Regional Travel Time and Transfer Impacts of the Noord/Zuidlijn using Interoperable Smart Card Data

S.M.H. van Hees

Technische Universiteit Delft

Gemeente Amsterdam

TUDelft

Regional Travel Time and Transfer Impacts of the Noord/Zuidlijn using Interoperable Smart Card Data

by

S.M.H. van Hees

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Thursday July 22, 2021

Student number:	4398807	
Project duration:	November, 2020 – July, 2021	
Thesis committee:	Prof. dr. ir. S.P. Hoogendoorn,	TU Delft, chair
	Dr. ir. N. van Oort	TU Delft, supervisor
	Dr. ir. S. van Cranenburgh	TU Delft, supervisor
	Dr. ir. T. Brands	TU Delft, daily supervisor
	Dr. ir. M. Kobus	Gemeente Amsterdam, external supervisor
	Ir. R. Veldhuijzen van Zanten	Gemeente Amsterdam, external supervisor

An electronic version of this thesis is available at http://repository.tudelft.nl/.

Cover image obtained from: https://www.instagram.com/noordzuidlijn/

Preface

Dear reader,

In front of you lies the thesis that marks the end of my 7-year journey at the TU Delft. During this period I learned a lot and grew as a person. It was a period in which I met a lot of amazing new people, some of which became very good friends, and made wonderful memories in Delft.

I would like to thank my graduation committee for their supervision during my thesis. It has not always been easy to write my graduation thesis during the COVID-19 pandemic. The lack of a sparring partner, hours and hours behind a screen, the absence of informal meetings and most importantly not being able to blow off steam as 'normal' life stood still. Despite that, you reading this preface, means that I have managed to finalize my MSc degree successfully. This would not have been possible without the supervision of my graduation committee. I would like to thank Martijn and Rutger for providing the opportunity to undertake this research at the Gemeente Amsterdam. Martijn, thanks for helping me find a pragmatic solution every time I got stuck. Without your help I would have kept rethinking all the choices that had to be made over and over. Rutger, thanks for your interesting questions from a public transport policy perspective. This has kept me motivated. Ties, thank you for introducing me to this topic and guiding me through the unexplored world of interoperable smart card data. The in-depth discussions we had were very valuable. Niels, thank you for your optimism. I always ended up more positive after a meeting with you. Sander, thank you for your critical review of my work. It improved the structure of my report drastically.

Additionally, I would like to thank my friends for listening to me whenever I had something to complain about, for proofreading my thesis and for relaxing during coffee or lunch breaks, when having dinner or while drinking a beer at the terrace.

Last, I would like to thank my father, mother, sister and girlfriend for their support during the writing of my thesis and the rest of my studies.

I'm grateful for the time I had in Delft. At the moment I do not know what the future will bring but the knowledge I gained and the friends and memories I made will stay with me forever.

Have fun reading!

S.M.H. van Hees
Delft, July 2021

Summary

On the 22nd of July 2018, the Noord/Zuidlijn became operative. This meant a large-scale change in the public transport network of Amsterdam. A metro line crossing the city center and connecting the northern and southern part of Amsterdam replaced a network of direct bus and tram lines to Amsterdam Centraal station. Many of the bus and tram lines that previously ran to Amsterdam Centraal became feeder lines to the new metro line.

The expected impacts of such large-scale overhauls are numerous and include among others environmental, economic, spatial and transportation impacts. However, the actual impacts often are not evaluated thoroughly. Especially given the fact that ex-ante evaluations are often too optimistic (Van Wee and Tavasszy, 2008), investigating the actual impacts provides useful insights in the use of public transport funds. This is valuable for the appraisal and planning of future large-scale network overhauls.

However, because public transport network overhauls of this proportion do not occur often, the opportunities to assess the impacts ex-post are limited. Therefore, the Vervoerregio Amsterdam, the Gemeente Amsterdam and several scientific institutions collaborated to seize the opportunity to assess the impact of the Noord/Zuidlijn. This thesis contributes to that study by investigating the impacts of the Noord/Zuidlijn on regional travel times and transfer patterns. It is a follow-up of a study by Brands et al. (2020) that investigates the transportation impacts on the urban network level.

To gain insight in the travel time and transfer impacts on a regional scale interoperable smart card data is necessary. This data captures the complete public transport journey and travel behavior across services from multiple operators. This data source is promising, but relatively new and unknown.

This research has a main, societal and an additional, methodological objective. The first is to investigate the impact of the Noord/Zuidlijn on regional travel times and transfer patterns. The second is to explore the strengths and weaknesses of interoperable smart card data. The societal objective results in the following research question:

What impact does the Noord/Zuidlijn have on travel times and transfer patterns on regional journeys in the Vervoerregio Amsterdam and what does this imply for public transport planning?

Network changes accompanying the Noord/Zuidlijn

The Vervoerregio Amsterdam consists of four concession areas: Amsterdam, Amstelland-Meerlanden, Waterland and Zaanstreek. Concession Amsterdam consists of the urban public transport network and the latter three entail the regional public transport network. This thesis focuses on journeys between the concession Amsterdam and the regional concession areas. In these regional concession areas bus lines that previously ran to Amsterdam Centraal station were rerouted after the Noord/Zuidlijn. A schematic overview of the of the regional bus network after the Noord/Zuidlijn is given in Figure 1.

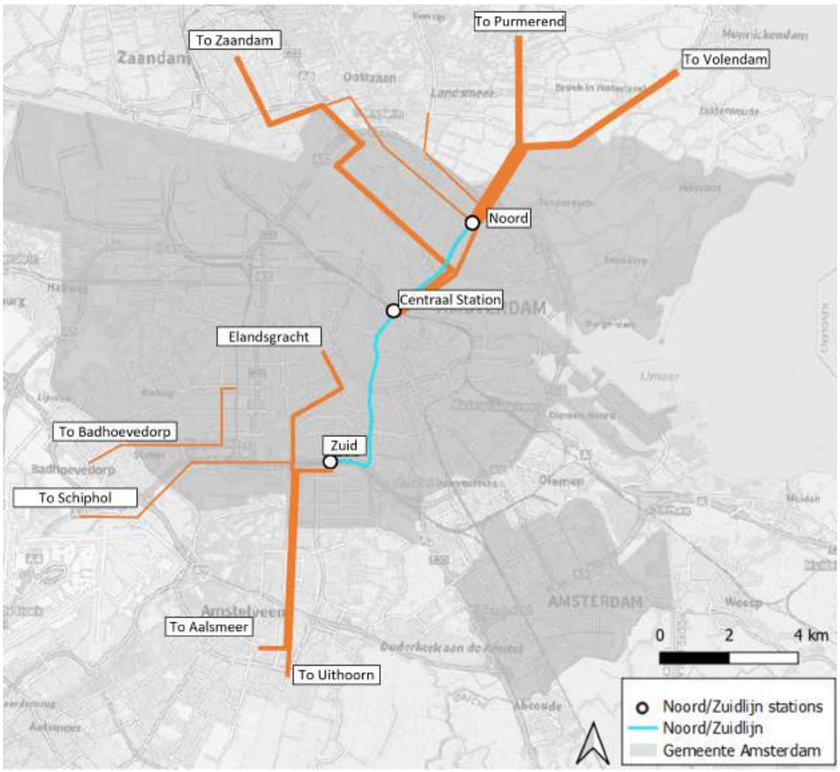


Figure 1: Configuration of the regional bus network to Amsterdam after the Noord/Zuidlijn. Adapted from Stadsregio Amsterdam (2014).

Data and methodology

In total, four data sets have been used for this study. Two contain the travel times between origin-destination (OD) relations over all half hour blocks in a day and two contain all routes (up to five trips), traveled between OD-relations. Of each type of data sets two sets have been used to compare the situation before and after the Noord/Zuidlijn.

An important limitation of these data sets is that because of privacy reasons the number of travelers is binned if it is less than 300. This has three important consequences. First, for the vast majority of records the number of travelers is not known exactly. Second, for all records where the number of travelers falls in the lowest bin (1-50 travelers) the representativeness of the travel time or route is uncertain. The share of records this applies to is over 90%. Three, it is hard to distinguish between the weights assigned to records for which the number of travelers falls in the same bin.

The first limitation has been overcome by estimating and validating an average value for each bin based on the trip totals from external data of the Vervoerregio Amsterdam. Thereafter, this average value has been split in a lowerbound and estimated number of travelers. The second limitation has been coped with by categorizing the results in different levels of representativeness. The idea behind the level of representativeness for a certain OD-relation is that if at least a certain lowerbound number of travelers has traveled, more value can be assigned to the average travel time or number of transfers on an OD-relation.

Travel time results

Overall, 1,350 hours of travel time are saved on regional journeys on a working day. On average the daily travel time benefits per traveler are around 50 seconds. The travel time savings and losses are 2,350 and 1,000 hours per day respectively. A larger positive impact than negative impact can also be seen in Figure 2 and Figure 3. Besides, it becomes clear that the travel time impacts are of larger magnitude and are more often negative or positive when analyzed on OD-level than one the individual level.

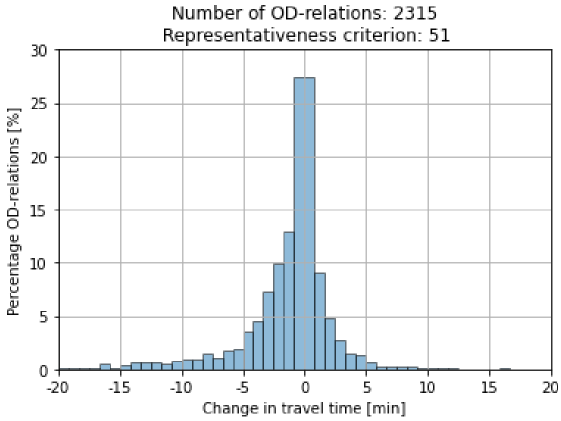


Figure 2: Distribution of the absolute change in travel time over OD-relations.

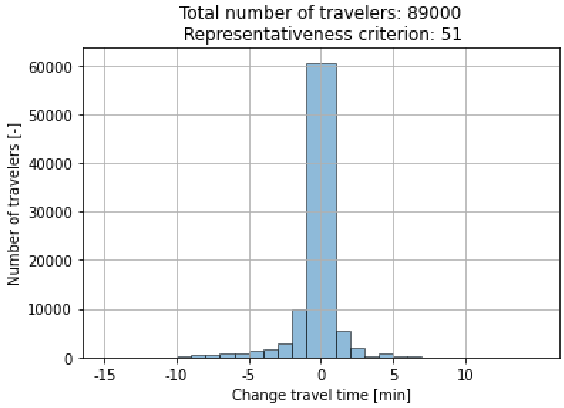


Figure 3: Distribution of the absolute change in travel time over travelers.

The largest decreases in travel time occur on OD-relations where a large part of the journey is replaced by a trip with the Noord/Zuidlijn. Travel times increase on OD-relations to areas in the city center that lose a direct connection after the Noord/Zuidlijn. To these areas, the metro trip is not sufficiently long enough to make up for the additional transfer time.

Transfer results

On average the number of transfers on a regional journey increases from 0.54 to 0.66 after the Noord/Zuidlijn. As a result of the Noord/Zuidlijn, the number of transfers made increases with 2,500 transfers per day. This is equal to 0.03 transfers extra per traveler per day. On a working day, on average the transfer losses are 7,000 transfers extra while the transfer gains are 4,500 transfers less. In Figure 4 it can be seen that on most OD-relations the impact of the Noord/Zuidlijn on the number of transfers is minimal. Besides, the share of OD-relations that experiences negative transfer impacts is larger than the share that experiences positive transfers impacts. Figure 5 shows that the share of travelers affected is smaller.

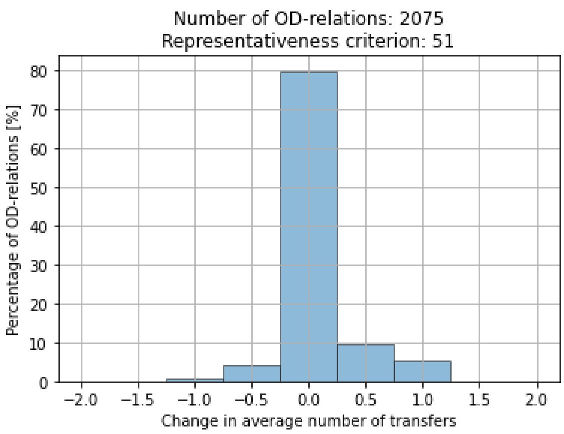


Figure 4: The distribution of the change in number of transfers over OD-relations.

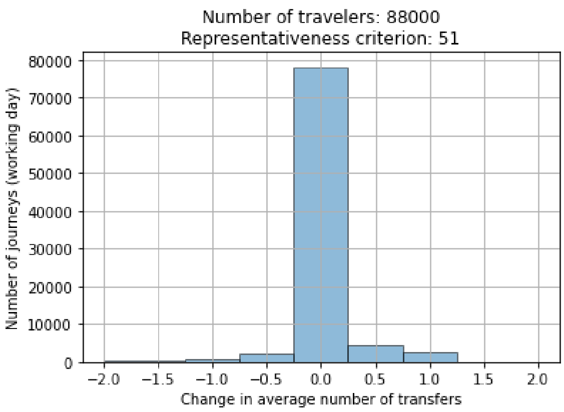


Figure 5: The distribution of the change in number of transfers over travelers.

The increases in the number of transfers occur on OD-relations where a direct connection to the city center is lost. The decreases in number of transfers can be ascribed to the new bus stop IJtunnel and the seemingly direct new metro journeys from Amstelveen to the city center.

Conclusion, discussion, recommendations

The net travel time and transfer benefits on regional journeys as a result of the Noord/Zuidlijn are €20,500 per day. The composition thereof can be found in Table 1.

Table 1: The regional societal travel time and transfer benefits of the Noord/Zuidlijn per working day.

	Quantity	**Monetary value per unit**	**Societal benefits per day**
Travel time savings	1,350 hours	18 €/hour	€24,000
Transfer savings	-2,500 transfers	1.4 €/transfer	€-3,500
Total	-	-	**€20,500**

From this research three important implications regarding public transport planning can be concluded. First, in a network overhaul that requires a transfer to a faster mode in the after situation, the trip leg on the faster mode should be sufficiently long to make up for the time lost by an additional transfer. Second, the distribution of the impacts could be different from an aggregate and disaggregate viewpoint. Third, the impacts differs if assessed on OD-level or individual level.

Regarding the methodological objective, the main strength of interoperable smart card data is that the whole public transport journey is recorded including transfers between operators. Besides, it provides insights in the distribution and exchange of travelers between operators serving overlapping areas. This enables a better understanding of travel behavior in multi-operator public transport networks. However, because of privacy reasons the number of travelers given in bins below a certain threshold value which results in three complications. First, if the number of travelers falls in the lowest bin, the representativeness of the record is uncertain. Second, the number of travelers the results apply to remains unknown. Third, it is not possible to distinguish between the weights assigned to records for which the number of travelers falls in the same bin. A methodology that enables working with interoperable smart card data despite these weaknesses has been developed in this thesis.

The data used results in several limitations of this research. Smart card data in general has some limitations: it does not contain all trips made, it does not include access and egress time, it does or does not include waiting time and no intramodal transfers are visible for metro and train. Another limitation is that the data before and after the Noord/Zuidlijn is aggregated over different months. Furthermore, the methodology to work with interoperable smart card data contains some limitations. The estimation and validation of the average values and the equal weighting thereof are inaccuracies that affect the results. Last, external factors that could have affected public transport demand and supply have not been excluded from the research.

A recommendation both for practice and further research is to investigate the possibilities to reduce the number of records that ends up in a bin, especially the same and lowest bin. Besides, regarding public transport planning it is advised to connect the feeder lines to the trunk line as early as possible. Additionally, the trip on the metro leg should be long enough to make up for an additional transfer. Furthermore, the distributive effects of a public transport overhaul should be incorporated in ex-ante evaluations.

This research could be extended by including the effects of the network overhaul on access and egress times. Next, the perception of travelers on the impacts could be included and compared to the measured impacts. Third, the impacts of the overhaul on national scale and on reliability of travel times should be included. Last, the passenger characteristics could be included to analyze the impact of the overhaul from an equity perspective.

Contents

List of Figures

List of Tables

List of Abbreviations

abs	absolute
AFC	Automatic fare collection
APC	Automatic passenger count
AVL	Automatic vehicle location
BTM	Bus, tram and metro
CXX	Connexxion
EBS	Egged Bus Systems Public Transportation BV
est	estimated
GTFS	General Transit Feed Specification
GVB	Gemeentelijk Vervoerbedrijf
KiM	Kennisinstituut voor Mobiliteitsbeleid (Eng: Netherlands Institute for Transport Policy Analysis)
lb	lower bound
MRA	Metropoolregio Amsterdam (Eng: Metropolitan Area Amsterdam)
NS	Nederlandse Spoorwegen (Eng: Dutch Railways)
NZL	Noord/Zuidlijn
OD	Origin-destination
OV	Openbaar vervoer (Eng: Public transport)
rel	relative
TLS	Translink Systems
VRA	Vervoerregio Amsterdam (Eng: Transport Authority Amsterdam)

List of Symbols

abs	Absolute
B	Benefits
d	Destination
Δ	Change
est	Estimated
h	Half-hour block
lb	Lower bound
np	Number of passengers
nt	Number of transfers
o	Origin
p	Period
r	Route
R	Set of routes
rel	Relative
tt	Travel time
y	Year
$\bar{\ }$	Average

1 Introduction

After decades of decision-making and construction issues, the Noord/Zuidlijn (see Figure 1.1) finally became operative on July 22nd, 2018. The new metro line connects the southern part of Amsterdam with the northern part of the city. It runs from train station Amsterdam Zuid via train station Amsterdam Centraal to metro station Noord. It crosses the city center and the IJ, the river that separates the center and northern part of Amsterdam, in a tunnel. However, the addition of the Noord/Zuidlijn to the public transport network of Amsterdam came with a major change in other public transport services. Several direct bus and tram services to Amsterdam Centraal that had become duplicate, were removed or rerouted. Besides, feeder services to the Noord/Zuidlijn were added (Stadsregio Amsterdam, 2014).

The costs of the Noord/Zuidlijn were huge, namely 3.1 billion euros (Vaillant, 2019). However, the expected impacts of the network change on the Amsterdam region should compensate for this. On forehand these were plenty and can be categorized based on the 5E-framework as proposed by Van Oort et al. (2017). The Noord/Zuidlijn connects the Northern and Southern part of the Amsterdam region effectively. It adds metro capacity to the public transport network and led to a decrease in the amount of bus and tram services in the city center resulting in a more efficient city (Witteveen, 2018). Besides, the Noord/Zuidlijn was expected to positively affect the average travel times, reliability, accessibility and comfort levels for the public transport user (Van Dijk et al., 2009). However, these are expected impacts and because a positive bias is often prevalent in ex-ante evaluation of public transport projects (Van Wee and Tavasszy, 2008) it is interesting to investigate if the goals have indeed been met.

The Noord/Zuidlijn and accompanying changes to Amsterdam's public transport network offer a unique opportunity to measure the transportation, spatial and economic impacts of a large-scale public transport network overhaul. To investigate these impacts a research collaboration between public and scientific institutions was initiated focusing on four themes: public transport, mobility and accessibility, public space and liveability and spatial economics. This thesis contributes to the public transport theme, but is intertwined with the other themes as well.

Regarding transportation impacts, media reports indicate that not all passengers benefit from the Noord/Zuidlijn and that some experience longer travel times and/or more transfers (Khaddari, 2019). From a policy perspective, it is interesting to investigate whether these negative impacts occur often or how they relate to the positive impacts. Brands et al. (2020) study the direct transportation impacts of the Noord/Zuidlijn and accompanying changes. However, their research is limited to the urban public transport network of Amsterdam while it can be expected that the transportation impacts are not limited to that area. This thesis is a follow-up to that research.

To add to the findings by Brands et al. (2020) on urban scale, this research focuses on the regional transportation impacts. It entails the travel time and transfer impacts of the Noord/Zuidlijn because the network overhaul is expected to have affected these. To assess the regional travel time and transfer impacts interoperable smart card data is necessary. This is a relatively new, but promising data source. It enables analysis of a complete public transport journey including trips made with services from multiple operators. This provides insights in the exchange of travelers between services from different operators as a result of the Noord/Zuidlijn.

Insights in the transportation impacts of the Noord/Zuidlijn could be used as input for future, similar public transport investments like the extension of the Noord/Zuidlijn to Schiphol Airport or an Oost/Westlijn in Amsterdam. Furthermore, policy-makers and researchers worldwide could learn important lessons of an ex-post evaluation of a change in public transport services. Because such large-scale network changes do not happen often, evaluations of these are uncommon as well. In scientific literature only a limited amount of ex-post evaluations of large changes in public transport services can be found.

1.1. Societal relevance

The insights obtained from this study are valuable as they will give an indication on the magnitude and distribution of the transportation impacts of a large-scale network overhaul. The outcomes could lead to valuable considerations regarding the planning and assessment of future large-scale network overhauls. Besides, the results of this study can partially determine whether the investments, made using public funds, were spent well or if the expected impacts have indeed occurred. This could lead to new or adjusted policy measures if necessary.

1.2. Scientific relevance

The scientific relevance of this research is two-fold. First, it explores the strengths and weaknesses of interoperable smart card data. This data source is unique because it combines trips by different operators into journeys. This offers huge opportunities because it enables analysis of multi-operator journeys. Currently, little is known about the pros and cons of interoperable smart card data. To the best of the author's knowledge, this research is one of the first using aggregated, interoperable smart card data. Second, this research adds to the limited amount of ex-post analysis of large-scale public transport network overhauls.

1.3. Scope

The scope of this research can be divided in a spatial and temporal scope.

Spatial scope

This research focuses on regional public transport journeys by bus, tram and metro (BTM) and train in the Vervoerregio Amsterdam (VRA) in the Netherlands. A regional journey is defined as a journey from concession Amstelland-Meerlanden, Waterland or Zaanstreek to concession Amsterdam or vice versa. It contains unimodal and multimodal trips operated by multiple service providers. A visualization thereof can be found in Figure 1.2.

Temporal scope

On July 22, 2018 the Noord/Zuidlijn became operative. To assess, the impact thereof a period before and after that date are compared. Additionally, the focus is on working days (excluding school vacations) to exclude large differences in demand and travel times between weekends and weekdays. Weekdays are expected to be most representative and interesting because public transport demand is then highest and working days have the largest share of all week days. A distinction in part of the day is made because demand and supply of public transport differ strongly during the day.

Figure 1.1: The trajectory of the Noord/Zuidlijn (Claus, 2019).

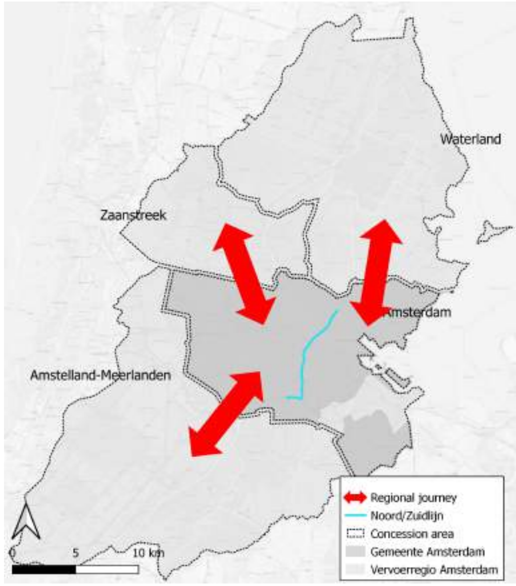


Figure 1.2: Visualization of regional journeys

1.4. Research objective and questions

This research has a main, societal objective and an additional, methodological objective. The societal objective is to assess the impact of the Noord/Zuidlijn on regional travel times and transfer patterns. Interoperable smart card data is used for that. Because that data source is quite unique but becoming more common, the methodological objective is to explore and understand the strengths and limitations of aggregated, interoperable smart card data. The societal objective leads to the following research question:

What impact does the Noord/Zuidlijn have on travel times and transfer patterns on regional public transport journeys in the Vervoerregio Amsterdam and what does this imply for public transport planning?

To answer this research question several sub-questions have been formulated. The first two sub-questions belong to the methodological objective, the other sub-questions to the societal objectives. The answer to the methodological sub-questions affects the methods and depth of the analysis to answer the other sub-questions.

- SQ1 - What are the strengths and weaknesses of aggregated, interoperable Dutch smart card data compared to disaggregated, single operator Dutch smart card data?
- SQ2 - How can the weaknesses of aggregated, interoperable Dutch smart card data be overcome or coped with?
- SQ3 - What impact did the Noord/Zuidlijn have on the travel times for regional public transport journeys in the Vervoerregio Amsterdam?
- SQ4 - What impact did the Noord/Zuidlijn have on the transfer patterns for regional public transport journeys in the Vervoerregio Amsterdam?
- SQ5 - What do the outcomes imply for public transport planning?

The Noord/Zuidlijn in these questions does not only refer to the metro line, but to changes in the bus and tram network that took place at the same time as well. These are described in Chapter 3.

1.5. Methodology

The methodology describes how the proposed research enables answering the research questions. This research is twofold as it has a methodological and societal goal. A schematic presentation of the methodology can be found in Figure 1.3. The overarching method of this research is the case study of the Noord/Zuidlijn. This is the common thread in this thesis. In the end, the findings regarding the data and transportation impacts of the Noord/Zuidlijn are generalized.

To measure the impact of the Noord/Zuidlijn on travel times and transfer patterns aggregated, multi-operator smart card data from before and after the Noord/Zuidlijn is compared. Two different kinds of data sets of public transport journeys are used. The first is a data set that contains travel times on origin-destination level. The second contains all routes traveled including the origin, destination, mode and operator of all individual trips that make up a journey. For both of these a before and after Noord/Zuidlijn data set is available. This allows comparing the before and after situation. So, in total four main data sets are used in this thesis. A detailed description of the data sets is given in Chapter 4

The first two sub-questions are about the unique data sets that are used. It is hard to determine on forehand what weaknesses and strengths are present in the data. While exploring, processing and analyzing the data several strengths and weaknesses are encountered. These are documented along the way and solutions are sought to overcome the weaknesses of the data sets. This gradual process of encountering and solving problems leads to the answer to the first two sub-questions. By answering the first and second sub-question the methodological objective is met.

To enable answering the third and fourth sub-question a methodology is developed to overcome the weaknesses encountered in the interoperable smart card data. Using this methodology the travel time and transfer impacts of the Noord/Zuidlijn are determined. For that, a hierarchical approach is used. This approach starts with the overall impacts, then the distribution of the impacts are investigated and last specific cases are examined to identify patterns in the impacts. Generalizing the findings for sub-question three and four leads to the implications that enable answering sub-question five. Answering sub-questions three, four and five ensure that the societal objective is reached.

An important limitation of the methodology is that the findings regarding the strengths and weaknesses of the data sources used affect the depth and outcome of the rest of the research as well.

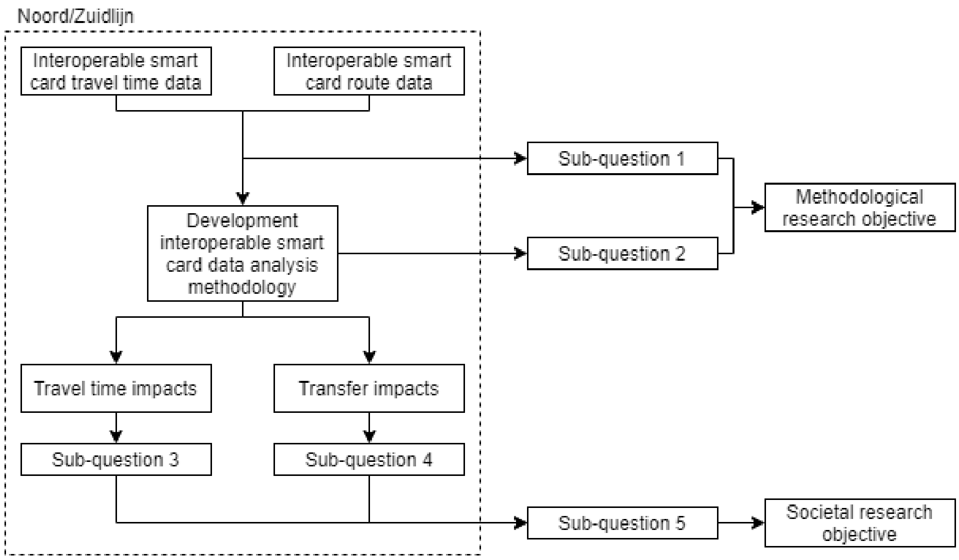


Figure 1.3: Schematic presentation of the methodology for this research

1.6. Report outline

The outline of this report is as follows. Chapter 2 contains an overview of literature on the transportation impacts of large-scale public transport overhauls. Besides, it discusses the state-of-the-art in smart card data analysis. Chapter 3 describes public transport in the Amsterdam region and the Noord/Zuidlijn. In Chapter 4 the data, processing thereof and the analysis are described. The results are presented in Chapter 5. After the results the discussion and conclusion and recommendations follow in Chapter Chapter 6 and 7.

2

Literature study

As described in the introduction this research investigates the transportation impacts of the Noord/Zuidlijn using interoperable smart card data. Therefore, the literature study is twofold. The first part contains research on the transportation impacts of large-scale public transport network overhauls. It aims to present the state of the art and caveats of the limited amount of research on this topic. Besides, it shows the difficulties of this field of research. The second part describes smart card data analysis and why interoperable smart card data and the analysis thereof is quite unique. The goal is to explain why studies do not focus on interoperable smart card data and why this is an interesting research gap. The literature search strategy and the keywords used are presented in Appendix B.

2.1. Transportation impacts of large-scale public transport network overhauls

This section starts with a brief introduction on the potential impacts of public transport. Thereafter, studies focusing on the transportation impacts are described in more detail. The aim is to present the state of the art of research on the impact of large-scale public transport network overhauls on travel times and transfer patterns. Besides, it identifies gaps in this field of scientific research and enables placing the proposed research in context.

The possible impacts of public transport are categorized by Van Oort et al. (2017) using the 5E framework. Impact studies often focus on one of these 5E's which are effective mobility, efficient city (Knowles, 2012; Nguyen et al., 2018), economy (Cervero and Kang, 2011; Dorantes et al., 2011), environment (Bel and Holst, 2018; Gramsch et al., 2013)) and equity (Manaugh and El-Geneidy, 2012; Venter et al., 2018). This thesis and thus the literature study as well focuses on the impacts of a large-scale public transport network overhaul on effective mobility. These are also called the direct operational effects (Ingvardson and Nielsen, 2018) or transportation impacts (Brands et al., 2020) and include among others ridership, travel time, transfers and modal shift.

This thesis is a follow-up on research done by Brands et al. (2020). They investigate the impact of a new metro line through the city center of Amsterdam using smart card and corresponding automatic vehicle location (AVL) data. This new metro line was accompanied by a restructuring of the existing bus, tram and metro network from a network with radial bus and tram lines to a trunk-feeder network with a metro line on the trunk section and feeder buses and tram. They find that in areas with new metro stations more trips are produced while in areas close to the new metro line a decrease in produced trips can be seen. This could indicate that travelers have changed routes including access and egress patterns. Besides, they notice an increase in metro trips after the introduction of the new metro line at the cost of bus and tram trips. Regarding travel time, Brands et al. (2020) find that 70% of the travelers do not experience a significant change (more than -1 or +1 minutes) in travel time. For the 30% of travelers that is affected by the network changes, the travel time gain is 3 minutes on average. However, this average gain is unevenly distributed over people that experience travel time savings and losses. Looking at the geographical distribution of changes in travel time, travel time savings are largest close to the new metro line. Areas right next to the new line are not affected or slightly negatively affected. In the outskirts of the city some areas that are negatively affected in terms of travel time can be found. Brands et al. (2020) investigate the impact on transfers in the network as well. On network level a slight decrease in transfers can be seen. However, this is caused by the attraction of new passengers, especially on relations with a direct metro connection in the after situation. When only travelers that were already present in the before situation are included, the average number of transfers per passengers increase slightly. The geographical distribution of the change in the number transfers is that along the new metro line the number of transfers decreases but increases in feeder bus areas. An area for further research that Brands et al. (2020) mention is to incorporate multiple operators in the analysis as they only incorporated Amsterdam's urban public transport operator.

Weckström et al. (2019) assess the impact of the transformation of Helsinki's direct, radial bus network to a trunk-feeder system with metro's on the trunk and feeder buses. They used open timetable for this. On a network wide level, they find a small increase in travel times and a small decrease in transfers. Besides, on zone level they find that areas with new and old metro stations generally benefit most as the number of transfers and travel time both decrease. For feeder bus areas the travel times and number of transfers generally increase. Trips that are outside of the area where the overhaul took place are not affected. The changes in travel time range from -15 to +15 minutes and the change in number of transfers ranges from -1 to +1. However, Weckström et al. (2019) also show that outcomes differ strongly on stop level. Besides, an important note regarding the research by Weckström et al. (2019) is that the speed at the trunk section did not increase as a result of the new metro line as the metro merely changed the alignment of the trunk section. As a result, the metro is not fast enough to make up for lost time in the feeder system, longer alignment of the trunk section and the increased number of stops. The fact that the metro does not lead to an increase in speeds on the trunk section and the Helsinki case approach make the results difficult to generalize. The findings by Weckström et al. (2019) indicate that balancing between aggregate and disaggregate effects is an important dilemma in public transport planning. Besides, the choice between a trunk-branch or trunk-feeder network is not binary and reintroduction of direct bus routes might alleviate the consequences of the network overhaul for negatively affected areas. Investigating effects for selected locations can help to identify and overcome neglected aspects during the planning process. However, as the selection will heavily affect the outcomes, a wide variety of different stops should be selected.

Arbex et al. (2019) combine General Transit Feed Specification (GTFS), AVL and smart card data to assess the impacts of an improvement of São Paulo's late night bus network. The bus network was restructured from a non-integrated network with long and circuitous itineraries to an integrated trunk-feeder network with better spatial coverage. Arbex et al. (2019) find that the new network which is more reliable and offers more frequent services, attracts more passengers. Besides, the network improvement led to longer distances traveled. As a result, travel times and number of trips became higher. A limitation of the study by Arbex et al. (2019) is that it only assesses the impact on urban scale. Besides, the smart card data used does not contain information on check-out location and time.

Fu and Gu (2018) investigate the impact of a new metro line in Nanjing, China based on smart card data. On network level, they find an increase in ridership after opening of the line. Besides, they find that the number of metro stations within 30 minutes of the central business district increases. Furthermore, on certain old metro lines the travel times increase as well because the new metro line enables faster journeys resulting from new transfer opportunities. The study by Fu and Gu (2018) is limited as it only assesses the impact of the overhaul on the metro network.

A study by Vuk (2005) investigates the transportation impacts of the Copenhagen metro along a certain corridor. Based on traffic counts, panel interviews and model forecasts he finds that 70% of the modal shift to metro can be attributed to bus travelers. Around 10% of travelers derives from cars users and around 15% is induced traffic.

Reinhold (2008) analyzed the results of a large restructuring of Berlin's public transportation network from an operator's perspective. This entailed improving the core network by adding high-frequent and intensively marketed bus and tram services on major traffic axes. On the other hand services on parallel lines were canceled and services on lines with a low utilization rate were reduced. As a result, the public transport operator reduced its annual volume of operations by 4 million kilometers a year resulting in savings of 9.5 million euros a year. Despite, the reduction in vehicle-kilometers, the total demand increased with 2% which is equal to 24 million trips a year. A drawback of this study is that it only takes the effects on one operator in Berlin's public transportation network into account. The S-bahn, Berlin's rail rapid transit sytem, was not considered.

Ingvardson and Nielsen (2018) review literature on the effects of new bus and rail rapid transit systems. Regarding transportation impacts, they find positive effects on travel times and ridership and significant modal shifts. However, the results depend strongly on the extent to which the system improves the existing situation and how the system is implemented locally.

Table 2.1 summarizes the literature on ex-post evaluation of the transportation impacts of large-scale public transport network overhauls and places the proposed research in the context thereof.

Table 2.1: Overview table of existing literature on ex-post evaluation and how the proposed research fits in there

Network change	Indicators of interest	Data source	Network level	Modes	Source
Transformation of a direct bus network to a metro system with feeder buses	Travel time Transfers	Timetable data	Urban	Bus Metro	(Weckström et al., 2019)
Late-night bus network implementation	OD-matrices Travel time Transfers Accessibility	Disaggregate smart card data	Urban	Bus	(Arbex et al., 2019)
Commissioning of a new metro line	Passenger flows Travel time Travel time reliability	Disaggregate smart card data	Urban	Metro	(Fu and Gu, 2018)
Implementation of a metro network with feeder buses	Traffic growth Modal shift Destination choice	Traffic counts Panel interviews Model forecasts	Corridor	Bus Car Metro	(Vuk, 2005)
Commissioning of a new metro line accompanied by changes in the existing bus and tram network	Ridership Modal shift Travel time Transfers Reliability Societal benefits	Disaggregate smart card data	Urban	Bus Tram Metro	(Brands et al., 2020)
Restructuring of the urban public transport network	Ridership Revenues Savings	Unknown	Urban	Bus Tram Metro	(Reinhold, 2008)
Commissioning of a new metro line accompanied by changes in the existing bus and tram network	Travel time Transfers	Aggregated, multi-operator smart card data	Regional	Bus Tram Metro Train	Proposed research

From the studies described above, several things become clear. First, studies that investigate the transportation impacts of a large-scale public transport network overhaul are scarce. Second, a couple of studies investigate the transportation impacts of a new metro line and feeder buses, but variation remains. Third, the scope and the methods used to determine the impact differ strongly. Last, because of these variations no clear conclusion can be drawn about the transportation impacts of large-scale public transport network overhauls. In general, an increase in demand and modal shift are likely to occur. Changes in travel times and transfer patterns are likely to occur as well albeit unevenly distributed among users of the network. The proposed research adds a more holistic view to the existing literature by investigating the impact of the overhaul on a multimodal, multi-operator and regional public transport network.

2.2. Interoperable smart card data analysis

This section starts with a brief general overview of smart card systems and smart card data. Thereafter, it briefly discusses the state-of-the-art of research into the application and analysis of smart card data for public transport planning. Then, the emergence, the potential and challenges regarding interoperable smart card data are presented.

The main purpose of smart card or AFC systems is to collect revenue. In addition, they provide large amounts of transaction data as well. This can be very useful for strategic, tactical and operational public transport planning (Pelletier et al., 2011). For a general overview on smart card data and its applications see Appendix C.

Smart card systems and thus the data they produce differ strongly. First, depending on the fare system of the operator users are only required to check-in or they are required to check-in and check-out (Hussain et al., 2021; Sun et al., 2016; Zannat and Choudhury, 2019). Second, validation of the card could take place in a vehicle or at a station. These two design choices of the smart card systems determine whether the destination of the public transport trip is known and whether routes are observed or not (Dixit, 2021).

Compared with other big data sources, smart card data is most often used in public transport planning and its usage has only been growing the last couple of years (Welch and Widita, 2019). However, research conducted with smart card data varies strongly. As mentioned above Pelletier et al. (2011) distinguish smart card usage for strategical, tactical and operational planning. The first relates to long-term network planning, analysis of customer behavior and demand forecasting. The second entails service adjustment and inference of longitudinal and individual trip patterns. Studies on operational level focus on calculation of performance indicators and operation of the smart card system itself. Fu and Gu (2018) categorize the following research areas: analysis of passenger flow patterns, transit network performance and investigation of travel patterns and travel behavior. (Covic and Voß, 2019) only distinguish the last two categories. Tzika-Kostopoulou and Nathanail (2021) explore smart card data usage in transport modeling. They find that only few studies deal with multiple modes of public transport. Besides, most studies using smart card data focus on deriving important attributes for transport modeling. Examples are trip reconstruction, journey reconstruction and transfer inference.

Recently, interoperability and standardization of smart card systems have gained interest from academic, public and industry stakeholders (Covic and Voß, 2019). First of all, interoperable smart card systems increase convenience for the traveler (National Academies of Sciences, Engineering, and Medicine, 2006; Yoh et al., 2006). Besides, interoperable smart card systems could enable more comprehensive insights in public transport travel (Yoh et al., 2006). Data from a single operator is not able to capture transfers between operators and thus does not contain all complete journeys. Besides, exchange of travelers between operators is not visible in single operator data. In some cases, a modal shift can therefore not be observed. Although several benefits of interoperable smart card systems can be identified, institutional issues hinder implementation thereof (Yoh et al., 2006). First, variation in goals and customers per operator make it difficult to identify common goals and usage of smart card systems. Second, operators are not willing to hand over control over fare policies and collection. Third, decentralized and weak decision-making structures compound coordination, facilitation, implementation and enforcement of interoperable smart card systems. Besides, uncertainty about the costs and benefits of these systems for both operator and users and uncertainty about the development of alternatives hinder the implementation of interoperable smart card systems.

Once an interoperable smart card system has been successfully implemented, data availability is another concern. The smart card data is owned by the operators and these are not willing to share their confidential company information with competitors (Van Oort et al., 2015). Usage of interoperable smart card data requires clear agreements between all parties involved. Normally, the most conservative approach is taken. This means that each operator owns the transaction data generated in its own system and transaction data related to another operator is only available to the other operator involved in the transaction (National Academies of Sciences, Engineering, and Medicine, 2006).

The implementation and data availability issues regarding interoperable smart card systems described above, explain the lack of research into interoperable smart card data analysis. However, the recent increase in such systems and the potential of the data they produce, clarifies why this is an interesting research gap.

2.3. Wrap up literature study

The literature study shows that research on the transportation impacts of public transport network overhauls is limited. In addition, there is a lot of variation in the overhauls studied, the indicators included, the methods used and the modes included. This results in varying impacts. An increase in demand and modal shift are likely to occur. Besides, it is important to be aware of a possible uneven distribution of travel time and transfer impacts.

Regarding interoperable smart card data analysis, the literature study explains that implementation and data availability issues are the reason for the very limited amount of research on interoperable smart card data. However, the potential of interoperable smart card data is an important reason to investigate the opportunities thereof.

3

Public transport in the Amsterdam region

The literature study showed that the transportation impacts of a large-scale change in public transport depend strongly on the network before and after the overhaul. Therefore, this chapter provides necessary background knowledge on the Noord/Zuidlijn. This is necessary to understand and explain the impacts in the following chapters.

This chapter starts with general information about public transport in the region of Amsterdam like the modalities and the concession areas. In addition, the emergence, the situation before the Noord/Zuidlijn and the future of public transport in the region are discussed. Besides, the change in infrastructure and services that occurred when the Noord/Zuidlijn became operative, is described extensively.

3.1. Concession areas in the Amsterdam region

Public transport in the Amsterdam region is organized by the Vervoerregio Amsterdam (VRA). It is a governmental body that consists of fifteen municipalities. An important task of the VRA is tendering public transport in the region. To that end the municipalities in the VRA are subdivided into four concession areas: Amstelland-Meerlanden operated by Connexxion, Amsterdam operated by GVB, Waterland operated by EBS and Zaanstreek operated by Connexxion. In Appendix D an overview of the municipalities per concession area is presented.

Figure 3.1 presents the public transport network in Amsterdam after the Noord/Zuidlijn. As can be seen it consists of trains, metros, trams and buses. Only concession Amsterdam has a metro and tram network. These are operated by GVB. However, the tram and metro lines penetrate the concession area Amstelland-Meerlanden partially.Concession Amstelland-Meerlanden, Waterland and Zaanstreek are mainly served by bus connections. These include bus connections to Amsterdam. Besides, Hoofddorp, Schiphol, Purmerend and Zaandam have train stations.

In the remaining of this research, regional journeys are defined as journeys from concession area Amsterdam to Amstelland-Meerlanden, Waterland or Zaanstreek and vice versa. It is therefore necessary to keep in mind that concession area Amstelland-Meerlanden is the only regional concession area that has tram and metro connections to Amsterdam.

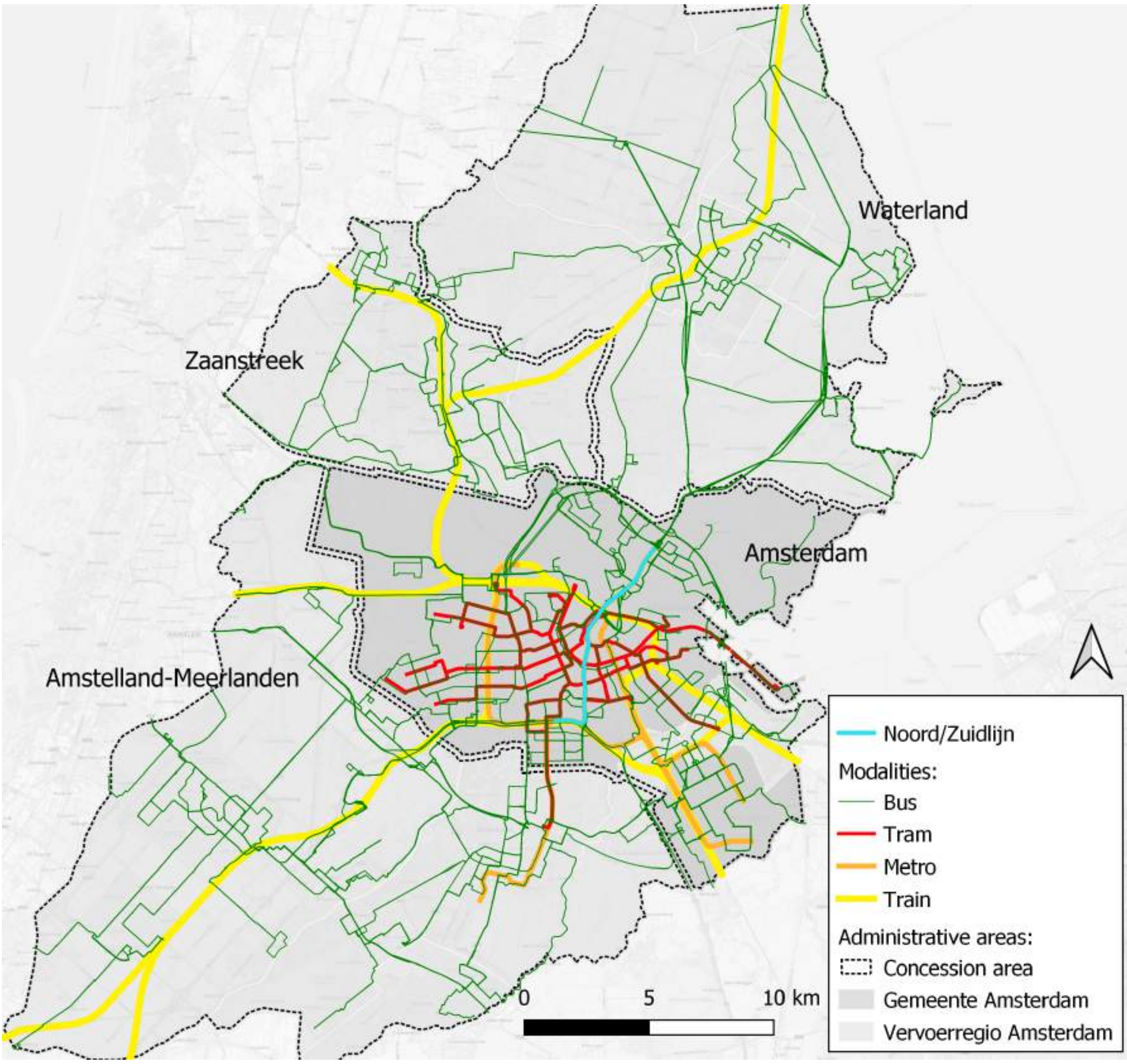


Figure 3.1: Public transport network in the VRA after the Noord/Zuidlijn

3.2. The emergence and trends of public transport in Amsterdam

This section offers necessary background knowledge on Amsterdam's old public transport network to understand the Noord/Zuidlijn network overhaul. Additionally, trends that created new challenges and opportunities are described. Last, the goals of Amsterdam's public transport network overhaul are presented.

