pandemic period with a post-pandemic one, when the emergency may be considered over. In this work, we focused on public transport and route choice, although other aspects of mobility can be analysed. For instance, analysing activity-based travel patterns and daily routines, one can observe how the pandemic affected the daily behaviour of different users. Finally, we remark that we analysed the effects on travel behaviour of the pandemic in general, while further research is needed to observe the consequences of specific restrictions or how passengers' perception of safety affect their behaviour.

ACKNOWLEDGEMENTS

The authors wish to thank all study participants.

REFERENCES

Abdullah, M., C. Dias, D. Muley, and M. Shahin (2020). Exploring the impacts of COVID-19 on travel behavior and mode preferences. *Transportation Research Interdisciplinary Perspectives* 8, 100255.

Aloi, A., B. Alonso, J. Benavente, R. Cordera, E. Echaniz, F. Gonzalez, C. Ladisa, R. Lezama-Romanelli, A. Lopez-Parra, V. Mazzei, L. Perrucci, D. Prieto-Quintana, A. Rodriguez, and R. Sanudo (2020). Effects of the COVID-19 Lockdown on Urban Mobility: Empirical Evidence from the City of Santander (Spain). *Sustainability* 12.9, 3870.

Anderson, M. K., O. A. Nielsen, and C. G. Prato (2017). Multimodal route choice models of public transport passengers in the Greater Copenhagen Area. *EURO Journal on Transportation and Logistics* 6.3, 221.

Apple (2021). Covid-19 mobility trends reports - `https://www.apple.com/covid19/mobility`.

Awad-Núñez, S., R. Julio, J. Gomez, B. Moya-Gómez, and J. S. González (2021). Post-COVID-19 travel behaviour patterns: impact on the willingness to pay of users of public transport and shared mobility services in Spain. *European Transport Research Review* 13.1, 20.

Badr, H. S., H. Du, M. Marshall, E. Dong, M. M. Squire, and L. M. Gardner (2020). Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *The Lancet Infectious Diseases* 20.11, 1247.

Bhadane, C. and K. Shah (2020). Clustering Algorithms for Spatial Data Mining. *Proceedings of the 2020 3rd International Conference on Geoinformatics and Data Analysis*, 5.

Bhaduri, E., B. S. Manoj, Z. Wadud, A. K. Goswami, and C. F. Choudhury (2020). Modelling the effects of COVID-19 on travel mode choice behaviour in India. *Transportation Research Interdisciplinary Perspectives* 8, 100273.

Bovy, P. H. L. (2009). On Modelling Route Choice Sets in Transportation Networks: A Synthesis. *Transport Reviews* 29.1, 43.

Bovy, P. H. L., S. Bekhor, and C. G. Prato (2008). The Factor of Revisited Path Size: Alternative Derivation. *Transportation Research Record* 2076.1, 132.

Calabrese, F., M. Diao, G. Di Lorenzo, J. Ferreira, and C. Ratti (2013). Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation Research Part C: Emerging Technologies* 26, 301.

Cartenì, A., L. Di Francesco, and M. Martino (2021). The role of transport accessibility within the spread of the Coronavirus pandemic in Italy. *Safety Science* 133, 104999.

Cats, O. (2011). Dynamic Modelling of Transit Operations and Passenger Decisions. PhD thesis. KTH Royal Institute of Technology.

Chai, X., X. Guo, J. Xiao, and J. Jiang (2020). Analysis of Spatial-temporal Behavior Pattern of the Share Bike Usage during COVID-19 Pandemic in Beijing. *arXiv:2004.12340 [physics]*.

Chen, C., L. Bian, and J. Ma (2014). From traces to trajectories: How well can we guess activity locations from mobile phone traces? *Transportation Research Part C: Emerging Technologies* 46, 326.

Duncan, L. C., D. P. Watling, R. D. Connors, T. K. Rasmussen, and O. A. Nielsen (2020). Path Size Logit route choice models: Issues with current models, a new internally consistent approach, and parameter estimation on a large-scale network with GPS data. *Transportation Research Part B: Methodological* 135, 1.

Ester, M., H.-P. Kriegel, J. Sander, and X. Xu (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 226.

Garcia-Martinez, A., R. Cascajo, S. R. Jara-Diaz, S. Chowdhury, and A. Monzon (2018). Transfer penalties in multimodal public transport networks. *Transportation Research Part A: Policy and Practice* 114, 52.

Google (2021). COVID-19 Community Mobility Reports - `https://www.google.com/covid19/mobility`.

Huang, J., H. Wang, M. Fan, A. Zhuo, Y. Sun, and Y. Li (2020). Understanding the Impact of the COVID-19 Pandemic on Transportation-related Behaviors with Human Mobility Data. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 3443.