Amsterdam's public transport network organically grew to a fine-meshed network focused on the city center (Stadsregio Amsterdam, 2011). The interurban tram network offering short, local movements was the basis of Amsterdam's public transport network. A fine bus network evolved in areas where demand was too low for tram services. These areas were mostly at the outskirts of the city. With the city's gradual growth, the existing network grew as well. Existing public transport lines were extended, stops were added and frequencies increased. This growth resulted in a public transport network that was fine-meshed and focused on the urban traveler. All public transport was directed to the city's main hub: Amsterdam Centraal. Tram lines were slow and unreliable due to a lot of interaction with other traffic in the city center while buses were operating inefficiently as a result of a large number of direct connections with high frequencies (Tiemersma and Govers, 2012).

However, recently several trends that do not fit to such a public transport network have emerged (Stadsregio Amsterdam, 2011). First, public transport demand keeps growing. This growth consists to a large extent of regional public transport and of access and egress movements to train stations. Besides, the bicycle has gained market share at the expense of short public transport trips. Another trend that should not be forgotten is that the Dutch government has cut subsidies for bus, tram and metro. As a result, the VRA is faced with the task to offer more and higher quality public transport while receiving less subsidy. Furthermore, Amsterdam grows at more locations than just in the city center around Amsterdam Centraal. These upcoming locations should be well connected by public transport as well (Tiemersma and Govers, 2012).

The trends mentioned above combined with characteristics of Amsterdam's public transport network led to challenges, but created opportunities as well. Public transport in Amsterdam had lost part of its passengers to the bicycle without taking advantage of the opportunities that have risen. The fine-meshed public transport network of Amsterdam was not able to attract the growth in regional public transport demand as a result of the slow and unreliable speeds in the city center. Furthermore, the increase in passenger numbers at Amsterdam's train stations led to a need for public transport offering access and egress options. Besides, subsidy cuts necessitated improved efficiency of public transport. Next to that, the development of Amsterdam at multiple locations required a public transport network less focused on Amsterdam Centraal.

An overview of these characteristics, trends and goals is listed below.

Characteristics old public transport network Amsterdam

- Fine-meshed
- Direct connections
- Direct to Amsterdam Centraal
- Lots of traffic in the city center
- Designed for the urban traveler
- Low occupancy in buses
- Slow and unreliable trams

Trends in Amsterdam affecting public transport

- Growth of agglomerative transport
- Growth of access and egress movements from and to train stations
- Growth of bicycle usage for short trips
- Polycentric development
- Reduction in public transport subsidies

Goals of public transport network overhaul

- More cost-effective public transport
- Faster and more reliable public transport
- Relieve pressure on Amsterdam Centraal
- Connecting multiple cores in Amsterdam
- More liveable city center
- Attract the agglomerative and train traveler

By identifying challenges and opportunities of public transport in Amsterdam a couple of goals regarding public transport in Amsterdam can be determined. First, public transport should be more cost-effective. Second, it should be faster and more reliable. This way the goal to attract more agglomerative and train travelers could be achieved. Besides, the network should be better connected to other developing cores in Amsterdam as well. Next, public transport should enable a liveable city center.

The Noord/Zuidlijn became an important mean to achieve these goals. What the Noord/Zuidlijn entails and how it should lead to achievement of the goals mentioned above, is described in the following section.

3.3. The Noord/Zuidlijn network overhaul

The commissioning of the Noord/Zuidlijn created a huge opportunity to revise the public transport network of Amsterdam. Therefore, the Noord/Zuidlijn was not the only change that was made to Amsterdam's public transport network on July 22nd 2018. Conceptually the network overhaul consists of two elements:

- Addition of the Noord/Zuidlijn to the metro network of Amsterdam. This created a high quality public transport connection through the city center that functions as the backbone of the regional public transport network.
- The tram and bus network is transformed from a radial, trunk-branch network directed to Amsterdam Centraal to a trunk-feeder network aimed at connecting to the metro network.

The Noord/Zuidlijn (see Figure 3.2) is 9.7 kilometers long of which 7.1 kilometers are below ground level. It connects eight metro stations directly to each other. Centraal Station and Zuid offer transfers to train and other metro lines. Besides, the Noord/Zuidlijn meant that Amsterdam Noord was connected to Amsterdam's metro network. Together with the Oostlijn (metro line 53 and metro line 54) and Ringlijn (metro line 50) the Noord/Zuidlijn should become the backbone of Amsterdam's public transport network. The Amstelveenlijn (metro line 51) is less important and is closed between in March 2019 and reopened as a tram line in December 2019. Bus, tram and train lines should complement the metro network. The vision for 2050 is that the metro will be used for movements within the agglomeration, the bus and tram should connect and provide access to neighborhoods (Gemeente Amsterdam, 2021).

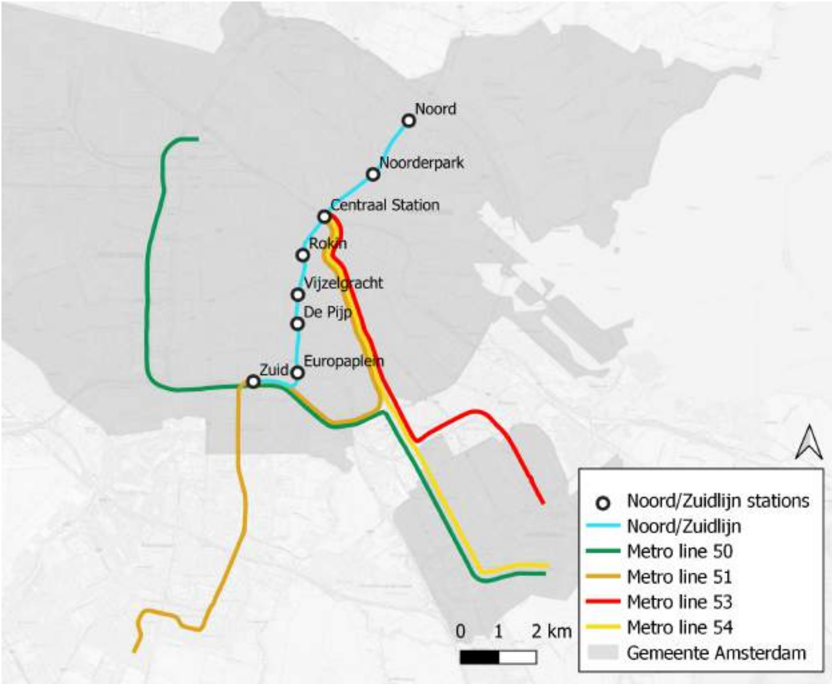


Figure 3.2: Metro network in Amsterdam after construction of Noord/Zuidlijn

The Stadsregio Amsterdam, the predecessor of the VRA, saw the addition of the Noord/Zuidlijn to Amsterdam's public transport network as a huge opportunity to revise the rest of the public transport network as well. Therefore, they set up a framework within which the public transport operators in the VRA should adjust their networks simultaneously with the commissioning of the Noord/Zuidlijn. This framework is called the Lijnennetvisie. It has been subdivided in three areas:

- Regional network north of Amsterdam
- Regional network south of Amsterdam
- Urban network in the city center

The configuration of the public transport networks in these areas as proposed by the VRA is described below.

Connection of the regional bus network north of Amsterdam to the Noord/Zuidlijn

Concession Waterland
After July 22nd 2018 around 70% of the buses in concession Waterland will end at station Noord. Passengers with destinations close to Noord/Zuidlijn stations or in the western part of the city center will have a shorter travel time. Travel times to destinations in the eastern part of Amsterdam should be similar to those before July 22nd 2018. As the travel time increases for passengers with Amsterdam Centraal as destination, 30% of the buses from Waterland will drive directly to Amsterdam Centraal without stopping at station Noord or Noorderpark. A visualization of the bus network in Waterland after opening of the Noord/Zuidlijn can be found in Figure 3.3. In the evaluation of the Lijnennetvisie the distribution of 70%/30% has been adjusted to 65%/35%. (Vervoerregio Amsterdam, 2020)

Concession Zaanstreek
For passengers from the concession Zaanstreek with a destination beyond Amsterdam Centraal a transfer to the Noord/Zuidlijn should positively affect their travel times. The earlier this transfer is made, the more beneficial it will be. Therefore, one bus line from Zaanstreek connects to the Noord/Zuidlijn at station Noord. However, passengers traveling to Amsterdam Centraal do not benefit from a transfer to the Noord/Zuidlijn. For those reasons, two bus lines from Zaanstreek will end at Amsterdam Centraal. It was planned that a new stop at Noorderpark would create an opportunity to transfer to the Noord/Zuidlijn from these bus lines would. However, this new stop has not been built yet due to high costs and complex construction (Vervoerregio Amsterdam, 2020).

Figure 3.3: Proposed configuration of connecting bus lines for concession Waterland. Adapted from Stadsregio Amsterdam (2014).

Figure 3.4: Proposed configuration of new bus network for concession Zaanstreek. Adapted from Stadsregio Amsterdam (2014).

Connection of the regional network south of Amsterdam to the Noord/Zuidlijn

Concession Amstelland-Meerlanden

Passengers traveling from Amstelland-Meerlanden to Amsterdam Centraal and its surroundings should benefit from a transfer to the Noord/Zuidlijn. Besides, passengers traveling to Amsterdam Zuid benefit from direct lines to Amsterdam Zuid as well. However, passengers that have their destination close to the Museumplein or Leidseplein will not benefit from a transfer to the Noord/Zuidlijn. For the reasons mentioned before, at least 50% of the buses from Amstelland-Meerlanden stops at Amsterdam Zuid. The other 50% will end at station Elandsgracht instead of Amsterdam Centraal after opening of the Noord/Zuidlijn. Besides, the bus lines that run through the center till Elandsgracht will be evenly spread over the cores in the region like Uithoorn, Aalsmeer en Amstelveen. This way all these towns have a direct connection to the city center and Amsterdam Zuid. The configuration of bus lines from Amstelland-Meerlanden to the city center can be found in Figure 3.5.

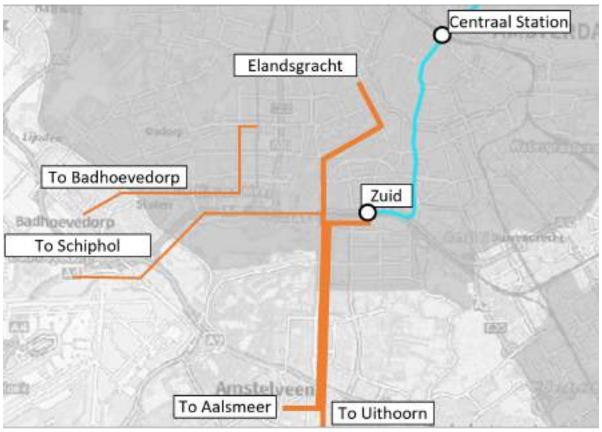


Figure 3.5: Proposed configuration of connecting bus lines for concession Amstelland-Meerlanden. Adapted from Stadsregio Amsterdam (2014).

Connection of the urban network to the Noord/Zuidlijn and the rest of the metro network

Generally speaking, the choices should result in a decrease of trips made by buses and trams in the city center. The expected percentage decrease can be found in Figure 3.6 and Figure 3.7. As can be seen the decrease in bus and tram trips should lead to an increase in metro trips in the city center.

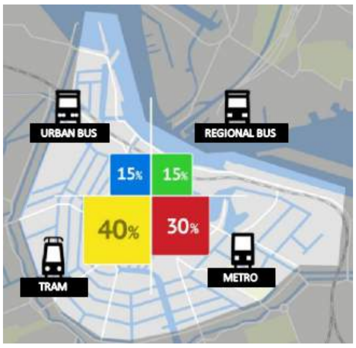


Figure 3.6: Distribution of modalities from and to city center before Noord/Zuidlijn. Adapted from Stadsregio Amsterdam (2014).

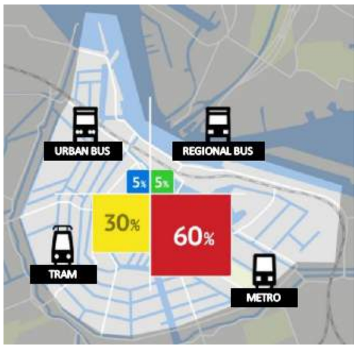


Figure 3.7: Distribution of modalities from and to city center after Noord/Zuidlijn. Adapted from Stadsregio Amsterdam (2014).

In the city center the tram network will be oriented less towards Amsterdam Centraal and more towards the metro network. This results in a decrease of tram routes on three tram corridors. The capacity that results from it, will be used to increase the amount of trams on east-west connections (see Figure 3.8). These changes will lead to a higher cost recovery rate for both tram and metro. It is expected that as a result of the network change 50% of the bus and tram passengers will transfer to the Noord/Zuidlijn.

Figure 3.8: Proposed change in tram network in the city center. Adapted from Stadsregio Amsterdam (2014). Orange = tram, blue = metro.

Figure 3.9: Proposed configuration of bus lines in Amsterdam Noord after opening of the Noord/Zuidlijn. Adapted from Stadsregio Amsterdam (2014). Orange = tram, blue = metro.

The urban bus network in Amsterdam-Noord will be restructured to connect to the Noord/Zuidlijn (see Figure 3.9). The new public transport network in Noord has two main goals. The first is to improve the connections within Noord. Besides, the network should offer a good connection to the Noord/Zuidlijn for passengers traveling from and to Amsterdam Noord. This results in high frequency east-west lines in Amsterdam Noord that stop at one of the two metro stations. After July 22nd 2018 no buses operated by GVB cross the river IJ anymore. Passengers with a destination close to Noord/Zuidlijn stations will have the largest benefit. However, passengers that have Amsterdam Centraal as their destination will experience a longer travel time but some have the option to choose for one of the regional buses that will still be directly connected to Amsterdam Centraal.

Referring back to the goals of the overhaul, the Noord/Zuidlijn is expected to offer a more reliable and faster public transport service in the city center of Amsterdam by substituting a bus and tram network by a metro line. This should be more attractive for the agglomerative traveler. Besides, it offers a high quality connection to the city center for the train traveler. Additionally, by restructuring the bus and tram network Amsterdam's public transport network is not solely directed to Amsterdam Centraal anymore. Next to that, this creates a more liveable city center as the amount of trams and buses driving to Amsterdam Centraal decreases. Last, by canceling parallel and direct bus and tram lines the cost-effectiveness of the remaining bus, tram and metro lines increases as more demand is generated for these lines.

3.4. Regional public transport network per concession after the Noord/Zuidlijn

This section describes the most important characteristics of the public transport network in each regional concession area after the Noord/Zuidlijn. Besides, it discusses the main changes between before and after the Noord/Zuidlijn. A side-by-side comparison of the changes per concession can be found in Appendix E

3.4.1. Public transport in concession Waterland after the Noord/Zuidlijn

The public transport network in Waterland after the Noord/Zuidlijn is described below. For a detailed description the reader is referred to the transport plan of EBS (EBS, nd). In Appendix E the public transport network in Waterland before and after the Noord/Zuidlijn are shown side-by-side.

In Figure 3.10 it can be seen that after the Noord/Zuidlijn bus line 301, 304, 307, 308, 312, 315 and 319 end at metro station Noord. Compared to the network before the Noord/Zuidlijn Purmerend-Oost, Purmerend-West, Landsmeer, Monnickendam and Marken have thus lost a direct connection to Amsterdam Centraal. Bus lines 305, 306, 314 and 316 offer a direct connection between the central parts of western and eastern Waterland and Amsterdam Centraal. These connections serve most passengers. Bus lines 301 and 319 have a connecting function in Amsterdam Noord as well. Bus stop IJtunnel is added to offer passengers with a destination in the east and center of Amsterdam an attractive alighting stop. Besides, for the travelers that lose a direct connection to Amsterdam Centraal attractive transfer opportunities to buses to Amsterdam Centraal are created. Furthermore, for the regional buses from Noord to the region the timetable is aligned with the timetable of the Noord/Zuidlijn because the differences in frequency could result in long waiting times. Besides, the transport plans have been aligned with Connexxion and GVB.

Figure 3.10: The public transport network in Waterland after the Noord/Zuidlijn (EBS, 2018)

3.4.2. Public transport in concession Zaanstreek after the Noord/Zuidlijn

This section describes the most important characteristics of the public transport network in Zaanstreek afer the Noord/Zuidlijn and compares it to the network before the Noord/Zuidlijn. For a detailed overview the reader is referred to the transport plan by CXX (Ouwehand and Duivis, 2016). In Appendix E the public transport network in Zaanstreek before and after the Noord/Zuidlijn are shown side-by-side.

Figure 3.11 shows that after the Noord/Zuidlijn bus line 392 ends at metro station Noord instead of Amsterdam Centraal. Bus line 391 and 394 do not stop at metro station Noorderpark. This is an important deviation from the Lijnennetvisie (Stadsregio Amsterdam, 2014). Furthermore, because a lot of direct bus connections between Amsterdam Noord and Amsterdam Centraal disappear as a result of the network revision by GVB and EBS, the occupancy of bus lines 391 and 394 is expected to increase. Therefore, the shared frequency of these buses is increased from 8 to 12 vehicles per hour between bus stop de Vlinder in Zaandam and Amsterdam Centraal. Connexxion expects that the Noord/Zuidlijn affects passenger flows on bus line 395 (from Zaandam to Amsterdam Sloterdijk) and 398 (from Zaandam to Amsterdam IBM). However, no changes in the timetable or routing have been made.

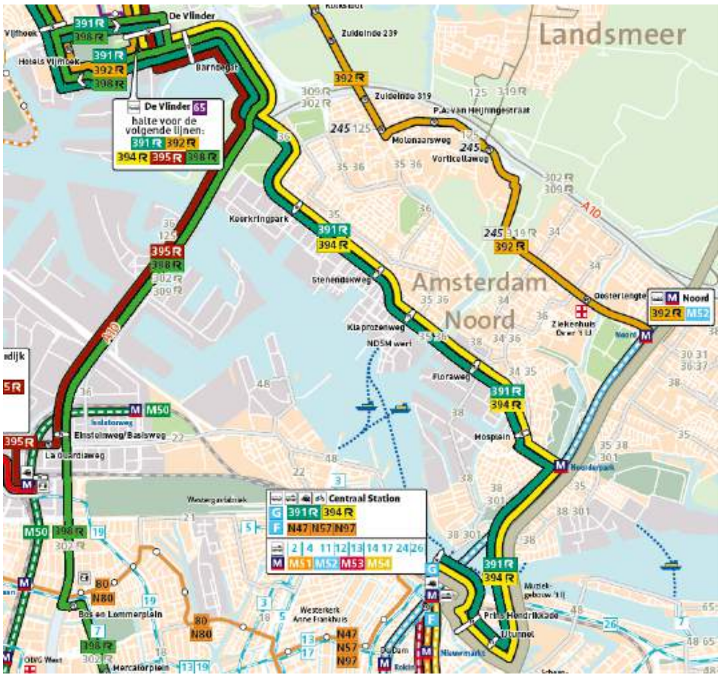


Figure 3.11: The public transport network in Waterland after the Noord/Zuidlijn (EBS, 2018)

3.4.3. Public transport in concession Amstelland-Meerlanden after the Noord/Zuidlijn

The main characteristics of the regional public transport network in Amstelland-Meerlanden after the Noord/Zuidlijn are described below. For a more detailed overview of the changes in public transport after the Noord/Zuidlijn the reader is referred to the transport plan for Amstelland-Meerlanden (Connexxion, 2017). Additionally, a side-by-side comparison of the public transport networks before and after the Noord/Zuidlijn is shown in Appendix E

The main change in the concession Amstelland-Meerlanden after the Noord/Zuidlijn is that buses from Uithoorn and Aalsmeer do not drive all the way to Amsterdam Centraal after the Noord/Zuidlijn. Instead, 50% of the buses ends at bus stop Elandsgracht, while the other 50% ends at Amsterdam Zuid. This can be seen in Figure 3.12. Before the Noord/Zuidlijn there were no direct bus connections between Aalsmeer, Kudelstaart and Uithoorn to Amsterdam Zuid. To meet the guidelines from the Lijnennetvisie, bus line 347 from Uithoorn has been shortened and ends at bus stop Elandsgracht instead of Amsterdam Centraal. Bus line 348 connects Uithoorn and Amsterdam Zuid after the opening of the Noord/Zuidlijn. Bus line 357 from Aalsmeer ends at bus station Elandsgracht after the Noord/Zuidlijn. It does not drive to Amsterdam Centraal anymore. Bus line 358 is added after the Noord/Zuidlijn and connects Kudelstaart to Amsterdam Zuid. Bus line 358 follows the same route as bus line 357 from Aalsmeer. At bus stop Sportlaan in Amstelveen bus line 347 and 348 join bus line 357 and 358. From there the four bus lines are bundled till De Boelelaan in Amsterdam Zuid. There, bus lines 347 and 357 drive straight into the city center of Amsterdam. Bus lines 348 and 358 turn right to Amsterdam Zuid. The number of buses to Elandsgracht and Amsterdam Zuid is evenly distributed over Aalsmeer and Uithoorn. Last, for some buses the reliability, frequency and speeds are increased to meet the R-net guidelines.

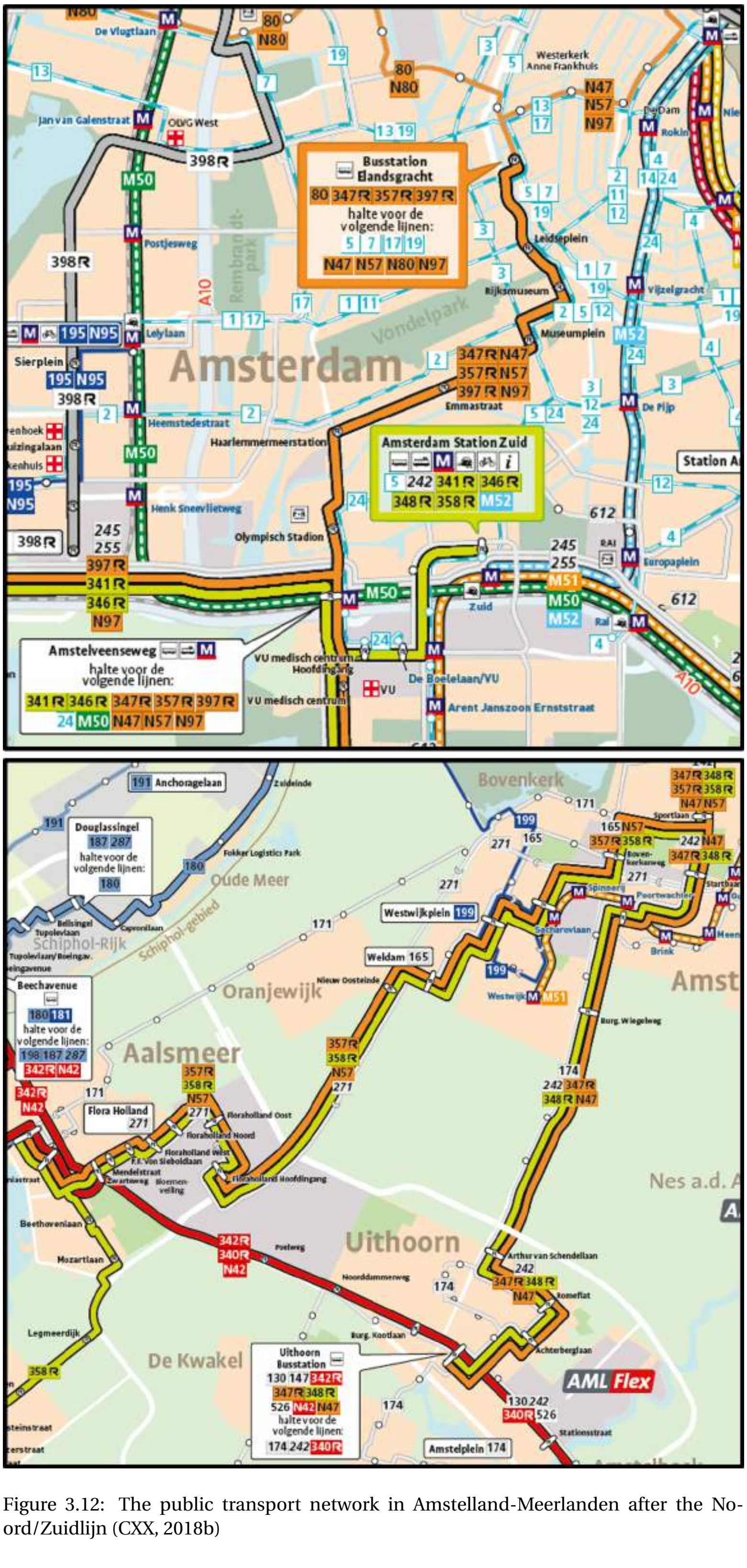


Figure 3.12: The public transport network in Amstelland-Meerlanden after the Noord/Zuidlijn (CXX, 2018b)

3.5. Expected travel time and transfer impacts of the Noord/Zuidlijn

Based on the network changes described in Section 3.3 and Section 3.4 some expectations can be determined regarding the impact of the Noord/Zuidlijn on the travel time and transfer impacts.

The Noord/Zuidlijn and the decrease in (direct) bus and tram lines likely results in more trips by metro and more transfers. A decrease in bus and tram trips is expected.

The impacts are expected to be larger for Waterland and Amstelland-Meerlanden than for Zaanstreek because in Zaanstreek only bus line 392 is affected. The travel times from Waterland and Oostzaan to areas near the southern Noord/Zuidlijn stations are expected to decrease. For the rest of Zaanstreek the travel time impacts are expected to be smaller because bus line 391 and 394 only offer transfer connections at Amsterdam Centraal. From Oostzaan, Purmerend-West, Purmerend-Oost, Landsmeer, Marken en Monnickendam the travel times to Amsterdam Centraal are expected to increase because a transfer is necessary after the Noord/Zuidlijn.

From Amstelland-Meerlanden the travel times to Centraal Station and Amsterdam Noord are expected to decrease. The travel time and number of transfers to reach areas between bus station Elandsgracht and Centraal Station is expected to increase as a result of the loss of direct connections.

Additionally, the Noord/Zuidlijn offers an attractive connection to the city center from Amsterdam Zuid. Therefore, a shift in train travelers from Amsterdam Centraal to Amsterdam Zuid can be expected.

3.6. Wrap up public transport in the Amsterdam region

From this chapter it is concluded that the network changed drastically when the Noord/Zuidlijn became operative. In the concession area Amsterdam, the metro line replaced direct buses and trams to Amsterdam Centraal. The replaced buses and trams were rerouted to feeder lines to the Noord/Zuidlijn. From the regional concession areas direct buses to Amsterdam Centraal were rerouted to connect to metro station Noord or Amsterdam Zuid. As a result for some travelers an extra transfer is necessary. Only a few lines from the region north of Amsterdam kept their direct connections to Amsterdam Centraal.

The Noord/Zuidlijn network overhaul is expected to result in a decrease in travel times on OD-relations where a long part of the journey can be made with the Noord/Zuidlijn. OD-relations where a direct connection is lost likely experience a deterioration in the number of transfers and travel times.

4

Data, methods and analysis

As already mentioned in the introduction aggregated, interoperable smart card data is used to investigate the travel time and transportation impacts of the Noord/Zuidlijn. From the literature study, it became clear that this data source is quite unique and thus no studies working with similar data have been found. Therefore, this chapter describes the data. Besides, the methods used to cope with the weaknesses are described and what analysis are performed to investigate the travel time and transfer impacts of the Noord/Zuidlijn.

The chapter starts with an overview of the raw data sets used and their contents. Next, the main analysis choices to make a selection of the data are motivated. Then, the data processing steps are described. Last, the analysis approach and elaboration thereof are presented.

4.1. Data

In total four main data sets have been used for this thesis:

- Travel time data before the Noord/Zuidlijn
- Travel time data after the Noord/Zuidlijn
- Route data before the Noord/Zuidlijn
- Route data after the Noord/Zuidlijn

In this list two types of data sets can be distinguished, namely a data set containing travel times and a data set containing the routes people travel. Besides, to compare the situation before and after the Noord/Zuidlijn of each of the data set types a before and after version is used in this thesis.

The data has been requested by the municipality of Amsterdam as part of the larger study on the impact of the Noord/Zuidlijn. Therefore, the data was already existent and has not been requested or collected specifically for this thesis. It is pre-processed by Translink Systems (TLS) based on the inquiry by the municipality and additional competition and privacy regulations. Pre-processing consisted of aggregation over space and time and coupling individual transactions into journeys.

An important feature of all data sets is that the number of travelers is not given exactly for values below 300. For these values the number of journeys is split into bins of 50. This means that the exact number of travelers is not known. An example to clarify this: if the number of travelers on a record in the data is 89 travelers, then the number of travelers is given as 51-100. Similarly, 21 travelers will be presented as 1-50 travelers. However, 301 travelers is presented as 301 travelers.

4.1.1. Raw travel time data set

The first data set that has been used contains the median and average of travel times between 133 stop clusters in the metropolitan area of Amsterdam. The travel times have been aggregated by the supplier based on the half hour block in which the check in took place and the type of day. The data has thus been averaged already. An overview of the attributes, number of categories and a description thereof can be found in Table 4.1. This data set will be referred to as the travel time data set.

Table 4.1: Overview of the contents of the raw travel time data set

Attribute	Number of categories	Description of categories
Quarter of the year	1 per data set	Before data set: 2018Q2 After data set: 2019Q1
Day	4	Working day (excluding school vacations) Working day (including school vacations) Saturday Sunday and public holidays
Check in stop cluster	133	88 BTM stop clusters 19 train stations 26 train station clusters
Check out stop cluster	133	88 BTM stop clusters 19 train stations 26 train station clusters
Time period	48	Time in a day divided into half hour blocks ranging from 00.00-00.30 till 23.30-00.00
Number of journeys	≤300: 6 categories >300: No categories	≤300: 1-50, 51-100, 101-150, 151-200, 201-250, 251-300 >300: All integers possible
Median travel time [s]	No categories	All integers possible
Average travel time [s]	No categories	All integers possible

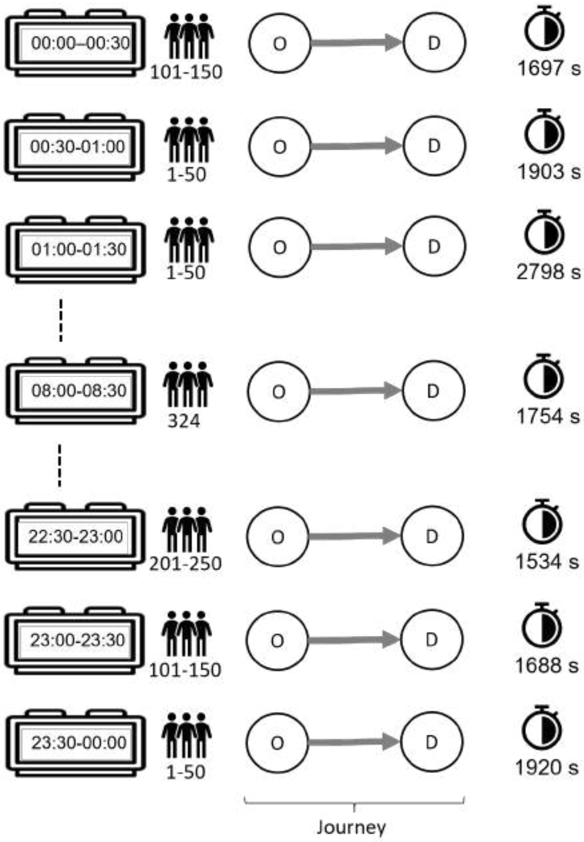


Figure 4.1: Visualization of the travel time data

4.1.2. Raw route data set

To investigate the impact the Noord/Zuidlijn had on transfer patterns a data set containing all routes traveled between each origin and destination has been used. For each route up to a maximum of five trips of that route are known. Per trip the origin, destination, modality and operator are known. Furthermore, the amount of travelers traveling that route in the quarter of interest is given as well. This data is aggregated over space, time and routes. The contents of this data set are shown in Table 4.2. This data will be referred to as the route data set.

Table 4.2: Overview of the contents of the raw route data set

Attribute	**Number of categories**	**Description of categories**
Quarter of the year	1 per data set	Before data set: 2018Q2 After data set: 2019Q1
Check in stop cluster	133	88 BTM stop clusters 19 train stations 26 train station clusters
Check out stop cluster	133	88 BTM stop clusters 19 train stations 26 train station clusters
Day and time period	7	Working day (excluding school vacations) morning peak Working day (excluding school vacations) evening peak Working day (excluding school vacations) rest day Working day (excluding school vacations) whole day Working day (including school vacations) Saturday Sunday and public holidays
Trip x* check in stop cluster	133	88 BTM stop clusters, 19 train stations, 26 train station clusters
Trip x* check out stop cluster	133	88 BTM stop clusters, 19 train stations, 26 train station clusters
Trip x* modality	4	Train, bus, tram, metro
Trip x* operator	4	NS, GVB, EBS, CXX
Number of journeys on route	≤300: 7 categories >300: No categories	≤300: 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, ≤300** >300: All integers possible

* The data contains up to five trips per route. For each trip the attributes are given.
** Only applies to time period 'Working day (excluding school vacations) full day' and 'Working day (including school vacations) full day'

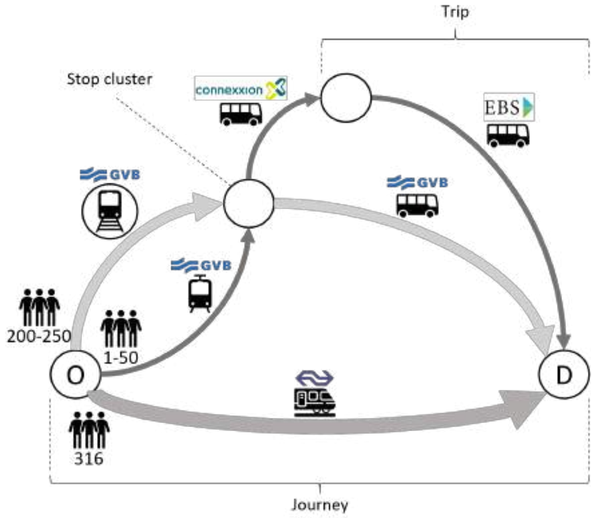


Figure 4.2: Visualization of the route data

4.1.3. Stop clusters

Both the travel times and routes data sets consist of 133 stop clusters. In a stop cluster multiple public transport stops are aggregated. 88 out of 133 stop clusters consist of bus, tram and metro stops. The other 45 stop clusters contain train stations. 19 of those contain one single train station. The other 26 train stop clusters are composed of multiple train stations. This categorization is based on the wind direction and the train services to Amsterdam. In Appendix G more information regarding cleaning of the stop clusters in the data can be found.

4.2. Analysis choices

Not all available, raw data has been processed and analyzed to answer the research questions. This section motivates the choices made to analyze the data. It contains the substantiation of making a selection of all available raw data. It describes which origin-destination (OD) relations are included in the research. Besides, it substantiates which months of the year before and after the Noord/Zuidlijn are compared and why these months have been chosen. Besides, it discusses the choice for the analyzed day of the week and hours of the day.

4.2.1. Selection of regional OD-relations

Because this thesis focuses on the regional impact of the Noord/Zuidlijn, only regional journeys have been included in the analysis. For that reason, only OD-relations between the region and Amsterdam are selected. Therefore, each stop cluster is assigned to a specific administrative and concession area. Amsterdam's stop clusters lie within the concession area Amsterdam (see Figure D.1). Regional stop clusters lie within the VRA, but not in the concession area Amsterdam. These criteria result in 62 stop clusters in Amsterdam and 40 regional stop clusters. In total, 62 times 40 stop clusters are analyzed in two directions. Hypothetically, the maximum amount of OD-relations is 4,960. An overview of all regional and Amsterdam stop clusters can be found in Appendix H.

4.2.2. Selection of the quarter before and after the Noord/Zuidlijn

Several raw travel time and route data sets that have been aggregated over different time periods, are available for analysis. To allow comparison between before and after the Noord/Zuidlijn one period before and one period after the Noord/Zuidlijn is selected. The choice for the quarter before opening of the Noord/Zuidlijn was easy as only data for one quarter is available, namely the second quarter of 2018. However, for the choice of a quarter after opening more options were available. Ideally the quarter after opening is chosen such that the combination of quarters satisfies all following criteria:

- The same quarters of different years are analyzed. The reason for this is that travel times and demand show a strong variation over a year. By choosing the same quarters for both years the effect of this

variation on the results is minimized.
- No external factors that could affect travel times or transfer patterns have taken place between the two quarters. If such changes have taken place, it will be difficult to determine which change affected the outcome to what proportion.
- There should be sufficient time between the opening of the Noord/Zuidlijn and the start of the quarter after opening of the Noord/Zuidlijn. The reason for this is that travelers need time to get used to changes in public transport services.

As no data after the Noord/Zuidlijn meets all criteria, a trade-off between these criteria had to be made. Regarding the quarter after opening of the Noord/Zuidlijn, the first quarter of 2019 has been chosen. The substantiation of this choice is twofold. First, excluding external factors, like the conversion of the Amstelveenlijn and the increase in parking fees, gives a clearer picture of the impact of the Noord/Zuidlijn only. Second, the magnitude of the impact of comparing different quarters before and after opening is probably easier to estimate than the magnitude of the impact of external factors.

The selection of these periods means that there is a difference between which quarter is analyzed for each year. This has to be accounted for when interpreting the results

To repeat the main outcome of this choice, the data set that will be used to measure the travel times before the opening of the North/Southline covers the second quarter of 2018 and is referred to as the 2018Q2 data or the before data. This period runs from the 16th of June 2018 till the 21st of July 2018. The other data set entails 2019Q1 and runs from the 8th of December 2018 till the 2nd of March 2019. This data is referred to as the 2019Q1 data or the after data.

4.2.3. Selection of the days and hours

Travel times can vary strongly over days in the week and over hours in the day. Therefore, in this research the focus is on working days and does not include school holidays. Regarding variation over the day, a distinction is made between travel times in the morning and evening peak and the rest of the day. However, the average travel times over the whole day are analyzed as well. The before data set contains 25 working days and the after data set 45 working days.

4.3. Occurence of ranges in the data

As shown in Table 4.1 and Table 4.2, the number of travelers is not exactly known for data records where less than 300 people have traveled during the aggregation period. For values under 300 the data is given in bins of 50. This applies to both the travel time and route data.

The total number of records including the distribution over ranges in both travel time data sets is presented in Table 4.3. It becomes clear that bin 1-50 occurs by far most often. The difference in the totals can be explained by the number of working days in both data sets, namely 25 for 2018Q2 and 45 for 2019Q1.

In Appendix P the contribution of the ranges and exact values to the total number of travelers on a working day are presented.

Table 4.3: Counts and share of the number of half hour blocks with a certain number of travelers in the travel time data sets

Travel time data set	**1-50 travelers** [-]	**51-100 travelers** [-]	**101-150 travelers** [-]	**151-200 travelers** [-]	**201-250 travelers** [-]	**251-300 travelers** [-]	**>300 travelers** [-]	**Total records** [-]
2018Q2	102,153 (91.9%)	5,198 (4.5%)	1,762 (1.6%)	884 (0.8%)	455 (0.4%)	267 (0.2%)	817 (0.7%)	111,536 (100%)
2019Q1	113,513 (84.2%)	10,178 (7.6%)	4,082 (3.0%)	2,246 (1.7%)	1,373 (1.0%)	843 (0.6%)	2,544 (1.9%)	134,779 (100%)

In the route data records with a number of travelers that falls in the lowest bin occur most often as well. This is shown in Table 4.3. Again, the difference in the totals between the before and after data can be explained by the longer period over which the after data was aggregated.

In Appendix P the contribution of the range and exact values to the total number of travelers are presented.

Table 4.4: Counts and share of the number of routes with a certain number of travelers in the route data sets

Route data set	1-50 travelers [-]	51-100 travelers [-]	101-150 travelers [-]	151-200 travelers [-]	201-250 travelers [-]	251-300 travelers [-]	>300 travelers [-]	Total records [-]
2018Q2	99,436 (94.7%)	2,287 (2.2%)	1,006 (1.0%)	546 (0.5%)	351 (0.3%)	223 (0.2%)	1129 (1.1%)	104,978 (100%)
2019Q1	163,790 (93.8%)	4,226 (2.4%)	1,783 (1.0%)	991 (0.6%)	664 (0.4%)	459 (0.3%)	2,686 (1.5%)	174,599 (100%)

For both data sets these ranges lead to a couple of issues. First, it makes it hard to quantify the impact of the Noord/Zuidlijn as the amount of people that have been affected is not known. Second, for bin 1 to 50 it is impossible to determine whether the recorded travel time or route is representative. Besides, it complicates further aggregation as it is not known how often each half hour block or route should be weighed.

4.4. Interoperable smart card data analysis methodology

A methodology has been developed to cope with the limitations of interoperable smart card data as a result of the large amount of binned values, especially in the lowest range, and the data analysis challenges this leads to. This section describes the methodology generally. In the following sections the processing steps for both data sets are described in detail.

The first limitation is that the number of travelers is not known exactly. This has been overcome by using the OV-dashboard (Vervoerregio Amsterdam, 2021) from the VRA to approximate the total number of travelers per day. Using that, for each range an average value has been estimated to ensure that the total number of travelers in the data sets approaches the total number of travelers per day as obtained from the OV-dashboard. The estimates for the average values are thus determined and validated using the OV-dashboard.

To keep track of the uncertainty of the estimates, each estimated average value for the number of travelers is split into a lower bound number of travelers and an additionally estimated number of travelers. To clarify, if it turns out that the validated average value for range 51 to 100 should be 70, then the lower bound number of travelers is 51 and the estimated number of travelers is 19. This adds up to the total number of 70 travelers.

Another limitation is that the representativeness of the recorded travel time or route is uncertain for records where the number of travelers is between 1 and 50. This complicates calculating the average travel time over multiple half hour blocks or the average number of transfers on and OD-relation with many different routes. To account for this in the results, when the data is further aggregated to calculate the averages on OD-level a representativeness criterion is added. This describes on how much travelers the calculated average is at least based. If that number of travelers is higher the representativeness criterion is higher.

An example is given to clarify this. Let's say there are three half hour blocks in the travel time data from origin O to destination D. Two of these half hour blocks are traveled by 1 to 50 travelers and one is traveled by 51 to 100 travelers. The lower bound number of travelers is than 1+1+51 which equals 53. 53 is the representativeness criterion of the average travel time on this OD-relation. Now, let's say there are fifteen routes from origin O to destination D in the route data that are all traveled by 1 to 50 travelers. The lower bound number of travelers is than 15*1 which equals 15. 15 is the representativeness criterion of the average number of transfers on this OD-relation.

The approach to add a representativeness criterion is substantiated by the following assumptions. First, if it is certain that a larger number of travelers has made a journey, then the other records with 1 to 50 travelers are likely traveled more as well. This makes these records more representative. Second, if a larger number of journeys is certainly made, the amount of not representative records should be larger to affect the calculated average.

The representativeness criterion enables exclusion of OD-relations that are minimally traveled. Additionally, it allows to distinguish between the value assigned to the average travel time or number of transfers calculated for each OD-relation. Besides, it makes it possible to track the results that are based on OD-relations that are really uncertain.

The representativeness criteria for the travel time and route data sets are further specified in Section 4.5.2 and Section 4.6.2.

4.5. Processing of travel time data

This section describes the steps that have been undertaken to get from raw travel time data to data that can be used to measure the travel time impacts of the Noord/Zuidlijn. First, the steps of processing the travel time data are presented. Then, the representativeness criteria are motivated.

4.5.1. Travel time data processing steps

Processing the data to enable travel time analysis can be divided in three phases. In the first phase the average value for each range is estimated and validated for both data sets simultaneously. In the second phase the average travel time and number of travelers is calculated per OD-relation. In the third phase the output of the second phase for the before and after data is combined and compared. The Python code for the processing of the travel time data can be found in Appendix Y.

Phase 1: Estimation and validation of average for each range

The first phase is executed for both travel times data sets simultaneously. The flowchart of the first phase can be found in Figure 4.3.