Intervista AG (2021). Mobilitäts-Monitoring COVID-19 - Online report published by Statistical Office of Kanton Zürich - `https://www.intervista.ch/mobilitaets-monitoring-covid-19`. de-DE.

Jenelius, E. and M. Cebecauer (2020). Impacts of COVID-19 on public transport ridership in Sweden: Analysis of ticket validations, sales and passenger counts. *Transportation Research Interdisciplinary Perspectives* 8, 100242.

Kung, K. S., K. Greco, S. Sobolevsky, and C. Ratti (2014). Exploring Universal Patterns in Human Home-Work Commuting from Mobile Phone Data. *PLOS ONE* 9.6, e96180.

Lau, H., V. Khosrawipour, P. Kocbach, A. Mikolajczyk, H. Ichii, M. Zacharski, J. Bania, and T. Khosrawipour (2020). The association between international and domestic air traffic and the coronavirus (COVID-19) outbreak. *Journal of Microbiology, Immunology and Infection* 53.3, 467.

Liu, X., Q. Huang, and S. Gao (2019). Exploring the uncertainty of activity zone detection using digital footprints with multi-scaled DBSCAN. *International Journal of Geographical Information Science* 33.6, 1196.

Marra, A. D., H. Becker, K. W. Axhausen, and F. Corman (2019). Developing a passive GPS tracking system to study long-term travel behavior. *Transportation Research Part C: Emerging Technologies* 104, 348.

Marra, A. D. and F. Corman (2020). Determining an efficient and precise choice set for public transport based on tracking data. *Transportation Research Part A: Policy and Practice* 142, 168.

Meyer de Freitas, L., H. Becker, M. Zimmermann, and K. W. Axhausen (2019). Modelling intermodal travel in Switzerland: A recursive logit approach. *Transportation Research Part A: Policy and Practice* 119, 200.

Molloy, J., T. Schatzmann, B. Schoeman, C. Tchervenkov, B. Hintermann, and K. W. Axhausen (2021). Observed impacts of the Covid-19 first wave on travel behaviour in Switzerland based on a large GPS panel. *Transport Policy* 104, 43.

Molloy, J., B. Schmid, F. Becker, and K. W. Axhausen (2019). mixl: An open-source R package for estimating complex choice models on large datasets. Working Paper. IVT, ETH Zurich.

Montini, L., C. Antoniou, and K. W. Axhausen (2017). Route and mode choice models using GPS data. *TRB Annual Meeting Online*. Vol. 1204, 17.

Nassir, N., M. Hickman, A. Malekzadeh, and E. Irannezhad (2015). Modeling Transit Passenger Choices of Access Stop. *Transportation Research Record: Journal of the Transportation Research Board* 2493, 70.

Nielsen, O. A., M. Eltved, M. K. Anderson, and C. G. Prato (2021). Relevance of detailed transfer attributes in large-scale multimodal route choice models for metropolitan public transport passengers. *Transportation Research Part A: Policy and Practice* 147, 76.

Phithakkitnukoon, S., Z. Smoreda, and P. Olivier (2012). Socio-Geography of Human Mobility: A Study Using Longitudinal Mobile Phone Data. *PLOS ONE* 7.6, e39253.

Prato, C. G. (2009). Route choice modeling: past, present and future research directions. *Journal of Choice Modelling* 2.1, 65.

Prato, C. G., T. K. Rasmussen, and O. A. Nielsen (2014). Estimating Value of Congestion and of Reliability from Observation of Route Choice Behavior of Car Drivers. *Transportation Research Record* 2412.1, 20.

Stadt Zürich (2021). Mikrozensus Mobilität und Verkehr 2015 - `www.stadt-zuerich.ch/ted/de/index/taz/verkehr/webartikel/webartikel_kennzahlen_verkehrsentwicklung.html`.

Tirachini, A. and O. Cats (2020). COVID-19 and Public Transportation: Current Assessment, Prospects, and Research Needs. *Journal of Public Transportation* 22.1.

Weis, C., M. Kowald, A. Danalet, B. Schmid, M. Vrtic, K. W. Axhausen, and N. Mathys (2021). Surveying and analysing mode and route choices in Switzerland 2010–2015. *Travel Behaviour and Society* 22, 10.

Xiong, X., S. Qiao, N. Han, F. Xiong, Z. Bu, R.-H. Li, K. Yue, and G. Yuan (2020). Where to go: An effective point-of-interest recommendation framework for heterogeneous social networks. *Neurocomputing* 373, 56.

Yap, M. and O. Cats (2021). Taking the path less travelled: Valuation of denied boarding in crowded public transport systems. *Transportation Research Part A: Policy and Practice* 147, 1.