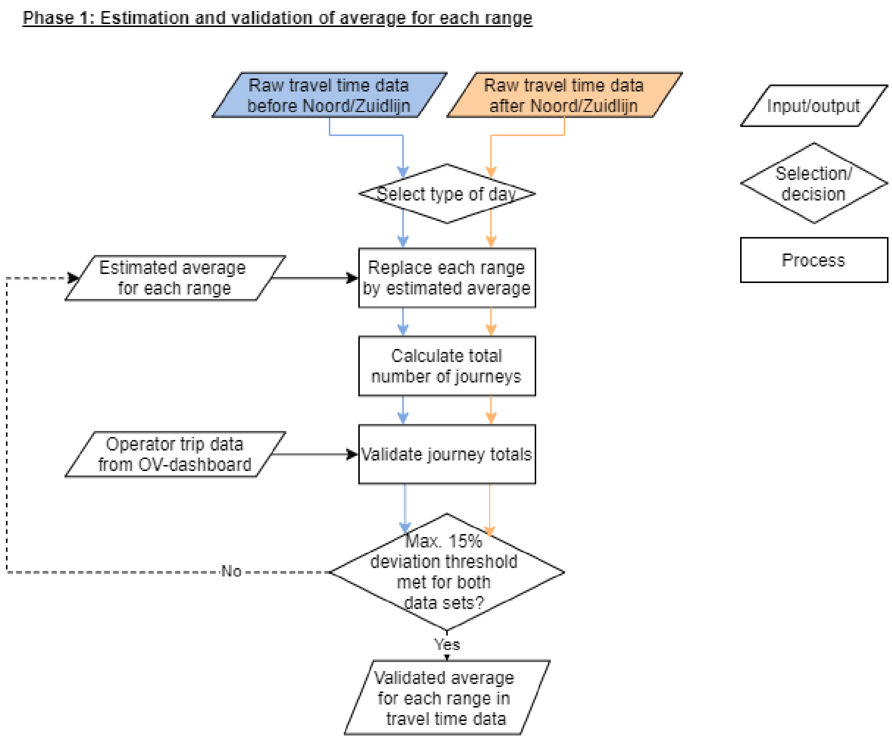


Figure 4.3: Flowchart of the first phase of processing the travel time data

To enable data processing the first phase aims to determine an average value per range. The same estimated averages are used for both data sets and for that reason the the first phase is executed for both data sets simultaneously.

First, for the raw travel time data sets before and after the Noord/Zuidlijn only data on working days (excluding school vacations) is selected. Then, the ranges are replaced by the estimated average. These values can be found in Table 4.5. These estimates are based on the sharply decreasing distribution of ranges in both data sets presented in Table 4.3. Besides, when the median and average travel time are equal it is very likely that the number of travelers is one or two because the median and average travel times are measured in seconds. The value one has been chosen as it likely occurs more often than two and because working with an integer is more convenient.

Table 4.5: Overview of validated average per range for the travel time data

Range [travelers]	Validated average number of travelers [travelers]
1-50	If median travel time [s] = average travel time [s]: 1 else: 13
51-100	65
101-150	120
151-200	175
201-250	225
251-300	275

When the ranges have been replaced by estimates, the total number of journeys is calculated. This value is then validated using individual operator data. This ensures that the estimates lead to a plausible number of journeys. If the calculated journey totals for the before and after data differ less than 15% from the journey totals based on individual operator data, then the estimated average are valid and are used in the second phase. If the threshold of maximum 15% deviation is not met for both data sets, then new estimated averages are tested. This is an iterative process. The results of the validation can be found in Table 4.6. A more extensive explanation on the validation can be found in Appendix J.

Table 4.6: The number of journeys on a working day (excluding school vacations) based on dashboards data and travel time data

Data set	Estimated daily journeys travel time data [journeys]	Daily journeys OV-dashboard [journeys]	Absolute difference [journeys]	Factor [-]
2018Q2	772,000	774,000	-2,000	1.0
2019Q1	950,000	1,001,000	-51,000	0.95

Phase 2: Calculate average travel time and number of travelers per OD-relation
The second phase entails calculation of the average travel time and number of travelers per OD-relation. This phase is executed for both data sets independently. The steps of the second phase are visualized in the flowchart in Figure 4.4

First, either the raw before or after travel time data set is selected. Then, the working days (excluding school vacations) and regional OD-relations are selected. Next, the validated averages from the first phase are used to replace the ranges. Thereafter, the median travel time per OD-relation over all half hour blocks in a day is calculated. The median travel time over the day has been obtained by adding the median travel time for each half hour block in the day to a list as many times as the number of travelers in that half hour block. Thereafter, that list has been ordered and the median of that list was taken to obtain the median travel time over the day.

Next, the outliers are removed. To be labeled as an outlier three criteria should be met:

- The number of journeys should be between 1 and 50.
- The median travel time in that half hour block should differ 50% from the median travel time over the whole day.
- The median travel time in that half hour block should differ at least 30 minutes from the median travel time over the day.

These criteria have been determined by observing the outliers in the data. The number of removed outliers can be found in Appendix L.

When the outliers are removed, the weighted average travel time per OD-relation over each of the four distinct time periods (morning peak, evening peak, rest day and whole day) is calculated. Formula 4.1 presents the calculation of the weighted average travel time per OD-relation per time period.

$$\overline{tt}_{o,d,p,y} = \frac{\sum_{h\in p} \overline{tt}_{o,d,h,y} * np_{o,d,h,y}}{\sum_{h\in p} np_{o,d,h,y}} \tag{4.1}$$

in which $\overline{tt}$ is the average travel time, np is the number of travelers, o is the origin, d is the destination, h is a half hour block, p is the time period and y stands for the year (i.e. before or after Noord/Zuidlijn).

Then, the number of travelers in that time period has been split in a lower bound and estimated number of travelers. This has been done using the following rules for each half hour block:

- If the number of journeys is above 300, then the lower bound number of travelers is equal to the number of journeys.
- If the number of journeys is 1 to 50 and the median travel time in a half hour block equals the average travel time in an half hour block, then the lower bound number of travelers is 1.
- If the number of journeys is 1 to 50 and the median travel time in a half hour block is not equal to the average travel time in an half hour block, then the lower bound number of travelers is 3.
- If the number of journeys is larger than 50, but smaller than 300, then the lower bound number of travelers is equal to the lower bound of a range.
- The estimated number of travelers is equal to the validated average value minus the lower bound of the bin that the validated average value replaces.

The total number of travelers in a half hour block is equal to the sum of the lower bound number of travelers and the estimated number of travelers (see Formula 4.2). To obtain the number of travelers on an OD-relation per time period, the lower bound and estimated number of travelers have been summed over all half hour blocks in the time period. This is done using Formula 4.3 and Formula 4.4.

$$np_{o,d,h,y} = np_{lb,o,d,h,y} + np_{est,o,d,h,y} \tag{4.2}$$

$$np_{lb,o,d,p,y} = \sum_{h \in P} np_{lb,o,d,h,y} \tag{4.3}$$

$$np_{est,o,d,p,y} = \sum_{h \in P} np_{est,o,d,h,y} \tag{4.4}$$

In these formulas lb stands for lower bound and est for estimated.

The lower bound number of travelers for each OD-relation are used to determine the representativeness of the travel time. This is described in detail in Section 4.5.2. The last step of the second phase is to convert the total number of travelers on an OD-relation during the aggregation phase to the number of travelers per working day. Therefore, the lower bound and estimated number of travelers has been divided by the number of working days.

The outcome of the second phase is a data set that consists of the weighted average travel time and the number of total, lower bound and estimated travelers for all regional OD-relations per time period.

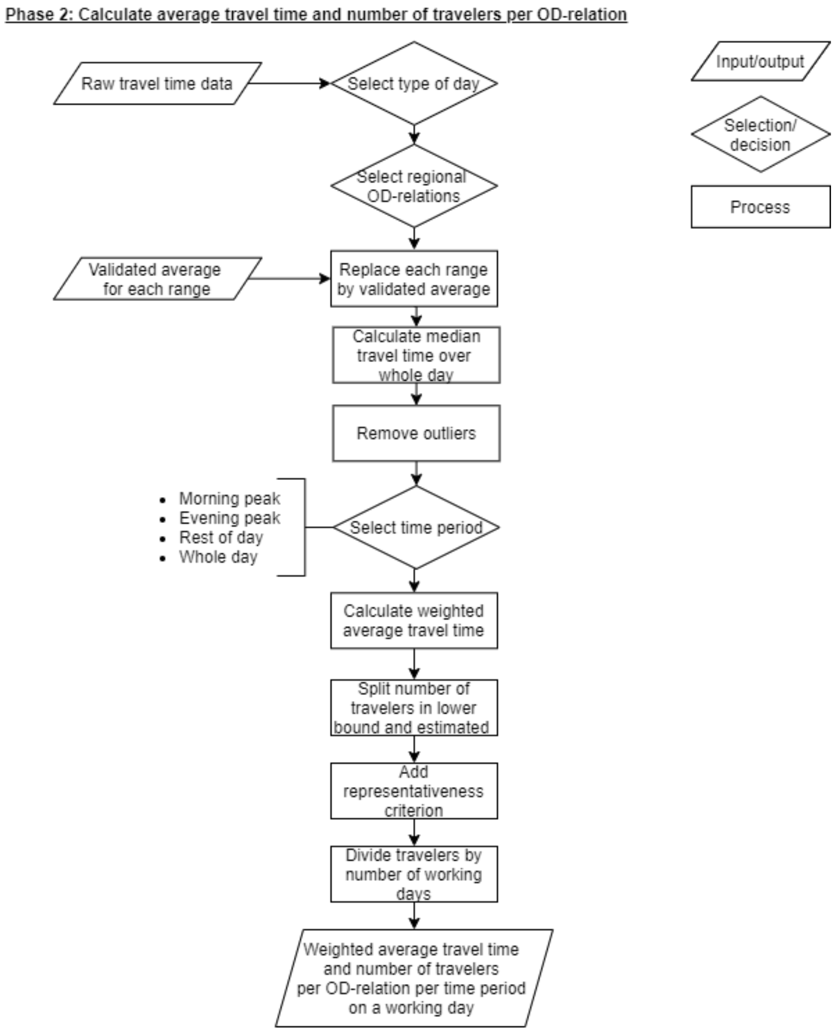


Figure 4.4: Flowchart of the second phase of processing the travel time data

Phase 3: Combine before and after data and calculate changes in travel time and number of travelers
The steps of the third phase have been visualized in Figure 4.5. In that phase, the before and after data set resulting from the second phase are merged on the origin, destination and time period. Afterward, the absolute and relative change in travel time and number of travelers are calculated. Formula 4.5 and Formula 4.6 are used to calculate the absolute and relative change in travel time respectively. The absolute and relative change in number of travelers are calculated using Formula 4.7 and Formula 4.8 respectively.

$$\Delta tt_{abs,o,d,p} = tt_{after,o,d,p} - tt_{before,o,d,p} \tag{4.5}$$

$$\Delta tt_{rel,o,d,p} = \frac{\Delta tt_{abs,o,d,p}}{tt_{before,o,d,p}} * 100 \tag{4.6}$$

$$\Delta np_{abs,o,d,p} = np_{after,o,d,p} - np_{before,o,d,p} \tag{4.7}$$

$$\Delta np_{rel,o,d,p} = \frac{\Delta np_{abs,o,d,p}}{np_{before,o,d,p}} * 100 \tag{4.8}$$

In these formulas abs stands for absolute and rel for relative.

Next, the coordinates, type and name of each check in and check out cluster are added to the table. The output of phase 3 is a data set with the weighted average travel time and number of travelers per OD-relation and time period before and after the Noord/Zuidlijn and the difference between them. In Appendix M the columns of the data set resulting from the processing phase are presented.

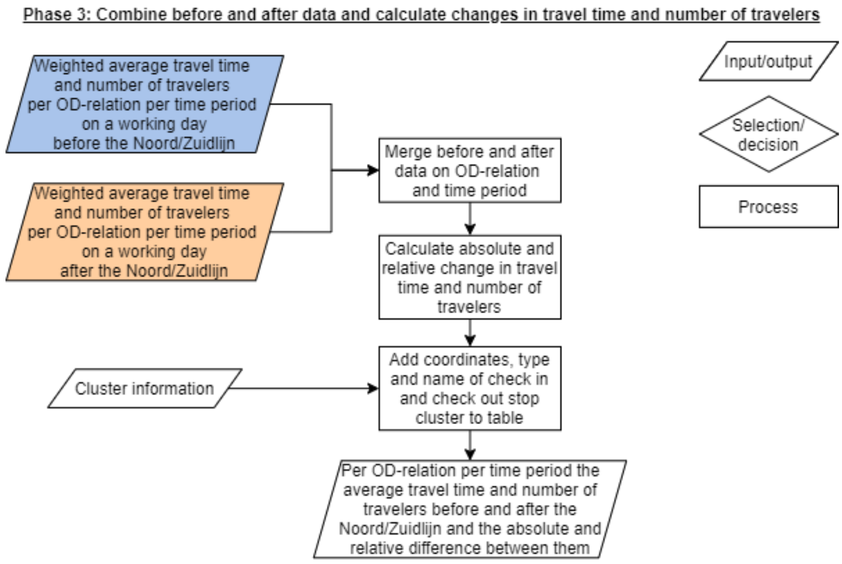


Figure 4.5: Flowchart of the third phase of processing the travel time data

4.5.2. Representativeness criteria for the travel time impacts

This section clarifies the representativeness criteria for the change in travel time per OD-relation.

If the number of travelers is between 1 and 50, the representativeness of the recorded travel time is uncertain. More than 51 travelers is assumed to be enough to obtain a representative travel time. Therefore, a distinction is made between different categories of representativeness of the average travel time. These categories are:

- The average travel time is based on at least 51 travelers on an OD-relation both before and after the Noord/Zuidlijn in a certain time period.
- The average travel time is based on at least 12 travelers on an OD-relation both before and after the Noord/Zuidlijn in a certain time period.
- The average travel time is based on at least 4 travelers on an OD-relation both before and after the Noord/Zuidlijn in a certain time period.
- The average travel time is based on at least 1 traveler on an OD-relation both before and after the Noord/Zuidlijn in a certain time period.

To determine those categories, it is assumed that if the median travel time is not equal to the average travel time in a half hour block then at least 3 journeys have been made. So, if at least 51 travelers have traveled in a time period, this means that over all half hour blocks in that time period bin 51-100 occurs at least one time or that bin 1-50 occurs at least 17 times, namely 17 times 3. The cut at 12 is based on at least 4, the number of half hour blocks in the morning and evening peak, times 3 journeys. The cut at 4 comes from 4 times 1 journey.

The number of OD-relations and the number of travelers before and after the Noord/Zuidlijn per representativeness category of the travel time data are presented in Table 4.7.

Table 4.7: Number of OD-relations and travelers per day for each representativeness criterion in the travel time data

Representativeness criterion	Number of OD-relations [OD-relations]	Total travelers per day before (lower bound | estimated) [travelers]	Total travelers per day after (lower bound | estimated) [travelers]
1-3	308	40 (30 | 10)	130 (60 | 70)
4-11	478	300 (150 | 150)	720 (270 | 450)
12-50	1,522	5,600 (1,800 | 3,800)	7,500 (2,700 | 4,800)
≥51	2,315	82,300 (55,300 | 27,000)	95,400 (76,700 | 18,700)
Total	**4,623**	**88,300** **(57,300 | 31,000)**	**103,700** **(79,700 | 24,000)**

4.6. Processing of route data

This section describes the steps that have been undertaken to get from raw route data to data that can be used to measure the transfer impacts of the Noord/Zuidlijn. It describes the data processing steps in detail. Thereafter, the representativeness criteria are described.

4.6.1. Route data processing steps

Data processing of the route data can be divided into two phases. In the first phase an average value for each range is estimated and validated for both data sets at the same time. During the second phase the average number of transfers and travelers per route and OD-relation are calculated. The third phase combines and compares the before and after data. The Python code for the processing of the travel time data can be found in Appendix Y.

Phase 1: Estimation and validation of average for each range

The first phase of route data processing is executed for both route data sets simulteneously. In Figure 4.6 the flowchart of the first phase can be found.

First, routes made on one of the following days and period on the day are selected: working days (excluding school vacations) morning peak, working days (excluding school vacations) evening peak and working days (excluding school vacations) rest day. Hereafter, the ranges for the number of travelers on a route are replaced by estimates averages. The estimates can be found in Table 4.8.

Table 4.8: Overview of validated average per range for the route data

Range [travelers]	Validated average number of travelers [travelers]
1-50	3
51-100	65
101-150	120
151-200	175
201-250	225
251-300	275

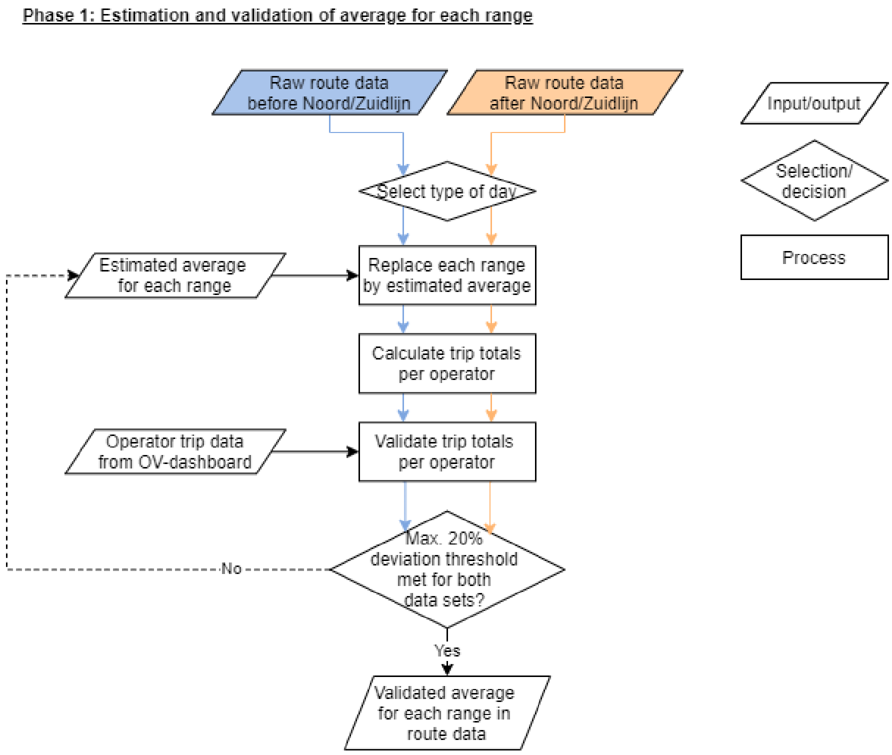


Figure 4.6: Flowchart of the first phase of processing the route data

After replacing the ranges by estimates, the trip totals per operator have been calculated. These are compared to the operator trip totals from the OV-dashboard (Vervoerregio Amsterdam, 2021). If the trip totals per operator deviate less than 20% for both the before and after data, then the estimated averages are valid. If the threshold is not met, new estimates averages are tested. This is an iterative process. The results of validation can be found in Table 4.9 and Table 4.10. A detailed explanation of the validation is set out in Appendix K.

Table 4.9: Trip totals on a working day (excluding school vacations) based on dashboards data and route data for 2018Q2

Operator	**Estimated daily trips route data [trips]**	**Daily trips OV-dashboard [trips]**	**Absolute difference [trips]**	**Factor [-]**
CXX	101,720	104,000	-2,280	0.98
EBS	23,640	26,000	-2,360	0.91
GVB	560,480	508,000	+52,480	1.1

Table 4.10: Trip totals on a working day (excluding school vacations) based on dashboards data and route data for 2019Q1

Operator	**Estimated daily trips route data [trips]**	**Daily trips OV-dashboard [trips]**	**Absolute difference [trips]**	**Factor [-]**
CXX	144,690	130,000	+14,690	1.11
EBS	35,730	35,000	+730	1.02
GVB	733,780	697,000	+36,780	1.05

Phase 2: Calculate average number of transfers and travelers per route and per OD-relation
The second phase is executed for both data sets independently. The flowchart of the second phase can be found in Figure 4.7.

This phase starts with selecting either the before or after data. Then, the route made on working days and regional OD-relations are selected. The ranges are then replaced by the validated average for each bin obtained during phase 1. Next, the number of transfers per route is added. The next step is to remove the outliers. A route is labeled as an outlier when one of the following four criteria is met:

- The number of travelers on the route should be between 1 and 50 and in the route a transfer from train to BTM and back should take place. Such a transfer pattern is illogical because of the hierarchy in the public transport network of Amsterdam.
- The number of travelers on the route should be between 1 and 50 and four or more transfers are made on that route. A route consisting of five trips and four transfers is not logical considering the regional public transport network.
- The number of travelers on the route should be between 1 and 50 and three or more transfers are made on that route while a direct route exists as well. If a direct route exists on an OD-relation, a route with at least three transfers more is illogical.
- The number of travelers on the route should be between 1 and 50 and the exact same route is not made in any of the other periods of the day and at least three or more transfers are made on that route. The combination of the fact that the route is not made during another time period and the relatively large number of transfers indicate that this route is not the most preferred route.

These criteria are based on expected and logical routes for regional public transport journeys in Amsterdam. Besides, the removal criteria are quite broad to prevent that logical and illogical routes are weighted the same in the analysis.

After removal of outliers, the number of transfers is calculated for each route by multiplying the number of transfers by the number of travelers. Afterwards, the number of transfers and travelers has been split in a lower bound and additional estimated number of travelers. The number of transfers and travelers are then divided by the number of working days. The resulting data set consists of all routes on regional OD-relations on working days and the lower bound and estimated number of transfers and travelers per route. This is the first output of phase 2 of the processing of route data.

The next step is to calculate the average number of travelers and transfers per OD-relation. Therefore, the number of travelers and transfers is summed over all routes on an OD-relation. Formula 4.9 is used to calculate the total number of transfers made on an OD-relation. With Formula 4.10 the number of travelers per OD-relation is calculated. The average number of transfers per OD-relation is then calculated using Formula 4.11.

$$nt_{o,d,y} = \sum_{r \in R_{o,d,y}} nt_{o,d,r,y} * np_{o,d,r,y} \tag{4.9}$$

$$np_{o,d,y} = \sum_{r \in R_{o,d,y}} np_{o,d,r,y} \tag{4.10}$$

$$\overline{nt}_{o,d,y} = \frac{nt_{o,d,y}}{np_{o,d,y}} \tag{4.11}$$

In these formulas nt is the number of transfers, np is the number of travelers, r is a route, R is a set of routes, o is the origin, d is the destination and y is the year.

Last, to each OD-relation a representativeness criterion is added that indicates the lower bound number of travelers on that OD-relation.

The second output of phase 2 consists of the weighted average number of transfer and the number of travelers per OD-relation on a working day.

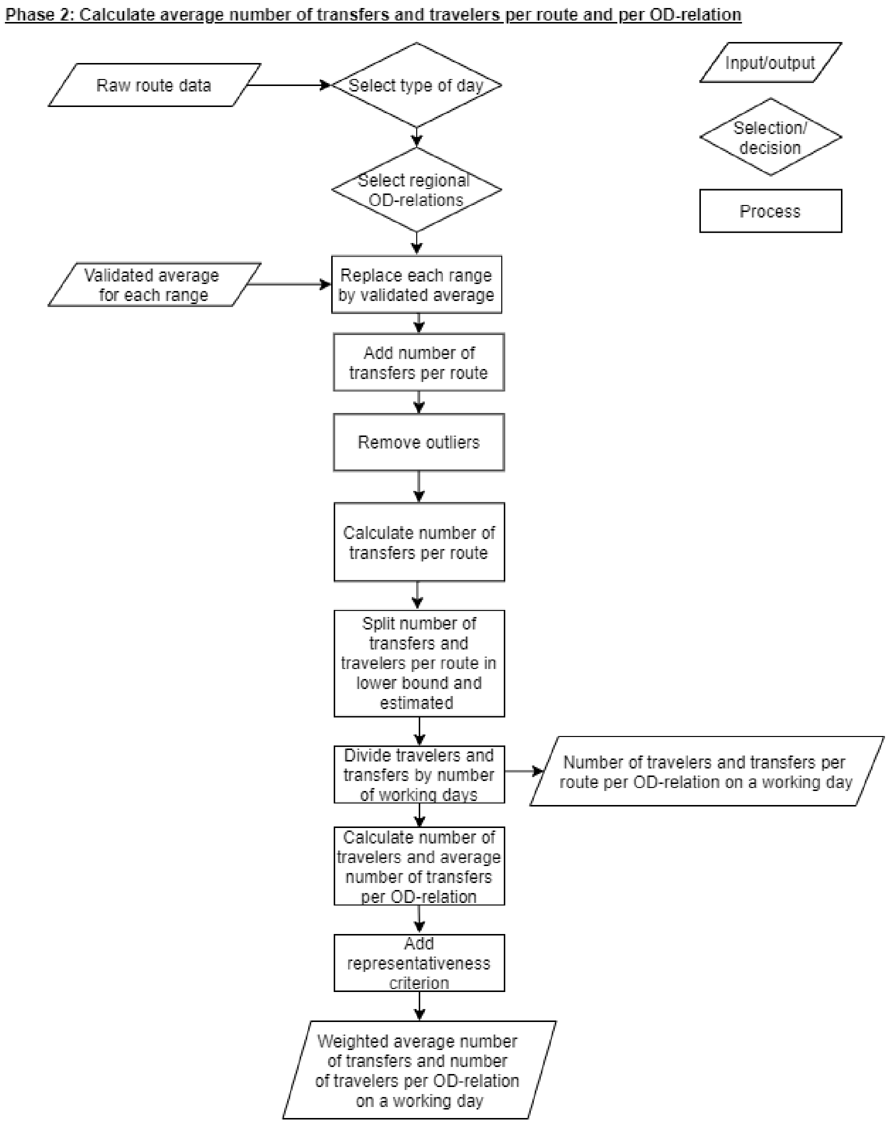


Figure 4.7: Flowchart of the second phase of processing the route data

Phase 3: Combine before and after data and calculate changes in number of transfers and travelers
The third processing phase consists of comparing the before and after situation on OD-level. The flowchart thereof is shown in Figure 4.8.

First, the before and after data are merged per OD-relation. Thereafter, the absolute and relative difference in number of travelers and average number of transfers between before and after is obtained for each OD-relation. Last, information on the coordinates, type and name of the check in and check out stop cluster is added. The output of the third phase is a data set that contains the average number of transfer and travelers on working days on regional OD-relations before and after the Noord/Zuidlijn. Besides, the difference in number of transfers and travelers is known.

$$\Delta \overline{nt}_{abs,o,d} = \overline{tt}_{after,o,d} - \overline{tt}_{before,o,d} \tag{4.12}$$

$$\Delta np_{abs,o,d} = np_{after,o,d} - np_{before,o,d} \tag{4.13}$$

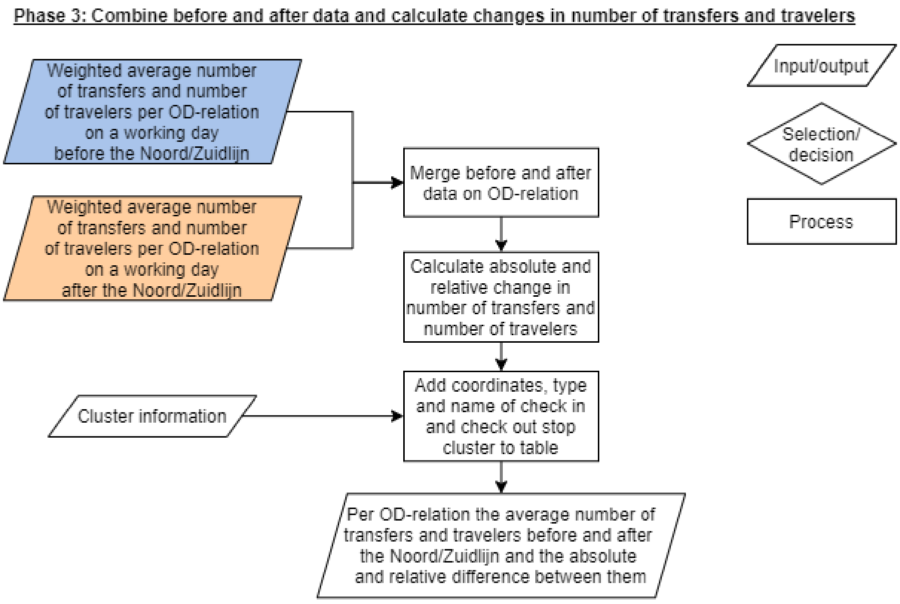


Figure 4.8: Flowchart of the third phase of processing the route data

4.6.2. Representativeness criteria for the transfer impacts

For the transfer analysis a distinction has been made between OD-relations that only have routes with 1-50 travelers and OD-relations that have at least one route with more than 51 travelers. This results in two representativeness criteria. The substantiation of this criterion is that in the calculation of the average number of transfers for that OD-relation the route traveled by at least 51 travelers, is weighed more than routes with only 1 to 50 travelers. Additional information on the representativeness of the route, for example the travel time, is lacking.

The number of OD-relations, the average number of travelers and the number of travelers before and after the Noord/Zuidlijn per representativeness category of the route data are presented in Table 4.11.

Table 4.11: Number of OD-relations and travelers per day for each representativeness criterion in the route data

Representativeness criterion	**Number of OD-relations [OD-relations]**	**Total travelers per day before (lower bound | estimated) [travelers]**	**Total travelers per day after (lower bound | estimated) [travelers]**
1-50	2,569	3,700 (1,300 | 2,400)	6,900 (4,300 | 2,600)
≥51	2,041	78,500 (69,800 | 8,700)	93,800 (86,100 | 7,700)
Total	**4,610**	**82,200 (71,100 | 11,100)**	**100,700 (90,400 | 10,300)**

4.7. Impact analysis

This section describes what analyses are performed, why these analyses are performed and how these analyses are performed. First, the approach of the analysis is discussed. Thereafter, the travel time and transfer analyses are described in detail.

4.7.1. Impact analysis approach

For both the travel time and transfer analyses a hierarchical approach is used. First, the overall impacts are analyzed. These give a general insight in the impacts and are easy to interpret from a policy perspective. However, they do not give any insights in the distribution of the impacts. Therefore, the distribution of the impacts is analyzed. This entails splitting the overall impacts in gains and losses and investigating the distribution of the impacts over all OD-relations and travelers. Then, the impact is analyzed on OD-level. This way useful insights are obtained in where the largest effects occur and regarding negative impacts, where additional measures are required. The most important OD-relations are those where both the impact and the number of travelers affected is large. However, OD-relations where a large impact occurs but that are not so heavily traveled are of interest as well. Last, a couple of cases are selected and visualized on a map to substantiate the findings.

In the analysis phase, the representativeness criteria and the lower bound and estimated number of travelers are recorded as well. This determines how much value can be assigned to the results. Below, the performed travel time and transfer analysis are described extensively.

4.7.2. Travel time impact analysis

The travel time analysis follows the hierarchical approach described above. Besides, the effect of selecting a different time period or direction is analyzed. This helps interpretation of the results later on.

Analysis of the time period on the travel time impacts

Because of differences in supply and demand of public transport during the day, a different time period (i.e. morning peak, evening peak and rest of day) might result in different travel times and thus changes in travel time as well. Considering that travel demand is unevenly spread over the day, different impacts per time period therefore affect the overall impacts. The overall impacts namely are a product of the change in travel time and the number of people it occurs to.

The influence of the time period on the results is analyzed by selecting all OD-relations for which a representative (i.e. the travel times in both the before and after situation are based on at least 51 travelers) change in travel time for all time periods can be obtained. The whole day time period is left out of this analysis because it is a weighting of the other time periods. For these OD-relations a cumulative distribution function of the absolute change in travel time against the percentage of OD-relations is plotted for each time period. This should give insight in whether the travel time impacts differ over time periods.

Analysis of the effect of travel direction on the travel time impacts

Because different frequencies on both ends of a transfer result in unequal waiting times and because travelers might take different routes in both directions, it is interesting to research the effect of the travel direction on the change in average travel time. A large difference in the impacts between both directions namely might affect ridership.

Direction effects are analyzed by selecting all OD-relations that are representative in both directions for the whole day. Afterwards for all OD-relations it is counted whether the travel time change is larger from Amsterdam to the region or vice versa. Next to that, for all representative OD-relations a cumulative distribution function of the difference in absolute travel time change between both directions is plotted for all time periods. Last, the top 20 OD-relations with the largest difference in absolute change in travel time have been selected. This helps finding explaining factors.

Analysis of the aggregated travel time impacts

Per OD-relation per time period the travel time benefits per day are calculated by applying Formula 4.14. The total travel time benefits can be calculated by applying that formula to all OD-relations over the whole day. It is possible to split the benefits in the lower bound and estimated travel time benefits by using either the lower bound or estimated number of travelers. To assess the travel time benefits for different representativeness criteria, Formula 4.14 is filled in for all OD-relations that meet the representativeness criterion.

$$B_{traveltime,o,d,p} = np_{before,o,d,p} * \Delta tt_{abs,o,d,p} + \frac{1}{2} * \Delta np_{abs,o,d,p} * \Delta tt_{abs,o,d,p} \tag{4.14}$$

In Formula 4.14 B stands for benefits. It uses the 'rule of half' which assumes that a new traveler experiences half of the benefits and a traveler that was already traveling experiences the full benefits.

The travel time benefits can be monetized by multiplying it by the value of time. A study by Knockaert and Koster (2020) determines the valuation of time and transfers in Amsterdam. For the value of time for bus, tram, metro and train they find 15.6, 13.9, 24.5 and 17.1 €/hour respectively. The average of these valuations is 18 €/hour. However, because the main network change is the addition of a new metro line, the valuation of the benefits could be assumed to be a little higher. However, waiting and walking time are not included and valuated at 9.5 €/hour and 14.3 €/hour respectively. For that reason, the value of time used to monetize the travel time is 18 €/hour.

Analysis of the distribution of the travel time impacts

Applying Formula 4.14 only to travelers that experience a decrease in travel time, gives the travel time savings. Similarly, calculating the benefits over all travelers that experience an increase in travel time results in the travel time losses.

Another method applied to determine the distribution of the travel time impacts is to create a histogram of the absolute change in travel time over the OD-relations and travelers. A change in travel time ranging from -1 to +1 minute is labeled as no change. These boundaries are chosen to exclude noise in the data, but they ensure that minor changes in travel time are visualized.

Analysis of the largest travel time impacts on OD-level

When considering the impact of the Noord/Zuidlijn on travel times on OD-level, the most interesting OD-relations are where the change in travel time and the number of travelers affected is large. However, the less frequently traveled OD-relations where travel time has changed strongly are interesting as well. For that reason, the analysis of the largest travel time impacts on OD-level focuses on both.

To obtain the top 10 OD-relations with the largest travel time savings and losses, Formula 4.14 has been applied to all OD-relations. The number of travelers before and the change in travelers have been summed over both directions. For the absolute change in travel time the average over both directions has been used. As a result the travel time savings and losses differ slightly from when the travel time savings or losses would have been determined for each direction separately and summed later. For the OD-relations with the largest travel time savings and losses it is assumed that the representativeness criterion of 51 is met. The substantiation thereof is that the number of travelers should be sufficiently large to lead to large travel time savings or losses.

Next to the largest travel time savings and losses, the top 20 OD-relations where the largest travel time changes occur have been investigated. This entails both decreases and increases in travel time. This analysis has been done for the relative and absolute change in travel time because small absolute changes might have a large impact on short journeys. This is not visible if only the large absolute changes are presented. This analysis does not depend on the number of travelers, except for the representativeness criterion which should meet 51 travelers. Opposite directions have not been added as this would lead to contamination of the data. Besides, there might be differences between both directions.

Spatial analysis of the travel time impacts

Last, the results are visualized geographically to explain and find patterns in the results.

First, the OD-relations with a large increase and or decrease in travel time have been visualized. The time period used is the whole day and the representativeness criterion is 51. Only OD-relations where the travel time has decreased with at least 10 minutes are shown. For OD-relations where the travel time increases only relations with at least a 5 minute decrease are visualized. This choice has been made to ensure that the most important patterns are still visible in the visualization. Additionally, the weighted average change in travel time per check in cluster is visualized. It is calculated using Formula 4.15. For that, all OD-relations with a representativeness criterion of more than 12 travelers have been selected. This choice is made to ensure that each stop cluster has a sufficient amount of OD-relations to calculate the weighted average change in travel time.

$$\overline{\Delta tt}_{o,p} = \frac{\sum_{d \in D} \Delta tt_{o,d,p} * \overline{np}_{o,d}}{\sum_{d \in D} \overline{np}_{o,d}} \tag{4.15}$$

In Formula 4.15 $\overline{\Delta tt}$ is the weighted average change in travel time, o is the origin cluster, p is the time period, d is the destination and $\overline{np}$ is the average number of travelers.

Besides, a selection of cases has been visualized to identify patterns. Therefore, a couple of many-to-one and one-to-many relations have been selected either from the region to Amsterdam or vice versa. This has resulted in the following cases:

- From all regional stop clusters to Centraal Station
- From all regional stop clusters to the Jordaan
- From all regional stop clusters to the Spui
- From all regional stop cluster to Nieuwmarkt en Lastage
- From Nieuwmarkt en Lastage to all regional stop clusters
- From Purmerend Oost to all stop clusters in Amsterdam
- From Uithoorn to all stop clusters in Amsterdam
- From Zaandam Midden to all stop clusters in Amsterdam

These cases have been selected to illustrate the change in travel time from and to different stop clusters in the area and region. For these cases the absolute change in travel time is visualized. The absolute change in travel time is chosen because it is more comprehensible than the relative change in travel time. Besides, it enables seeing the change in travel time propagate through the network. The change in travel time is visualized using a color gradient from green to red. Green stands for a decrease in travel time and red for an increase in travel time. A travel time change between -1 minute and +1 minute is visualized by a grey color. Besides, the average number of travelers per day is presented by the size of the circle or triangle of the stop cluster. A circle indicates a BTM-cluster and a triangle a train cluster.

4.7.3. Transfer impact analysis

The transfer analysis follows the same hierarchical approach as the travel time analysis.

Analysis of the aggregated transfer impacts

The aggregated analysis of the transfer impacts entails the calculation of the total amount of transfers per day in the before and after data set. This is calculated by applying Formula 4.16 to all regional OD-relations and taking the sum afterwards.

$$nt_y = \sum_{o \in O} \sum_{d \in D} np_{o,d,y} * \overline{nt}_{o,d,y} \tag{4.16}$$

In Formula 4.16 nt is the number of transfers per day, np is the number of travelers per day, $\overline{nt}$ is the average number of transfers and y is the year, either before or after the Noord/Zuidlijn.

However, because the aggregation period differs between the before and after data sets, it is likely that the daily number of journeys per day differs as well. As a result, part of the difference in the daily number of transfers calculated is not a result of the Noord/Zuidlijn, but a result of the different aggregation periods among both data sets. To account for this the average number of transfer per journey is calculated. Therefore, first the total number of journeys per day has been calculated by applying Formula 4.17 to all OD-relations and taking the sum afterwards. Then, the average number of transfers per journey can be calculated by Formula 4.18.

$$np_y = \sum_{o \in O} \sum_{d \in D} np_{o,d,y} \tag{4.17}$$

$$\overline{nt}_{journey,y} = \frac{\sum_{o \in O} \sum_{d \in D} nt_{o,d,y}}{\sum_{o \in O} \sum_{d \in D} np_{o,d,y}} \tag{4.18}$$

Formula 4.16 till Formula 4.18 are applied to all OD-relations with at least 1 traveler before and after the Noord/Zuidlijn. For the number of transfers and journeys per day the total amount is split in a lower bound and additionally estimated number of transfer and journeys respectively.

Another analysis performed to determine the aggregated impact of the Noord/Zuidlijn on transfer patterns is to determine the share of each possible combination of modalities over all journeys made before and after the Noord/Zuidlijn. The possible combinations are presented in Table 4.12

Table 4.12: Modal combinations to determine transfer impacts

One mode	Two modes	Three modes	Four modes
Metro only	Metro + Bus	Metro + Bus + Tram	Metro + Bus + Tram + Train
Bus only	Metro + Tram	Metro + Bus + Train	
Tram only	Metro + Train	Metro + Tram + Train	
Train only	Bus + Tram	Bus + Tram + Train	
	Bus + Train		
	Tram + Train		

The number of times a mode is used on a journey is not included in this analysis. This means that unimodal combinations also include journeys with intramodal transfers. There are two reasons to exclude duplicate modes in a journey. First, the possible number of combinations is kept small. Second, intramodal transfers for train and metro are not visible in smart card data. Next to exclusion of duplicates, the order of modalities is not important. This allows analysis of the effect of the Noord/Zuidlijn on transfer patterns in both directions simultaneously. For each data set the number of journeys with one of the enumerated combinations of modalities is counted and divided by the total number of journeys. This results in the modal splits of each combination of modes over all regional journeys. To account for the increase in the number of journeys as a result of the differing aggregation periods, the percentage share of each combination of modes over all journeys is calculated.

Next, the transfer benefits per OD-relation per day are calculated using Formula 4.19. To obtain the total transfer benefits the transfer benefits per OD-relation should be summed over all OD-relations. A distinction is made between both representativeness criteria.

$$B_{transfer,o,d} = np_{before,o,d} * \Delta nt_{abs,o,d} + \frac{1}{2} * \Delta np_{abs,o,d} * \Delta nt_{abs,o,d} \tag{4.19}$$

The transfer benefits are monetized by multiplying the number of transfers by 1.4 €/transfer. This valuation of a transfers has been calculated by Knockaert and Koster (2020) for Amsterdam specifically.

Analysis of the distribution of the transfer impacts

To investigate the distribution of the transfer benefits, the total transfer benefits have been split in transfer gains and losses. The transfer gains are the transfer benefits (see Formula 4.19) for all OD-relations where the average number of transfers decreases. The transfer losses occur at OD-relations where the average number of transfers increases.

Thereafter, the distribution of the transfer impact has been plotted in a histogram for both representativeness criteria. It is interesting to know which part of the OD-relations is affected and in what way. Besides, it interesting to know how what proportion of travelers undergoes a particular change. Therefore, the distribution is shown over all regional OD-relations and all regional travelers. A large binsize of 0.5 transfers is taken because the results are expected to be noisy because of the large amount of routes that fall in bin 1 to 50.

Spatial analysis of the transfer impacts

Once insights in the distribution of the benefits are obtained, the results are visualized geographically to explain and find patterns in the results. Only OD-relations with at least one route with 51 travelers and an increase or decrease of the average number of transfers of at least 0.75 are shown. This choice has been made to ensure that the visualized OD-relations are representative and a large change in the number of transfers has occurred on that OD-relation.

To visualize the effect of the Noord/Zuidlijn a selection of cases has been visualized. A case from each concession has been selected. Besides, the cases should be traveled frequently and the Noord/Zuidlijn should have affected the route travelers take. This has resulted in the following three cases:

- Amstelveen - Dam
- Purmerend Oost - Nieuwmarkt en Lastage
- Oostzaan - Centraal Station

To enable clear visualization only the top 3 most frequently used routes have been selected both before and after Noord/Zuidlijn. Besides, at least 51 people should have traveled that route.

4.7.4. Sensitivity analysis

The goal of the sensitivity analysis is to determine the impact of different estimated averages for the ranges on the results. The averages per range affect the results in two ways. First, the total number of travelers changes. Second, weighting of different half hour blocks and route to calculate the average travel time and number of transfers per OD-relation changes.

For the travel time the sensitivity of the results is tested for an underestimation and overestimation of the average value for each range. The input values should lead to a minimum deviation of 15% with the journey totals obtained from the OV-dashboard. 15% is the threshold value used to determine the validated averages for each range in Section 4.5. If the criterion of a minimum 15% deviation requires an estimated value smaller than the lower bound or larger than the upper bound of a range, then the lower or upper bound of a range is used respectively. The input values for the sensitivity analysis of the travel time results are presented in Table 4.13. The underestimated and overestimated journeys totals resulting from these input values can be found in Appendix Q.

The sensitivity of the transfer impacts is investigated using the values presented in Table 4.14. Ideally, these values result in an underestimation and overestimation of at least 20% of the trip totals per operator. 20% is the threshold value used to determine the validated averages for each range in Section 4.6. However, for the underestimated values this threshold could not be met. Therefore, the lower bound of each range is taken. The underestimated and overestimated trip totals totals for the sensitivity analysis can be found in Appendix Q.

The sensitivity is only investigated for the aggregated results. This entails the total travel time and transfer benefits.

Table 4.13: Input values for the sensitivity analysis of the travel times

Range [travelers]	**Underestimated average for each range [travelers]**	**Overestimated average for each range [travelers]**
1-50	2*	40
51-100	51	85
101-150	101	150
151-200	151	200
201-250	201	250
251-300	251	300

*If median travel time [s] = average travel time [s] then a value of 1 is used.

Table 4.14: Input values for the sensitivity analysis of the transfer results

Range [travelers]	**Underestimated average for each range [travelers]**	**Overestimated average for each range [travelers]**
1-50	1	6
51-100	51	90
101-150	101	140
151-200	151	190
201-250	201	240
251-300	251	290

A limitation of the sensitivity analysis is that only one set of underestimated averages and one set of overestimated averages is tested. A change in estimates changes the weights assigned to each half hour block (travel time data) or route (route data). This affects the share of a half hour block or route when calculating the average travel time or number of transfers on an OD-relation. This share depends on the set of averages chosen.

4.8. Wrap up data, methods and analysis

In total, four smart card data sets are used. Two thereof contain the travel times and number of travelers on regional OD-relations both before and after the Noord/Zuidlijn. The other two contain the routes traveled between regional OD-relations. Of these route data sets a before and after Noord/Zuidlijn data set is used as well. The analyses focus on working days and exclude school vacations.

The fact that records with less than 300 journeys are binned, complicates analysis because the exact number of travelers remains unknown. Besides, as most of the records fall into the lowest bin, the representativeness of the travel time or route taken is unknown. These difficulties have been overcome by estimating an average value for each bin using an external data source (Vervoerregio Amsterdam, 2021). Additionally, after calculating the average travel time or number of transfers a representativeness criterion has been added that indicates the lower bound number of travelers that traveled on this OD-relation.