Zhu, S., D. Levinson, H. X. Liu, and K. Harder (2010). The traffic and behavioral effects of the I-35W Mississippi River bridge collapse. *Transportation Research Part A: Policy and Practice* 44.10, 771.

Zimmermann, M. and E. Frejinger (2020). A tutorial on recursive models for analyzing and predicting path choice behavior. *EURO Journal on Transportation and Logistics* 9.2, 100004.

Zürich Statistic Office (2021). Statistik Stadt Zürich - `www.stadt-zuerich.ch/prd/de/index/statistik.html`.

7

CONCLUSIONS

7.1 MAIN FINDINGS

This thesis focuses on understanding passengers' behaviour in public transport from GPS tracking. A particular attention is given to public transport disturbances, their impact on passengers, and how passengers react to them. We formulated the objective of this thesis in the following research question:

How can tracking technology be exploited to understand passengers' behaviour in public transport, and in particular during different disturbances in the network?

We answered this question developing several algorithms and methods that, starting from raw GPS data of passengers and Automatic Vehicle Location data (AVL) of the operations, can derive precise information on passengers' travel behaviour. Moreover, analysing real-life network disturbances and observing passengers' route choices in those cases, we were able to understand how passengers' behave in case of disturbances.
Given the complexity of the research question, we divided it in different sub-questions, to each of which we answered in a different chapter. Below, we report each sub-question, with a summary of the proposed methods and the results obtained.

1) How to derive travel diaries from long-term GPS tracking, based on a smartphone application?

In Chapter 2, we developed a smartphone application, able to continuously collect GPS data, without a significant burden on the tracked users. In particular, the application is able to track users for a long period, without requiring any interaction other than its installation, and without significantly affecting the battery consumption. To achieve this, the GPS data are collected with a low sampling frequency. Therefore, specific algorithms need to be used to derive travel diaries from GPS data, able to deal with low-frequency data.
In particular, we developed algorithms for activity and trip identification,

trip segmentation, and mode detection. This latter is the main focus of our research, given its higher complexity. The main characteristic of the mode detection algorithm, which makes it stand out from other works in literature, is the ability to identify the exact vehicle used in each public transport stage (therefore mode, line, vehicle, and exact departure and arrival times). In fact, the algorithm exploits the AVL data of public transport operations, comparing the GPS points of the tracked user with the locations of all public transport vehicles running in the network. If a matching is identified at the origin and at the destination of the user's stage, the used public transport vehicle is detected. The mode detection algorithm is therefore unsupervised, since it is not based on an inference model, which would require a training set, based on an extensive manual labelling of users' movements. Instead, only an eventual tuning of the parameters is needed. An additional novel aspect of the proposed algorithm is the usage of past travel information. In fact, previously visited locations of a user are exploited to identify possible transfers in future public transport trips.
The smartphone application and the proposed algorithms were tested in Zürich and Basel. In particular, the mode detection algorithm obtained an average accuracy of 86.14%. The combination of a long-term tracking system not affecting the battery consumption, highly accurate algorithms for collecting travel diaries, and a mode detection algorithm identifying also the public transport vehicles, makes the proposed methodology particularly suitable to observe long-term travel behaviour of public transport passengers.

2) How to identify the available alternatives for a public transport trip, which more realistically were considered by a passenger, given a certain knowledge of current network conditions?

Chapter 3 answers this research question, proposing a novel choice set generation algorithm for route choice in public transport. The proposed algorithm is based on constrained enumeration, which means it identifies all available routes, given certain constraints on the maximum duration and the maximum number of transfers.
We evaluated the algorithm on public transport trips from a large-scale tracking study, based on three different aspects, in which the algorithm obtained outstanding performance. First, the computation time to generate one choice set is significantly short, with a median of 7 seconds on a standard computer. Second, the algorithm has a coverage above 94%, which

represents a great ability to identify the passengers' routes in the generated choice sets. Third, the algorithm is evaluated based on the quality of the estimation of a route choice model. In this regard, the estimation of a Path Size Logit model leads to significant estimated parameters, in-line with the values expected from the literature.
An important characteristic of the proposed algorithm is the modelling of the public transport network based on AVL data, which makes it possible to generate the alternatives based on different information provisions on the network conditions to the passengers. We evaluated three different information provisions (full information, partial information and no information) on the tracked passengers, to identify the information representing best their choices. On average, the absence of information resulted as the best fit, despite we identified high heterogeneity among the passengers.
An important aspect of route choice modelling rarely considered in literature is the choice set size. In fact, a choice set must contain all and only the relevant alternatives considered by the passenger. In this regard, we identified the right size as the minimum number of alternatives, which does not decrease significantly the performances (i.e. including more alternatives does not change the understanding of passengers' choices).
Finally, a further aspect we analysed is the distance travelled walking by passengers. In fact, we identified a large variation of it among different users and different trips of each user. This is particularly relevant, to define the maximum walking distance in choice set generation, since people may be willing to walk more or less for different trips. In this regard, considering both alternatives based on short walks and those based on long walks leads to a higher coverage, at the cost of more non-relevant alternatives included in the choice set.