The analysis then follows a hierarchical approach. It starts with the overall impacts. Afterwards, the distribution of the impacts over all OD-relations and travelers is investigated. Then, the OD-relations where the largest changes occur are visualized. Last, some cases are visualized to identify patterns and substantiate the earlier findings.

In the next chapter the travel time impacts are presented first. Thereafter, the transfer impacts follow.

5
Results

After obtaining the necessary background knowledge on the Noord/Zuidlijn in Chapter 3 and after the description of the data, methods and analysis in Chapter 4, the results regarding the travel time and transfer impacts are presented and explained in this chapter. What analyses are performed, why these analyses are done and how these are performed can be found in Section 4.7.

The chapter is divided into two main sections, namely results regarding the impact of the Noord/Zuidlijn on travel times and results regarding the impact of the Noord/Zuidlijn on transfer patterns. The travel time impacts are discussed first and then the transfer impacts are presented. This is done in order from the impact over all regional journeys to specific cases.

5.1. Impact of the Noord/Zuidlijn on regional travel times

This section presents the impact of the Noord/Zuidlijn on regional travel times. The impacts are presented in order from the overall impact on regional network level to the impacts on selected OD-relations. Last, the effect of selecting a different time period or direction has been investigated. This way the reader understands the general effect of a choosing a different time period or travel direction on the impact of the Noord/Zuidlijn.

5.1.1. Aggregated travel time impacts

Table 5.1 presents the travel time benefits per representativeness criterion. It distinguishes the lower bound and estimated benefits. From Table 5.1 it becomes clear that the travel time benefits are around 1,350 hours per day. This is equal to 0.84 minutes or 50 seconds per average traveler. Using a value of time of 18 €/hour (Knockaert and Koster, 2020) the daily travel time benefits are 24,000 euros. This can be split in lower bound benefits of 13,000 euros and 11,000 euros based on estimated travelers.

The lower bound travel time benefits on OD-relations for which the travel time changes are representative, are 605 hours. Another 501 hours of travel time benefits occur to travelers that are additionally estimated. A less conservative representativeness criterion, namely 12 to 50, results in an additional 220 hours of travel time benefits.

The largest part of the total benefits comes from the most representative OD-relations. This is because these OD-relations are most heavily traveled as well. The share of the less conservative representativeness criteria is minimal. Additionally, in Table 5.1 it can be seen that almost half of the travel time benefits are based on estimated travelers.

Table 5.1: Travel time benefits over all OD-relations over the whole day distinguished in different representativeness categories and split in the lower bound and estimated benefits

Representativeness criterion	Total travel time benefits [h] (lower bound | estimated)	Average number of travelers per day [travelers]	Average travel time benefits per traveler [min]
1-3	6 (3 | 3)	100	4.4
4-11	18 (9 | 10)	500	2.1
12-50	220 (94 | 126)	6,500	2.02
≥51	1,105 (605 | 501)	88,800	0.75
Total	**1,349** **(710 | 639)**	**95,600**	**0.84**

The aggregated travel time impacts, however, do not give any information about the magnitude of the travel time savings and losses separately. Therefore, these are discussed in the following section.

5.1.2. The distribution of the travel time impacts

Table 5.2 presents the travel time savings and losses. Per day, 2,350 hours of travel time are saved while 1,000 hours are lost. The travel time savings are thus more than two times as large as the travel time losses. Additionally, with a total of more than 1,000 hours of travel time losses the negative impact of the Noord/Zuidlijn on regional travel times is substantial. Similar as for the overall travel time impacts (see Section 5.1.1) the OD-relations with a representativeness criterion larger than 12 travelers contribute most to the travel time savings and losses. The share of the less representative OD-relations is very small.

Table 5.2: Travel time savings and losses

Representativeness criterion	Net travel time benefits [h]	Travel time savings [h]	Travel time losses [h]
1-3	6	12	6
4-11	18	39	21
12-50	220	353	133
≥51	1,105	1,953	848
Total	**1,349**	**2,357**	**1,008**

In Figure 5.1 it can be seen that for around 50% of the OD-relations the travel time decreases. For around 27.5% of the OD-relations the travel time is minimally affected by the Noord/Zuidlijn and around 22.5% of the OD-relations experiences an increase in travel time. In Appendix U it can be seen that selection of a lower representativeness criterion results in a slightly larger percentage of OD-relations for which the travel time has increased.

However, analyzing the travel time impacts on travelers a less spread pattern is visible. Figure 5.2 shows that 60,000 passengers do not experience a significant change in travel time. This is around 67% of the total number of passengers on representative OD-relations. Around 10% of the travelers experience an increase in travel time, while around 20% of the travelers experience a decrease in travel time. The high peak and short and thin tails of Figure 5.2 compared to Figure 5.1 indicate that most people travel on OD-relations where no or minor changes occur.

Another interesting observation in Figure 5.2 and Figure 5.1 is that the tails in the positive direction are longer than in the negative direction. This means that the the positive changes in travel time are of a larger magnitude than the negative changes.

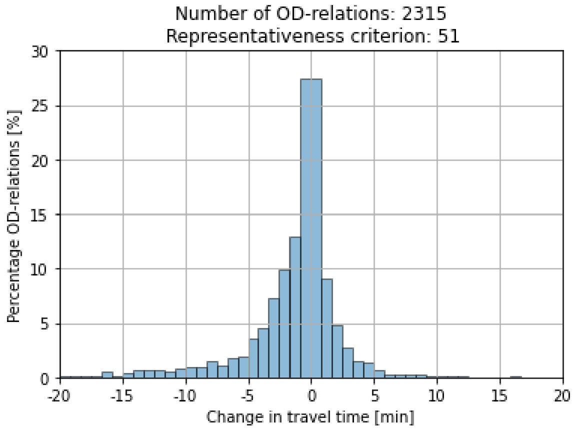


Figure 5.1: Distribution of the absolute change in travel time against the percentage of OD-relations. Results are plotted for the average travel time over the whole day.

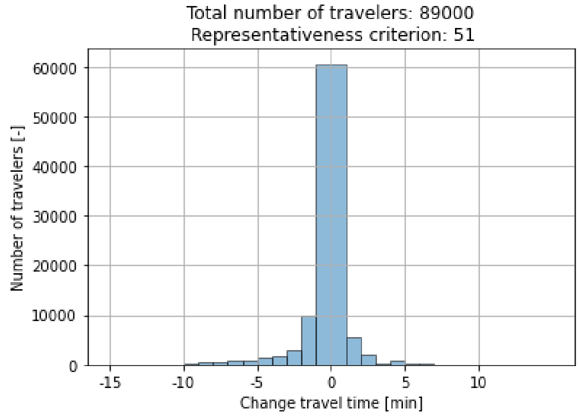


Figure 5.2: Distribution of the absolute change in travel time against the number of travelers. Results are plotted for the average travel time over the whole day.

5.1.3. OD-relations with the largest total travel time savings and losses

In this subsection the top 10 OD-relations with the largest travel time savings and losses are presented. For that, both directions of the OD-relation have been summed. The representativeness criterion that the average travel time is based on at least 51 travelers before and after the Noord/Zuidlijn is met.

In Table 5.3 the top 10 OD-relations with the largest travel times savings over a whole working day are shown. The largest travel time savings occur on OD-relations either from or to areas near the Noord/Zuidlijn. This matches the expectations. For these relations the travel time has decreased strongly and an increase in demand can be observed. The latter is probably the result of passengers shifting from bus and tram to metro and these metro stops being assigned to different stop clusters.

However, the OD-relations between Amstelveen and Bijlmer can not be explained by the commissioning of the Noord/Zuidlijn. As can be seen in Table 5.3 this can be explained by the fact that many travelers experience a small decrease in travel time on these OD-relations.

Again, it can be observed that the lower bound values contribute most to the total travel time savings.

Table 5.3: Top 10 bidirectional OD-relations with largest travel time savings over a whole working day

From/to	**From/to**	**Travel time savings [h] (lower bound | estimated)**	**Travelers before per day [travelers/day]**	**Change in travelers per day [travelers/day]**	**Change in travel time [min]**
Pijp	Schiphol NS	65 (58 | 7)	172	+274	-12.7
Amstelveen Zuid	Spui	34 (30 | 4)	28	+303	-11.3
Metrostation Vijzelgracht	Schiphol NS	29 (23 | 6)	29	+175	-15.0
Purmerend Zuid-Oost	Zuid	24 (20 | 4)	64	+102	-12.4
Amstelveen Centrum	Bijlmer ArenA	23 (22 | 1)	1,006	+165	-1.3
Purmerend West	Zuid	18 (14 | 3)	44	+72	-13.0
Amstelveen Zuid	Bijlmer-West	17 (16 | 1)	391	+209	-2.1
Purmerend Zuid-Oost	Spui	17 (15 | 2)	43	+254	-6.0
Amstelveen Zuid	Noord midden	17 (11 |6)	57	+26	-14.4
Amstelveen Zuid	Centraal Station	17 (15 | 2)	411	+13	-2.4

Table 5.4 shows that the travel time losses are large between the center of Amsterdam, around Centraal Station and Nieuwmarkt en Lastage, and Purmerend West, Purmerend Oost and Purmerend Zuid-Oost. This can be explained by the fact that from these areas the direct bus connections to Amsterdam Centraal disappeared. This results in an additional transfer to a bus that runs directly to Amsterdan Centraal or to the Noord/Zuidlijn. However, the short metro leg does not compensate the time lost by the extra transfer because the time in the metro is not long enough.

Besides, two out of ten OD-relations with the largest travel time losses are between Amstelveen and the southwestern part of the city center, namely between Amstelveen and Leidseplein en Leidsetraat and between Amstelveen and Museumkwartier where a slight increase in travel time affects a lot of travelers. This can be explained by the fact that the distribution of travelers over the buses from Amstelland-Meerlanden to Amsterdam is not 50%-50% as forecasted in the Lijnennetvisie (Stadsregio Amsterdam, 2014). It turns out that buses to Elandsgracht are preferred by travelers because these buses stop at metrostation Amstelveenseweg. For a part of the travelers a transfer to the metro at Amstelveenseweg is quicker than a transfer at Amsterdam Zuid. The travel time to Amstelveenseweg from the point where the buses to Zuid and Elandsgracht diverge is namely shorter than to Amsterdam Zuid. Besides, the walking distance between the bus stop and the metro platform is shorter at Amstelveenseweg. This preference for buses to Elandsgracht causes that theses buses have a higher occupancy resulting in longer boarding and alighting times eventually resulting in longer travel times. As a result, buses to station Amsterdam Zuid have a lower occupancy. However, because the buses from Amstelland-Meerlanden run over the same route for a large part in Amstelveen without any overtaking locations this slows down coming successive buses as well.

Comparing Table 5.3 and Table 5.4 shows that the magnitude of the top 10 travel time savings is slightly larger than the magnitude of the top 10 travel time losses. Besides, it can be observed that for the travel time savings the change in travel time and the change in travelers often determines a large part of the total travel time savings. For the travel time losses, however, the number of travelers before the change occurred contributes to a large part to the total travel time losses. In line with that, a decrease in the number of travelers can be observed on most OD-relations in Table 5.4. This indicates that the Noord/Zuidlijn leads to a shift in demand over stop clusters.

Table 5.4: Top 10 bidirectional OD-relations with largest travel time losses over a whole working day

From/to	**From/to**	**Travel time losses [h] (lower bound | estimated)**	**Travelers before per day [travelers/day]**	**Change in travelers per day [travelers/day]**	**Change in travel time [min]**
Nieuwmarkt en Lastage	Purmerend West	28 (22 | 5)	252	-187	+10.7
Centraal Station	Purmerend West	27 (24 | 3)	300	-75	+5.1
Centraal Station	Purmerend Oost	26 (23 | 3)	280	1	+4.7
Centraal Station	Landsmeer/ Den Ilp	22 (18 | 4)	161	10	+6.4
Nieuwmarkt en Lastage	Purmerend Oost	21 (16 | 5)	211	-152	+9.7
Amstelveen	Museumkwartier	21 (19 | 2)	452	10	+2.8
Centraal Station	Purmerend Zuid-Oost	20 (19 | 1)	538	67	+2.0
Amstelveen	Leidseplein en Leidsestraat	18 (16 | 2)	439	-135	+3.0
Landsmeer/ Den Ilp	Nieuwmarkt en Lastage	17 (11 | 6)	123	-95	+14.2
Nieuwmarkt en Lastage	Purmerend Zuid-Oost	16 (15 | 2)	480	-164	+2.6

5.1.4. OD-relations with the largest absolute decrease and increase in travel time

This subsection presents the OD-relations where the largest absolute changes in travel time occur. The changes are independent of the number of travelers on the OD-relation.

Table 5.5 shows that the largest travel time savings occur at OD-relations with a long part of the journey on the Noord/Zuidlijn. These OD-relations are from Amstelland-Meerlanden to Amsterdam Noord and from Waterland to the southern part of Amsterdam.

An interesting observation is that the impact on travel time for relations from Amstelland-Meerlanden to Noord is stronger than vice versa. A reason for this could be that the frequency of the Noord/Zuidlijn is higher than the frequency of the Amstelveenlijn. This leads to larger waiting times for a transfer from the Noord/Zuidlijn to the Amstelveenlijn than vice versa.

No OD-relations between Zaanstreek and Amsterdam are present in the top 20 of OD-relations with the largest decrease in travel time. This can be explained by the fact that from that area only bus line 392 is connected to the Noord/Zuidlijn at Noord. The proposed stop at station Noorderpark has not been built because of the high costs and large construction efforts (Vervoerregio Amsterdam, 2020).

Because there's quite some overlap between the top 20 OD-relations with the largest relative and absolute decrease in travel time, the top 20 OD-relations with the largest relative decrease in travel time can be found in Appendix T. As can be expected OD-relations covering a relatively large distance experience a large absolute decrease more often. OD-relations over short distances, for example between concession Waterland and Amsterdam Noord or Oostelijke Eilanden undergo a large relative decrease in travel time.

Table 5.5: Top 20 OD-relations with largest absolute decrease in travel time over a whole working day. The representativeness criterion is 51 travelers.

Check in cluster	Check out cluster	Change travel time [min]
Amstelveen Busstation	Noord midden	-24.6
Amstelveen Centrum	Noord midden	-20.4
Amstelveen	Noord midden	-19.4
Rivierenbuurt	Monnickendam	-18.4
Amstelveen Centrum	Noord oost	-17.6
Amstelveen Zuid	Noord midden	-17.0
Rivierenbuurt	Purmerend Oost	-16.6
Rivierenbuurt	Ilpendam/Broek in Waterland	-16.6
Rivierenbuurt	Purmerend West	-16.6
Zuid	Purmerend West	-16.5
Ilpendam/Broek in Waterland	Rivierenbuurt	-16.3
Ilpendam/Broek in Waterland	Oostelijke Eilanden en Kadijken	-16.2
Uithoorn	Noord midden	-16.2
Schiphol NS	Metrostation Vijzelgracht	-16.1
Monnickendam	Rivierenbuurt	-16.1
Rivierenbuurt	Purmerend Centrum	-16.1
Zuid	Purmerend Centrum	-16.0
Noord midden	Aalsmeer/Kudelstaart	-16.0
Zuid	Ilpendam/Broek in Waterland	-15.8
Aalsmeer/Kudelstaart	Noord midden	-15.3

In Table 5.6 it can be seen that travelers between Nieuwmarkt en Lastage or Centraal Station and Landsmeer/-Den Ilp, Purmerend West and Oost or vice versa experience a large increase in travel time. This is caused by the disappearance of direct connections between those stop clusters.

Additionally, in Table 5.6 it can clearly be seen that from Amstelland-Meerlanden to stop clusters in the city center, like Jordaan en Dam, the direct connections have disappeared. As a result, the travel times have increased.

By comparing Table 5.5 and Table 5.6 it can be seen that the absolute decreases in travel time are larger than the absolute increases. The top 20 absolute travel time savings range from -25 minutes to -15 minutes, while the top 20 absolute increases in travel time range from around +17 minutes to +6 minutes.

There's quite some overlap between the OD-relations that are most heavily affected absolutely and relatively speaking. Therefore, only the results for the OD-relations with the largest absolute change in travel time are presented here. The results for the relative change can be found in Appendix T. The relations between Nieuwmarkt en Lastage or Centraal Station and Landsmeer/Den Ilp, Purmerend West and Oost or vice versan are visible for the relative change in travel time as well. The direct connections to Noord midden from Oostzaan, Landsmeer/Den Ilp and Monnickendam have disappeared as well after the Noord/Zuidlijn. However, because of the smaller total travel times these are only present in Appendix T.

Table 5.6: Top 20 OD-relations with largest absolute increase in travel time over a whole working day. The representativeness criterion is 51 travelers.

Check in cluster	**Check out cluster**	**Change travel time [min]**
Nieuwmarkt en Lastage	Landsmeer/Den Ilp	16.6
Landsmeer/Den Ilp	Nieuwmarkt en Lastage	11.7
Nieuwmarkt en Lastage	Purmerend West	11.6
Nieuwmarkt en Lastage	Purmerend Oost	10.4
Centraal Station	Landsmeer/Den Ilp	10.3
Leidseplein en Leidsestraat	Aalsmeer/Kudelstaart	10.1
Purmerend West	Dam	10.0
Purmerend West	Nieuwmarkt en Lastage	9.9
Purmerend Oost	Nieuwmarkt en Lastage	9.1
Aalsmeer/Kudelstaart	Jordaan	9.0
Nieuwmarkt en Lastage	Marken	8.6
Marken	Nieuwmarkt en Lastage	8.4
Centraal Station	Marken	8.4
Jordaan	Uithoorn	8.3
Buitenveldert	Aalsmeer/Kudelstaart	8.2
Noord midden	Landsmeer/Den Ilp	7.9
Jordaan	Aalsmeer/Kudelstaart	7.8
Centraal Station	Purmerend West	7.6
Amstelveen Centrum	Dam	7.2
Dam	Amstelveen	7.1

5.1.5. Spatial patterns of the travel time impacts

In this subsection spatial patterns of the travel time impacts are shown. First, the weighted average travel time per check in cluster is shown. Next, the largest increases and decreases in travel time from the region north and south of Amsterdam are visualized. Last, a selection of cases is presented.
In Figure 5.3 the weighted average travel time change from each check-in cluster to all destinations is shown. For the areas inside the red polygon this means the weighted average change in travel time to all stop clusters outside the polygon. For the regional areas (i.e. outside the red polygon) the weighted average change in travel time is calculated over journeys to all stop clusters in the center of Amsterdam.

The first thing that stands out is that the number of stop clusters that experiences a decrease is much larger than the number of stop clusters for which the travel time deteriorate. For another substantial part of the stop clusters the weighted average travel time does not change substantially.

Focusing on the stop clusters in Amsterdam, a decrease around the northern and southern Noord/Zuidlijn stations is visible. Where the decrease is larger for the stop clusters of the southern Noord/Zuidlijn stations. Besides, the eastern and western part of Amsterdam Noord have become better accessible in terms of travel times as well. An increase in travel times to areas near Amsterdam Centraal without a metro station is visible. This is the result of the disappearance of direct connections to Amsterdam Centraal.

Focusing on the regional stop clusters, for most stop clusters in Waterland and Zaanstreek the travel times decrease after the Noord/Zuidlijn. Only for Assendelft/Westzaan the travel times to Amsterdam increase slightly.

In Appendix V the absolute change in travel time between all stop clusters is shown for varying representativeness criteria. It becomes clear that using a representativeness criterion of 1 results in unlikely changes in travel time. Choosing 51 as the representativeness criterion leads to exclusion of many OD-relations.

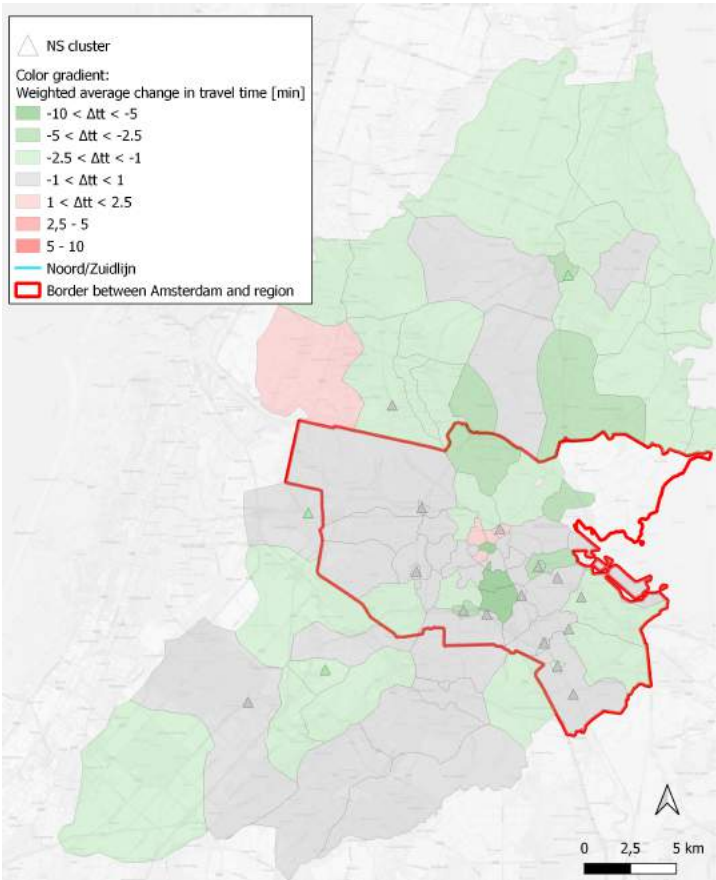


Figure 5.3: Weighted average change in travel time [min] per check-in cluster. The representativeness criterion is 12 and the time period is the whole day.

Figure 5.4 indicates that the largest positive travel time impacts for OD-relations between the region north of Amsterdam and Amsterdam occur mainly on OD-relations where a large part of the journey can be made using the Noord/Zuidlijn. Besides, the most positive travel time impacts occur to stop clusters in Purmerend. The number of OD-relations with large positive travel time impacts between Zaanstreek and Amsterdam is smaller. This could be explained by the fact that bus line 391 and 394 only offer transfer opportunities at Centraal Station.

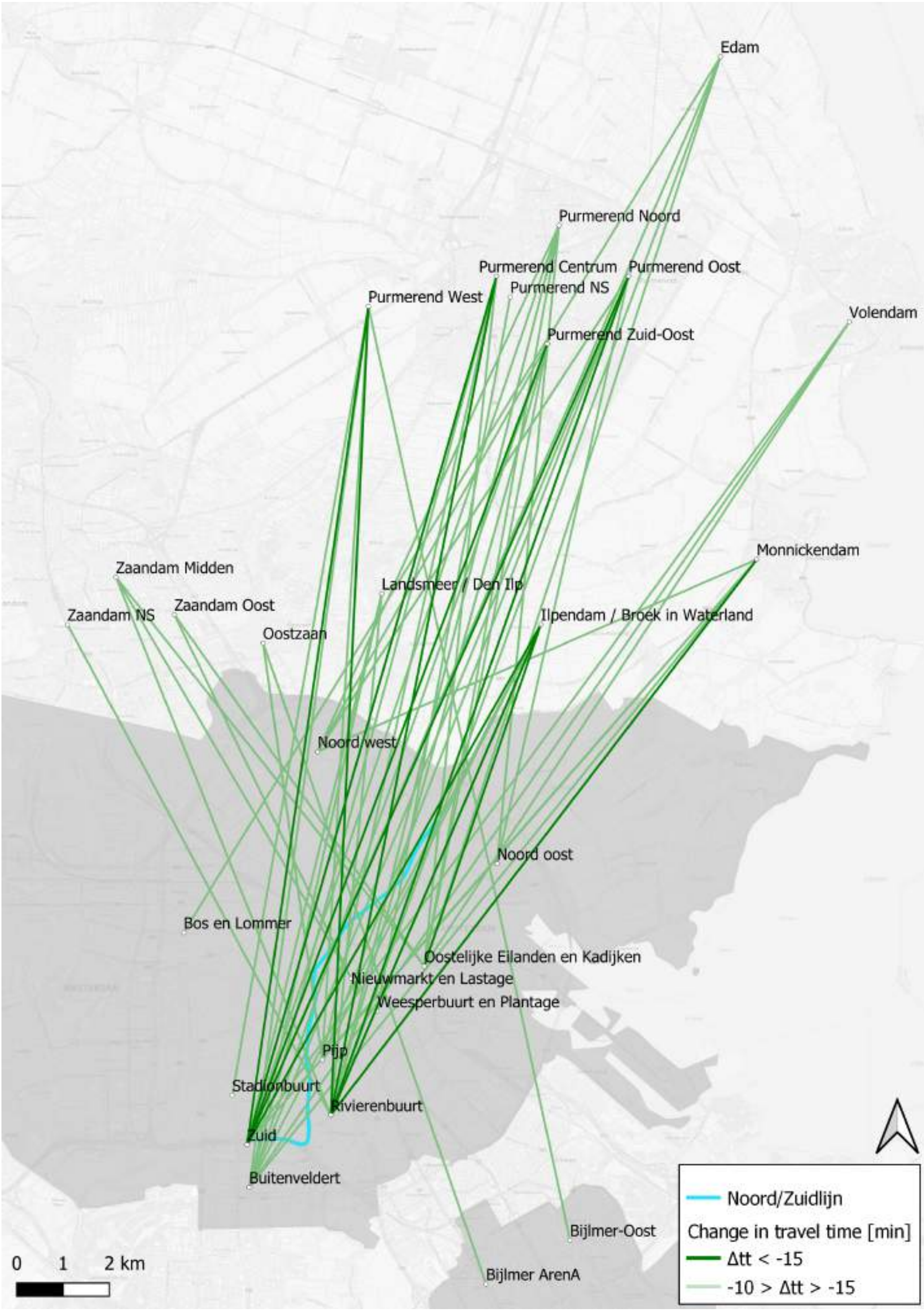


Figure 5.4: The OD-relations between the region north of Amsterdam and Amsterdam with a decrease in travel time of more than 10 minutes. The representativeness criterion is 51 and the time period is the whole day.

Figure 5.5 shows that from the region south of Amsterdam the largest travel time decreases occur on OD-relations from and to the northern part of Amsterdam. This is explained by the fact that a long part of the journey is made with the Noord/Zuidlijn. This is considerably faster and more direct than before the Noord/Zuidlijn. It is striking that on only few OD-relations from and to the southern Noord/Zuidlijn stations the decrease in travel time is larger than 10 minutes. This is explained by the fact that the travel time gain of the metro trip is nullified by an additional, possibly longer transfer. The decrease in travel time between Hoofddorp and Schiphol and Amsterdam is caused by the replacement of a bus or tram leg in the city center by a metro leg. On those OD-relations the Noord/Zuidlijn has not led to an extra necessary transfer.

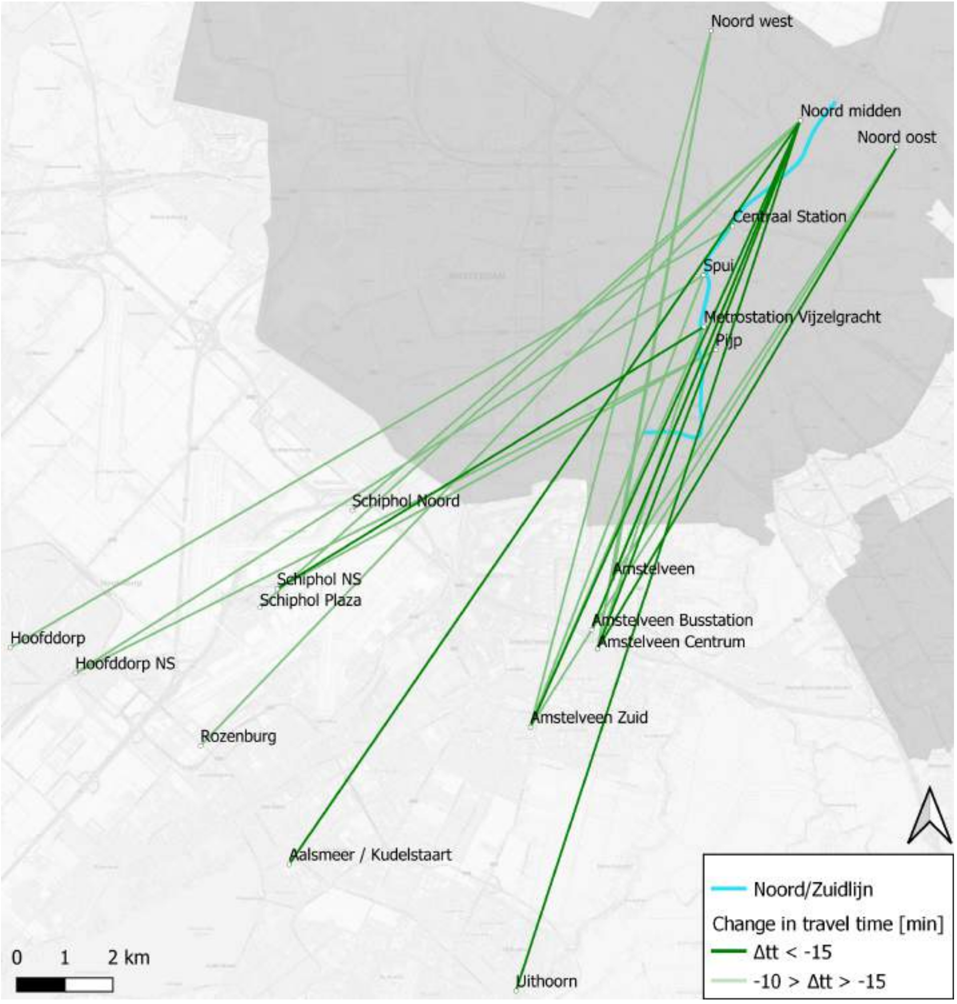


Figure 5.5: The OD-relations between the region south of Amsterdam and Amsterdam with a decrease in travel time of more than 10 minutes. The representativeness criterion is 51 and the time period is the whole day.

Figure 5.6 presents the OD-relations between Amsterdam and the region north of Amsterdam on which the travel time has increased at least 5 minutes. Most increases in travel time occur on OD-relations from and to Centraal Station, Dam and Nieuwmarkt en Lastage. This increase in travel time is caused by the disappearance of direct connections to Amsterdam Centraal. This is substantiated by the observation that OD-relations from Purmerend Centrum, Purmerend Noord and all clusters in Zaanstreek except Oostzaan do not experience an increase in travel time to those clusters. The increase in travel time to Amsterdam Noord can be explained by the fact that these direct connections ran through Amsterdam Noord as well.

It should be noted that for the increase in travel time the threshold value in Figure 5.6 and Figure 5.7 lies at an increase of 5 minutes. The threshold value for the decrease in Figure 5.4 and Figure 5.5 is -10 minutes.

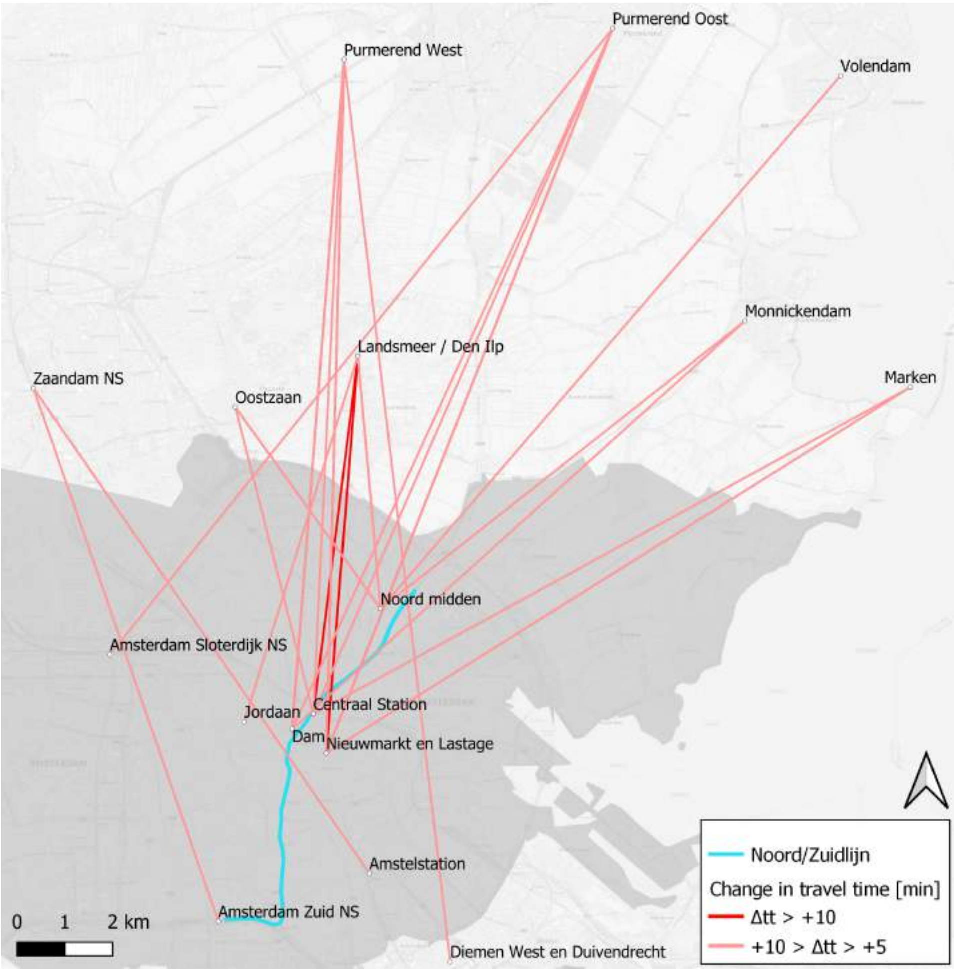


Figure 5.6: The OD-relations between the region north of Amsterdam and Amsterdam with an increase in travel time of more than 5 minutes. The representativeness criterion is 51 and the time period is the whole day.

In Figure 5.7 the OD-relations from the region south of Amsterdam to Amsterdam with an increase in travel time of at least 5 minutes are shown. The relations between Uithoorn, Aalsmeer and Amstelveen to the city center can be explained by the loss of direct connections and by the preference of travelers for the bus to Elandsgracht.

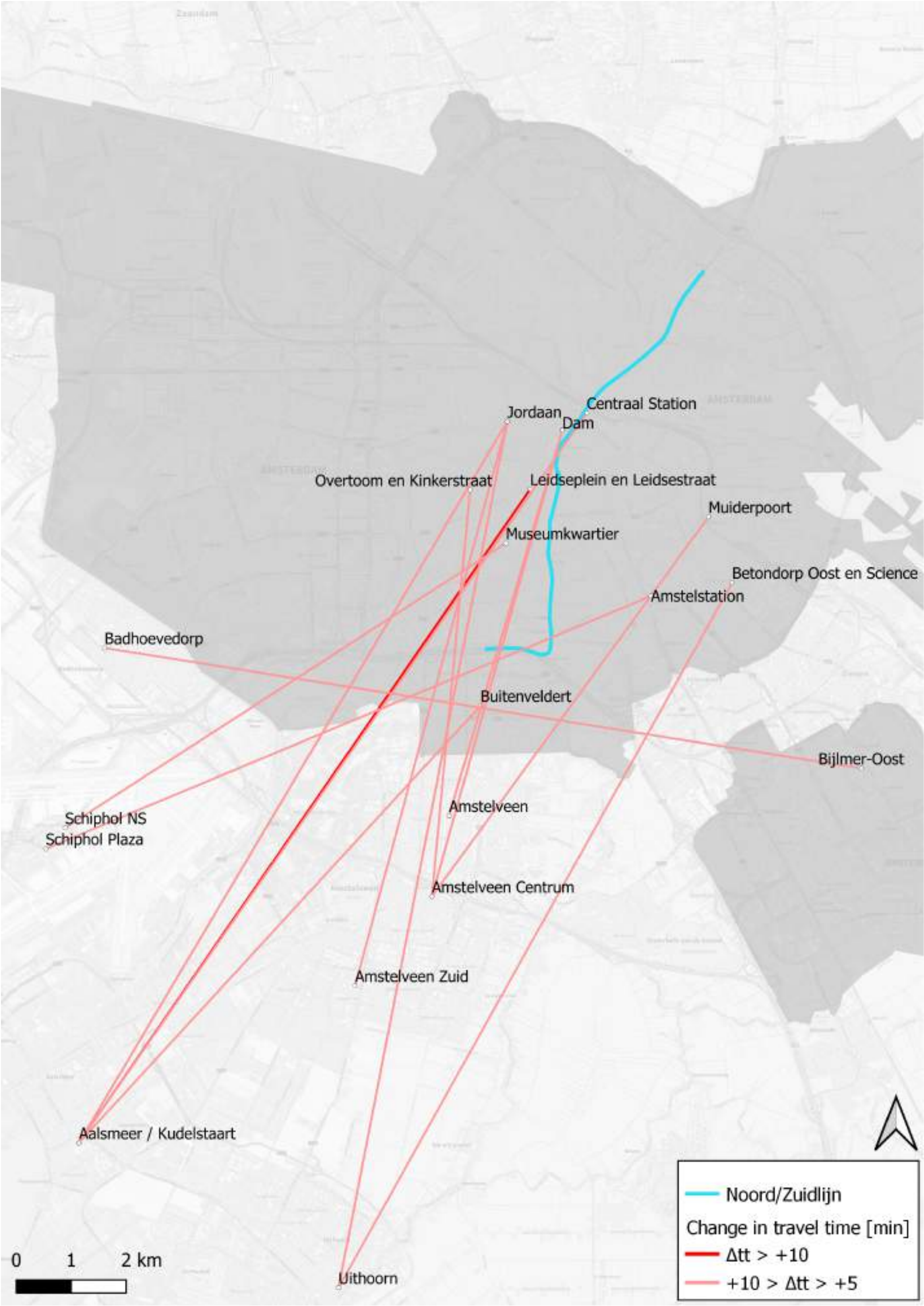


Figure 5.7: The OD-relations between the region south of Amsterdam and Amsterdam with an increase in travel time of more than 5 minutes. The representativeness criterion is 51 and the time period is the whole day.

In Figure 5.8 the change in travel time from regional stop clusters to a bus, tram or metro stop around Centraal Station is visualized. Journeys that end with a train trip to Amsterdam Centraal are not included in this figure because the data distinguishes between journeys that start or end with a train trip and journeys that start or end with a bus, tram or metro trip. It is clear that for the stop clusters north of Amsterdam where a direct connection to Amsterdam Centraal disappears, the travel time increases. From Amstelveen and Aalsmeer the travel time to Centraal Station decreases slightly. The length of the trip by the Noord/Zuidlijn is long enough to make up for the additional transfer, but the change in travel time is not very large because before the Noord/Zuidlijn the Amstelveenlijn offered a direct connection between Amstelveen and Amsterdam Centraal. From Hoofddorp and the train stations Schiphol and Hoofddorp the travel time clearly decreases. The Noord/Zuidlijn offers a much faster route to Centraal Station. It stands out that for stop clusters in Waterland that kept direct bus connection to Amsterdam Centraal no significant change in travel time occurs. An explanation for this is that the largest share, around 75% of the travelers from Purmerend Centrum and Noord, uses a direct bus to reach Amsterdam Centraal station after the Noord/Zuidlijn. Reason for this could be that the travelers prefer direct connections over routes requiring a transfer or that the bus and metro combination is not faster than a direct bus.

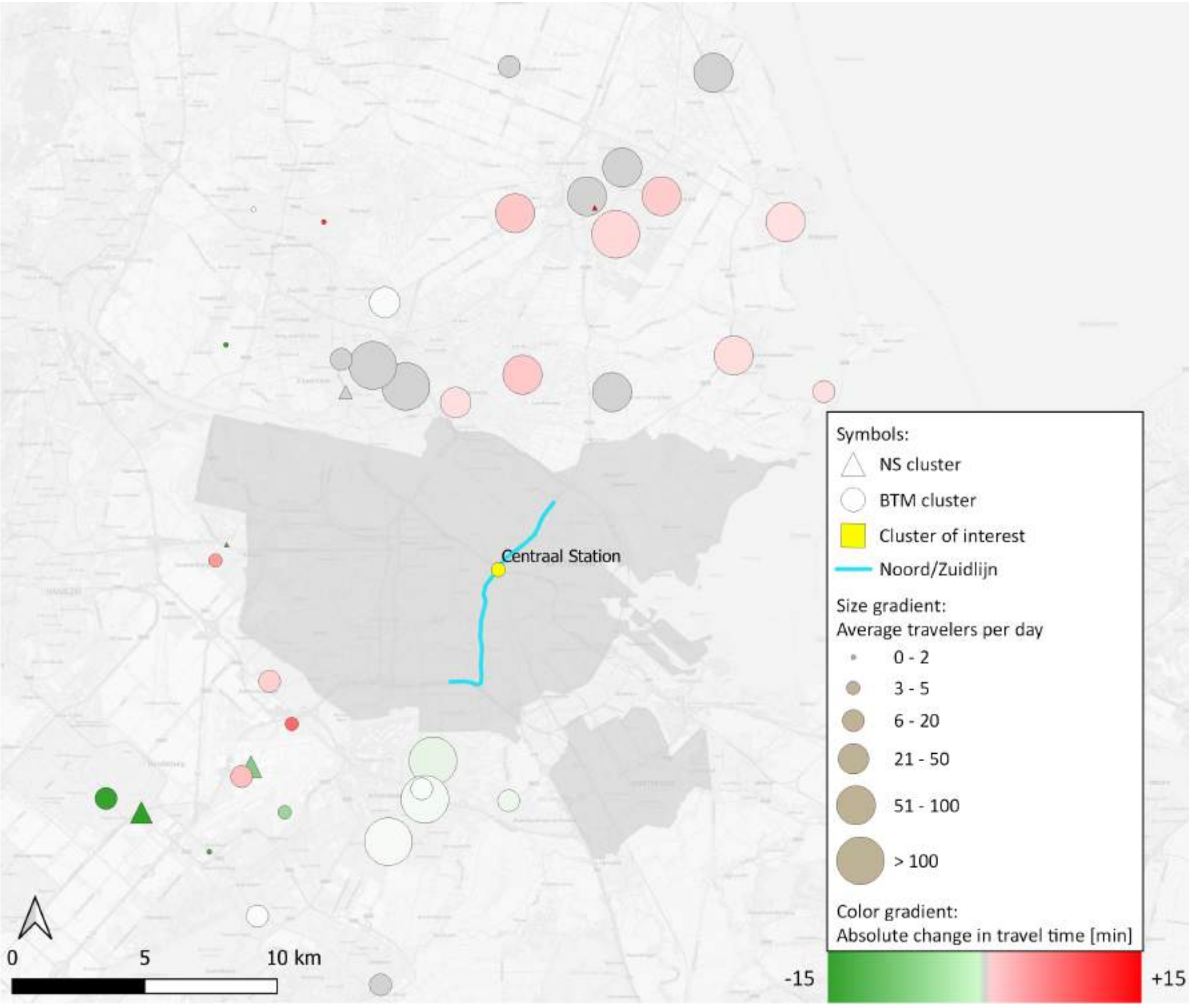


Figure 5.8: The absolute change in travel time from the region to Centraal Station

Figure 5.9 shows that from the region to stop cluster Spui, where Noord/Zuidlijn station Rokin is located, the travel times decrease. This indicates that for stop cluster Spui the length of the trip on the Noord/Zuidlijn is large enough to make up for the additional transfer time. It should be kept in mind that travelers from the region north of Amsterdam had to make a transfer to reach Spui before the Noord/Zuidlijn as well. The decrease in travel time can thus be attributed to the faster metro leg. From the region south of Amsterdam the impacts are positive as well. This holds for BTM and train stop clusters. This indicates that the Noord/Zuidlijn offers a faster connection into the city center than the buses and trams did.

Figure 5.9: The absolute change in travel time from the regional stop clusters to Spui

From Figure 5.10 it becomes clear that the direct connection to the Jordaan has disappeared. As a result, the travel times from Amstelveen, Uithoorn and Aalsmeer have increased. Besides, it is visible that the change in travel time increases in the direction of stop clusters further south. This is an indication that the travel time increases in Amstelveen. An explanation for this is that the buses through Amstelveen to Elandsgracht have become slower as a result of an uneven distribution of passengers over the buses driving to Elandsgracht and Amsterdam Zuid. From Zaanstreek some slightly positive effects are visible. No difference in routes taken have been found. Therefore, these slightly positive effects might be explained by the fact that the increase in frequency of bus line 391 and 394 results in shorter transfer times. From Waterland a slight decrease is visible for some clusters as well. A possible explanation for this is a slight decrease in travel times to Amsterdam Centraal for the direct buses as a result of the cancellation of many other buses.

Figure 5.10: The absolute change in travel time from the regional stop clusters to Jordaan

In Figure 5.11 the change in travel time from the region to Nieuwmarkt en Lastage (left) and vice versa (right) is shown. It is clear that from the region to Nieuwmarkt en Lastage the connection between Waterland and Nieuwmarkt en Lastage deteriorates. The largest decline is visible for stop clusters where the direct connection to Centraal Station was canceled. When comparing the two directions it stands out that a couple of stop clusters in Zaanstreek has no or a small change in travel time in the direction from region to Amsterdam. However, in the other direction the travel time decreases strongly. This is explained by a new, direct route that occurs after the Noord/Zuidlijn from Nieuwmarkt en Lastage to Zaanstreek. This is caused by the new IJtunnel and Prins Hendrikkade bus stops. Bus lines from Zaandam to Amsterdam do not stop in stop cluster Nieuwmarkt en Lastage. They stop at the bus stop IJtunnel which is situated in stop cluster Oostelijke Eilanden en Kadijken. In the other direction they do stop in stop cluster Nieuwmarkt en Lastage namely at bus stop Prins Hendrikkade. As a result the decrease in travel time between Nieuwmarkt en Lastage en Zaandam is only visible in the direction from Nieuwmarkt en Lastage to Zaandam.