3) What is the impact of public transport disturbances on passengers? What are the main characteristics of disturbances affecting their impact?

Chapter 4 answers this question, proposing a novel metric of disturbance impact, which quantifies the effect of any disturbance (named disruption in Chapter 4) from the passengers' point of view. This metric evaluates the impact of a disturbance on a certain origin-destination pair (OD), it is based on random utility theory and it represents the difference between the expected travel cost in case of disturbances with the expected one without disturbances. The major advantages of this metric are the applicability

to any disturbance (from small to large ones), and the evaluation of the impact on a specific OD. Therefore, the impact of a disturbance is assessed in relation to the origin and destination of the affected passenger.
To identify the main characteristics of disturbances affecting their impact, we based our analyses on real disturbances occurred in two Swiss cities, Zürich and Bern. We considered a disturbance as a combination of delays and cancelled arrivals of public transport vehicles, near each other in time and space. This flexible definition includes both small and large disturbances, allowing the comparison of disturbances with different characteristics. Therefore, we identified around 2000 disturbances from long-term AVL data, applying the clustering algorithm ST-DBSCAN. Analysing more than 25000 ODs starting in disturbed areas, based on random forest regression and feature importance metrics, we identified the main characteristics of disturbances affecting their impact on passengers.
Two of them are the service frequency and the choice set size (number of alternatives available). In fact, high values of these features are related with a lower impact of disturbances. Network metrics of the disturbed area, such as betweenness and closeness centralities, also play a key role in determining the disturbance impact. In particular, we identified that a disturbance in a peripheral area or in a hub has higher impact compared to one in an intermediate zone. The disruption impact increases also with higher delays, but until a certain value ($\approx$ 17 minutes in Zürich), after which it does not increase further.
Finally, among the less important features, there are those related to the destination. This shows the disturbance impact is mainly determined by the disturbed area. In that sense, the number of events involved in a disturbance was also found less important than the features describing the disturbed area.

4) How different network disturbances affect route choice of public transport passengers?

Chapter 5 answers this question, based on real observations of passengers' behaviour during network disturbances. To observe passengers' route choices in case of disturbances, we combined the methodology proposed in the three previous chapters. We derived travel diaries of several users from a large-scale tracking study, as described in Chapter 2. We identified the available alternatives of each public transport trip, in planned conditions and in case of disturbances, as described in Chapter 3. We evaluated

the service degradation during passengers' trips, adapting to the new context the metric of disturbance impact proposed in Chapter 4.
We observed the effects of disturbances on passengers in terms of both route choice and increased travel cost. As main outcome, we identified high heterogeneity among the effects of disturbances on the observed trips. Despite this, we observed the following general trend. A higher service degradation corresponds to an increased travel cost, compared to a condition of no disturbances. Similarly, a negative service degradation, referring to variations from the timetable leading to less costly alternatives, such as early departures, corresponds also to an increased travel cost. Regarding the route choice, with higher service degradation, passengers choose more likely the less costly alternative. In contrast, with negative service degradation, passengers do not exploit the new available alternatives.

5) Is it possible to understand travel behaviour during the COVID-19 pandemic from tracking data? How the route choice criteria differ from those in a pre-pandemic period?

Chapter 6 answers this research question, based on two tracking studies in Zürich. In particular, the same group of participants was tracked in both studies, one in 2019, before the pandemic, and the other in 2020, during the first pandemic wave. The availability of data of the same people, before and during the pandemic, gives a unique occasion to observe how travel behaviour changed. In fact, collecting travel diaries and understanding their route choice criteria, based on the methodology described in Chapter 2 and Chapter 3, we were able to compare the travel behaviour between the two different periods, and understand the effects of the pandemic on it.
The main observed effect of the pandemic is a strong reduction of the travelled distance, especially for public transport trips. The modal split was also affected, with the share of public transport decreasing, in favour of walk and private modes. These changes varied during the different phases of the pandemic. In fact, the lowest travel distance and public transport usage were observed with the strongest restrictions implemented by the authorities. In contrast, with the easing of the restrictions, these two values increased again.
Users' behaviour related to work was also particularly affected by the pandemic. In fact, we observed a strong reduction of the time spent at the workplace, while an increase of the time spent at home. Moreover, among the regular trips, commuting trips were the most affected, observing a

strong decrease of public transport usage, in favour of private modes. Estimating two route choice models, we identified important differences between the route choice criteria in 2019 and 2020, during the pandemic. The travel time in train, which is the least preferred mode in 2019, is perceived with a much lower cost in 2020. The transfer time is also perceived better in 2020, while the number of transfers has a higher cost. Moreover, heterogeneity was found among the users regarding perception of additional transfers. These identified differences clearly show that the pandemic influenced the route choice in public transport. Among the possible explanations, there is the importance given to safety and its perception by travellers. Anyhow, identifying how each specific consequence of the pandemic affected travel behaviour is left for a future work.

7.2 IMPLICATIONS FOR PRACTICE

Based on the analyses and results of this dissertation, in this section, we complement the contributions to society mentioned in Section 1.5.2, with the following implications and recommendations to the public transport industry.

In this thesis, we showed how automatic procedures and methods can derive from tracking data a variety of information on travel behaviour. In particular, we analysed passengers' behaviour in case of normal network conditions, network disturbances and during the COVID-19 pandemic. Regarding network disturbances, understanding passengers' behaviour required a series of steps, including: mode detection, choice set generation, route choice model estimation and evaluating the service degradation. The application of these methods, as proposed in this thesis, allowed observing passengers' choices over a long period and during different disturbances. The high level of detail of passengers' choices allows a better understanding of passengers' reaction to disturbances, and therefore a better planning of disturbance management by service operators.

It follows, as major implication for public transport industry, the proposed methods are a valid tool for transport analysts, to automatically understand passengers' behaviour. In fact, the proposed data-driven approach supersedes traditional surveys in several aspects, such as the duration of the data collection and the level of details of the collected information. Moreover, each method can be adapted individually to different contexts and the specific needs of the transport analyst, without affecting the other methods. For instance, different algorithms can be used to derive travel diaries, if Automatic Fare Collection data (AFC) are used, instead of GPS data; or a different choice set can be generated, if private modes should be considered.

Given the heterogeneity of the proposed methods, each analysis we performed has important implications on its own.

In Chapter 2, we showed how GPS data can be collected and exploited to understand passengers' behaviour in public transport. In particular, we tested a smartphone application and a mode detection algorithm to automatically collect long-term travel dairies, without a significant burden on the users. The high accuracy of the mode detection algorithm, and the realistic and significant results achieved in behavioural analyses (in the following chapters) show the validity of our methods in practice. Moreover, the algorithms proposed in Chapter 2 are directly applicable by public transport operators, to collect long-term travel diaries. In fact, the mode detection algorithm is not bound to our case study, and it can be applied in any city providing locational data of the operations. Only a tuning of the parameters may be required. We see the conduction of regular tracking studies by public transport operators as a potential application of our methods. Such studies allow to monitor the passengers in the system, and therefore to better understand their needs and adapt the provided service consequently. Moreover, the costs of these studies would be limited, since the data processing can be fully automatized, and the participants may be recruited proposing offers or discounts on the public transport service.

In Chapter 3, we proposed a choice set generation algorithm. Besides finding application for research and modelling purposes, the proposed algorithm can be applied by public transport operators for route recommendation. In fact, the high coverage obtained (above 94%) shows a high precision in identifying the route chosen by the users. Moreover, the very low computation time makes possible using the algorithm in route recommender systems in daily life. We analysed choice sets based on different information provisions of network conditions, to identify the one representing best the passengers' behaviour. Such analysis is particularly relevant for service providers, since it indicates based on what information passengers consider their alternatives. In our test case, a condition of no information performed best, which may indicate a need of more information for passengers or better recommendations to be provided. In this regard, the same analysis can be performed independently for each user, identifying which one is less informed or might benefit from targeted recommendations.

Regarding public transport disturbances, in Chapter 4, we proposed a metric to quantify their impact on passengers, and we identified the characteristics of disturbances related to a higher impact. Knowing these characteristics is particularly important for service providers, both for transportation planning, disturbance management and reliability analysis. In our test case, we identified the service frequency, the number of alternatives available, and the position in the network, as the main factors reducing the impact of potential disturbances. Therefore, service providers may adapt their transport service in accordance with this information, to increase its reliability. For instance, in stations considered particularly important or vulnerable, they may increase the service frequency, the connections with other stops, or the number of public transport lines running. Moreover, with the proposed methodology, the effects of such measures on passengers can be quantified, allowing the comparison of different measures for disturbance management. Among the less important characteristics of a disturbance, we identified the number of disturbed events (arrival of a vehicle at a stop). This highlights to service providers that the size of a disturbance is not particularly important, while a major attention must be given to poorly connected locations, providing few alternative routes and are therefore more vulnerable to disturbances. Hence, to provide a more resilient service, providers should increase the number of alternatives available in less connected locations.