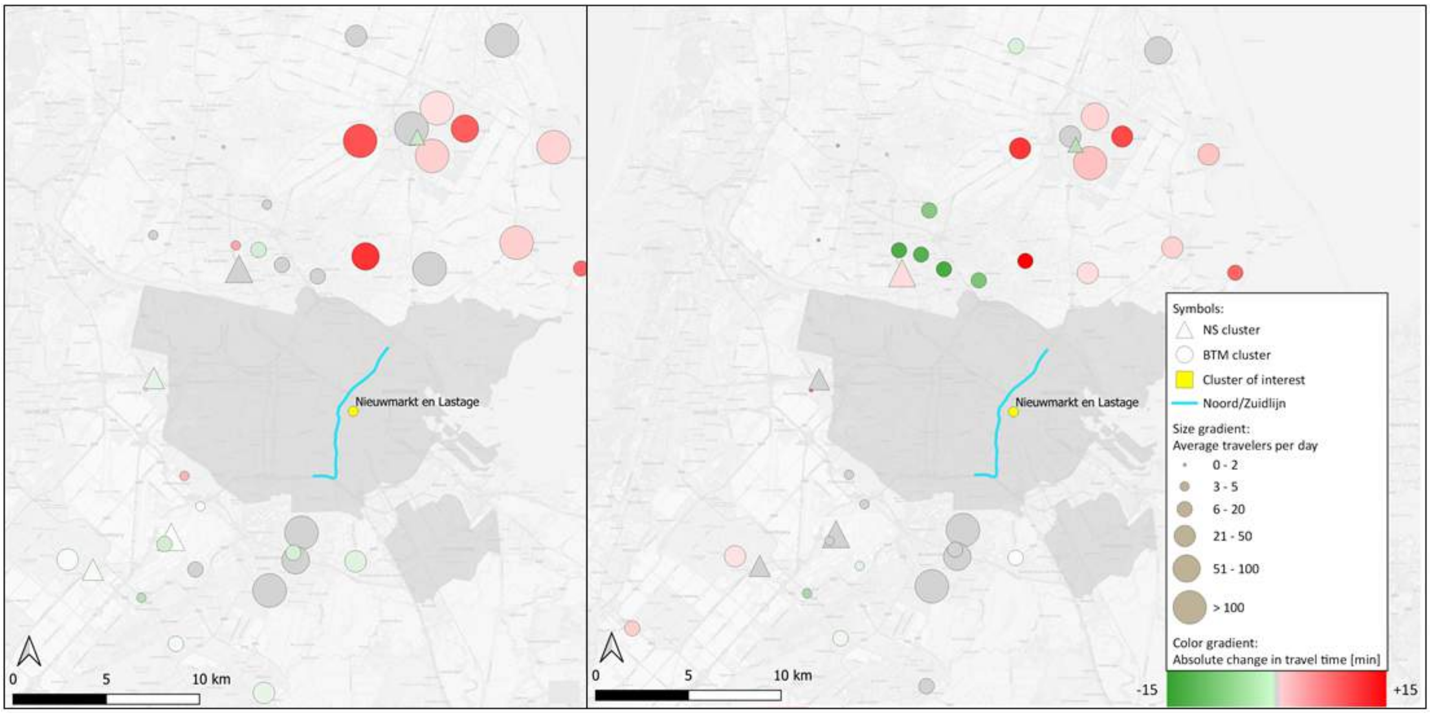


Figure 5.11: The absolute change in travel time from the regional stop cltuers to Nieuwmarkt en Lastage (left) and from Nieuwmarkt en Lastage to the regional stop clusters (right)

In Figure 5.12 the change in travel time from Purmerend Oost to all stop clusters in Amsterdam is visualized. The travel time to Noord/Zuidlijn stations south of Amsterdam Centraal decreases. However, the travel time to Amsterdam Centraal increases. This indicates that the trip on the Noord/Zuidlijn only makes up for the extra transfer to the Noord/Zuidlijn if that trip passes Amsterdam Centraal. The increase in travel time to stop clusters in southeast Amsterdam are a result of the propagation of the additional travel time to Amsterdam Centraal. Besides, an additional intramodal metro transfer is necessary to reach destinations along the southeastern metro corridor. Furthermore, it can be seen that travel time to stop clusters in Amsterdam Noord decreases. The explanation for this is that the bus network in Noord has been redesigned to connect several areas in Noord. As a result, the frequencies of and transfer opportunities to bus connections to the eastern and western part of Amsterdam Noord have improved.

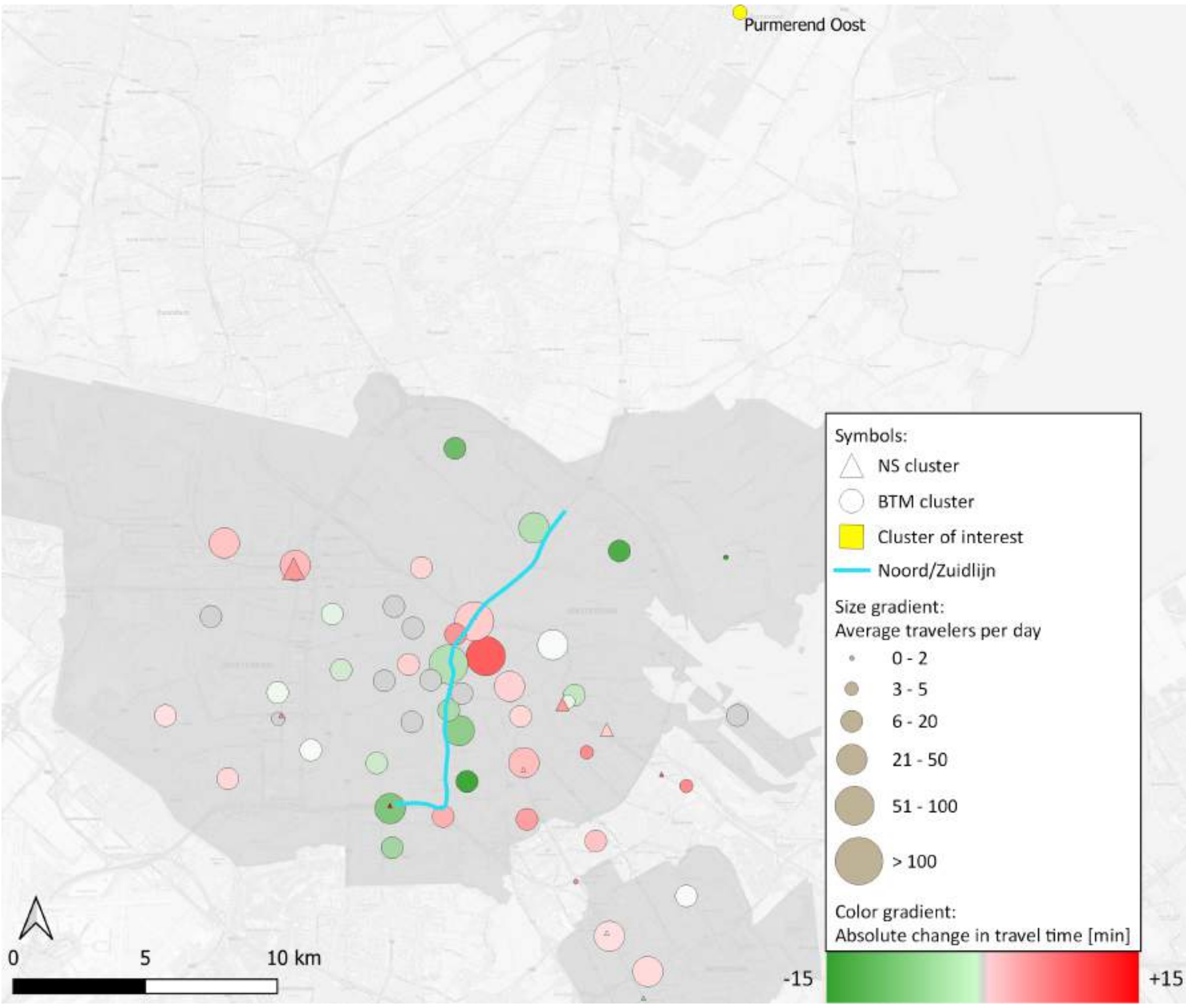


Figure 5.12: The absolute change in travel time from Purmerend Oost to stop clusters in Amsterdam

Figure 5.13 visualizes the change in travel times from Uithoorn to Amsterdam. It stands out that the travel time to stop clusters west of the Noord/Zuidlijn has increased. This can be explained by the fact that buses from Amstelland-Meerlanden to Elandsgracht are busier than the buses from Amstelland-Meerlanden to Amsterdam Zuid. The buses to Elandsgracht namely offer a more attractive transfer opportunity to the metro network at Amstelveenseweg. Next, it can be seen that a longer trip on the Noord/Zuidlijn results in a larger decrease in travel time. Therefore, the travel time to Amsterdam Noord decreases most. For short trips on the Noord/Zuidlijn, for example to metro station De Pijp, the decrease in travel time is only small. The fact that the travel time to Amsterdam Centraal does not change significantly can be explained by the fact that in the situation after the Noord/Zuidlijn a direct metro connection exists to Amsterdam Centraal, namely metro line 51. From Uithoorn people transfer to the Amstelveenlijn in Amstelveen and directly travel to Centraal Station.

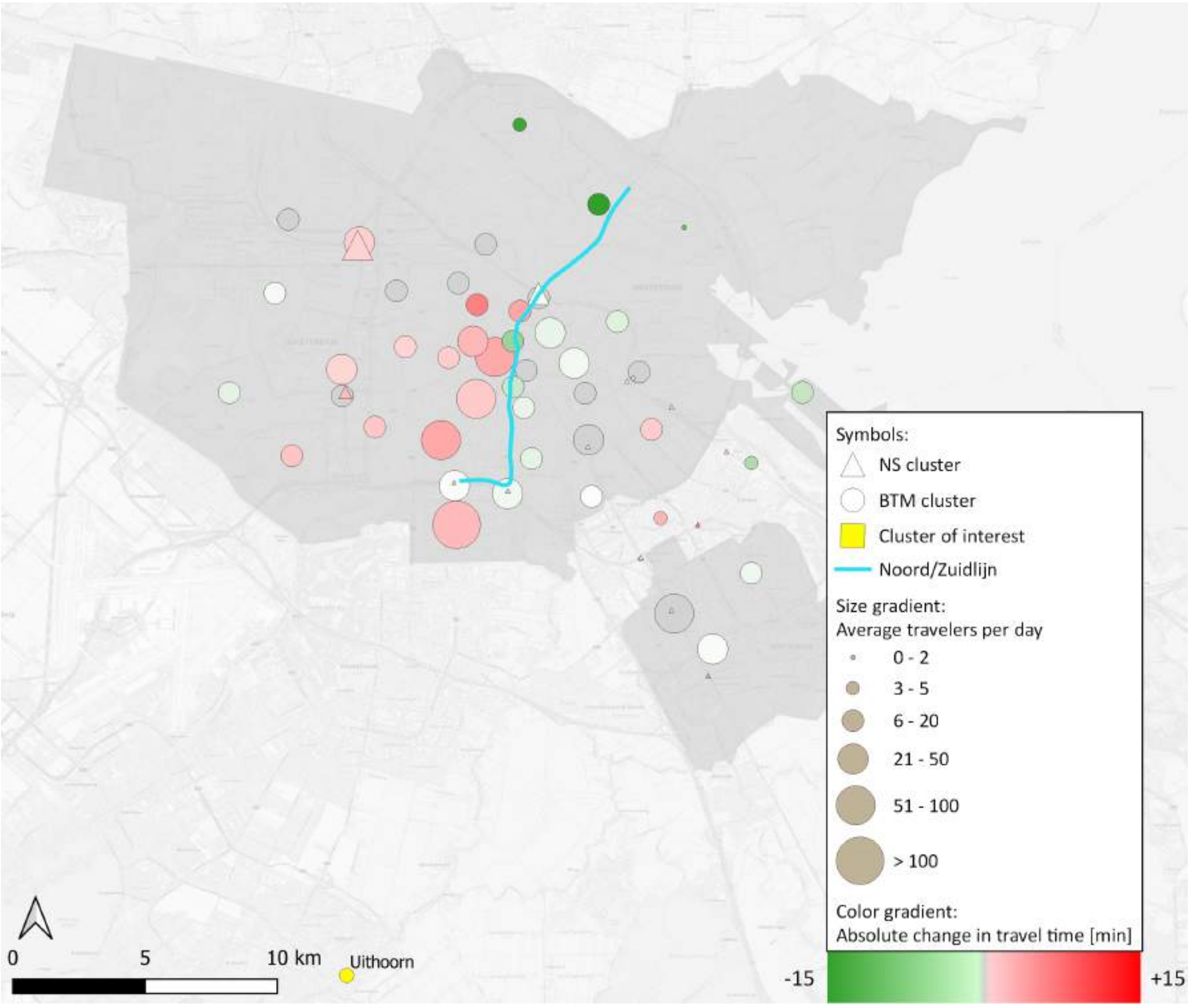


Figure 5.13: The absolute change in travel time from Uithoorn to stop clusters in Amsterdam

From Zaandam Midden to Amsterdam no effect on travel times to Amsterdam Centraal Station is visible in Figure 5.14. This is explained by the fact that bus lines 391 and 394 from Zaandam are not connected to the Noord/Zuidlijn. For Noord/Zuidlijn stations south of Centraal Station a slight decrease in travel time is visible. The new metro leg is thus faster than the bus and tram options before the Noord/Zuidlijn. Besides, slight decreases in travel times to Amsterdam West are visible. This is probably the result of increased frequencies and changed connections between Amsterdam Sloterdijk and the Leidseplein.

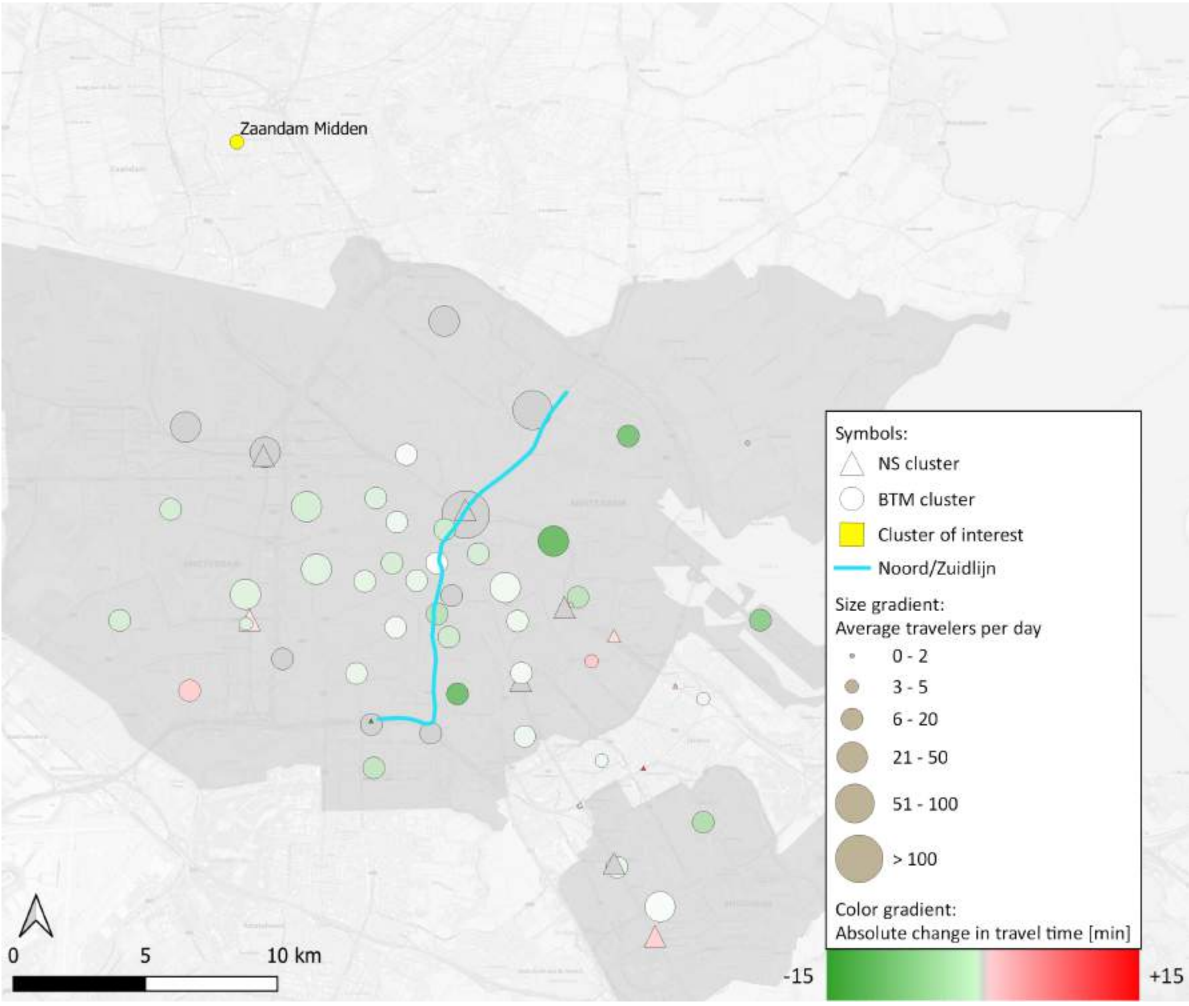


Figure 5.14: The absolute change in travel time from Zaandam Midden to stop clusters in Amsterdam

5.1.6. Effect of time period on the travel time impacts

This section investigates whether the travel time impacts differ among time periods. Figure 5.15 shows that the travel time impacts are slightly less positive for the morning peak and evening peak than for the rest of the day. This can be concluded because the cumulative distribution function (CDF) for the rest of the day is above the CDF for the evening and morning peak. For this, only the OD-relations that have at least 51 travelers in each time period have been selected.

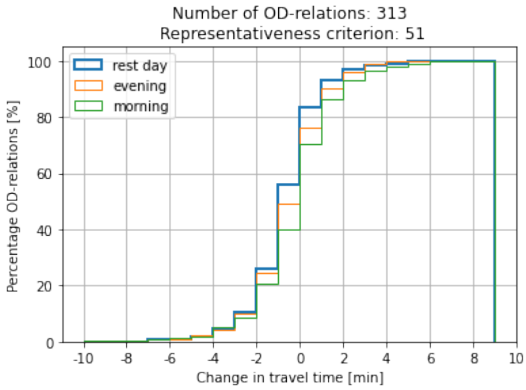


Figure 5.15: Cumulative distribution function of the change in travel time for each time period. Only OD-relations that are representative for all time periods are plotted.

5.1.7. Effect of direction on the travel time impacts

This section describes whether large differences in the travel time impacts are observed for OD-relations in opposite direction. In Figure it can be seen that for around 30% of the OD-relations the difference in travel time change between both directions is less than 1 minute. For an additional 60% the difference in travel time change is between 1 and 5 minutes. It can thus be concluded that there are differences in the change of travel time depending on the travel direction. However, for a large part of the OD-relations these are not large. Figure 5.17 shows that the change in travel time is often larger (i.e. less positive impact) for the direction from Amsterdam to the region than for the opposite direction. This can be explained by different waiting times because the frequencies of the services before and after a transfer differ.

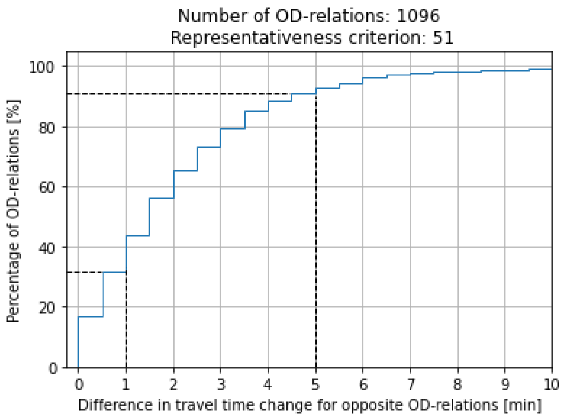


Figure 5.16: Difference in travel time change over the whole day for OD-relations in opposite directions.

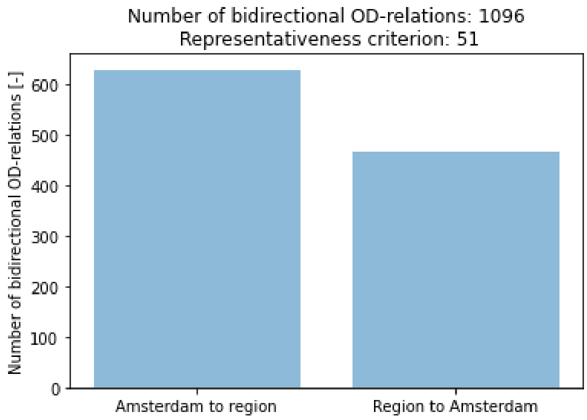


Figure 5.17: Count of which direction on an OD-relation has a less positive travel time impact over the whole day.

In Appendix S the 20 OD-relations with the largest difference in travel time changes for opposite directions are presented. The new bus stop IJtunnel explains 8 out of 20 OD-relations. Only buses driving from the region to Amsterdam Centraal stop there. Explanations for a difference in travel time change for opposite directions could be a better or worse transfer connection or more peaked demand and thus busier vehicles in one of the directions.

5.2. Impact of the Noord/Zuidlijn on regional transfer patterns

This section describes the impact of the Noord/Zuidlijn on transfer patterns. It starts with the aggregated impacts. Thereafter, the distribution of the impacts is discussed. Last, spatial patterns of the transfer impacts are presented.

5.2.1. Aggregated transfer impacts

In Table 5.7 it can be seen that the total number of transfers on a working day increased by 48.3% after the Noord/Zuidlijn. This increase should be put in perspective of two factors. First, the number of journeys increases by 22.3%. A large share of this increase is explained by the fact that summer months are compared to winter months. A smaller part could have resulted from newly generated demand. Second, the number of transfers per journey increases by 22.2%. This is likely an effect of the Noord/Zuidlijn. So, part of the increase in the number of transfers is caused by an increase in the number of journeys and another part of that increase is caused by the Noord/Zuidlijn.

Table 5.7: The total number of transfers per day, the total number of journeys per day and the average number of transfers per journey in each data set and the relative difference between both data sets.

	Total number of transfers per day (lower bound | estimated) [transfer/day]	**Total number of journeys per day (lower bound | estimated) [journey/day]**	**Number of transfers per journey [transfer/journey]**
2018Q2	44,700 (32,700 | 12,000)	82,400 (71,100 | 11,300)	0.54
2019Q1	66,300 (54,400 | 11,900)	100,800 (90,300 | 10,500)	0.66
Relative change	+48.3%	+22.3%	+22.2%

Figure 5.19 shows that before the Noord/Zuidlijn 55.3% of the regional journeys is direct, 35.5% of the journeys consists of one transfer and on 9.2% of the journeys two or more transfers are made. After the Noord/Zuidlijn the share of direct trips decreases by 8%-point and is 47%. The share of journeys with at least one transfer increases. These findings match the expectations of transforming a direct network to a trunk-feeder network.

What stands out as well is that the share of estimated journeys becomes larger if the number of transfers increases. This indicates that routes with more transfers are traveled less compared to direct routes. For less traveled routes the share of estimated travelers is larger.

Figure 5.18: The share of journeys with zero, one or two or more transfers before the Noord/Zuidlijn.

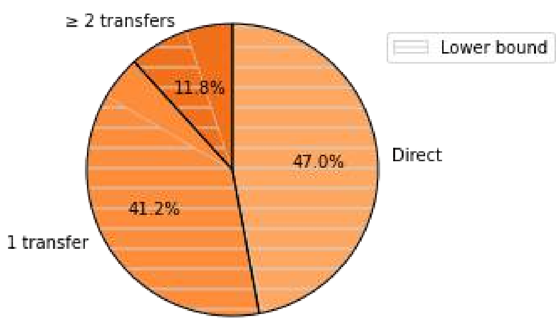


Figure 5.19: The share of journeys with zero, one or two or more transfers after the Noord/Zuidlijn.

In Table 5.8 the transfer benefits per representativeness criterion are shown. Overall, after the Noord/Zuidlijn an increase of 2,500 transfers per day is observed. Taking the average over all 91,400 travelers this is an average increase of 0.027 transfers per journey. Multiplying these extra transfers by 1.4 €/transfer (Knockaert and Koster, 2020) results in societal costs of 3,500 euros per day as a result of the transfer impacts of the Noord/Zuidlijn. The lower bound number of travelers contributes for 3,200 euros to that. The remaining 300 euros are based on the additionally estimated number of passengers.

Furthermore, Table 5.8 shows that the largest part of the transfer impacts can be attributed to the lower bound of travelers. Besides, it can be observed that the passengers traveling on less representative OD-relations experience a decrease in the number of transfers while on the OD-relations with at least one route with than 51 travelers an increase in transfers occurs.

Table 5.8: Transfer benefits over all OD-relations over the whole day distinguished in different representativeness categories

Representativeness criterion	**Total extra transfers per day (lower bound | estimated) [transfers/day]**	**Average number of travelers per day [travelers/day]**	**Average transfer benefits per traveler [transfers/traveler]**
1-50	-433 (-391 | -42)	5,200	-0.083
≥51	2,893 (2,644 | 249)	86,200	+0.034
Total	**2,460** **(2,253 | 208)**	**91,400**	**+0.027**

However, the aggregated transfer impacts tell nothing about the distribution thereof. The distribution of the transfer impacts can be found in Section 5.2.2.

Figure 5.20 presents the modal split over all regional journeys before and after the Noord/Zuidlijn. Besides, the relative change for each combination is shown.

It stands out that the share of journeys made (partially) by metro grows. Especially, the share of journeys by metro and bus increases. Besides, an increase in unimodal metro journeys is visible. The Noord/Zuidlijn offers new unimodal routes by metro between the city center and Amstelveen.

A decrease in unimodal journeys, except for metro, is visible. This can be attributed to the decrease of direct connections to Amsterdam Centraal. It should be noted that the columns with one modality in Figure 5.20 contain unimodal journeys with a transfer as well. An explanation for the decrease of unimodal train journeys could be that travelers with a destination close to Amsterdam Centraal shifted from going to Amsterdam Centraal by train and walking the last part to taking a train to Amsterdam Zuid and transferring to the metro. This is supported by the increase of journeys by metro and train.

Looking at the bimodal journeys, the tram and metro combination loses share. This combination is only possible from or to Amstelland-Meerlanden. For a large part of that area a journey by bus and metro or metro only is preferred. The bus and tram combination loses share as well. This is probably the result of a decrease in transfer possibilities between bus and tram at Amsterdam Centraal. The share of journeys made using three or more modalities remains relatively small, but increases slightly. Again, the combinations including metro show an increase.

When comparing the modal split between before the Noord/Zuidlijn and after the Noord/Zuidlijn in Figure 5.20 it stands out that overall the share of unimodal journeys decreases, but remains most common. This could indicate a preference for direct journeys for travelers who can choose between a direct, longer journey and a shorter journey with a transfer. However, another explaining factor could be the fact that for OD-relations from and to Waterland a bus to bus transfer is an attractive option to reach Amsterdam Centraal. A transfer from bus to bus is assigned to bus only in Figure 5.20. Another observation in Figure 5.20 is that the share of journeys with a transfer increases. This is mostly caused by a strong increase of the metro and bus combination.

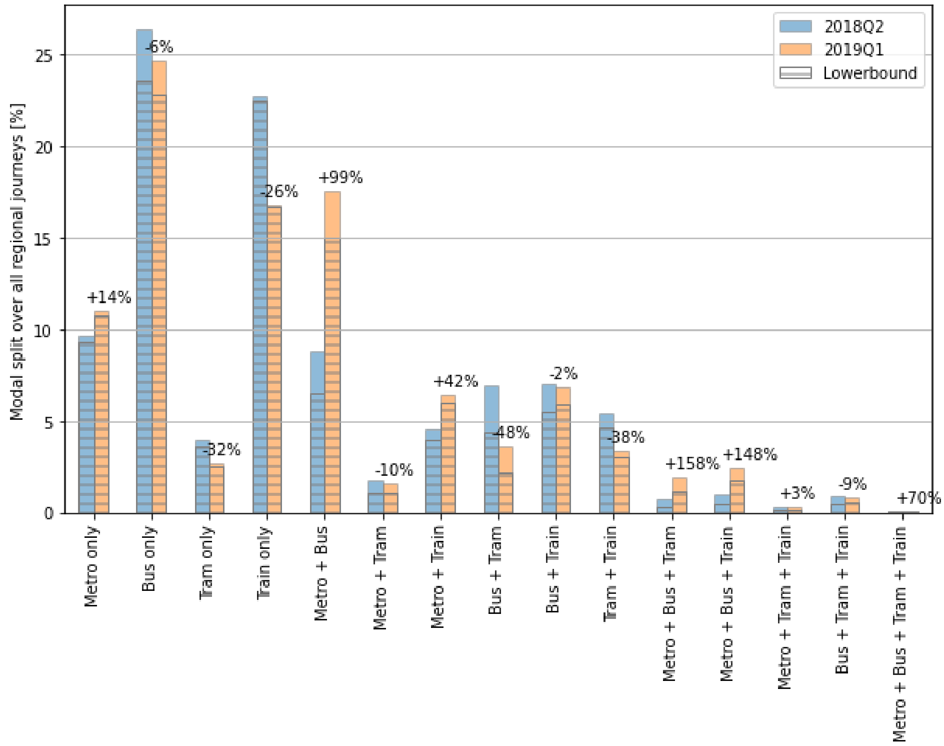


Figure 5.20: The share of journeys with a certain modal split over all regional journeys before (blue) and after (orange) the Noord/Zuidlijn and the relative change thereof.

5.2.2. The distribution of the transfer impacts

Table 5.9 shows that the total transfer losses are 7,000 transfers. This means that over all OD-relations where the average number of transfers increased after the Noord/Zuidlijn, on an average working day 7,000 extra transfers are made. The transfer savings are 4,500 transfers. The decrease in number of transfers can be explained by the fact that people traveling from Amstelveen to Noord/Zuidlijn stations have to make a transfer between metro line 51 and the Noord/Zuidlijn instead of between metro and bus or tram. However, a transfer from metro to metro is not visible in the smart card data. The increase in number of transfers is the result of a decrease in direct regional connections to the city center.

Furthermore, for the representativeness criterion of 1 to 50 the transfer gains are larger than the losses while for the more conservative representativeness criterion the transfer losses are larger than the gains.

Table 5.9: Transfer benefits over all OD-relations over the whole day split into transfer gains and losses for different representativeness criteria

Representativeness criterion	**Total extra transfers [number of transfers]**	**Transfer gains [number of transfers]**	**Transfer losses [number of transfers]**
1-50	-455	1191	736
≥51	2,966	3,342	6,308
Total	**2,511**	**4,533**	**7,044**

Figure 5.21 and Figure 5.22 show the distribution of the change in number of transfers over all OD-relations for the representativeness criterion 1 and 51 respectively. In both figures it can be seen that for a large part of the OD-relations the change in average number of transfers is minimal, namely around 60% and around 80%. Besides, more OD-relations experience an increase in the number of transfers than a decrease. However, a difference in magnitude between the results for both representativeness criteria can be observed. This could indicate that the OD-relations that are less frequently traveled, experience larger impacts.

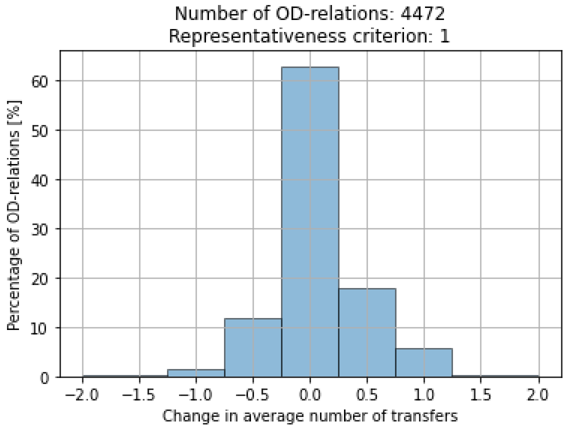


Figure 5.21: The distribution of the transfer impacts over OD-relations.

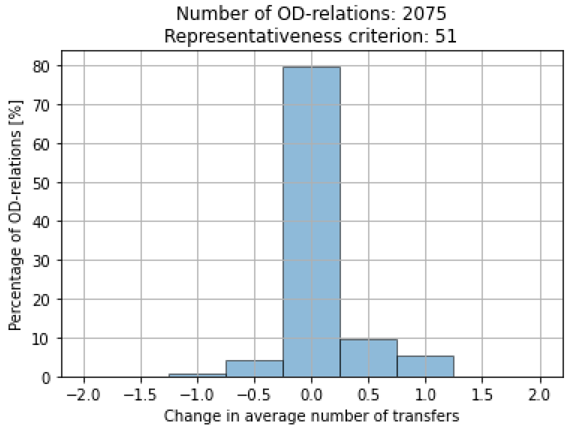


Figure 5.22: The distribution of the transfer impacts over OD-relations.

In Figure 5.23 and Figure 5.24 the distribution of the change in average number of transfers over all regional travelers is shown. Most of the travelers experience a minimal change for both representativeness criteria. The number of travelers that experiences a negative impact is slightly larger than the number of travelers that experiences a positive impact. Comparing the distribution of the transfer impacts over OD-relations and travelers shows that the percentage of OD-relations for which the number of transfers increases is larger than the percentage of travelers for which the number of transfers increases. The same holds for a decrease in the number of transfers. This suggests that the impacts are larger on OD-relations that are less frequently traveled.

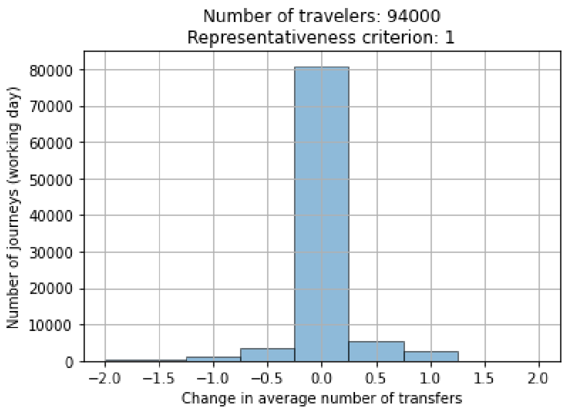


Figure 5.23: The distribution of transfer impacts over travelers.

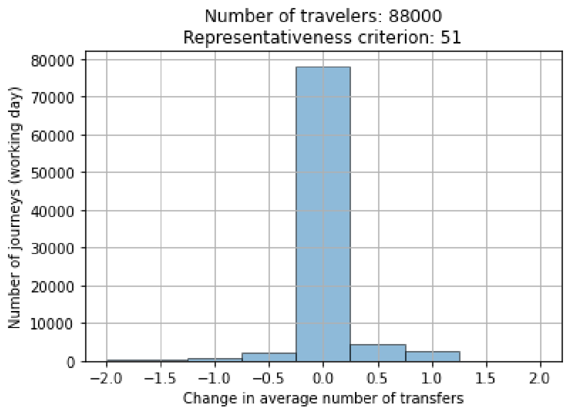


Figure 5.24: The distribution of transfer impacts over travelers.

5.2.3. Spatial patterns of the transfer impacts

In Figure 5.25 the OD-relations between the region north of Amsterdam and Amsterdam where the number of transfers changes by at least 0.75 are visualized. It is clear that the negative transfer impacts are much larger than the positive transfer impacts. Especially, Oostzaan, Purmerend West, Landsmeer/Den Ilp and Purmerend Oost experience large negative impacts. All these areas have lost their direct connection to Amsterdam Centraal after the Noord/Zuidlijn.

Additionally, in Figure 5.25 it stands out that the city center of Amsterdam has become less accessible from the region north of Amsterdam. This could be explained by the fact that the new routes to Amsterdam Centraal require an extra transfer. The negative transfer impacts propagate beyond Amsterdam Centraal as well. This effect is not visible for OD-relations from or to areas where a new Noord/Zuidlijn station is built. On these OD-relations both before and after the Noord/Zuidlijn one transfer is necessary. Similarly, Purmerend Noord and Purmerend Centrum are not visible in Figure 5.25 because these areas have kept a direct bus connection to Amsterdam Centraal.

Furthermore, it is interesting to see that the clusters in Amsterdam south of Museumkwartier do not experience a negative impact from the Noord/Zuidlijn. This could indicate that on these OD-relations new routes have emerged that do not require an additional transfer.

Regarding the positive impacts, the relations between the region and Oostelijke Eilanden en Kadijken are striking. The explanation thereof is that bus stop IJtunnel was added for buses from Zaanstreek and Waterland that end at Amsterdam Centraal.

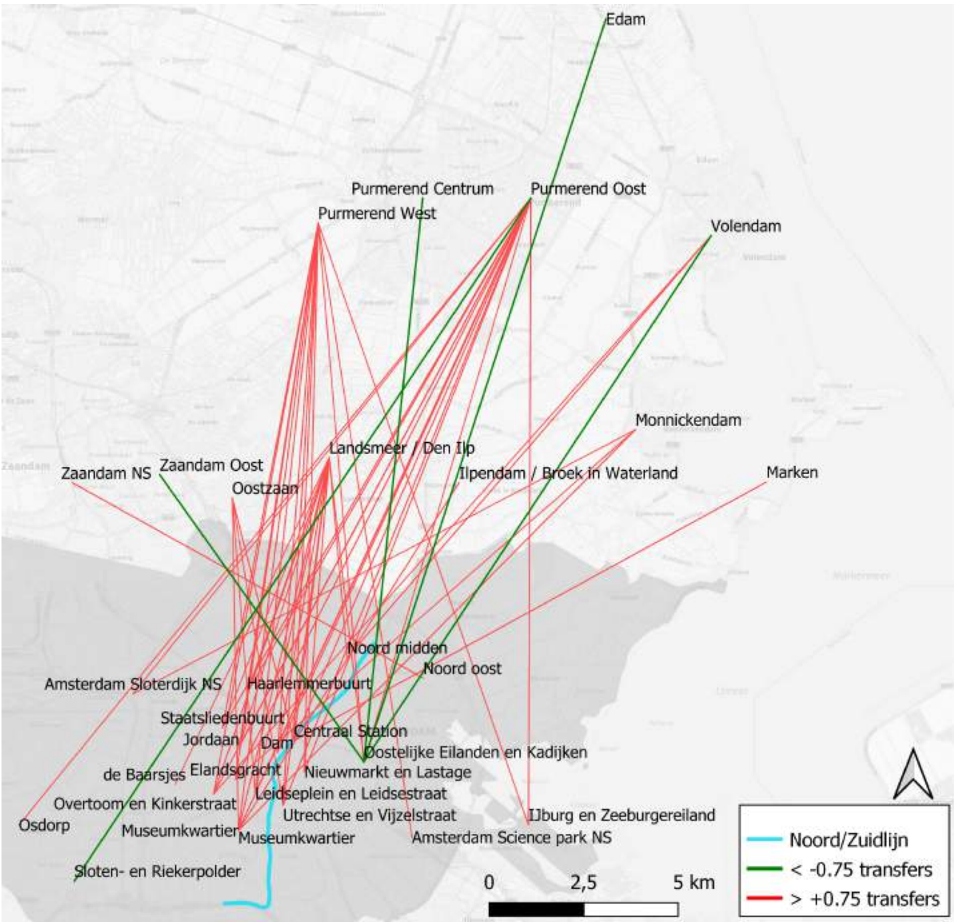


Figure 5.25: The OD-relations between the region north of Amsterdam and Amsterdam where the average number of transfers changes with more than 0.75 transfer per journey. The representativeness criterion is 51.

In Figure 5.26 it stands out that from Amstelveen, Aalsmeer/Kudelstaart and Uithoorn the connections to the Dam and Jordaan have deteriorated in terms of number of transfers. This is the result of regional buses stopping at Elandsgracht and the relocation of the final stop of tram line 5 from Centraal Station to the Westergasfabriek. Furthermore, on two connections from train station Halfweg-Zwanenburg and Schiphol to Amsterdam Noord oost the number of transfers increases. This is caused by the disappearance of a direct connection between Centraal Station and Noord oost.

The positive effects from the region south of Amsterdam are caused by the seemingly direct connection to Noord/Zuidlijn stations from Amstelveen. In reality, a transfer between metro line 51 and the Noord/Zuidlijn is necessary. For that reason it is remarkable that other clusters that contain a Noord/Zuidlijn station, like Noord midden and Metrostation Noord, are not visible in Figure 5.26. Two potential explanations for that are that the representativeness criterion is not met or that the stop cluster contains more stops than just the Noord/Zuidlijn station. For those stops a transfer might be necessary. Another interesting finding in Figure 5.26 is that cluster Staatsliedenbuurt is better accessible after the Noord/Zuidlijn. This could be explained by the rerouting of tram line 5.

Comparing the transfer impacts north and south of Amsterdam it is apparent that the largest part of the negative impacts occurs on OD-relations between Waterland and Amsterdam. Besides, the number of OD-relations that deteriorate is much larger for the region north of Amsterdam than for the region south of Amsterdam.

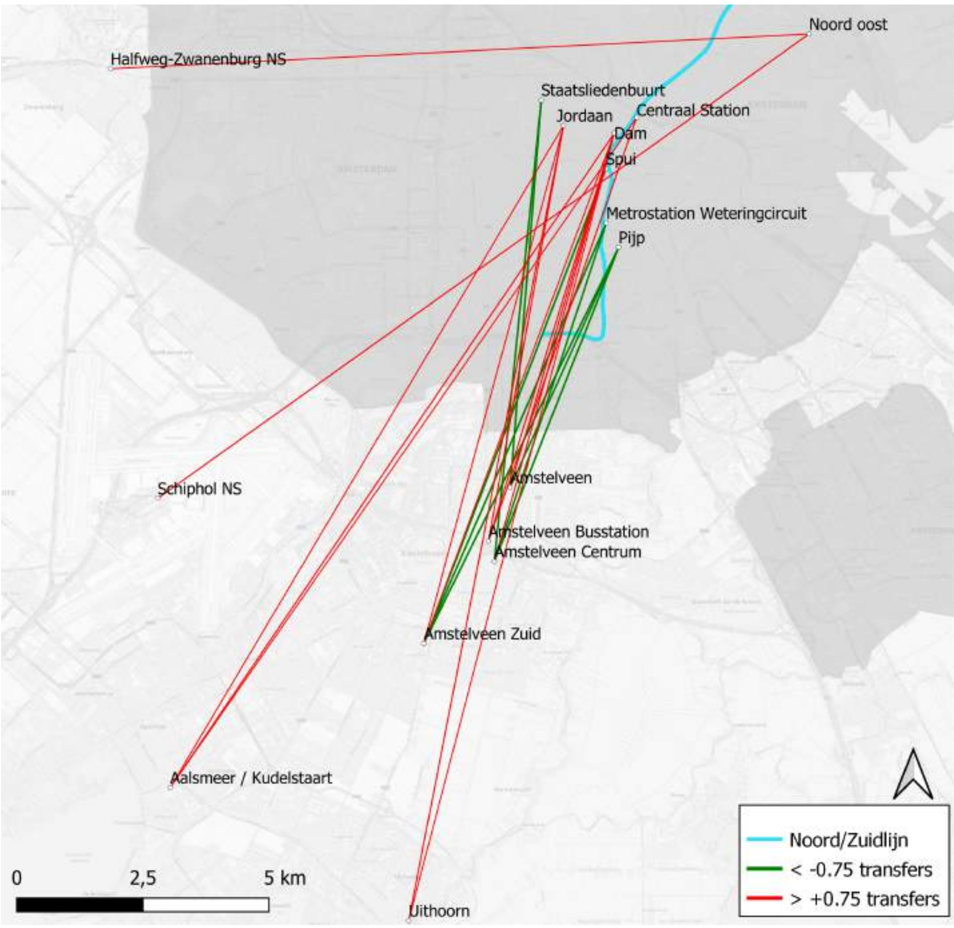


Figure 5.26: The OD-relations between the region south of Amsterdam and Amsterdam where the average number of transfers changes with more than 0.75 transfer per journey. The representativeness criterion is 51.

In Figure 5.27 till Figure 5.29 three cases that illustrate the impact of the Noord/Zuidlijn on transfer patterns are visualized. Only the main routes are shown. It stands out that the number of travelers per day after the Noord/Zuidlijn is smaller than before the Noord/Zuidlijn. This could have several causes. First, travelers could have changed their activity location. Second, travelers could change their alighting stop. Third, travelers could change to private modes.

Figure 5.27 shows that before the Noord/Zuidlijn a direct bus line connected Oostzaan and Centraal Station. After the Noord/Zuidlijn the most traveled journey consists of a bus to Metrostation Noord, a transfer to the Noord/Zuidlijn and the last part of the journey to Centraal Station by metro. This matches the expected changes as reported in the Stadsregio Amsterdam (2014).

In Figure 5.28 it can be noticed that the direct connection between Purmerend Oost and Nieuwmarkt en Lastage has disappeared. As a result two new routes appear. The first route consists of a trip by bus to metro station Noord. There the travelers transfer to the Noord/Zuidlijn. At Centraal Station, another transfer is necessary from the Noord/Zuidlijn to metro line 51, 53 or 54. However, transfer from metro to metro is not visible in the data. The second route starts with a bus trip from Purmerend Oost to stop cluster Ilpendam/Broek in Waterland. In that stop cluster, a transfer is made to a direct bus to Centraal Station. The strong decrease in travelers in Figure 5.28 is likely causes by the new bus stop at IJtunnel. This leads to a change in alighting stop for a part of the travelers to Nieuwmarkt en Lastage.

The last case is shown in Figure 5.29. It visualizes the journeys between stop cluster Amstelveen and stop cluster Dam. Before the Noord/Zuidlijn a direct route by bus and a direct route by tram exists. After the Noord/Zuidlijn, these direct routes are replaced by routes with a transfer. Two routes start with a trip by tram or bus to stop cluster Museumkwartier. There, a transfer to tram is made. Another route that appears consists of a metro trip to Centraal Station. This is probably with an additional, 'invisible' transfer to the Noord/Zuidlijn at Amsterdam Zuid, but a direct trip is possible as well using metro line 51. At Centraal Station a transfer to tram is then made. It should be kept in mind that journeys to Centraal Station and/or Rokin by metro with the last part of the journey by foot are not visible. This is an attractive option as well.