Following this analysis, in Chapter 5 we observed the effects of different network disturbances on route choice of public transport passengers. One of the main results we identified is a high heterogeneity among the effects of disturbances, in terms of increased travel cost for passengers. This is particularly relevant for disturbance management, since this heterogeneity must be taken into account to evaluate the effects of disturbances and plan mitigation actions accordingly. Analysing "good disturbances", which we refer as deviations from the timetable leading to less costly alternatives, such as early departures, we identified an increased travel cost for passengers. This type of disturbance is rarely considered in practice and in scientific literature, although our analysis shows it should be considered as much as a normal disturbance. In this regard, our analysis shows passengers choose the less costly alternatives in case of normal disturbances, while more costly ones in case of "good disturbances". Therefore, a recommendation for service providers resulting from this analysis is that in case of normal disturbances, operations should be regularized and there is not a particular need of informing the passengers; in contrast, in case of "good

disturbances", either passengers should be informed of the new available connections or operations should be regularized.
Finally, in Chapter 6, we analysed travel behaviour during the COVID-19 pandemic, observing important variations in the travel distance, mode share and route choice, compared to a pre-pandemic period. Given the small and recent literature on travel behaviour during a pandemic, this study contributes to its understanding, and helps decision makers on planning and evaluating policies. For example, the observed reduction in travel distance during the pandemic helps understanding the effects of restrictions on mobility. Moreover, the observed variations in the route choice criteria helps predicting the traffic flow during the pandemic, and thus adapting the transport service, for instance to reduce crowding.

7.3 LIMITATIONS OF THE RESEARCH

In addition to the research boundaries and assumptions described in Section 1.3, the proposed research has further limitations, which must be considered by researchers or experts interested in reproducing the results.
In this research, we conducted several tracking studies in a university environment, informing participants about data collection and usage. Therefore, practitioners who wish to apply similar methods must consider regulations on privacy and transparency.
Regarding the data collection, different smartphones provided GPS data with different quality, due to the large heterogeneity of smartphones available on the market and the relative software installed. It follows that the data collection (and the following mode detection) might not work properly for certain users.
To observe travel behaviour in public transport, we proposed a series of methods to apply in sequence (GPS tracking, mode detection and choice set generation). Despite the high precision of the proposed methods, each one has a small percentage of error, which can be accumulated during the data processing. Therefore, a high accuracy in the first steps of data processing is necessary to have reliable results on the following analyses. In the proposed mode detection algorithm, the accuracy for train detection is lower than for other modes, due to the lower GPS quality inside trains (68.3%). Despite train trips represent a minority of the observed public transport trips in this work (10%), we recommend to improve the train detection in future works.
To study route choice in public transport, in particular in Chapter 3 and

Chapter 5, we relied on a series of assumptions. First, we assumed the route choice process starting at the ending time of the previous activity (i.e. the last GPS point). This may not always be true, since people may wait in their current location a scheduled public transport vehicle. Second, despite constraints in a choice-set generation algorithm (e.g. maximum number of transfers) are necessary to drastically reduce the computation time (at the expense of discarding $\approx 5\%$ of observations, as subject to unmodelled situations), they may create endogeneity and bias the coefficients during the model estimation. Third, we did not take into account the trip purpose of passengers, since it was not available in the data. This information may improve our analyses, since different trip purposes may correspond to different choice criteria for passengers.
Finally, we remark that our analysis of travel behaviour during disturbances is conducted in Zürich, a city with a very reliable public transport service, which rarely faces large disturbances. In fact, during the three tracking studies, no large disturbances occurred, affecting several participants. We therefore believe that repeating the same analyses in a city with an unreliable service may provide more insights about it.

7.4 FUTURE RESEARCH

GPS tracking is a relatively modern technology, which only in the last years has been applied to collect travel diaries. In fact, with the broad diffusion of smartphones, the collection of GPS data became much easier, given that people were already carrying a GPS device with them. For this reason, there is plenty of room for research in the analysis of travel behaviour based on this technology. Moreover, given the heterogeneity of the proposed methodology in this dissertation, we identified several future research directions, which can extend and improve the current work:

Exploiting long-term travel diaries
With the increasing quality of GPS data and of the methods for deriving travel diaries from them, there is plenty of room for research on analysing long-term travel diaries with detailed travel information. In fact, in this thesis, we focused on travel behaviour and in particular on route choice in case of disturbances. Nevertheless, the same smartphone application and data processing proposed in Chapter 2 can be exploited for other studies, requiring long-term travel diaries. A possible research direction is given by analysing sequences of activities. Observing which activities users per-

form and in which order has several potential applications, ranging from urban planning (Jiang et al., 2012) to understanding human behaviour (Horanont et al., 2013) or health management (Chiang et al., 2014). Related to this, long-term travel diaries can be used to study the regularity of human mobility (Mucelli Rezende Oliveira et al., 2016), and therefore to identify anomalies in users' behaviour in case of special events.

Improving the mode detection algorithm
The mode detection algorithm presented in Chapter 2 stands out among the ones available in literature for being able to identify the exact public transport vehicle used, and for not being primarily based on machine learning. Therefore, it is worth to study a new mode detection algorithm, which integrates the two methodologies. Such an algorithm may exploit GPS and AVL data to identify the public transport vehicles (as the one proposed), and machine learning based on previous information about the user to improve the detection. For instance, if a user travels regularly to a certain destination, information of previous trips can be used to adapt the likelihood function described in Section 2.7.1. In any case, the mode detection algorithms described in literature are all heuristics, whose performance depends on the test case and the quality of the GPS data. Therefore, we encourage further research on this field, exploring alternative methods.

Exploiting longitudinal data to predict route choice
In this thesis, we analysed route choice in public transport using random utility models, as the Path Size Logit. While such a model is useful to understand the route choice criteria, machine learning approaches may be better suited for route prediction. In this context, the travel diaries collected in this study can be used as a training set for a route prediction model. This model may exploit the longitudinal nature of the dataset (e.g. the previous choices of the same user), socio-demographic information, or external information (e.g. the weather conditions). This stream of research is particularly new and it is emerging in the last years in the context of mode choice, while not yet in route choice. For instance, Wang et al. (2020) proposes a neural network with alternative-specific utility for mode choice. Similarly, Sifringer et al. (2020) proposes a neural network, including a data-driven parameter in the utility, which outperforms traditional choice models.

Relevance of paths in a choice set
In Chapter 3, we analysed the size of a choice set and when a path can

be considered relevant and included in it. We identified the right size of a choice set based on a trade-off with its performance (i.e. the minimum size that guarantees high performance). Nevertheless, we acknowledged the relevance of a path is very subjective and the number of alternatives considered can vary among users and among trips with different characteristics. We are not aware of any work investigating the choice set size in detail, despite it can affect the quality of models' estimation (Zimmermann and Frejinger, 2020), and its importance in practice (for instance in route recommender systems). Therefore, a future work may focus on identifying which alternatives the users consider relevant. Such work can also be based on tracking, including regular surveys asking about the alternatives considered for each trip. In this regard, the surveys can be visually integrated in the smartphone application, to encourage a response.

Exploring the impact of disturbances in different contexts
In Chapter 4, we analysed the impact of public transport disturbances on passengers and identified their main characteristics. In particular, we focused on small disturbances in multi-modal networks in Switzerland. Given the heterogeneity of the various public transport networks and the possible types of disturbances, we recommend for future work to perform a similar analysis in different contexts. In particular, in case of railway networks, we expect different results, such as a higher importance identified to the location of the disturbance in the network. In fact, the lower accessibility of the stations in a railway network makes more difficult for passengers to move from a station with disturbances. Moreover, the types of railway disturbances can be significantly different from the ones in a multi-modal network. For similar reasons, we expect different results in public transport networks providing a significantly different quality of service, compared to the ones analysed.

Inference of the information available to passengers
Until now, research works model passengers' behaviour assuming a certain information available to them (Gentile and Noekel, 2016; Leng, 2020). This is particularly useful, for instance, to estimate the effects of informing (or not) passengers of a disturbance. In practice, it is not known if passengers are informed (or not), if the right information is provided to them, if they use it, and how they use it. In this regard, a novel stream of research may focus on understanding the information available to (or used by) passengers from realized observations. Knowing if passengers are cor-

rectly and promptly informed about disturbances in the network, and if they use this information, is particularly useful for service providers, to understand the effectiveness of their information systems. The analyses in Chapter 3 and Chapter 5 go in this direction, comparing the passengers' choices with those expected knowing or not about the disturbances. Therefore, a possible future work may be based on a large-scale tracking dataset, where information is provided in different ways, and passengers' choices are analysed to understand if and how they use the provided information.

Analysis of large disturbances
In this work, we focused on small disturbances in a public transport network, while we did not analyse large disturbances, such as a big service interruption. We believe the main obstacle for such analysis is the data collection of realized data of passengers, given the rarity of large unplanned disturbances. In fact, disturbances with higher impact are typically rarer (Yap, 2020). For this reason, the literature focusing on large disturbances is based either on simulations (Leng, 2020), on one or few disturbances (Sun et al., 2016), or on planned disturbances (Yap et al., 2018). Therefore, for a future work on passengers' behaviour during large disturbances, a methodology similar to that of this thesis can be applied, while a different attention must be paid to data collection. This last one should involve a large amount of passengers, and span several months (or years), to observe a sufficient number of unplanned disturbances. Moreover, the burden on passengers and the economic costs should be minimized. AFC systems are a technology that meets these requirements. Unfortunately, they are not available in all cities and they monitor the passengers only within the transport system. On the contrary, GPS tracking can be used potentially everywhere and can observe also walks and activities, but at the cost of a more difficult recruitment of participants. Finally, we remark privacy issues are possible for such a long study period, such as the need to anonymize AFC data, which does not allow studying regular trips.