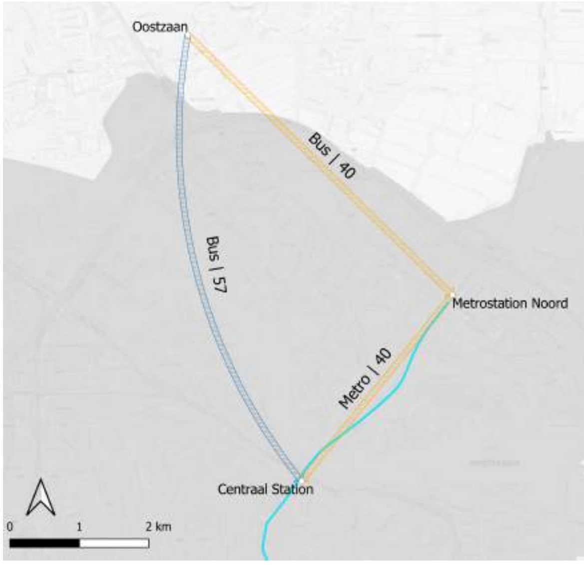


Figure 5.27: The main routes including mode and number of travelers per day from Oostzaan to Centraal Station before (blue) and after the Noord/Zuidlijn (orange)

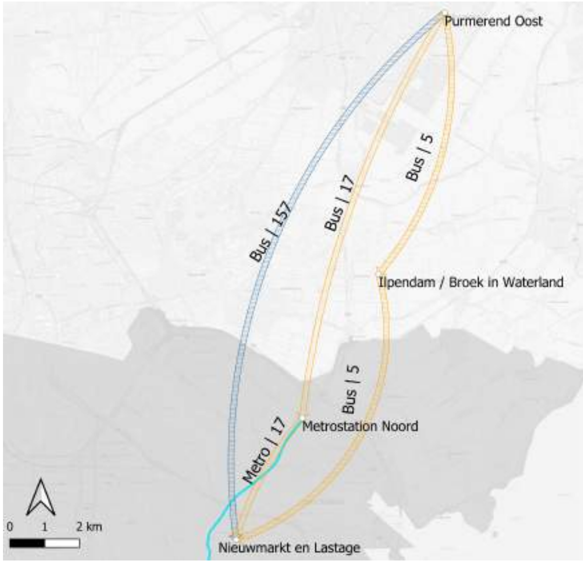


Figure 5.28: The main routes including mode and number of travelers per day from Purmerend Oost to Nieuwmarkt en Lastage before (blue) and after the Noord/Zuidlijn (orange)

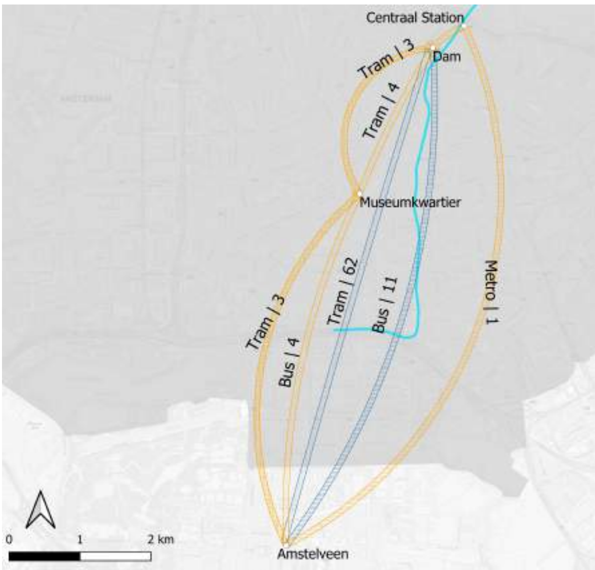


Figure 5.29: The main routes including mode and number of travelers per day from Amstelveen to Dam before (blue) and after the Noord/Zuidlijn (orange)

5.3. Wrap up results

This section presented the results. The total travel time benefits are 1,350 hours of travel time saved on an average working day. However, over all travelers 2,500 extra transfers are made per working day. The travel time benefits outweigh the transfer costs. However, the results show that the impacts are unevenly distributed. The largest travel time gains occur on journeys where a large part of the trip is replace by a metro leg. The largest travel time losses occur to travelers that have to make an extra transfer to make a short metro trip.

In Chapter 6 several nuances are made that are important to keep in mind while interpreting the results. Chapter 6 presents the sensitivity of the results. In addition, the findings are related to the expectations and similar research. Furthermore, it discusses the shortcomings of the research.

6
Discussion

How do the transportation impacts of the Noord/Zuidlijn relate to the literature presented in Chapter 2? Did the intended outcomes from the Lijnennetvisie as described in Chapter 3 actually come true? Did the data or methods as set out in Chapter 4 lead to any limitations of this research that should be kept in mind when reading the conclusion. Answers to these question are given in this chapter.

First, the sensitivity of the results is analyzed. Then, the findings are related to similar literature. This entails comparison to the Lijnennetvisie (Stadsregio Amsterdam, 2014) and to the literature presented in Chapter 2. Thereafter, the limitations of the research are described and if possible, the effect thereof as well.

6.1. Sensitivity analysis

In this section the results of the sensitivity analysis are presented and interpreted. How and why the sensitivity analysis has been performed is described in Section 4.7.4. First, the sensitivity of the travel time results are presented. Thereafter, the results of the sensitivity analysis of the transfers is discussed.

6.1.1. Sensitivity analysis of the travel time results

The most important finding in Table 6.1 is that the lower bound travel time benefits change when the averages for each range are underestimated. This is explained by a different weighting of the travel time per half hour block as a result of the changed estimates. Another interesting outcome is that the travel time benefits decrease more strongly than the number of travelers per day. This is another indication that the calculated average travel times per OD-relation change strongly when the average value per range is underestimated. Besides, this results in lower travel time benefits per traveler. Furthermore, it can be seen that the estimated travel time benefits become zero. This behavior is in line with the expectations as the underestimated average value equals the lower bound of each range.

Table 6.1: Change in travel time benefits for the underestimated averages for each range. The benefits are calculated over all OD-relations over the whole day and distinguished in different representativeness categories.

Representativeness criterion	**Total travel time benefits [h] (lower bound | estimated)**	**Average number of travelers per day [travelers/day]**	**Average travel time benefits per traveler [min/traveler]**
1-3	6 - 3 (3 + 0 | 3 - 3)	100 - 40	4.4 - 1.4
4-11	18 - 7 (9 + 3 | 10 - 10)	500 - 300	2.1 + 1.2
12-50	220 - 23 (94 + 103 | 126 - 126)	6,500 - 3,200	2.02 + 1.56
≥51	1,105 - 748 (605 - 247 | 501 - 501)	88,800 - 26,000	0.75 - 0.41
Total	**1,349 - 780** **(710 - 141 | 639 - 639)**	**95,600 - 29,600**	**0.84 - 0.32**

Table 6.2 presents the change in travel time benefits if the average value for each range is overestimated. It can be seen that the travel time benefits are very sensitive to an overestimation. The travel time benefits more than double and the number of travelers per day increases by almost 70%. This leads to an increase in the average travel time benefits per traveler.

Another interesting observation is that the increase in travel time benefits is largely caused by an increase in estimated travel time benefits. The explanation thereof is that the overestimation only affects the estimated part of the travelers. The lower bound number of travelers does not change. The effect on the lower bound benefits which is caused by a change in average travel time per OD-relation, is much smaller. This indicates that the average travel time is not so sensitive for an overestimation of the average value for each range.

Comparison of the travel time benefits for an underestimation and overestimation of the average value shows that the change in average travel time per OD-relation is more sensitive to an underestimation than to an overestimation. This is apparent from the difference in the effect on the lower bound travel time benefits.

Besides, it becomes clear that overestimation of the average value for each range leads to much larger effects than an underestimation.

Table 6.2: Change in travel time benefits for the overestimated averages for each range. The benefits are calculated over all OD-relations over the whole day and distinguished in different representativeness categories.

Representativeness criterion	**Total travel time benefits [h] (lower bound | estimated)**	**Average number of travelers per day [travelers/day]**	**Average travel time benefits per traveler [min/traveler]**
1-3	6 + 8 (3 + 0 | 3 + 8)	100 + 100	4.4 - 0.2
4-11	18 + 21 (9 - 1 | 10 + 22)	500 + 800	2.1 - 0.3
12-50	220 + 292 (94 - 7 | 126 + 299)	6,500 + 11,400	2.02 - 0.30
≥51	1,105 + 1,266 (605 + 44 | 501 + 1,221)	88,800 + 53,200	0.75 + 0.25
Total	**1,349 + 1,586** **(710 + 36 | 639 + 1,550)**	**95,600 + 65,500**	**0.84 + 0.25**

6.1.2. Sensitivity analysis of the transfer results

Table 6.3 presents the transfers benefits for an underestimation of the average value for each range. The number of transfers made increases strongly. This is largely attributed to a change in the average number of transfers made per OD-relation. An underestimation of the average values causes that the lower ranges are weighed relatively less heavily compared to the higher ranges. More transfers are made in the after situation that are now weighed even heavier. The difference in the number of transfers on an OD-relation between the before and after situation thus increases as a results of the underestimation. This leads to an increase of the number of extra transfers.

Next, as could be expected the estimated extra transfers go to zero because the underestimated average equals the lowerbound of each range.

Table 6.3: Change in transfer benefits for the underestimated averages for each range.

Representativeness criterion	**Total extra transfers (lower bound | estimated) [transfers]**	**Average number of travelers per day [travelers/day]**	**Average transfer benefits per traveler [transfers/traveler]**
1-50	-433 - 82 (-391 - 123 | -42 + 42)	5,200 - 2,500	-0.083 - 0.108
≥51	2,893 + 504 (2,644 + 753 | 249 - 249)	86,200 - 8,200	0.034 + 0.01
Total	**2,460 + 422** **(2,253 + 630 | 208 - 208)**	**91,400 - 10,700**	**0.027 + 0.009**

Table 6.4 shows that the lower bound and estimated transfer benefits are both quite sensitive to an overestimation of the average value for each range. The decrease in the lower bound number of transfers indicates that the overestimation leads to a decrease in the number of transfers per OD-relation. However, as the overall impacts remain negative (i.e. an increase in transfers) this is nullified by the additional number of travelers.

Table 6.4: Change in transfer benefits for the overestimated averages for each range.

Representativeness criterion	**Total extra transfers (lower bound | estimated) [transfers]**	**Average number of travelers per day [travelers/day]**	**Average transfer benefits per traveler [transfer/traveler]**
1-50	-433 + 166 (-391 + 92 | -42 + 75)	5,200 + 3,700	-0.083 + 0.053
≥51	2,893 - 136 (2,644 - 464 | 249 + 327)	86,200 + 11,400	0.034 - 0.006
Total	**2,460 + 30 (2,253 - 372 | 208 + 402)**	**91,400 + 15,100**	**0.027 - 0.004**

6.1.3. Conclusion sensitivity analysis

It can be concluded that the results are sensitive to a change in the estimated average value for each range. The change in benefits is caused by two factors. First, the number of estimated travelers changes. Besides, this change affects the calculated average travel time and number of transfers per OD-relation. As a result the change in travel time and number of transfers per OD-relation before and after the Noord/Zuidlijn is affected as well.

For the travel time results the effects on the estimated benefits are larger than the effects on the lower bound benefits. However, the latter cannot be ignored. Regarding the transfer benefits the change in the results is caused by a change in the number of transfers per OD-relation and to a change in the number of travelers. The magnitude of these effects is similar.

The lower bound benefits (and thus the average travel time and number of transfers per OD-relation) are more sensitive to an underestimation than an overestimation of the average value for each range.

6.2. Relation of the findings to literature

In this section the findings of this research are related to literature on large-scale public transport networks overhauls. First, the relation to the Lijnennetvisie and to the impact of the Noord/Zuidlijn on urban and concession level is discussed. Thereafter, the findings are compared to papers about ex-post evaluation of networks overhauls worldwide.

6.2.1. Relation of the findings to the Lijnennetvisie

In the Lijnennetvisie (Stadsregio Amsterdam, 2014) the expected impact of the Noord/Zuidlijn is that for 90% of the travelers the travel time decreases or stays the same. The findings of this research approximate this percentage as out of the total 90,000 travelers 60,000 are minimally affected and 15,000 experience a decrease in travel time. Thus, 85% experiences no or a positive effect of the Noord/Zuidlijn.

Next, as expected in Section 3.5 the impacts of the Noord/Zuidlijn are larger for Waterland and Amstelland-Meerlanden than for Zaanstreek.

For the concession Waterland, the expectation as described in the Lijnennetvisie is that the travel times to areas near the Noord/Zuidlijn decrease and to the city center increase. These predictions turn out to be true. The expectation for the eastern flank was that the travel times would stay the same. However, the travel times between some areas in Waterland and the eastern flank of Amsterdam have increased. It should be investigated whether additional measures could alleviate or overcome this effect.

For Zaanstreek, the prediction was that travelers benefit from a transfer to the Noord/Zuidlijn if the metro trip is long enough. This is confirmed by the results. However, bus line 391 and 394 only offer transfer connections to the Noord/Zuidlijn at Centraal Station. As a result, the impact of the Noord/Zuidlijn is smaller than expected.

Comparing the predictions and the actual impacts for Amstelland-Meerlanden shows that the travel times to the city center between Elandsgracht and Centraal Station have increased as expected. However, the travel times south of Elandsgracht have increased as well. This is the result of the preference of a part of the travelers for a transfer to the metro network at Amstelveenseweg. The distribution of buses, e.g. 50% to Amsterdam Zuid and 50% to Elandsgracht, is not in line with the distribution of travelers over these buses. Therefore, it is recommended to revise this distribution or account for this effect in the timetable.

6.2.2. Relation of the findings to the results on urban and concession level

Brands et al. (2020) found that each day the travel time savings on urban level are more than 6,000 hours. Additionally, they find transfer savings of around 2,000 transfers per day. The travel time impacts on urban level are thus more than 4 times larger than the impacts on regional level. These were 1,350 hours of travel time benefits per working day. This is largely explained by the difference in number of urban and regional journeys. The travel time impacts per traveler and the distribution thereof are more similar. The transfer impacts are negative on regional level while on urban level the transfer impacts are positive. The reason for this is that on urban scale the Noord/Zuidlijn has triggered an increase in direct metro trips. This probably even includes travelers using a direct regional bus before the Noord/Zuidlijn.

Brands et al. (2021) investigated the effect of the Noord/Zuidlijn on passenger flows, travel times and service reliability for passengers from Waterland and Amstelland-Meerlanden. From Waterland they observe that the travel time to Centraal Station has increased as a result of the extra transfer. This is in line with the findings of this research. However, for the direction of Amstelland-Meerlanden the impact found by Brands et al. (2021) is more positive than the travel time impacts from the region south of Amsterdam in this research. The difference in data aggregation period for the after situation could be the reason for this. Brands et al. (2021) use data from June 2019 while this research uses data from December 2018 and January and February 2019. An important difference between these periods is that metro line 51 or the Amstelveenlijn was operative until the 3rd of March 2019. This offered fast connections from Amstelveen to Amsterdam Zuid. Besides, it ran directly from Westwijk to Centraal Station. An additional explaining factor could be that in the winter months the number of travelers is larger which aggravates the effects of uneven distribution of travelers over buses from Amstelland-Meerlanden.

6.2.3. Relation of the findings to ex-post evaluations of public transport networks worldwide

Regarding transfer patterns and travel times Arbex et al. (2019) find that the overhaul of São Paulo's late night bus network leads to an increase in travel times and transfers. However, the increase in travel time is ascribed to the fact that people undertake activities at locations further away that were not accessible before. Such a phenomenon can not be concluded from this research as a shift in activity location has not been investigated.

Fu and Gu (2018) conclude that after the construction of a new metro line in Nanjing the travel time changes for many stations. The Noord/Zuidlijn has led to travel time changes as well, but a large part of the OD-relations is minimally affected. The fact that Fu and Gu (2018) find that many stations are affected is probably because they only analyze the travel time impacts in the metro network.

Weckström et al. (2019) find that the transition from a direct bus network to a trunk-feeder network results in more transfers for passengers relying on feeder buses, while the service level near metro stations improves. These distributive impacts are in line with the findings from this study.

Besides, Reinhold (2008), Arbex et al. (2019), Fu and Gu (2018) and Brands et al. (2020) all report an increase in ridership after the network overhaul. In this research that effect is visible as well. However, a significant part of the increase in travelers is probably the result of a difference in aggregation period.

6.3. Limitations of the research

This section presents the limitations of the research. These are grouped in three, slightly overlapping categories: data limitations, methodological limitations and limitations due to external factors. If possible, the effect of a limitation is described as well. The most important limitations are described in detail first. Last, less important limitations are presented.

6.3.1. Data limitations

This section describes the limitations related to the data used in this research. These include limitations of the data specifically used for this research and limitations of smart card data in general.

General weaknesses smart card data

Smart card data has some general limitations. These are set out below.

First, it does not contain all public transport journeys made. Journeys made using single-use tickets (CT-tickets), paper tickets or tickets purchased in the vehicle are not apparent in the data delivered by Translink. The share of CT-tickets and paper tickets is presented in Appendix X. Over all trips made in the VRA single-use and paper tickets add up to 17.2% of the total trips. However, there are large differences between concession areas and modalities. The lack of these ticket types results in an underestimation of the number of travelers. Additionally, because these tickets are often used by tourists the underestimation is stronger for the summer months.

Second, smart card data does not capture the door-to-door travel time but only the time between the first check in and last check out. It does thus not include access and egress time. It can be expected that the overhaul has affected the access and egress location. Therefore, not all travel time impacts have been captured because access and egress time could not be included in the research.

Another weakness is that in buses and trams the smart card is validated in the vehicle, while for train and metro the card is validated at the station. This means that if the first trip of a journey is made by train or metro waiting time is included in the recorded travel time. Thus, if the mode for the first trip between before and after the network overhaul changes from bus or tram to metro or train, this means that after the network overhaul the recorded travel time contains waiting time. Therefore, part of the recorded change in travel time results from a different measurement instead of actual changes in travel time.

Another weakness is that intramodal transfers for train and metro are not visible in smart card data. The effect on the number of transfers therefore is underestimated because on trips from Amstelveen to Amsterdam it seems like the Noord/Zuidlijn resulted in a direct metro connection while these travelers need to make an intramodal transfer.

Next, smart card data only contains information on journeys that have been made. Therefore, OD-relations on which the travel time or number of transfers has increased so much after the Noord/Zuidlijn that travelers do not make a journey any more are not included in the analysis. This could result in a more positive picture of the impacts of the Noord/Zuidlijn.

Besides, smart card data does not contain information on access and egress (time). Because the network change affected the route choice of travelers and likely the origin and destination station as well, the absence of access and egress time makes it impossible to capture the full travel time impacts. The strong increase and decrease in the number of travelers for certain stop clusters indicates that a change in origin and destination stop cluster indeed occurred. In that case, not just the in-vehicle time has changed but the access and egress time as well. In addition, upgrading regional bus lines to R-net has resulted in the removal of bus stops. This results in lower in-vehicle time at the cost of longer access and egress times. However, the effect of this limitation cannot be described generally.

Different data aggregation periods

The difference between the aggregation periods of the before and after data sets is another main limitation of this research. This has led to a couple of external factors that should be taken into account when interpreting the results.

First, the aggregation period of the before data set consisted of summer months while the after data sets consisted of winter months. This likely affected the travel times and number of travelers. The number of travelers is higher in the winter than in the summer. In Appendix W it can be seen that the number of travelers on a working day is around 1.17 times higher for the smart card data before the Noord/Zuidlijn than after the Noord/Zuidlijn. Thus, the number of travelers affected by the Noord/Zuidlijn is overestimated. The travel times for bus and tram are expected to be longer in the winter months because of longer alighting and boarding times due to more ridership and more traffic on the road. For that reason, the resulting travel time impacts in this research are less positive than in reality.

Second, the number of working days in both data sets differ while the ranges are the same. This complicates comparison because it is likely that the number of travelers per half hour block or route is larger in the after data. However, this is not visible if the ranges and accompanying estimates are similar. Besides, a longer aggregation period likely results in more half hour blocks and routes traveled.

Third, the aggregation period after the Noord/Zuidlijn is quite close to the opening of the Noord/Zuidlijn. For travelers, especially for such a large-scale change as the Noord/Zuidlijn, it might take more time to get fully accustomed to the new network.

Many records in the data in the lowest bin

The fact that for most data records the number of travelers falls in the lowest bin, results in two complications. First, for a large part of the records the representativeness of the travel time or route taken cannot be determined with certainty. Second, in relative terms the estimated number of travelers is very uncertain. This leads to fragmentation of the results. In presenting the results a trade-off needs to be made between a representative and a full picture of the results. This holds especially for the results on OD-level.

It might even be argued that exclusion of the less representative OD-relations results in a positive bias on the outcomes. An OD-relation is less representative if the number of travelers on that OD-relation is low. A possible reason for this low number of travelers could be that the travel times and/or number of transfers on that OD-relation makes it an unattractive OD-relation. This way more unattractive OD-relations might be left out of the analysis.

6.3.2. Methodological limitations

In this section the limitations of the applied methodology are presented. This entails choices during data processing and to overcome the weaknesses of interoperable smart card data.

Equal weighting of similar ranges

One of the main limitations of this research is that in calculation of the average travel time and average number of transfers per OD-relation, half hour blocks (for the travel times data) and routes (for the route data) with a similar range are weighted the same. This affects the calculated average travel time or number of transfers on an OD-relation. The impact of this limitation is larger if the travel time or number of transfers differs strongly among half hour blocks or routes on an OD-relation. Besides, this impact is larger if the number of half hour blocks or routes on an OD-relation is smaller and if the total number of travelers on the OD-relation is smaller.

The reason for this limitation is that additional information on the number of travelers is lacking. As a result, records that have an actual number of travelers close to the lower bound of a range and records where the actual number of travelers is close to the upper bound, are weighted the same.

Valuation of travel time and transfers

To monetize the benefits a valuation of €18 for an extra hour of travel time is used. One additional transfer is assumed to cost €1.4. These valuations are based on Knockaert and Koster (2020) and calculated for the Amsterdam region specifically. Compared to the valuations used by the KiM (Kennisinstituut voor Mobiliteitsbeleid) for transport policy analysis (Kennisinstituut voor Mobilieitsbeleid, 2003) in the Netherlands these values are relatively high.

Besides, the data did not allow to distinguish between waiting, transfer and separate in-vehicle times for each mode. Additionally, one value is taken for the valuation of an extra transfer. Valuating a transfer differently depending on the mode or transfer environment (Schakenbos et al., 2016) would be more accurate.

Validation of the estimates is based on all trips in the VRA

Validation of the estimates is done using all trips in the concession areas instead of just the regional trips. Because the number of travelers in the concession area Amsterdam is much higher than the number of travelers in the other concession areas, validating the estimates using data for all trips likely resulted in an overestimation. It would have been better to validate the estimated average values using regional trip data only. However, this data was not available.

Estimation of the number of travelers

Another remark regarding the validation is that the estimated values for the travel times data lead to an underestimation of the total number of journeys after the Noord/Zuidlijn. This results in an overall overestimation of the travel time benefits because the change in number of travelers is smaller in reality.

For the route data the total number of daily trips is overestimated for the before and after data. The overall negative impact of the number of transfers is expected to be overestimated as a result of the overestimation of the number of travelers before the Noord/Zuidlijn. The explanation thereof is as follows. The overestimation of the number of travelers before the Noord/Zuidlijn results in a more negative outcome if the number of transfers increases. If the number of transfers decreases the overestimation leads to a more positive result. The overestimation of the number of trips after the Noord/Zuidlijn is expected to have a minimal effect because the number of trips before the Noord/Zuidlijn is overestimated as well. The change in number of travelers is thus neither underestimated nor overestimated.

Choice to analyze working days
In this thesis the analyses focus on working days. Travel demand is generally higher on these days compared to Saturdays and Sundays. The impact of the Noord/Zuidlijn is thus larger on working days than on Saturdays and Sundays.

Another difference between working days and weekend days is that public transport services operate at lower frequencies during the weekends. The travel time could then be larger for journeys that require a transfer. This holds in the situation before and after the Noord/Zuidlijn. It is likely that this affects the travel time impacts. However, the magnitude and direction thereof remain unknown. The number of transfers in a journey is expected to depend minimally on the type of day analyzed.

Ratio of estimated average values affects averages per OD-relation
For the travel time analysis and the transfer analysis one set of validated estimates is used to obtain the results. However, another set of estimates could have also led to a valid total number of trips and journeys. Another set of estimates would have changed the ratio of the estimated average values. This could have affected the average travel time or number of transfers per OD-relation as the half hour blocks or routes would then have been weighted differently.

Sporadically arguable representativeness criteria
The representativeness criteria allow to distinguish between different degrees of representativeness of the results. However, in some cases it could be that a lot of records (i.e. half hour blocks for the travel times data and routes for the route data) fall in the 1-50 bin and only few are larger than 51, the 1-50 bins then have a large effect.

6.3.3. External limitations

This section describes external factors that affect public transport supply and demand that have not been excluded from this research.

No exclusion of societal trends
An external limitation is that the effects of societal trends have not been separated from the effects of the Noord/Zuidlijn. Trends or factors that could have contributed to the impacts found in this study are for example a natural growth in demand, young families leaving Amsterdam, growth in the number of expats, Amsterdam's policy to decrease car use, companies relocating or an increase in working from home. However, because the time between the period before and after the Noord/Zuidlijn is around six months, the fact that these trends are not accounted for, is expected to be small.

Possible other causes for a change in travel time
Next to the Noord/Zuidlijn, timetable changes, traffic, weather or diversions could have affected the travel time changes as well. These factors could not be excluded from the analysis. This leads to noise in the results.

Other impacts of the Noord/Zuidlijn
It should be noted that impact of the Noord/Zuidlijn is absolutely not limited to the travel time and transfer impacts on regional journeys. Additional impacts include the impact on urban and national scale. Besides, transportation impacts (e.g. increased reliability), environmental impacts (e.g. decrease in emissions), spatial impacts (e.g. more liveable city center) and economic impacts (e.g. increase of real estate prices) have not been included.

7 Conclusion and recommendations

The first section of this chapter answers the sub-questions as stated in Chapter 1 one by one. This builds up to answering the main research question. Using the experiences of working with aggregated, interoperable smart card data as described in Chapter 4 the methodological research questions can be answered. The first two sub-questions are the methodological sub-questions. The results presented in Chapter 5 enable answering the societal research questions. These are the latter three sub-questions. Chapter 2 and Chapter 3 do not lead to answers to the sub-questions directly, but provided the necessary background knowledge. After the conclusion, the recommendations for practice and further research are presented.

7.1. Conclusion

In this section the main conclusions that are drawn from this research are presented. This is done by answering the sub-questions in order first. Thereafter, the main research question is answered.

Sub-question 1: **What are the strengths and weaknesses of aggregated, interoperable Dutch smart card data compared to disaggregated, single operator Dutch smart card data?**

Conceptually, aggregated, interoperable smart card data consists of a public transport journey between the access and egress stop of the public transport network. All trips are combined into a journey made by the traveler. Individual journeys are aggregated over space and time and this affects the level of detail of the number of travelers. The level of detail regarding the number of travelers depends on the level of aggregation over space and time. If the number of travelers is smaller than a certain threshold value, then the number of travelers is binned in ranges. This is done for privacy reasons. If the number of travelers exceeds the threshold value, then the number of travelers is exactly known.

The main strength of aggregated, interoperable smart card data compared to single operator data is that it provides insights in the whole public transport journey across services from multiple operators. Besides, aggregated, interoperable smart card data provides insights in the distribution and exchange of travelers between operators serving overlapping areas. Third, since interoperable smart card data gives insights in the whole picture regarding public transport travel, the exchange between public transport and private modes is better visible than when single operator data is used.

Because the whole journey is recorded and the exchange and distribution of travelers among multiple operators is known, interoperable smart card data facilitates comprehensive analysis of a multi-operator public transport network. This could improve comprehension of route and mode choice in a multimodal, multi-operator public transport network.

Better understanding of route and mode choice in a multimodal, multi-operator public transport network, could improve public transport planning and the alignment of transport plans between operators. This improves the efficient use of public transport funds. However, unwillingness of operators could be expected if cooperation results in a loss of revenue. This is a clear example of the recurring, opposing objectives for the traveler, operator and transport authority in public transport planning.

The main weakness of the aggregated, interoperable smart card data used in this thesis is that for most records the number of travelers is given in bins. The sharing of individual records is not allowed and desirable because the data contains privacy and commercially sensitive information.

The fact that the number of travelers is given in bins leads to several limitations regarding the analysis of interoperable smart card data. The most important limitation is that if the number of travelers for a record does not meet a certain threshold value, the reliability of the recorded variable is low. The more records this applies to, the more complicated analysis of aggregated, interoperable smart card data gets. Another drawback of not knowing the exact number of travelers is that it is uncertain to how many travelers the results of the analysis apply. A third limitation is that binning makes it harder to make a distinction in the value assigned to similar records. This becomes more problematic if a lot of records end up in the same bin.

Additionally, the fact that the data has already been aggregated, affects further processing in three ways. First, aggregation choices cannot be reverted (e.g. a certain transfer threshold or clustering of stops). Second, uncertainty propagates (e.g. average of averages). Third, it is hard to detect outliers in aggregated data. There are two reasons for this. First, the effect of outliers is evened out by aggregation. This results in slight deviations for the aggregated values which makes the outliers harder to detect. Second, because the number of travelers is not known it is hard to determine whether an unusual record is indeed an outlier or occurred to multiple travelers. In the latter case, the record should not be classified as an outlier.

These limitations especially complicate analysis of small flows. For these relations the less reliable records weigh more heavily. Besides, for these relations the estimated travelers make up a larger share of the total number of travelers.

Sub-question 2: **How can the weaknesses of aggregated, interoperable Dutch smart card data be overcome or coped with?**

The effect of binning and aggregation on the workability of aggregated, multi-operator smart card data can be decreased on forehand and while analyzing the data.

On forehand, the level of detail over space and time for the data analysis should be clear when submitting the data inquiry. This way, the level of detail for the number of travelers can be adjusted to the level of detail over space and time. The level of detail for the number of travelers is then never lower than necessary. Besides, if the resulting level of detail turns out to be insufficient, the analyst can tweak the two other variables. Second, the edges of the bins for the number of travelers should be selected in such a way that there is a clear distinction between records that are possibly unreliable and records that are certainly reliable. To clarify, if for example 25 travelers is assumed to be enough for a representative record, there should be a bin edge at 25. This way, the data in a higher bin can be classified as reliable. Outlier detection can then only focus on the lowest bin. If the bin edge would have been at for example 40, all records with an actual number of travelers between 25 and 40 could be classified as an outlier because to the analyst it is not known whether the actual value is 1 or 40 travelers. The value that is assumed to be enough depends on the data requested. Of course, this is only possible if this is allowed by the privacy and competition regulations.

When the aggregated, interoperable smart card data is acquired, the weaknesses can be coped with by estimating and validating an average value for each bin based on an external data source. Afterward, the average value should be split in a lower bound and estimated number of travelers. This allows the analyst to keep track of which part of the results is certain and which part is estimated on top of that. Additionally, it is recommended to distinguish between different representativeness categories based on the lower bound number of travelers in the analysis phase. This way, the insights and conclusion can be divided into different degrees of certainty.

Although the solutions mentioned above facilitate working with interoperable smart card data workable, it is hard to overcome the weaknesses of aggregated, multi-operator smart card data entirely.

Sub-question 3: **What impact did the Noord/Zuidlijn have on the travel times for regional public transport journeys in the Vervoerregio Amsterdam?**

Overall, the Noord/Zuidlijn results in 1,350 hours of travel time savings on regional journeys on a working day. This is equal to societal benefits of 24,000 euros per day. The travel time decreases by around 50 seconds per average traveler.

Regarding the distribution of the travel time impacts, it is found that the daily travel time savings are 2,350 hours while the losses are 1,000 hours. The travel time savings are more than twice as large as the travel time losses. This is explained by the fact that the magnitude of travel time decreases is larger than that of travel time increases. Besides, in Figure 7.2 it can be seen that more travelers are positively affected than negatively affected. From Figure 7.1 it can be concluded that the share of OD-relations that is not or minimally affected is much smaller than the share that is either positively or negatively affected by the overhaul.

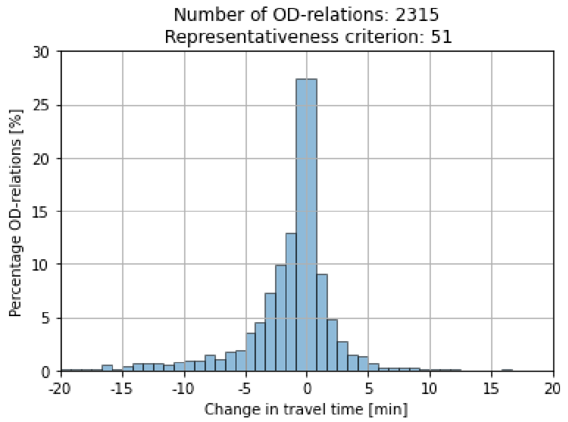


Figure 7.1: Distribution of the absolute change in travel time against the percentage of OD-relations. Results are plotted for the average travel time over the whole day.

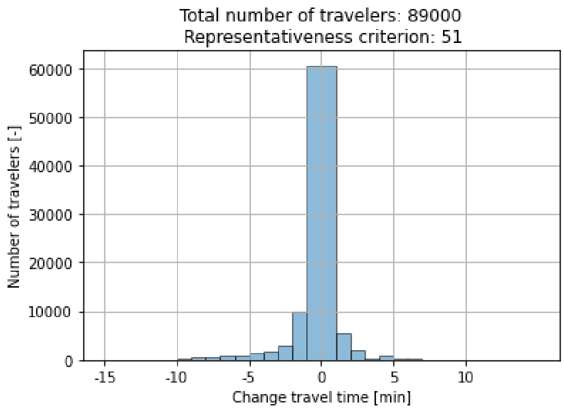


Figure 7.2: Distribution of the absolute change in travel time against the number of travelers. Results are plotted for the average travel time over the whole day.

Overall, only for stop clusters in the city center that have lost a direct connection the travel times deteriorate on average. The rest of the stop clusters experiences no effect or a positive effect on the travel time on regional journeys. The largest decreases in travel time occur on OD-relations from and to Noord/Zuidlijn stations if the length of the metro trip is long enough to make up for the additional transfer. The travel times to areas in the city center without a Noord/Zuidlijn station increase most. This is the result of losing a direct connection. The time lost by the extra transfer is then not compensated by the faster metro trip. Besides, the travel time impacts can differ strongly between stop clusters that are not far apart in a beeline. This is explained by different changes to the public transport networks in these clusters.

Last, there are indications that the travel time impacts are less positive in the morning peak and evening peak than over the rest of the day. Additionally, the travel time impacts are less positive when traveling from Amsterdam to the region than vice versa.

Sub-question 4: **What impact did the Noord/Zuidlijn have on the transfer patterns for regional public transport journeys in the Vervoerregio Amsterdam?**

As a result of the Noord/Zuidlijn, the average number of transfers per regional journey increased from 0.54 to 0.66. The share of multimodal journeys by metro and bus or train increases most after the Noord/Zuidlijn. On a working day, the overall transfer impacts comprise an increase of 2,500 transfers. This costs the society 3,500 euros per day. On average a traveler has to make 0.027 transfers extra per journey.

A large difference in transfer gains and losses can be observed. On a working day the transfer gains are at least 3,000 transfers and the transfer losses are at least 6,000 transfers. However, the 3,000 transfer gains solely result from seemingly unimodal metro journeys, while actually regional metro journeys require an intramodal transfer. In Figure 7.4 and Figure 7.3 it can be seen that the largest part of the travelers and OD-relations are not affected by the Noord/Zuidlijn. A larger share of OD-relations than travelers is affected by the Noord/Zuidlijn.

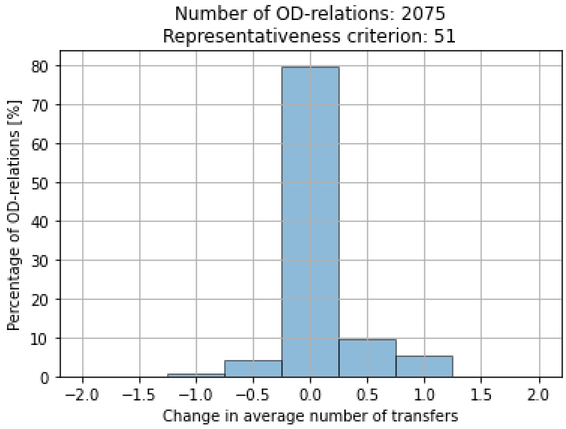


Figure 7.3: The distribution of the transfer impacts over OD-relations.

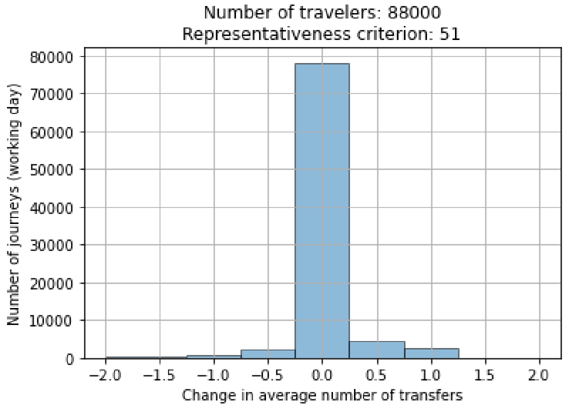


Figure 7.4: The distribution of transfer impacts over travelers.

For stop clusters where the direct connection to the city center disappears, an extra transfer is necessary to reach the city center. The decreases in transfers can be attributed to the addition of bus stop IJtunnel for regional buses and to metro journeys from Amstelveen to Noord/Zuidlijn stations.

Sub-question 5: **What do the outcomes imply for public transport planning?**

No public transport networks and thus no public transport network overhauls are the same. Therefore, the findings of this research cannot be copied exactly to other large changes in public transport networks worldwide. However, the findings do give ideas about what to bear in mind when planning a large-scale public transport network overhaul.

First, when transforming a regional public transport network consisting of direct bus lines to a trunk-feeder network consisting of a metro trunk line and feeder bus lines, the metro leg should be sufficiently long to make up for the time lost by the additional transfer. The main losses occur on OD-relations where a direct connection is replaced by a route with an additional transfer to a short metro leg. More general, for an overhaul that requires an additional transfer to a faster mode, the length of the trip with that mode should be long enough to make up for the time lost by the additional transfer.

Second, the effects of a large-scale public transport overhaul differ from an aggregate or disaggregate viewpoint. When planning a large-scale overhaul the authority should find the right balance between these.

Third, the impact of a large-scale network change is distributed differently over OD-relations than over travelers. Therefore, the transport authority deciding on the network overhaul should have a clear goal of the network overhaul. An overhaul targeting large flows over passengers likely scores better in a cost-benefit analysis. However, regarding equity assessment on OD-level is better suitable.

Next, transport planners should be aware that the impact might be different for other time periods and in opposite travel directions. This could affect public transport ridership.

Furthermore, the findings show that the impacts of a large-scale network overhaul are not limited to the area where the overhaul took place. Therefore, the travel time and transfer impacts on a larger scale should be incorporated in ex-ante evaluations.

Besides, the impact of a network overhaul is smaller if similar, competing routes remain prevalent after the overhaul. This is a substantiation for the removal of duplicate routes. Besides, the impacts on journeys from and to areas perpendicular to the network overhaul can be expected to be smaller.

Additionally, when assessing the expected effects of another large-scale network change, do not forget transferring passengers. These might opt for another transfer location than the seemingly most logical. Besides, if the journey leg to the transfer location is affected the rest of the journey is as well.

Another takeaway of this research is that areas that benefit and deteriorate of a large-scale network change could be close to each other. Therefore, improving connections between these areas could alleviate the negative impacts. An access or egress trip from or to an area that benefits, might be quicker than accessing or egressing the public transport network in an area that deteriorates.

Answering all sub-questions enables answering the main research question. The main research question is:

What impact does the Noord/Zuidlijn have on travel times and transfer patterns on regional public transport journeys in the Vervoerregio Amsterdam and what does this imply for public transport planning?

The Noord/Zuidlijn resulted in at least 1,350 hours or €24,000 of travel time benefits on regional journeys on a working day. This outweighs the negative transfer benefits of 2,500 transfers or €3,500 extra on a working day. The net impacts on regional travelers come down to €20,500 on a working day. On a yearly basis this is roughly €7.5 million. This amount solely consists of changes in travel time and number of transfers. It can thus be concluded that the transformation of the regional, direct bus network to a network consisting of a trunk metro line and feeder buses has resulted in net positive societal impacts when only considering the impact on travel times and transfers.

Table 7.1: The regional societal travel time and transfer benefits of the Noord/Zuidlijn per working day.

	Quantity	**Monetary value per unit**	**Societal benefits per day**
Travel time savings	1,350 hours	18 €/hour	€24,000
Transfer savings	-2,500 transfers	1.4 €/transfer	€-3,500
Total	-	-	**€20,500**

The main implications for public transport planning are that for a public transport overhaul that requires a transfer to a faster mode, the length of the trip leg should make up for the time lost by the transfers. Furthermore, in future large-scale public transport overhauls the different aggregate and disaggregate effects should be accounted for. Besides, public transport planners should take the distribution of the effects among travelers, OD-relations, time periods and travel directions of a large scale network overhaul in consideration.

7.2. Recommendations

The findings and limitations of thesis lead to several recommendations for practice and further research. The former consists of recommendations for the public transport sector. The latter are directed to public transport researchers. Additionally, a distinction can be made in recommendations regarding the data inquiry and processing and regarding public transport planning.

7.2.1. Recommendations for practice

The recommendations for practice have been categorized in recommendations on data inquiry and clustering, on data processing and on public transport planning.

Recommendations on interoperable smart card data inquiry and clustering of stops

In this section recommendations that improve data aggregation and delivery are presented. These recommendations should thus be considered before the data is requested. First, recommendations regarding the data request are presented. Thereafter, advice is given on the clustering of stops.

Regarding the data inquiry, the first recommendation is that the level of aggregation over space and time required for the analysis should be determined before the data is requested. When the aggregation level for the request and analysis of the data are aligned, further aggregation while processing the data is not necessary. This prevents propagation of uncertainty when further aggregating the data.

Second, the bin size should be as small as the privacy regulations allow it be. Additionally, the upper bound of the lowest bin for the number of travelers should be the number of travelers that is required to obtain representative journey characteristics. Then, all the records for which the number of travelers is above that upper bound can be considered as representative.

Third, it is advised to outsource outlier removal to the provider of the data. Before aggregation the data is more detailed and thus outliers can be identified more easily. However, this process should be monitored intensively and the outlier removal criteria should be clear. Besides, outlier removal by the provider might lead to longer data delivery times.

Fourth, when the analysis entails travel time, the travel time of the whole journey should be split in trip times and transfer times. This enables more complete analysis.

Fifth, the total number of travelers should be included in the request. This simplifies validation as an external data source is no longer needed.

Clustering of stops consists of a difficult trade-off between level of detail and enough observations. Besides, it depends strongly on the goal of the research. Despite that, some general recommendations on the clustering of stops can be given.

First, the public transport network should be studied thoroughly before assigning the stops to clusters. A good understanding of the public transport network prevents illogical assignment of stops to stop clusters. Furthermore, it is advised to ensure that the same stop in different directions is assigned to the same cluster. This prevents undesirable asymmetry in the results. Last, depending on the goal of the research, it might be valuable to distinguish between stop clusters per modality. This provides information on mode choice.

Recommendations on interoperable smart card data processing

This section consists of recommendations on how to work with interoperable smart card data to obtain useful results.

To work with interoperable smart card data it is recommended to estimate an average value for each binned number of travelers. This enables working with the binned values and leads to more plausible results than choosing the lower bound or upper bound of each bin. Ideally, this average value should be validated using the total number of travelers in the data. However, if these totals are not available, external data sources can be used to estimate an average value. In addition to estimating an average value for each bin, it is advised to keep track of the uncertainty. This can be done by splitting the average number of travelers in a lower bound and estimated number of travelers. To illustrate this, if validation using an external data source results in an estimated value of 65 travelers for range 51-100, the lower bound number of travelers is 51 and the estimated number of travelers is 14. This enables dividing the results in a certain and less certain part.

Another recommendation for data processing is to distinguish different categories of representativeness. This way, different parts of the results can be valued differently. This should prevent drawing conclusions from less reliable results. The reason behind this recommendation is that for records where the number of travelers falls in the lowest bin the characteristics of the journey could be very unusual. As these are then weighed the same as ‘usual’ journeys, this leads to noise in the results.

The last recommendation for practice is to use interoperable smart card data (as used in this thesis) mainly for analysis of large flows of passengers. This is advised because for smaller flows the estimated number of travelers can deviate relatively much from the actual number of travelers. Additionally, the representativeness of small flows is less certain.

Recommendations on public transport planning

The recommendations on public transport planning consist of recommendations on future ex-ante evaluations, recommendations on future large-scale network overhauls, especially transformation of a direct network to a trunk-feeder network, and recommendations on public transport in the Amsterdam region.

For future ex-ante evaluations, it should be kept in mind that the impacts might be distributed differently over OD-relations and travelers. When an overhaul should have much impact, it is advised to target heavily traveled OD-relations. From an equity perspective it is advised to assess the overhaul on OD-level.

Additionally, analysis of the door-to-door effects leads to results that approximate the actual effects better than only assessing the effects on the public transport journeys would do.

Furthermore, it is recommended to be aware of possible different travel time impacts of a large-scale network overhaul for opposite travel directions or different time periods. On OD-relations where the travel time in both directions differs strongly, public transport demand could decrease. Besides, it is advised to consider that the effects differ among time periods. It is desirable that the largest positive impacts occur for busy time periods.

It is advised to investigate travel behavior of transferring travelers thoroughly ex-ante. Here, a distinction should be made between travelers that need to make an extra transfer and another transfer after the overhaul. The former could be negatively affected by the overhaul while the latter is not negatively affected in most cases. Transferring travelers often have multiple, similar route options. Their behavior after a network overhaul is more difficult to determine. Besides, when the part of the journey to a transfer location deteriorates, the whole journey deteriorates.

From the perspective of the traveler, it is advised to prevent the negative impacts of a future large-scale public transport network overhaul. Therefore, cancellation of direct lines is not recommended from a passenger's perspective.

Regarding trunk-feeder networks it is recommended to connect the feeder lines to the trunk line as early as possible. This way, a larger part of the journey can be made on the faster trunk line and thus full advantage is taken thereof.

Besides, when a direct journey by bus is replaced by a multimodal bus and metro journey, the metro leg should be sufficiently long to make up for the time lost by an additional transfer. The time lost by an additional transfer is not compensated if the metro leg is shorter than 5 minutes. To prevent negative transportation impacts the overhaul could be designed in such a way that the trunk line and feeder lines partially overlap. This way an early transfer opportunity is given but travelers that do not benefit from a transfer can use a direct service.

Last, an important advice is to allow adjustments in public transport planning after the overhaul. It is impossible to plan a large-scale public transport network overhaul perfectly.

Regarding public transport in the region of Amsterdam, it is recommended to not build a bus station at station Noorderpark to offer transfer connections from bus line 391 and 394 from Zaanstreek to the Noord/Zuidlijn and vice versa. This station would only minimally improve travel times to southern Noord/Zuidlijn stations.

Besides, it is advised to solve the increase of travel times from Amstelland-Meerlanden to Amsterdam. This is caused by a preference by travelers for the buses to Elandsgracht because these offer transfer opportunities at Amstelveenseweg. As a result, the number of travelers is not evenly distributed over the buses. This leads to delay for the busy buses and operation ahead of schedule for the empty buses. A possible solution could be to reassess the distribution of buses from Amstelland-Meerlanden to Elandsgracht and to Amsterdam Zuid or to increase the frequency.

7.2.2. Recommendations for further research

This research has led to interesting topics for further research. These have been divided in further research on interoperable smart card data and on the transportation impacts of large-scale public transport network overhauls.

Recommendations for further research on interoperable smart card data

The applied methodology to overcome the weaknesses of interoperable smart card data has been developed in this thesis. Therefore, an interesting topic for further research is to further investigate this methodology developed in this thesis. This could focus on the sensitivity of the results for different estimated average values or on the validation using an external data source. A study that compares the use of interoperable and single operator smart card data to obtain the same results could provide interesting insights.

It is recommended to investigate possibilities to apply data fusion to distinguish between the weights assigned to routes or journeys with a similar range. A data source that might be used for this is route planner data. Likely travel times and number of transfers can be obtained using route planner data. Records that match these routes can then be valued higher.

Next, it is recommended to research the possibilities to decrease the number of records that ends up in a bin, especially the same and lowest bin. An expert on privacy and commercially sensitive data should be consulted to advise on this. Interesting directions for further research are to investigate whether the binsizes can be lowered or the threshold value can be reduced.

Recommendations for further research on the transportation impacts of large-scale public transport network overhauls

This research has resulted in interesting new topics for further research. These are recommendations that apply to large-scale public transport network overhauls in general but can be applied to the Noord/Zuidlijn case study as well.

An important recommendation for further research is to investigate the shift of passengers between stop clusters or to private modes. This would provide useful insights in whether the travel time and transfer impacts have triggered a change in route or mode choice of the traveler. This could provide additional, useful insights in change of access and egress times or a shift of travelers to bicycle or car. Such effects should be accounted for in the planning of the overhaul.

Furthermore, it is interesting to investigate how the impact is perceived by the traveler. There could be a difference between the actual measured impacts and the perception thereof. This could affect usage of the public transport network and the public opinion on the overhaul. The fact that travelers perceive and value waiting, access, egress, transfer and in-vehicle time, different modalities and transfers differently substantiates the recommendation for further research on the traveler's perception of an overhaul.

Next, because inclusiveness is an important goal of public transport, it is interesting to investigate the effects on different groups of passenger characteristics. To investigate that, this research could be extended by incorporating passenger characteristics. The main overhaul took place in the city center where people have different socio-economic characteristics than in the regional cores.

Additionally, this research has triggered the question how a large-scale public transport network overhaul affects walking distances. This entails access, egress and transfer walking distances. The walking distance to metro platforms could differ strongly from the walking distances to bus or tram platforms.

This research can be extended by investigating the impact of the Noord/Zuidlijn on a national scale. The Noord/Zuidlijn has created a new possibility to reach the city center from Amsterdam Centraal and Amsterdam Zuid. This likely has affected the route choice of travelers by train. This results in a change in the distribution of travelers over the main train stations. That a change in route choice has occurred, is substantiated by the increase of the share of journeys made by metro and train as can be seen in Section 5.2.1.

Another interesting direction for further research is to investigate the regional impact of the Noord/Zuidlijn on the reliability of travel time. It is expected that the Noord/Zuidlijn has affected the reliability of travel times because the travel time by metro is more reliable than buses and trams. Especially, on journeys that require a transfer unreliable travel times are problematic as this could result in missing a transfer.

References

Arbex, R., da Cunha, C. B., and Speicys, R. (2019). Before-and-after evaluation of a bus network improvement using performance indicators from historical smart card data. *Public Transport*.

Bagchi, M. and White, P. (2005). The potential of public transport smart card data. *Transport Policy*, 12(5):464–474.

Bel, G. and Holst, M. (2018). Evaluation of the impact of bus rapid transit on air pollution in mexico city. *Transport Policy*, 63:209–220.

Brands, T., Dixit, M., and van Oort, N. (2020). Impact of a new metro line in amsterdam on ridership, travel times, reliability and societal costs and benefits. *European Journal of Transport and Infrastructure Research*, 20(4):335–353.

Brands, T., Dixit, M., and van Oort, N. (2021). Transformation to a trunk and feeder network: effects on passenger flows, travel times and service reliability. *BIVEC/GIBET Transport Research Day 2021*, pages 1–12.

Cervero, R. and Kang, C. D. (2011). Bus rapid transit impacts on land uses and land values in seoul, korea. *Transport Policy*, 18(1):102–116.

Claus, S. (2019). De Amsterdamse Noord-Zuidlijn een halfjaar na opening: flop of succes? `https://www.trouw.nl/nieuws/de-amsterdamse-noord-zuidlijn-een-halfjaar-na-opening-flop-of-succes~babf7718/`.

Connexxion (2017). Vervoerplan 2018 Amstelland en Meerlanden. Technical report, Connexxion.

Connexxion (2021). Dienstregeling. `https://www.connexxion.nl/nl/resultaten/dienstregeling`.

Covic, F. and Voß, S. (2019). Interoperable smart card data management in public mass transit. *Public Transport*, 11(3):523–548.

CXX (2018a). Lijnennetkaart Amstelland-Meerlanden december 2017. Technical report.

CXX (2018b). Lijnennetkaart Amstelland-Meerlanden juli 2018. Technical report.

CXX (2018c). Lijnennetkaart Zaanstreek december 2017. Technical report.

CXX (2018d). Lijnennetkaart Zaanstreek juli 2018. Technical report.

Dixit, M. (2021). Lecture notes CIE5825 on transit data analytics - processing & fusion.

Dorantes, L. M., Paez, A., and Vassallo, J. M. (2011). Analysis of house prices to assess economic impacts of new public transport infrastructure: Madrid metro line 12. *Transportation research record*, 2245(1):131–139.

EBS (2017). Lijnennetkaart Waterland december 2017. Technical report.

EBS (2018). Lijnennetkaart Waterland juli 2018. Technical report.

EBS (n.d.). Vervoerplan Waterland 2018. Technical report, EBS.

Fu, X. and Gu, Y. (2018). Impact of a New Metro Line: Analysis of Metro Passenger Flow and Travel Time Based on Smart Card Data. *Journal of Advanced Transportation*, 2018.

Gemeente Amsterdam (2021). Ontwerp Omgevingsvisie 2050: Een menselijke metropool. Technical report, Amsterdam.

Gramsch, E., Le Nir, G., Araya, M., Rubio, M. A., Moreno, F., and Oyola, P. (2013). Influence of large changes in public transportation (transantiago) on the black carbon pollution near streets. *Atmospheric Environment*, 65:153–163.

Hussain, E., Bhaskar, A., and Chung, E. (2021). Transit OD matrix estimation using smartcard data : Recent developments and future research challenges. *Transportation Research Part C*, 125(August 2020):103044.

Ingvardson, J. B. and Nielsen, O. A. (2018). Effects of new bus and rail rapid transit systems–an international review. *Transport Reviews*, 38(1):96–116.

Kennisinstituut voor Mobilieitsbeleid (2003). De maatschappelijke waarde van kortere en betrouwbaardere reistijden. Technical report, Ministerie van Infrastructuur en Milieu.

Khaddari, R. (2019). Langer onderweg door de Noord/Zuidlijn: 'Ik merk niets van de voordelen'. `https://www.parool.nl/nieuws/langer-onderweg-door-de-noord-zuidlijn-ik-merk-niets-van-de-voordelen{~}bd92b144/`.

Knockaert, J. and Koster, P. R. (2020). Waardes van tijd en overstappen voor amsterdam. Technical report, Vrije Universiteit Amsterdam.

Knowles, R. D. (2012). Transit oriented development in copenhagen, denmark: from the finger plan to ørestad. *Journal of transport geography*, 22:251–261.

Manaugh, K. and El-Geneidy, A. (2012). Who benefits from new transportation infrastructure? using accessibility measures to evaluate social equity in public transport provision. In *Accessibility Analysis and Transport Planning*. Edward Elgar Publishing.

National Academies of Sciences, Engineering, and Medicine (2006). *Smartcard Interoperability Issues for the Transit Industry*. The National Academies Press, Washington, DC.

Nederlandse Spoorwegen (2019). Reizigersgedrag 2019. `https://dashboards.nsjaarverslag.nl/reizigersgedrag/`.

Nguyen, H. A., Soltani, A., and Allan, A. (2018). Adelaide's east end tramline: Effects on modal shift and carbon reduction. *Travel Behaviour and Society*, 11:21–30.

NU.nl (2020). NS vervoerde in 2019 recordaantal passagiers, grootste groei sinds 2008. `https://www.nu.nl/binnenland/6033694/ns-vervoerde-in-2019-recordaantal-passagiers-grootste-groei-sinds-2008.html#:~:text=In%202019%20werden%203%2C7,2%2C2%20procent%20per%20jaar.&text=%22We%20willen%20in%20Europa%20minder,meer%20met%20de%20trein%20reizen`.

Ouwehand, R. M. and Duivis, J. (2016). Zaanstreek vervoerplan 2018. Technical report, Connexxion.

Pelletier, M. P., Trépanier, M., and Morency, C. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, 19(4):557–568.

Reinhold, T. (2008). More Passengers and Reduced Costs-The Optimization of the Berlin Public Transport Network. *Journal of Public Transportation*, 11(3).

Schakenbos, R., La Paix, L., Nijenstein, S., and Geurs, K. T. (2016). Valuation of a transfer in a multimodal public transport trip. *Transport Policy*, 46:72–81.

Stadsregio Amsterdam (2011). Een nieuwe kijk op het Amsterdamse OV. Technical report, Stadsregio Amsterdam.

Stadsregio Amsterdam (2014). OV LIJNENNETVISIE 2018. Technical report, Stadsregio Amsterdam.

Sun, Y., Shi, J., and Schonfeld, P. M. (2016). Identifying passenger flow characteristics and evaluating travel time reliability by visualizing afc data: a case study of shanghai metro. *Public Transport*, 8(3):341–363.

Tiemersma, D. and Govers, B. (2012). Een robuuste openbaar vervoerstructuur in Amsterdam, subsidiëren of investeren? In *Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk 2012*.

Tzika-Kostopoulou, D. and Nathanail, E. (2021). Exploring the Big Data Usage in Transport Modelling. In *Advances in Intelligent Systems and Computing*, volume 1278, pages 1117–1126. Springer Science and Business Media Deutschland GmbH.

Vaillant, D. (2019). De Noord/Zuidlijn in 50 feiten en cijfers. `https://wijnemenjemee.nl/divers/nieuws/50-feitjes/`.

Van Dijk, W., Spit, W., Gille, J., Kok, R., van Bussel, B., Weck, J., and Albers, N. (2009). Review van de baten van de Noord/Zuidlijn. Technical report, ECORYS Nederland BV, INFRAM BV, Rotterdam, Marknesse.

Van Oort, N., Brands, T., and De Romph, E. (2015). Short-term prediction of ridership on public transport with smart card data. *Transportation Research Record*, 2535:105–111.

Van Oort, N., van der Bijl, R., and Verhoof, F. (2017). The wider benefits of high quality public transport for cities. In *European Transport Conference, Barcelona*.

Van Wee, B. and Tavasszy, L. A. (2008). Ex-ante evaluation of mega-projects: methodological issues and cost-benefit analysis. *Decision-making on mega-projects*, pages 40–65.

Venter, C., Jennings, G., Hidalgo, D., and Valderrama Pineda, A. F. (2018). The equity impacts of bus rapid transit: A review of the evidence and implications for sustainable transport. *International Journal of Sustainable Transportation*, 12(2):140–152.

Vervoerregio Amsterdam (2020). Evaluatie doelstellingen lijnennetvisie 2018. Technical report, Vervoerregio Amsterdam.

Vervoerregio Amsterdam (2021). OV-dashboard. `https://vervoerregio.nl/ov-dashboard`.

Vuk, G. (2005). Transport impacts of the Copenhagen Metro. *Journal of Transport Geography*, 13(3):223–233.

Weckström, C., Kujala, R., Mladenović, M. N., and Saramäki, J. (2019). Assessment of large-scale transitions in public transport networks using open timetable data: case of Helsinki metro extension. *Journal of Transport Geography*, 79:102470.

Welch, T. F. and Widita, A. (2019). Big data in public transportation: a review of sources and methods. *Transport Reviews*, 39(6):795–818.

Witteveen, L. (2018). Waarom de nieuwe Noord/Zuidlijn onze hoofdstad gaat redden. `https://www.nporadio1.nl/achtergrond/10874-waarom-de-nieuwe-noord-zuidlijn-onze-hoofdstad-gaat-redden`.

Yoh, A. C., Iseki, H., Taylor, B. D., and King, D. A. (2006). Interoperable transit smart card systems: Are we moving too slowly or too quickly? *Transportation Research Record*, 1986(1):69–77.

Zannat, K. E. and Choudhury, C. F. (2019). Emerging big data sources for public transport planning: A systematic review on current state of art and future research directions. *Journal of the Indian Institute of Science*, 99(4):601–619.

A

Scientific paper

Regional Travel Time and Transfer Impacts of the Noord/Zuidlijn using Interoperable Smart Card Data

S.M.H. van Hees

July, 2021

Abstract

On July 22nd 2018 the Noord/Zuidlijn, a metro line crossing the city center of Amsterdam, became operative. This entailed a transformation from a direct bus and tram network to a network consisting of a metro trunk line and bus and tram feeder lines. This study explores and uses interoperable smart card data from before and after the Noord/Zuidlijn became operative to measure the regional travel time and transfer impacts of this network overhaul. On a working day on average 1,350 hours of travel time are saved. The travel time savings and losses are 2,350 and 1,000 hours per day respectively. Additionally, 2,500 extra transfers are made per working day. The transfers gains are 4,500 transfers and the transfer losses are 7,000 transfers per working day. 20% of the travelers experiences a decrease in travel time of more than 1 minute, 10% experiences an increase in travel time of more than 1 minute. Furthermore, interoperable smart card data showed to be promising as it captures travel behavior across multiple operators. However, the fact that the number of travelers is given in bins complicates analysis. This study develops a methodology to work with interoperable smart card data. Additionally, as only few studies have evaluated the transportation impacts of a large-scale public transport network overhaul ex-post, the findings of this research could improve public transport planning and assessment.

Keywords: *Ex-post evaluation, public transport network overhaul, interoperable smart card data, metro, travel time, transfer, benefits*

Introduction

Compared to ex-ante evaluations, literature on ex-post evaluations of large-scale public transport networks overhauls is rare [1]. Considering the optimism bias often prevalent in ex-ante evaluations [2], ex-post evaluations could offer useful insights in the actual effects of large-scale public transport network overhauls. This is valuable for future public transport planning and assessment thereof.

Especially, literature on the transportation impacts of a large-scale public transport network overhaul on network level is limited. In Nanjing, China, the opening of a new metro line led to an increase in ridership, a decrease in travel times and a decrease in the reliability of travel times [3]. The improvement of the late-night bus network from a direct to a trunk-feeder network in São Paulo, Brazil led to an increase in ridership. Additionally, travelers undertake activities further away after the network improvement resulting in longer travel distances and times and an increase in the number of transfers [4]. In Helsinki, Finland, a direct bus network was transformed to a metro trunk line with feeder buses. Using timetable data it was found that this overhaul has different effects on the traveler when investigated from an aggregate or disaggregate viewpoint. In areas, relying on feeder buses the number of transfers increases while in areas close to metro stations the service level improves [5].

A study by Brands et al. [6] deserves special attention. It investigates the transportation impacts of the Noord/Zuidlijn on urban level. This led to an increase in ridership, more than 6,000 hours of travel time saved and a decrease in number of transfers made on an average working day. However, the effects are not distributed evenly over all travelers [6].

This study is a follow-up on the research performed by Brands et al. [6]. A limitation of that study is that it only incorporated smart card data from the urban public transport operator while regional operators serve the area as well. Therefore, not the whole journey by regional travelers has been investigated. Besides, a shift of travelers between operators could not be captured although it is likely that this effect occurred. The Noord/Zuidlijn network overhaul namely affected the travel times and route choices for the regional traveler as well. This study investigates the travel time and transfer impacts of the Noord/Zuidlijn on the regional traveler.

Because the regional traveler uses public transport services offered by multiple operators, interoperable smart card data is necessary to analyze their travel behavior. This namely gives a comprehensive overview of public transport travel. However, this data source is quite unique. Institutional issues hinder the implementation of interoperable smart card systems [7]. Besides, after implementation data availability is another concern. Operators that own the data are not willing to share their confidential company information with competitors [8]. Due to this lack of availability, literature for which interoperable smart card data is used, is limited. Therefore, the strengths and weaknesses of this promising data source are unknown.

The lack of research on the transportation impacts of large-scale network overhauls and the lack of research on interoperable smart card data, result in a societal and a methodological objective for this paper. The societal objective is to assess the impact of the Noord/Zuidlijn on regional travel times and transfer patterns. The methodological objective is to explore, understand and overcome the strengths and weaknesses of interoperable smart card data.

In the remaining of this paper first the public transport network overhaul of the Noord/Zuidlijn is presented. Thereafter, the data sources and methods are described. Then the results follow. After which the paper is concluded and the main limitations and recommendations for further research are discussed.

Case study Amsterdam

The Vervoerregio Amsterdam (VRA) is responsible for most public transport in the region of Amsterdam. It manages four different concession areas: Amsterdam, Amstelland-Meerlanden, Waterland and Zaanstreek. All concession areas are served by train and buses. In addition to that, the concession area Amsterdam has a tram and metro network as well. These tram and metro lines partially run through Amstelland-Meerlanden as well. An overview of the network after the Noord/Zuidlijn is given in Figure 1.

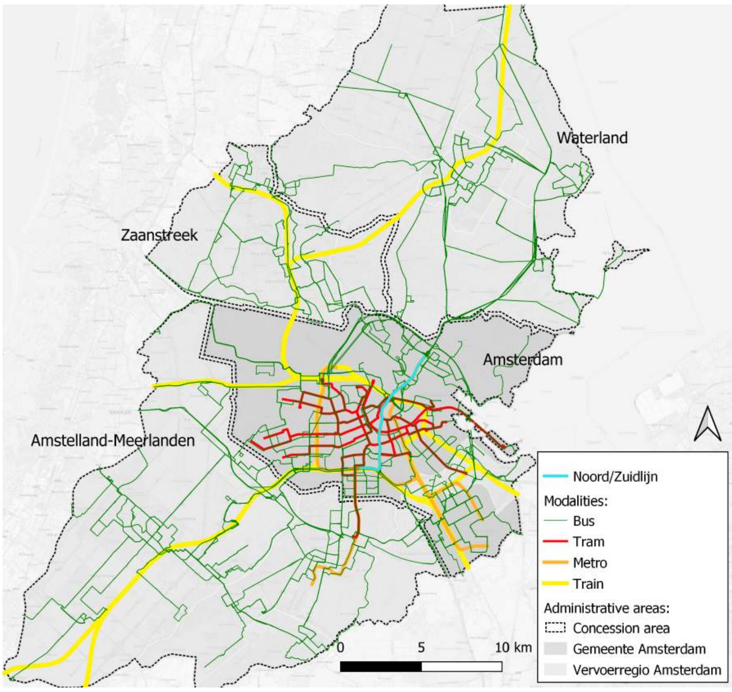
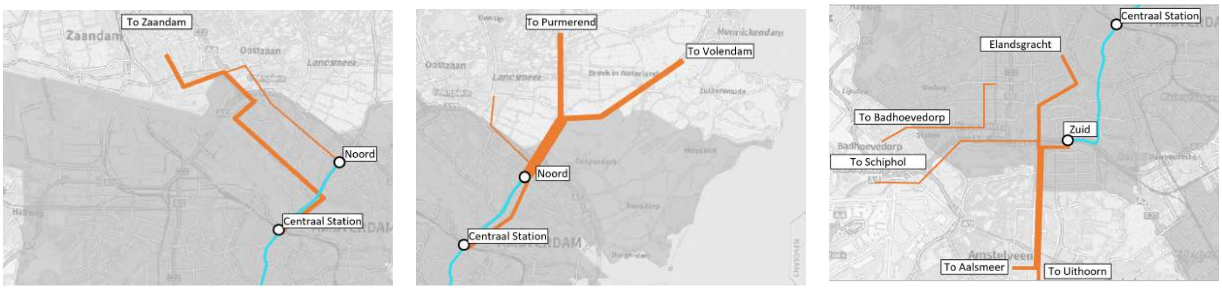


Figure 1: Public transport network in the VRA after the Noord/Zuidlijn

On July 22^{nd}, 2018 Amsterdam's public transport network changed radically. From that day, the Noord/Zuidlijn offered a fast and reliable connection between the northern and southern part of the city. Existing bus and tram lines were rerouted to feed to the new metro line or removed entirely [9]. Regarding the regional concession areas, this meant that part of the regional buses that previously ran to Amsterdam Centraal now connected to the Noord/Zuidlijn. The conceptual changes are visualized in Figure 2.



(a) Zaanstreek (b) Waterland (c) Amstelland-Meerlanden

Figure 2: The proposed configuration of the public transport networks for each concession area after the Noord/Zuidlijn.

Methodology

In this section the applied methodology is described. It starts with the data sources used and the contents thereof. Next, the processing steps are set out.

Data sources

In total four main data sets have been used for this research. Two types of data sets have been utilized. One that contains information on the travel times between all stop clusters per half hour block in a day and one that contains information on the routes traveled (including up to five trips) between the stop clusters. To measure the impact of the Noord/Zuidlijn a before and after Noord/Zuidlijn data set is used of the travel time and route data sets. The contents of both data sets are:

Contents travel time data

- Type of day
- Check in stop cluster
- Check out stop cluster
- Half-hour block of check in time
- Number of journeys
- Median travel time [s]
- Average travel time [s]

Contents route data

- Check in stop cluster
- Check out stop cluster
- Day and period on the day
- Trip x* check in stop cluster
- Trip x* check out stop cluster
- Trip x* modality
- Trip x* operator
- Number of journeys on route

* The data contains up to five trips per route. For each trip the attributes are given.

Because privacy regulations prohibit tracking of an individual's travel behavior, the number of journeys is grouped into classes of 50 for values lower than 300. Only above 300 the exact number of travelers is known.

Before the data was delivered, individual transaction were converted to journeys by Translink Sytems (TLS), the company behind the Dutch smart card system. Besides, the journeys were aggregated over time and space based on the inquiry by the VRA. Aggregation took place during different time periods before and after the Noord/Zuidlijn to exclude the effects of converting a metro line to a tram line from the analysis. The before data set covers the period from the 16^{th} of June 2018 till the 21^{st} of July 2018. The other data set runs from the 8th of December 2018 till the 2nd of March 2019. Both periods only contain working days and exclude school vacations.

Data processing

Data processing can roughly be divided into three phases. The first phase consists of estimation and validation of an average value for each range. In the second phase the average travel time or number of transfers per OD-relation is calculated. This is done for the data before and after the Noord/Zuidlijn separately. In the third phase the average travel time and number of transfers before and after the Noord/Zuidlijn is compared for each OD-relation.

The limitation of the binned values has been overcome by estimating an average value for each bin. The average value has been validated using a dashboard with trip totals per concession from the VRA [10]. The estimates reported in Table 1 gave the best validation results. For the travel time data this resulted in a 0% and -5% deviation from the number of travelers reported in the OV-dashboard [10] for the before and after data set respectively. The estimates for the route data result in a +7.5% and +6% overestimation of the number of journeys for the before and after data respectively.

Table 1: Estimated average values for the ranges in both data sets

Range [travelers]	**Travel time data [travelers]**	**Route data [travelers]**
1-50	13*	3
51-100	65	65
101-150	120	120
151-200	175	175
201-250	225	225
251-300	275	275

* If the median travel time [s] = average travel time [s] then the estimate is 1.

In the following processing steps, the estimated average value for each range has then been split in the lowerbound and an additionally estimated number of travelers to keep track of the uncertainty. To illustrate this, the estimated average of 120 travelers for range 101 to 150 consists of a lowerbound number of travelers of 101 and additionally estimated number of travelers of 19.

Thereafter, only regional journeys (e.g. from the region concession areas to Amsterdam and vice versa) are included in the analysis. Of all stop clusters 62 lay in the concession area Amsterdam and 40 in the other concession areas. Considering journeys in both directions, this means that hypothetically there are 4,960 origin-destination (OD) relations in the data.

Next, the outliers have been removed. For the travel time data this entailed 2.1% and 1.8% of the records for the before and after data respectively. 14.9% and 20.9% of the records in the before and after route data have been excluded from the analysis respectively.

Then, the average travel time and number of transfers per OD-relation has been calculated. The formula to calculate the average travel time on an OD-relation is presented in Formula 1. The average number of transfers per OD-relation has been calculated using Formula 2.

$$\overline{tt}_{o,d,p,y} = \frac{\sum_{h \in p} \overline{tt}_{o,d,h,y} * np_{o,d,h,y}}{\sum_{h \in p} np_{o,d,h,y}} \tag{1}$$

$$\overline{nt}_{o,d,y} = \frac{\sum_{r \in R_{o,d,y}} nt_{o,d,r,y} * np_{o,d,r,y}}{\sum_{r \in R_{o,d,y}} np_{o,d,r,y}} \tag{2}$$

in which $\overline{tt}$ is the average travel time, np is the number of travelers, o is the origin, d is the destination, h is a half hour block, p is the time period, y stands for the year (i.e. before or after Noord/Zuidlijn), nt is the number of transfers, r is a route and R is a set of routes.

However, as stated before the representativeness of records that fall in the lowest bin is uncertain. To account for this, for each OD-relation a representativeness criterion has been added that indicates the lowerbound number of travelers that the average travel time or number of transfers on an OD-relation is based on. For the travel time data four different representativeness categories have been distinguished: 1-3, 4-11, 12-50 and >51. For the route data only two: 1-50 and >51. At least one half hour block or route traveled by at least 51 travelers is assumed to lead to a representative average travel time or number of transfers.

Results

This section presents the results. They are divided in travel time impacts and transfer impacts.

Travel time impacts

Table 2 presents the travel time benefits per representativeness criterion and distinguishes between the lower bound and estimated benefits. From Table 2 it becomes clear that the travel time benefits are around 1,350 hours per day. This is equal to 0.84 minutes or 50 seconds per average traveler. Using a value of time of 18 €/hour [11] the daily travel time benefits are 24,000 euros.

The largest part of the total benefits comes from the most representative OD-relations. Additionally, in Table 2 it can be seen that almost half of the travel time benefits are based on estimated travelers.

Table 2: Travel time benefits over all OD-relations over the whole day distinguished in different representativeness categories and split in the lower bound and estimated benefits

Representativeness criterion	**Total travel time benefits [h] (lower bound | estimated)**	**Average number of travelers per day [travelers/day]**	**Average travel time benefits per traveler [min]**
1-3	6 (3 | 3)	100	4.4
4-11	18 (9 | 10)	500	2.1
12-50	220 (94 | 126)	6,500	2.02
≥51	1,105 (605 | 501)	88,800	0.75
Total	**1,349** **(710 | 639)**	**95,600**	**0.84**

In Table 3 it can be seen that per day around 2,350 hours of travel time are saved while 1,000 hours are lost. The travel time savings are thus more than two times as large as the travel time losses. Additionally, with a total of more than 1,000 hours of travel time losses the negative impact of the Noord/Zuidlijn on regional travel times is substantial.

Table 3: Travel time savings and losses

Representativeness criterion	**Net travel time benefits [h]**	**Travel time savings [h]**	**Travel time losses [h]**
1-3	6	12	6
4-11	18	39	21
12-50	220	353	133
≥51	1,105	1,953	848
Total	**1,349**	**2,357**	**1,008**

In Figure 3 it can be seen that for around 50% of the OD-relations the travel time decreases. For around 27.5% of the OD-relations the travel time is minimally affected by the Noord/Zuidlijn and around 22.5% of the OD-relations experiences an increase in travel time. From Figure 4 it becomes clear that 60,000 passengers do not experience a significant change in travel time. This is around 67% of the total number of passengers on representative OD-relations. For around 10% of the travelers the travel time increases and for around 20% of the travelers the travel time improves. The high peak and short and thin tails of Figure 4 compared to Figure 3 indicate that most people travel on OD-relations where no or minor changes occur.

Another interesting observation in Figure 4 and Figure 3 is that the tails in the positive direction are longer than in the negative direction. This means that the the positive changes in travel time are of a larger magnitude than the negative changes.

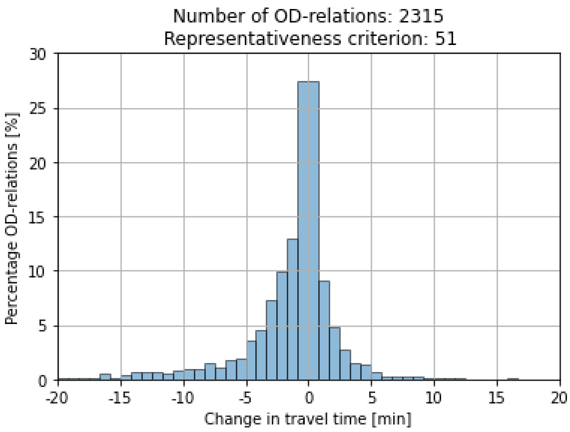


Figure 3: Distribution of the absolute change in travel time against the percentage of OD-relations. Results are plotted for the average travel time over the whole day.

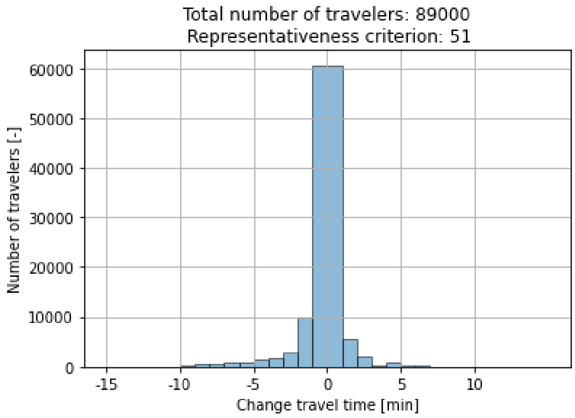


Figure 4: Distribution of the absolute change in travel time against the number of travelers. Results are plotted for the average travel time over the whole day.

Figure 12 and Figure 13 show that the positive travel time impacts are larger than the negative travel time impacts. Besides, it can be seen that the largest travel time impacts occur on OD-relations where a large part of the journey can be made on the Noord/Zuidlijn. OD-relations where a small part of the journey can be made using the Noord/Zuidlijn are not visible in both figures. This is explained by the fact that the metro leg should be sufficiently long to make up for the time lost by the additional transfer. The largest negative impacts occur on OD-relations that lose a direct connection to the city center. These can be explained by the disappearance of direct lines from Waterland to Amsterdam Centraal and buses from Amstelland-Meerlanden ending at bus station Elandsgracht instead of Amsterdam Centraal.

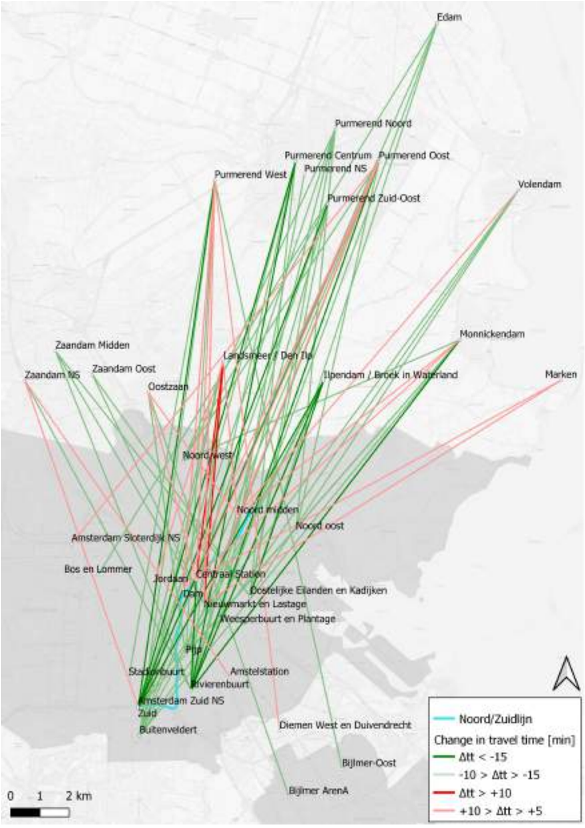


Figure 5: The OD-relations between the region north of Amsterdam and Amsterdam where the travel time changes most.*

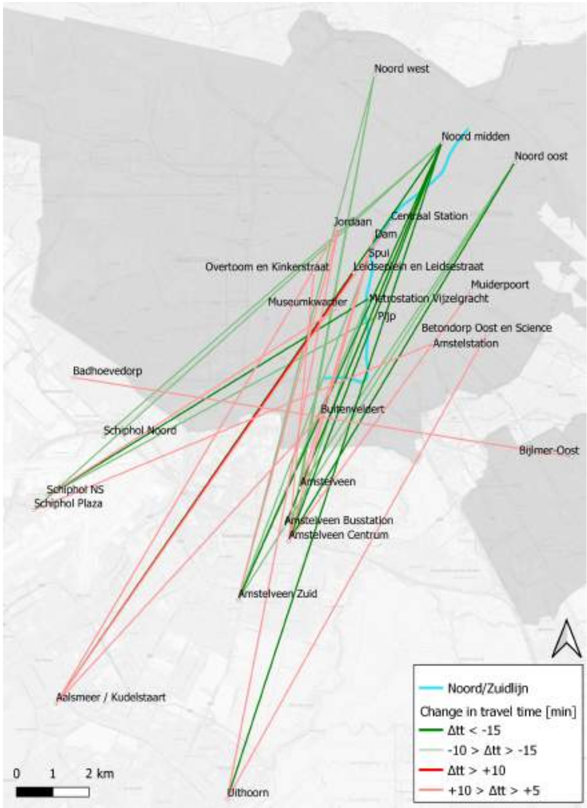


Figure 6: The OD-relations between the region south of Amsterdam and Amsterdam where the travel time changes most.*

*Note that the range for the positive impacts starts at -10 minutes and the range for the negative impacts start at +5 minutes.

Transfer impacts

Table 4 shows that after the Noord/Zuidlijn 2,500 extra transfers are made per day. This is equal to 0.027 transfers per journey. An extra transfer costs €1.4 [11]. This results in societal costs of 3,500 euros per day as a result of the transfer impacts of the Noord/Zuidlijn.

Table 4: Transfer benefits over all OD-relations over the whole day distinguished in different representativeness categories

Representativeness criterion	Total extra transfers [-] (lower bound | estimated)	Average number of travelers per day [travelers]	Average transfer benefits per traveler [transfers/traveler]
1-50	-433 (-391 | -42)	5,200	-0.083
≥51	2,893 (2,644 | 249)	86,200	0.034
Total	**2,460 (2,253 | 208)**	**91,400**	**0.027**

After the Noord/Zuidlijn the number of transfer per journey increased by 22% from 0.54 transfers per journey to 0.66 transfers per journey. In 7 and 8 it is visible that the share of direct journeys decreases at the expense of journeys with multiple trip legs.

Figure 7: The share of journeys with zero, one or two or more transfers before the Noord/Zuidlijn.

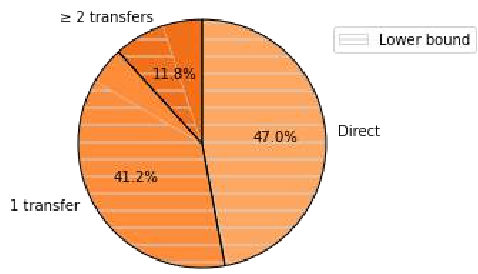


Figure 8: The share of journeys with zero, one or two or more transfers after the Noord/Zuidlijn.

From Figure 9 it becomes clear that the share of journeys made (partially) by metro grows. Especially, the share of journeys by metro and bus increases. The decrease in unimodal journeys, except for metro can be attributed to the decrease of direct connections to Amsterdam Centraal. An explanation for the decrease of unimodal train journeys could be that travelers with a destination close to Amsterdam Centraal shifted from going to Amsterdam Centraal by train and walking the last part to taking a train to Amsterdam Zuid and transferring to the metro. This is supported by the increase of journeys by metro and train.

Looking at the bimodal journeys, the tram and metro combination loses share. This combination is only possible from or to Amstelland-Meerlanden. For a large part of that area a journey by bus and metro or metro only is preferred. The bus and tram combination loses share as well. This is probably the result of a decrease in transfer possibilities between bus and tram at Amsterdam Centraal. The share of journeys made using three or more modalities remains relatively small, but increases slightly. Again, the combinations including metro show an increase.

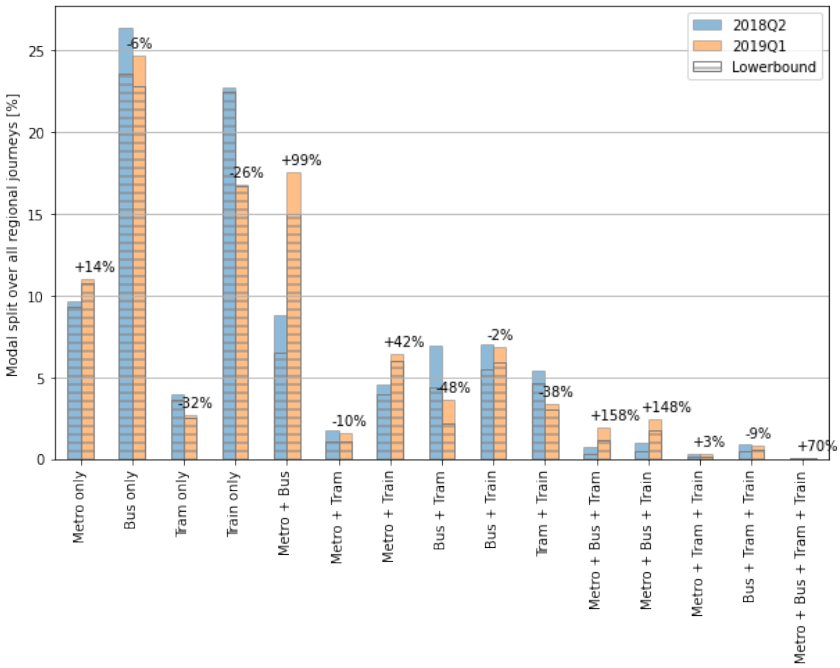


Figure 9: The share of journeys with a certain modal split over all regional journeys before (blue) and after (orange) the Noord/Zuidlijn and the relative change thereof. The columns with one modality contain unimodal journeys with a transfer as well.

Table 5 shows that the total transfer losses are 7,000 transfers. The transfer savings are 4,500 transfers. The decrease in number of transfers can be explained by a seemingly direct connection between the Amstelveenlijn and the Noord/Zuidlijn.

Table 5: Transfer benefits over all OD-relations over the whole day split into transfer gains and losses for different representativeness criteria

Representativeness criterion	**Total extra transfers [transfers]**	**Transfer gains [transfers]**	**Transfer losses [transfers]**
1-50	-455	1191	736
≥51	2,966	3,342	6,308
Total	**2,511**	**4,533**	**7,044**

Figure 10 shows that on 80% of the OD-relations with at least 51 travelers the number of transfers does not increase. 15% of these OD-relations experience an increase in number of transfers and for 5% of the OD-relations the number of transfers decreases. Figure 11 visualizes that the largest part of the regional travelers experiences no or a minimal change in the number of transfers. The number of travelers that experiences a negative impact is slightly larger than the number of travelers that experiences a positive impact. Comparing Figure 10 and Figure 11 shows that the percentage of OD-relations for which the number of transfers increases is larger than the percentage of travelers for which the number of transfers increases. The same holds for a decrease in the number of transfers. This suggests that the impacts are larger on OD-relations that are less frequently traveled.

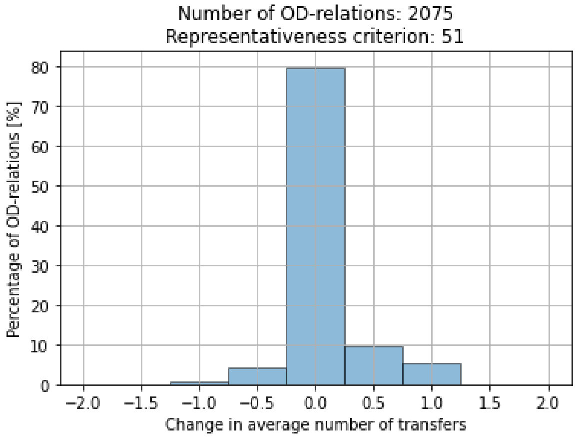


Figure 10: The distribution of the transfer impacts over OD-relations.

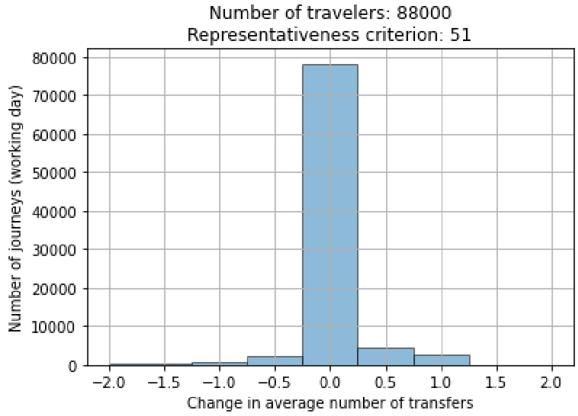


Figure 11: The distribution of transfer impacts over travelers.

From Figure 12 and Figure 13 it becomes clear that the number of OD-relations that experiences an increase in the number of transfers is much larger than the number of OD-relations for which the number of transfers decreases. For the region north of Amsterdam the negative impacts can be ascribed to the loss of direct connections to Amsterdam Centraal. This results in an increase in transfers on OD-relations from and to locations further south than Amsterdam Centraal as well. From Amstelland-Meerlanden the negative impacts concentrate around stop clusters Jordaan and Dam. Between Elandsgracht and Amsterdam Centraal the direct bus connections have disappeared. The positive impacts from Waterland and Zaanstreek are caused by the new bus stop IJtunnel. From the region south of Amsterdam the positive transfer impacts can be explained an intramodal metro transfer that is not visible in smart card data.

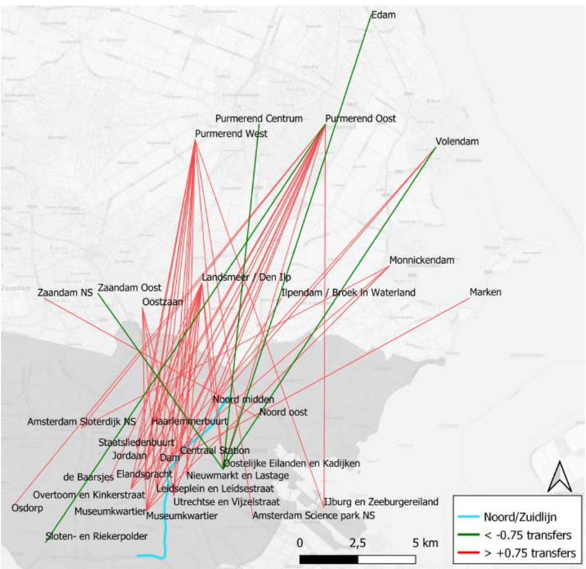


Figure 12: The OD-relations between the region north of Amsterdam and Amsterdam where the number of transfers changes by at least 0.75 transfers .

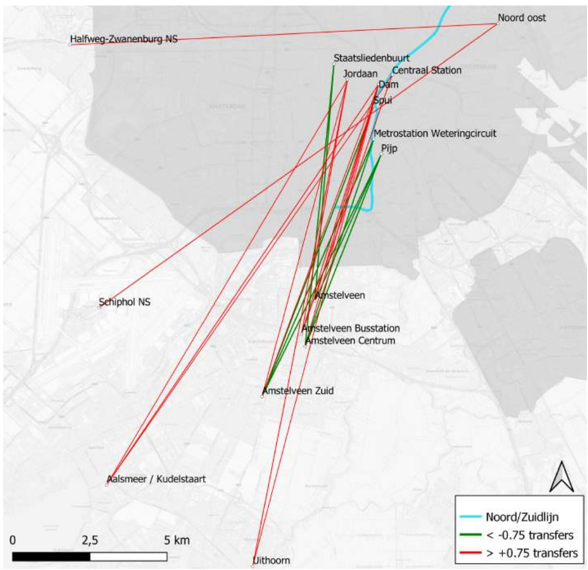


Figure 13: The OD-relations between the region south of Amsterdam and Amsterdam where the number of transfers changes by at least 0.75 transfers .