Further research on travel behaviour during a pandemic
The COVID-19 pandemic is an exceptional event that affected the travel behaviour of the world population. At the time of writing, the pandemic is not over yet, and we do not know when it will last and what its future phases will be. For this reason, and given the heterogeneity of how the pandemic evolved in different countries, there is not yet a clear understanding of travel behaviour during a pandemic. In this thesis, we shed some light

on travel behaviour during the first wave in Switzerland, with a focus on route choice. Therefore, for future works, we encourage exploring other aspects of travel behaviour during the different phases of a pandemic, such as the trip purpose, activity sequences and daily routines. In this regard, GPS tracking is a valid technology to collect highly detailed information on passengers. Finally, a possible future work can be based on a further tracking study, to compare a pre-pandemic period with a post-pandemic one, and to observe possible long-term changes in travel behaviour.

REFERENCES

Chiang, J.-H., P.-C. Yang, and H. Tu (2014). Pattern analysis in daily physical activity data for personal health management. *Pervasive and Mobile Computing* 13, 13.

Gentile, G. and K. Noekel, eds. (2016). Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems. Springer Tracts on Transportation and Traffic. Springer International Publishing.

Horanont, T., S. Phithakkitnukoon, T. W. Leong, Y. Sekimoto, and R. Shibasaki (2013). Weather Effects on the Patterns of People's Everyday Activities: A Study Using GPS Traces of Mobile Phone Users. *PLOS ONE* 8.12, e81153.

Jiang, S., J. Ferreira, and M. C. Gonzalez (2012). Discovering urban spatial-temporal structure from human activity patterns. *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*. UrbComp '12, 95.

Leng, N. (2020). The role of information to passengers in public transport disruptions. Doctoral Thesis. ETH Zurich.

Mucelli Rezende Oliveira, E., A. Carneiro Viana, C. Sarraute, J. Brea, and I. Alvarez-Hamelin (2016). On the regularity of human mobility. *Pervasive and Mobile Computing* 33, 73.

Sifringer, B., V. Lurkin, and A. Alahi (2020). Enhancing discrete choice models with representation learning. *Transportation Research Part B: Methodological* 140, 236.

Sun, H., J. Wu, L. Wu, X. Yan, and Z. Gao (2016). Estimating the influence of common disruptions on urban rail transit networks. *Transportation Research Part A: Policy and Practice* 94, 62.

Wang, S., B. Mo, and J. Zhao (2020). Deep neural networks for choice analysis: Architecture design with alternative-specific utility functions. *Transportation Research Part C: Emerging Technologies* 112, 234.

Yap, M. (2020). Measuring, Predicting and Controlling Disruption Impacts for Urban Public Transport. PhD Thesis. Delft University of Technology.

Yap, M., S. Nijënstein, and N. van Oort (2018). Improving predictions of public transport usage during disturbances based on smart card data. *Transport Policy* 61, 84.

Zimmermann, M. and E. Frejinger (2020). A tutorial on recursive models for analyzing and predicting path choice behavior. *EURO Journal on Transportation and Logistics* 9.2, 100004.

8 APPENDIX

This appendix contains additional details to specific passages in the Thesis, marked with the superscript [Thesis Appendix]. These comments were not added directly in the text, to not modify the published articles.

1. We used a different dataset for validation, given the availability of ground truth data.
2. The public transport operational data are assumed reporting correct information on actual arrival and departure times, for all chapters of the thesis.
3. We refer to the following constraints mentioned later in the same Section: walking distance is $\leq N$; the waiting time is $\leq TD$; for each possible transport line, only the best transfer is considered.
4. Fiorenzo-Catalano (2007) affirm that in route choice in transport networks it is expected travellers consider a choice set with a limited number of alternatives, since the number of routes that might be known, considered or used is limited.
5. In other terms, one minute in the tram is perceived as less penalizing than one minute in other vehicles.
6. We consider a characteristic of a disruption important, when it affects more the impact of the disruption compared to other characteristics.
7. The random forest model has in input the list of features in Table 4.1 and as target value the disruption impact.
8. The feature importance metrics should be used to identify which characteristics affect most the impact of a disruption.
9. In practice, the proposed framework can be used to estimate the impact of current or possible future disruptions. Instead, the feature importance metrics in Table 4.1 show a ranking of the features, in terms of their influence on the disruption impact. Based on those metrics,

operators can intervene on specific features to reduce the disruption impact (e.g. increasing the service frequency or the closeness centrality).