Conclusion

This study analyzed the regional travel time and transfer impacts of a large-scale change in Amsterdam's public transport network. Interoperable smart card data was necessary to capture regional travel behavior. First the main conclusions regarding the methodological objective are presented. Thereafter the conclusions on the societal objective follow.

The main strength of this relatively new and promising data source is that it provides insights in the whole public transport journey across multiple operators. This leads to improved understanding of route and mode choice and the exchange and distribution of travelers between operators in a multimodal, multi-operator public transport network.

However, the most important weakness of interoperable smart card data is that the number of travelers is given as a range below a certain threshold value. The reason behind this is that the data contains privacy and commercially sensitive information. This results in three main limitations. First, the representativeness of records with a low number of travelers remains unknown. Second, it is not known to how many travelers the results of the analysis apply. Third, it is not possible to distinguish between the weights assigned to records with the same number of travelers.

The impact of these weaknesses can be reduced on forehand and while analyzing the data. On forehand the bin size and the level of detail over space and time should be adjusted to decrease the number of records that end up in a bin, especially the lowest. When analyzing the data an average value for each bin should be estimated and validated using an external data source. Next, that average value should be split in a lower bound and estimated number of travelers. After that, the results should be divided in different degrees of certainty based on the lower bound number of travelers that the result is based on.

The regional travel time and transfer benefits of the Noord/Zuidlijn are presented in Table 6. On average a traveler gains 50 seconds of travel time but has to make 0.03 transfers extra per working day.

Table 6: The regional societal travel time and transfer benefits of the Noord/Zuidlijn per working day.

	Quantity	**Monetary value per unit**	**Societal benefits per day**
Travel time savings	1,350 hours	18 €/hour	€24,000
Transfer savings	-2,500 transfers	1.4 €/transfer	€-3,500
Total	-	-	**€20,500**

This study confirms findings by Weckström et al. [5] and Brands et al. [12] that the impacts of a large-scale public transport network overhaul are unevenly distributed. Additionally, the impacts of the overhaul are differently distributed over OD-relations and travelers. The largest part of the travelers does not experience large changes in travel times or number of transfers. However, focusing on OD-relations larger impacts are observed.

The largest travel time savings occur on journeys where a large part of the journey can be made with the Noord/Zuidlijn. Travel time losses occur on journeys that lose a direct connection and where the trip on the faster metro leg is not long enough to make up for the time lost by the additional transfer. Negative transfer impacts are ascribed to the loss of direct connections.

Limitations

Smart card data has some general weaknesses that limit the research. First, it does not contain all trips made. Second, it does not include access and egress time. Third, the measured travel time components differ among modes because card validation is done in the vehicle or at the station. Fourth, no intramodal transfers are visible for metro and train.

Another important limitation of this study is that number of days and period of the year during which the before and after data were aggregated differ. Besides, the period after the Noord/Zuidlijn is quite close to the Noord/Zuidlijn. These limitations regarding data aggregation have likely affected public transport supply and demand.

Furthermore, the fact that an average value has been estimated and validated for each range and these averages were then weighed the same is a simplification of reality which leads to noisy results.

Last, external factors that have affected public transport supply and demand like time table changes, weather impacts or societal trends could not be excluded from the research.

Recommendations

For further research on the transportation impacts of a large-scale public transport network overhaul it is recommended to investigate the impact of the overhaul on access and egress location choice. This choice affects access and egress time which have not been included in this research. Furthermore, the perception of the traveler should be investigated. The perceived impacts could deviate from the measured impacts which affects usage of the public transport network and the public opinion on the overhaul. Last, from an equity perspective it is interesting to investigate whether the overhaul affects different socio-economic groups differently. This could be done by incorporating passenger characteristics in the assessment of the overhaul. Next, different valuations should be used to distinguish between travel time and transfer types. Next, further research could focus on the effect of the overhaul on the reliability of the travel time. This is especially interesting because the overhaul led to an increase in transfers per journey and unreliable travel times could result in missing transfers.

Further research on interoperable smart card data should focus on improving the methodology to overcome the weaknesses of interoperable smart card data applied in this paper. A study that compares the use of interoperable and single operator smart card data to achieve the same goal could provide interesting insights. Besides, possibilities to fuse route planner data and interoperable smart card data to value the likely routes and travel times higher should be investigated. Besides, additional research can be done on decreasing the number of records that ends up in a bin. Privacy experts should be consulted whether the bin sizes can be reduced or the threshold value can be lowered.

Regarding public transport planning it is advised to incorporate the distribution of the negative and positive effects of a network overhaul in the assessment of the overhaul. Next, a different distribution of the effects on OD-relations and travelers should be accounted for. Additionally, when transforming a direct public transport network to a trunk-feeder network the length of the trip leg on the trunk line should compensate the time lost by the additional transfer required.

References

[1] Morten Skou Nicolaisen and Patrick A. Driscoll. An International Review of Ex-Post Project Evaluation Schemes in the Transport Sector. *Journal of Environmental Assessment Policy and Management*, 18 (1), 2016. ISSN 14643332. doi: 10.1142/S1464333216500083.

[2] Bert Van Wee and Lóránt A Tavasszy. Ex-ante evaluation of mega-projects: methodological issues and cost-benefit analysis. *Decision-making on mega-projects*, pages 40–65, 2008.

[3] Xiao Fu and Yu Gu. Impact of a New Metro Line: Analysis of Metro Passenger Flow and Travel Time Based on Smart Card Data. *Journal of Advanced Transportation*, 2018, 2018. ISSN 20423195. doi: 10.1155/2018/9247102.

[4] Renato Arbex, Claudio Barbieri da Cunha, and Roberto Speicys. Before-and-after evaluation of a bus network improvement using performance indicators from historical smart card data. *Public Transport*, 2019. ISSN 16137159. doi: 10.1007/s12469-019-00214-z. URL `https://doi.org/10.1007/s12469-019-00214-z`.

[5] Christoffer Weckström, Rainer Kujala, Miloš N. Mladenović, and Jari Saramäki. Assessment of large-scale transitions in public transport networks using open timetable data: case of Helsinki metro extension. *Journal of Transport Geography*, 79:102470, jul 2019. ISSN 09666923. doi: 10.1016/j.jtrangeo.2019.102470.

[6] Ties Brands, Malvika Dixit, and Niels van Oort. Impact of a new metro line in amsterdam on ridership, travel times, reliability and societal costs and benefits. *European Journal of Transport and Infrastructure Research*, 20(4):335–353, 2020. ISSN 15677141. doi: 10.18757/ejtir.2020.20.4.4084.

[7] Allison C. Yoh, Hiroyuki Iseki, Brian D. Taylor, and David A. King. Interoperable transit smart card systems: Are we moving too slowly or too quickly? *Transportation Research Record*, 1986(1):69–77, 2006. ISSN 03611981. doi: 10.3141/1986-11.

[8] Niels Van Oort, Ties Brands, and Erik De Romph. Short-term prediction of ridership on public transport with smart card data. *Transportation Research Record*, 2535:105–111, 2015. ISSN 03611981. doi: 10.3141/2535-12.

[9] Stadsregio Amsterdam. Een nieuwe kijk op het Amsterdamse OV. Technical report, Stadsregio Amsterdam, 2011. URL `http://bit.ly/1hzfKC0`.

[10] Vervoerregio Amsterdam. OV-dashboard. `https://vervoerregio.nl/ov-dashboard`, 2021.

[11] Jasper Knockaert and Paul R. Koster. Waardes van tijd en overstappen voor amsterdam. Technical report, Vrije Universiteit Amsterdam, 2020.

[12] Ties Brands, Malvika Dixit, and Niels van Oort. Transformation to a trunk and feeder network: effects on passenger flows, travel times and service reliability. *BIVEC/GIBET Transport Research Day 2021*, pages 1–12, 2021.

B

Literature study search strategy

This Appendix describes the search strategy to find relevant literature. It aims to substantiate that there is a limited amount of literature on the topics relevant for this thesis.

To find useful literature for this thesis several keywords have been used in search engines like Google Scholar, Web of Science and Scopus. The terms that have been used in the search included, but are not limited to the ones presented below. Different spellings and abbreviations of the same word have been used. Besides, operators (AND/OR/NOT), wildcards (*), phrase searches (" ") and nesting (()) have been used while searching literature. Backward and forward snowballing have been applied as well. The lists below are grouped per synonym. No studies that use interoperable smart card data similar to the data used in this thesis were found.

To find literature regarding the ex-post evaluation of large-scale public transport networks the search terms in the three lists below have been combined.

- Meta
- Impact
- Ex-post
- Before-and-after
- Pre-and-post
- Posteriori
- Effect

- Analysis
- Evaluation
- Assessment
- Review

- Overhaul
- Construction
- Transition
- Redesign
- Extension
- Restructuring
- Project
- Commissioning

To find literature relating to multi-operator smart card data combinations of the search terms in the following two lists have been used:

- Smart card
- Automated fare collection
- Validation
- Data

- Integrated
- Range
- Aggregated
- Grouped
- Binned
- Clustered
- Multi-operator
- Interoperable

C

Smart card data and its applications

This Appendix gives a general overview of what smart card data is and how it can be used to analyze public transport.

Pelletier et al. (2011) reviewed literature on smart card and its use in public transport. The smart card was invented in 1968 by Dethloff and Grotrupp. These two German inventors developed the idea of a plastic card containing a microchip. In the years that followed several smart card systems were developed further and started entering the market. Depending on the software and hardware on a smart card it could be used for applications such as identification, authorization and payment. Therefore, the smart card is used in many sectors. Examples of these sectors are health care, banking, government and transportation.

The main use of smart cards in public transport is to handle fares. But next to that, the data of these transactions creates huge opportunities for insights in the trips made. Bagchi and White (2005) mention the following opportunities of smart card systems for public transport operators:

- Access to larger volumes of personal travel data
- Be able to link this data to a passenger or a certain smart card
- Access to continuous trip data that covers longer periods of time compared to existing transport data sources
- Knowledge on their most frequent passengers

Pelletier et al. (2011) distinguishes three types of levels for studies using smart card data. First, strategic-level studies can be used for long-term network planning, analysis of customer behavior and for demand forecasting. Second, the studies on the tactical level focus on schedule adjustment and identification of temporal and individual patterns in public transport use. Third, on an operational level smart card data can be used to analyze and calculate the performance of the transit network. Fu and Gu (2018) reviewed existing literature on applications of smart card data. They categorize existing transport studies using smart card data in three areas: passenger flow patterns, travel behavior and public transport network performance. Studies on passenger flow analyze and visualize the spatial and temporal distribution thereof. Besides, based on historical data predictions for future passenger flows are made. Regarding travel behavior, studies focus on transit ridership, origin, destination, mode and route choices, transfer behavior, extraction of individual or collective spatial and temporal travel patterns and the regularity and variety of these travel patterns. The last category distinguished by (Fu and Gu, 2018) is network performance. Performance indicators that one can think of are travel times and service reliability.
Although smart card data can result in very valuable insights. two reasons affect the usage of this data in public transport planning. First, the nature of smart card systems and the data that comes with it raise major privacy concerns. The storage of huge amounts of data on transactions and card holder information makes users vulnerable to identity theft, or to the misuse of information regarding their (travel) behavior (Pelletier et al., 2011). Second, commercial reasons make operators not eager to share their smart card data.

D
Administrative and concession areas in the Amsterdam region

The municipalities in the region of Amsterdam consist of are presented in Figure D.1.

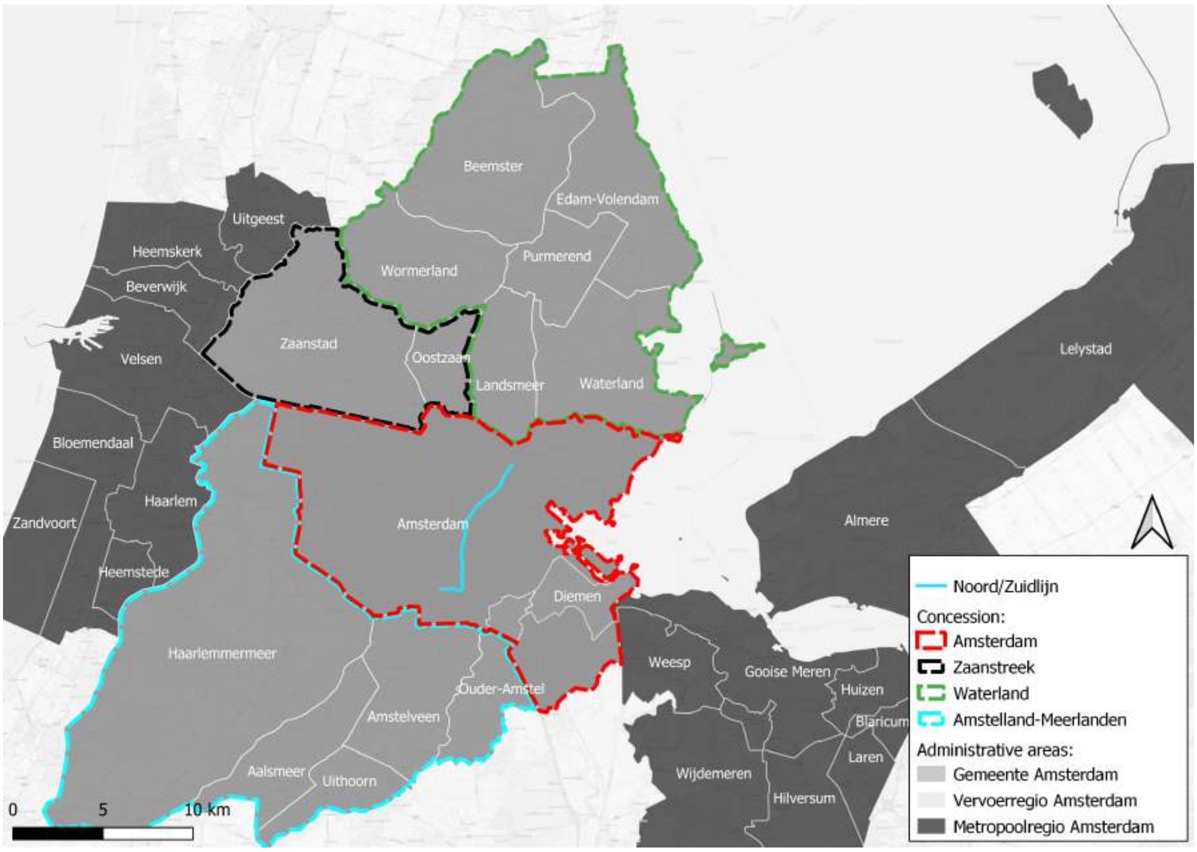


Figure D.1: Administrative and concession areas in the Amsterdam region

E
Changes in public transport network per concession

This Appendix presents the public transport network before and after the Noord/Zuidlijn per concession.

Figure E.1: The northern half of the public transport network in Waterland before (left) (EBS, 2017) and after (right) (EBS, 2018) the Noord/Zuidlijn.

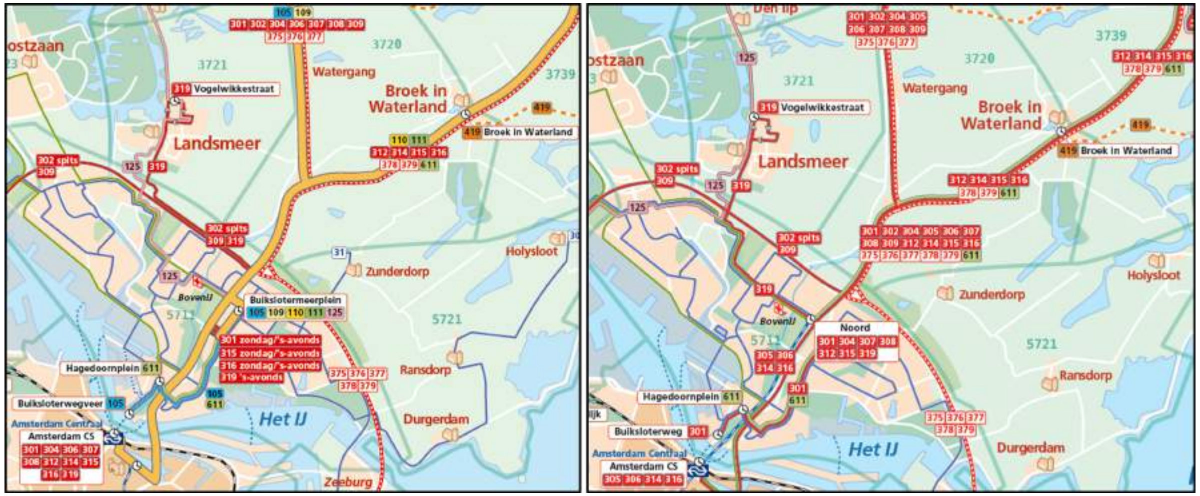


Figure E.2: The southern half of the public transport network in Waterland before (left) (EBS, 2017) and after (right) (EBS, 2018) the Noord/Zuidlijn.

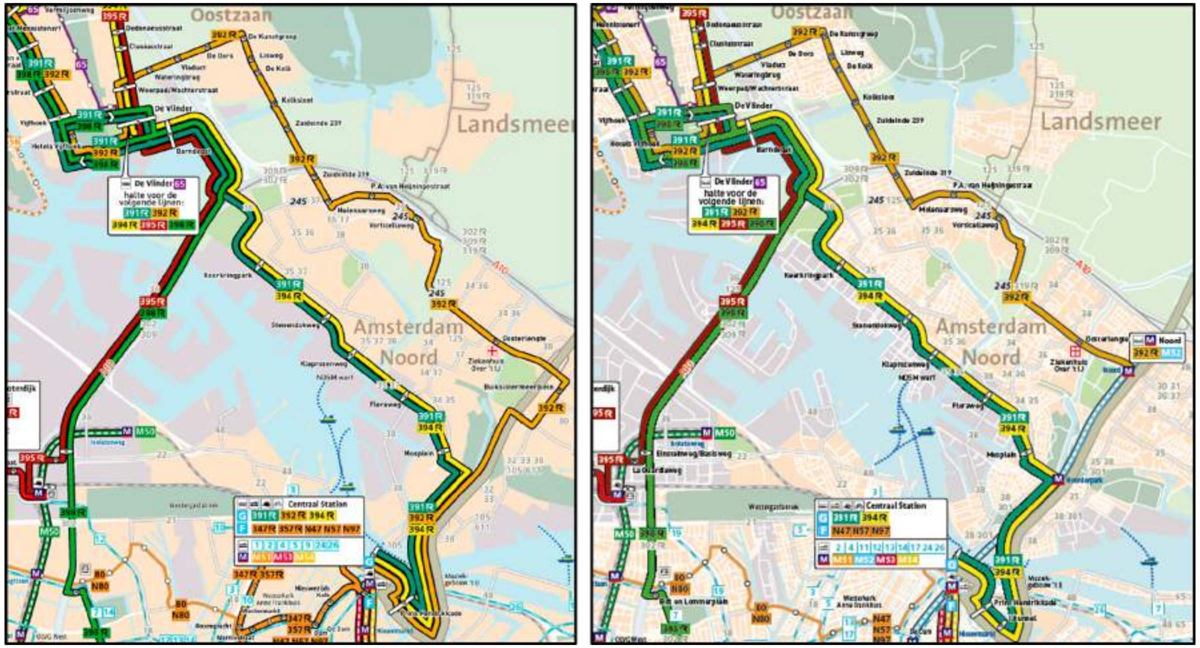


Figure E.3: The public transport network in Zaanstreek before (left) (CXX, 2018c) and after (right) (CXX, 2018d) the Noord/Zuidlijn.

Figure E.4: The public transport network in Amstelland-Meerlanden before (left) (CXX, 2018a) and after (right) (CXX, 2018b) the Noord/Zuidlijn.

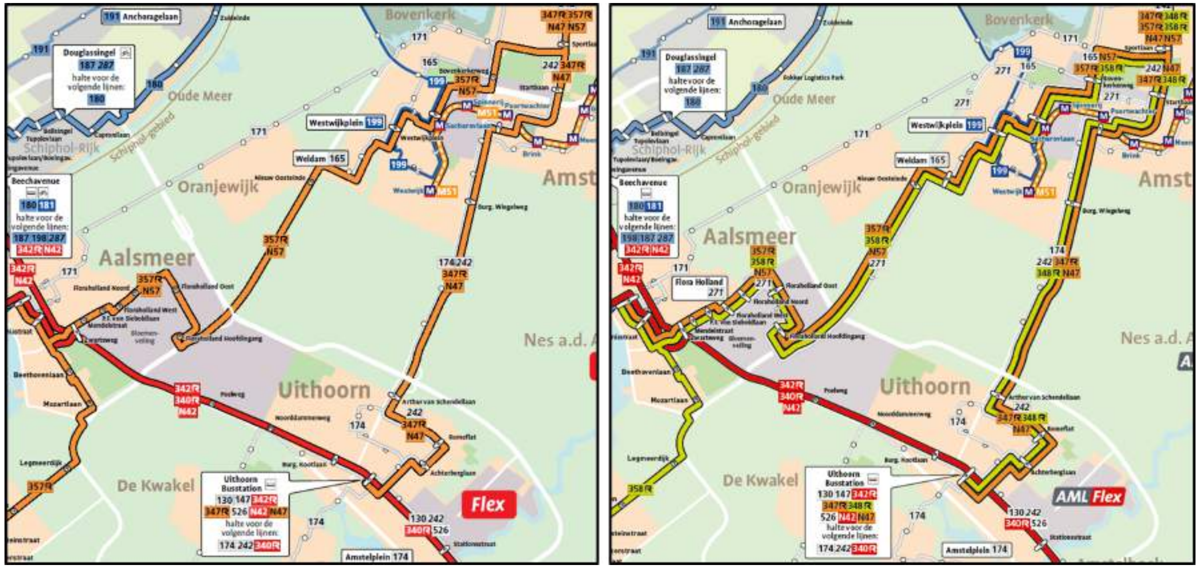


Figure E.5: The public transport network in Amstelland-Meerlanden before (left) (CXX, 2018a) and after (right) (CXX, 2018b) the Noord/Zuidlijn.

F
The Dutch smart card system

This research uses Dutch smart card data. For that reason a brief overview of the Dutch smart card system is presented in this Appendix.

The *OV-Chipkaart*, the smart card used in the Dutch system, became operative nation-wide in 2012. The system consists of several components which have been described by (Van Oort et al., 2015):

- Level O is the smart card itself. This smart card may include personal products, for example to get student discounts.
- Level 1 consists of check-in and check-out devices in-vehicle or at a station. Besides, card vending machines are part of this layer as well. These devices and machines communicate with the smart card using near-field communication technology. Via these machines smart card data is collected.
- Level 2 is a system that collects and temporarily stores data from the devices in level 1. These systems are located at a public transport operator.
- Level 3 is a central system that collects and stores all data from one public transport company. At this level data is prepared to be shared with Trans Link Systems (TLS). TLS is the national agency that collects data from public transport smart cards.
- Level 4 is the database of TLS. From here smart card transactions are verified, the required payments are determined and transaction history is provided to the smart card user.

Van Oort et al. (2015) extensively describe which data is collected by the Dutch smart card system. For all public transport trips the Dutch smart card system records the check-in and check-out station and time. Additionally, trips made with a smart card can be combined into journeys with a unique ID. This also allows for identifying transfers. Because the location of check-in-check-out devices differs among modes, additional information on tram and bus trips can be gathered. The reason for this is that in the train and metro, ticket validation takes place at the station, while for bus and tram the check-in-check-out devices are located in the vehicle. Therefore, the vehicle number or chosen line can be identified for tram and bus trips but not for trips made by metro or train. This enables tracing the whole route through the transport network and thus more is known about the route choice and the transfers made by the passenger. Next to that, smart cards may also collect data on the ticket type which could be used to predict the purpose of the trip.

G
Data cleaning stop clusters

This Appendix describes how the stop cluster data has been cleaned.

Each stop cluster consists of one or multiple public transport stops. In the corresponding reference data set these stops and to which stop cluster they belong can be found. Besides, it contains the coordinates of each stop. However, these tables contain several irregularities or duplicates. To account for these, several data cleaning steps have been performed. These are set out below:

- The BTM-clusters 'Zaandijk/Koog ad Zaan' and 'Zaandam West' have been merged because these have the same number but a different name.
- Several BTM-stops appear multiple times in the reference table. BTM-stops with the same value for the columns 'publicname', 'gebiedsnaam' and 'gebiedsnummer' have been removed. This resulted in the removal of 177 stops. The coordinates and other information of the stop that occured first are preserved.
- Eleven BTM-stops with 'stopplacetype'= 'ferryPort' have been removed.
- Six BTM-stops have been added manually because they belong to line 80 operated by Connexxion but where not included in the data set (Connexxion, 2021).
- Twenty NS stations have been removed because they were not assigned to a stop cluster and shouldn't be either. These were for example stations on the cancellede Hoeksche Lijn or stations that are not served by NS.
- For five NS stations the coordinates have been added manually.

After this the coordinates have been transformed from the WGS84-system to the Dutch Rijksdriekhoeksstelsel.

H
Cluster information

This Appendix provides the name, code and type of each stop cluster in Table H.1. Besides, it entails information on the concession area and administrative area the cluster belongs to and to which analysis area it has been assigned. The clusters are ordered alphabetically by analysis area and cluster name.

Table H.1: Information of all stop clusters in the data sets

Cluster name	**Cluster code**	**Cluster type**	**Concession area**	**Administrative area**	**Analysis area**
Amstelstation	45	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Amsterdam Amstel	asa	NS	Amsterdam	Gemeente Amsterdam	Amsterdam
Amsterdam Bijlmer ArenA	asb	NS	Amsterdam	Gemeente Amsterdam	Amsterdam
Amsterdam Centraal	asd	NS	Amsterdam	Gemeente Amsterdam	Amsterdam
Amsterdam Holendrecht	asdh	NS	Amsterdam	Gemeente Amsterdam	Amsterdam
Amsterdam Lelylaan	asdl	NS	Amsterdam	Gemeente Amsterdam	Amsterdam
Amsterdam Muiderpoort	asdm	NS	Amsterdam	Gemeente Amsterdam	Amsterdam
Amsterdam RAI	rai	NS	Amsterdam	Gemeente Amsterdam	Amsterdam
Amsterdam Science park	assp	NS	Amsterdam	Gemeente Amsterdam	Amsterdam
Amsterdam Sloterdijk	ass	NS	Amsterdam	Gemeente Amsterdam	Amsterdam
Amsterdam Zuid	asdz	NS	Amsterdam	Gemeente Amsterdam	Amsterdam
Betondorp Oost en Science	46	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Bijlmer ArenA	48	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Bijlmer-Oost	33	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Bijlmer-West	34	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Bos en Lommer	14	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam

Table H.1: Information of all stop clusters in the data sets

Cluster name	Cluster code	Cluster type	Concession area	Administrative area	Analysis area
Buitenveldert	44	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Centraal Station	10	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Dam	25	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
de Baarsjes	49	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Diemen	dmn	NS	Amsterdam	VRA	Amsterdam
Diemen Noord Centrum	32	BTM	Amsterdam	VRA	Amsterdam
Diemen West en Duivendrecht	31	BTM	Amsterdam	VRA	Amsterdam
Diemen Zuid	dmnz	NS	Amsterdam	VRA	Amsterdam
Duivendrecht	50	BTM	Amsterdam	VRA	Amsterdam
Duivendrecht	dvd	NS	Amsterdam	VRA	Amsterdam
Elandsgracht	23	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Geuzenveld	7	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Haarlemmerbuurt	13	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Hoofddorppleinbuurt	19	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
IJburg en Zeeburgereiland	35	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Indische Buurt	29	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Industriegebied Amstel	30	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Jordaan	24	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Leidseplein en Leidsestraat	22	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Lelylaan	39	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Metrostation Noord	87	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Metrostation Weteringcircuit	88	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Muiderpoort	47	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Museumkwartier	20	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Nieuwmarkt en Lastage	1	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Noord midden	9	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Noord oost	12	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Noord west	11	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam

Table H.1: Information of all stop clusters in the data sets

Cluster name	**Cluster code**	**Cluster type**	**Concession area**	**Administrative area**	**Analysis area**
Oostelijke Eilanden en Kadijken	36	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Osdorp	6	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Overtoom en Kinkerstraat	16	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Overtoomse Veld	17	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Pijp	21	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
RAI	40	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Rivierenbuurt	43	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Sloten- en Riekerpolder	18	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Sloterdijk	38	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Spui	26	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Staatsliedenbuurt	15	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Stadionbuurt	42	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Utrechtse en Vijzelstraat	27	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Waterland	3	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Weesperbuurt en Plantage	2	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Weesperzijde	28	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Westpoort	8	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Zuid	41	BTM	Amsterdam	Gemeente Amsterdam	Amsterdam
Aalsmeer / Kudelstaart	72	BTM	Amstelland-Meerlanden	VRA	Region
Amstelveen	5	BTM	Amstelland-Meerlanden	VRA	Region
Amstelveen Busstation	86	BTM	Amstelland-Meerlanden	VRA	Region
Amstelveen Centrum	81	BTM	Amstelland-Meerlanden	VRA	Region
Amstelveen Zuid	82	BTM	Amstelland-Meerlanden	VRA	Region
Assendelft / Westzaan	63	BTM	Zaanstreek	VRA	Region
Badhoevedorp	85	BTM	Amstelland-Meerlanden	VRA	Region
Beemster / De Rijp	59	BTM	Waterland	VRA	Region
Edam	54	BTM	Waterland	VRA	Region

Table H.1: Information of all stop clusters in the data sets

Cluster name	Cluster code	Cluster type	Concession area	Administrative area	Analysis area
Halfweg-Zwanenburg	hwzb	NS	Amstelland-Meerlanden	VRA	Region
Hoofddorp	73	BTM	Amstelland-Meerlanden	VRA	Region
Hoofddorp	hfd	NS	Amstelland-Meerlanden	VRA	Region
Ilpendam / Broek in Waterland	52	BTM	Waterland	VRA	Region
Krommenie	64	BTM	Zaanstreek	VRA	Region
Landsmeer / Den Ilp	55	BTM	Waterland	VRA	Region
Marken	0	BTM	Waterland	VRA	Region
Monnickendam	51	BTM	Waterland	VRA	Region
Nieuw-Vennep / Lisse	71	BTM	Amstelland-Meerlanden	VRA	Region
Oostzaan	68	BTM	Zaanstreek	VRA	Region
Ouderkerk ad Amstel	80	BTM	Amstelland-Meerlanden	VRA	Region
Purmerend	pmr	NS	Waterland	VRA	Region
Purmerend Centrum	61	BTM	Waterland	VRA	Region
Purmerend Noord	58	BTM	Waterland	VRA	Region
Purmerend Oost	57	BTM	Waterland	VRA	Region
Purmerend West	60	BTM	Waterland	VRA	Region
Purmurend Zuid-Oost	56	BTM	Waterland	VRA	Region
Rozenburg	77	BTM	Amstelland-Meerlanden	VRA	Region
Schiphol	shl	NS	Amstelland-Meerlanden	Land	Region
Schiphol Noord	78	BTM	Amstelland-Meerlanden	VRA	Region
Schiphol Oost	76	BTM	Amstelland-Meerlanden	VRA	Region
Schiphol Plaza	75	BTM	Amstelland-Meerlanden	VRA	Region
Uithoorn	79	BTM	Amstelland-Meerlanden	VRA	Region
Volendam	53	BTM	Waterland	VRA	Region
Wormer / Wormerveer	65	BTM	Waterland	VRA	Region
Zaandam	zd	NS	Zaanstreek	VRA	Region
Zaandam Midden	69	BTM	Zaanstreek	VRA	Region
Zaandam Oost	70	BTM	Zaanstreek	VRA	Region
Zaandijk / Koog ad Zaan / Zaandam West	66	BTM	Zaanstreek	VRA	Region
Zaanse Schans	67	BTM	Zaanstreek	VRA	Region
Zwanenburg	84	BTM	Amstelland-Meerlanden	VRA	Region

I Geographic representation of the stop clusters

This Appendix contains three figures in which the location of each stop cluster is visualized. The mean of all stops that make up a stop cluster, is calculated to present the stop cluster. For a clear presentation, the stop clusters have been presented in three separate figures. Figure I.1 contains all stop clusters belonging to Amsterdam. Figure I.2 consists of the regional stop clusters south of Amsterdam and of the concession area Amstelland-Meerlanden. The stop clusters north of Amsterdam and in the concession areas Zaanstreek and Waterland are presented in Figure I.3.

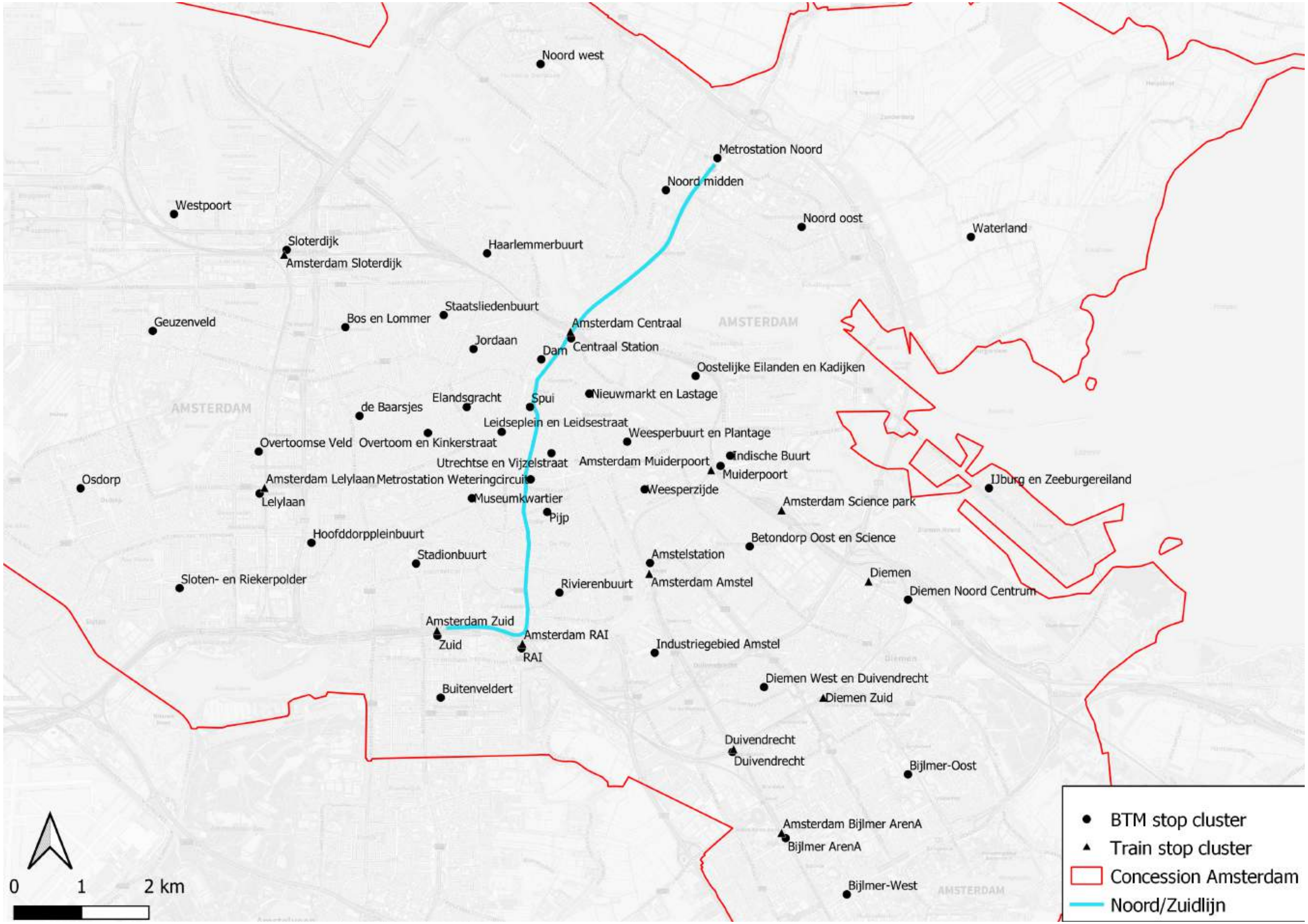


Figure I.1: Amsterdam stop clusters

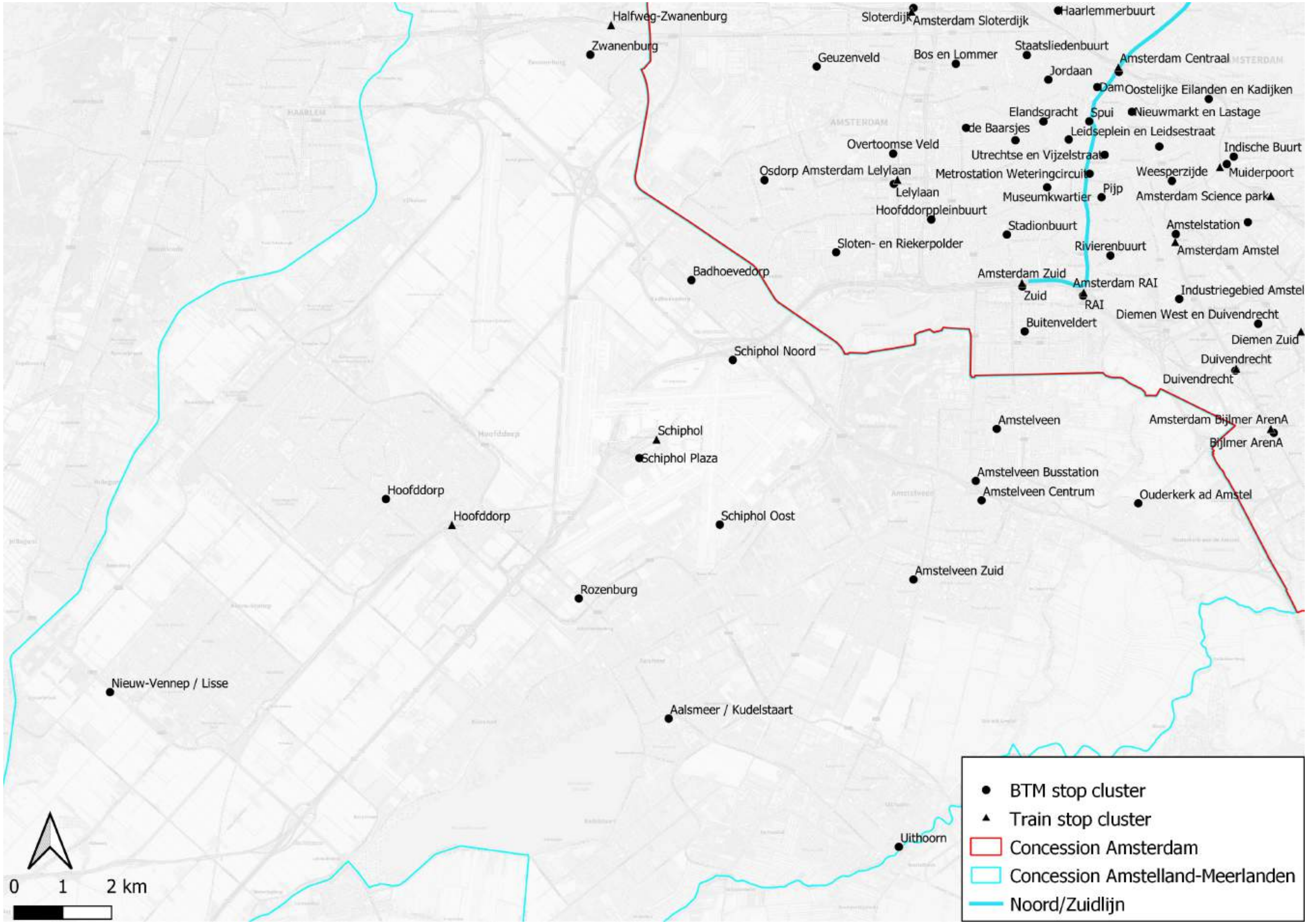


Figure I.2: Regional stop clusters south of Amsterdam

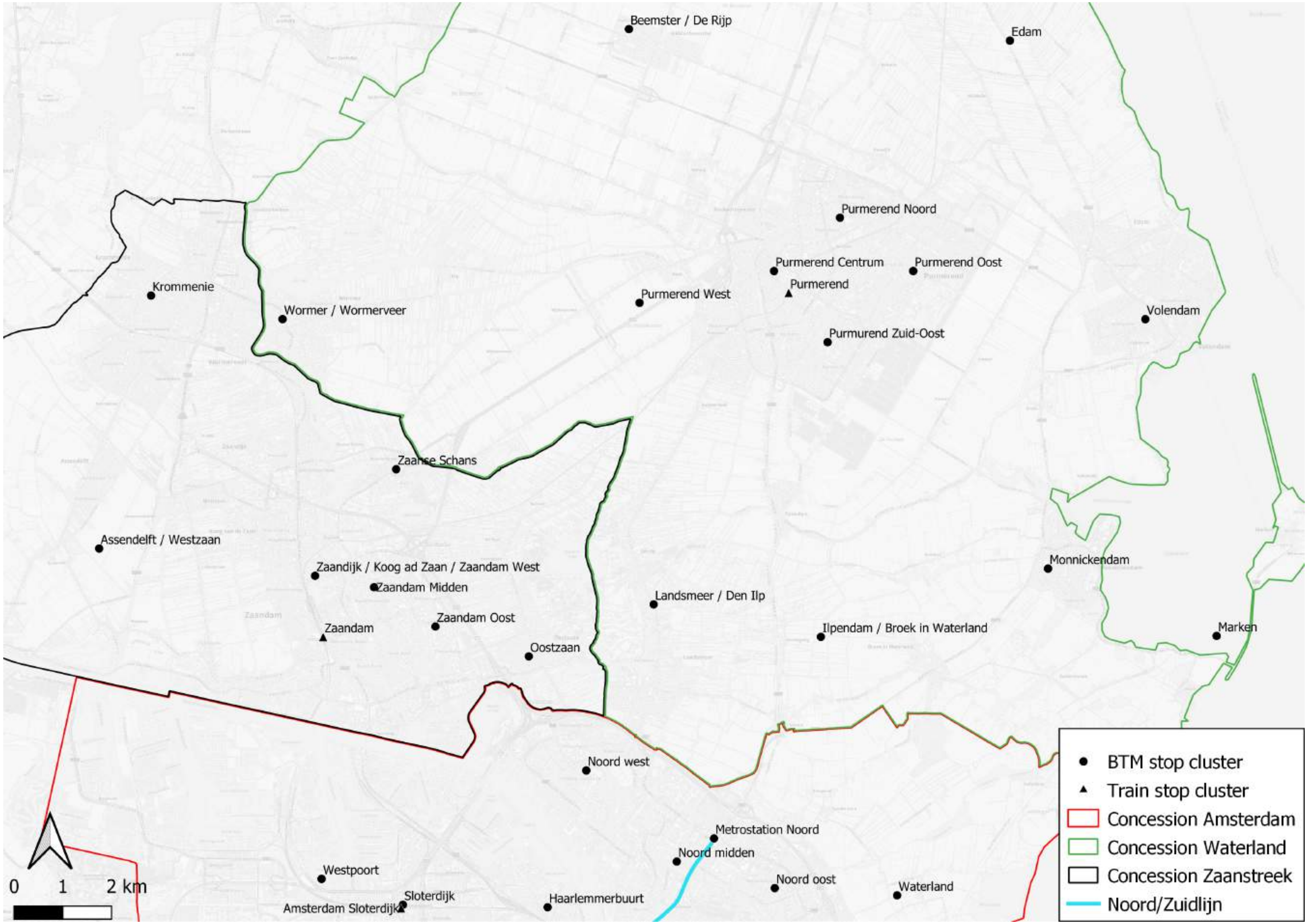


Figure I.3: Regional stop clusters north of Amsterdam

J

Validation of estimates for ranges in travel time data

This Appendix describes how the estimates used to replace the ranges in the OD travel times data set have been validated. The estimates should lead to a plausible number of journeys per day. To check this individual, openly available operator data has been used. Because the factor to convert trips to journeys and the number of train journeys have been assumed, the estimates for the ranges are valid if the total number of journeys obtained from the OV-dashboard and the OD travel times data set differ less than 15%.

Each month GVB, EBS and Connexxion report a selection of key performance indicators to the VRA. The VRA publishes these on an openly available dashboard (Vervoerregio Amsterdam, 2021). On this dashboard the number of travelers per operator, per type of day, per part of the day and ticket type can be found. For both periods 2018Q2 and 2019Q1 the number of travelers on a whole working day for ticket types 'saldo reizen' and 'abonnementen' have been selected. These ticket types should approximate all the transactions that are recorded by TLS. However, the data from the VRA dashboard and TLS do not match exactly. The dashboard shows monthly trip totals, does not distinguish between working days including or excluding school vacations and the data aggregation periods differ slightly. To account for this, monthly trip totals on a working day have been converted to the average number of journeys per working day (excluding school vacations) to ensure fair comparison. Besides, for each operator a factor to convert trips to journeys has been assumed. Figure J.1 shows how the monthly trip totals from the OV-dashboard have been converted to the number of BTM-journeys on an average working day (excluding school vacations).

		Saldo reizen			Abonnementen			Total journeys in VRA-dashboard	Working day factor	Total journeys on working days (excluding school vacation) in VRA-dashboard	Explanation
		GVB	**EBS**	**CXX**	**GVB**	**EBS**	**CXX**				
	Trip/journey factor	0,85	0,75	0,9	0,85	0,75	0,9				
2018Q2											Period from 16th of June 108 to 21th of July 2018, contains 25 working days (excluding school vacations)
June	Trips	6750000	255200	1119800	6037000	298500	1156100				
	Journeys	5737500	216920	951830	5131450	253725	982685	13274110	0,476	6321005	Convert OV-dashboard (all of June) to TLS data (16th of June - 30th of June)
July	Trips	4150200	204400	800700	2465200	176200	703400				
	Journeys	3527670	173740	680595	2095420	149770	597890	7225085	1	7225085	Include all days as July 2018 breaks on 21 July 2018 in dashboard
All working days										13546090	
Average working day										541844	
2019Q1											Period from 9th of December 2018 to 2nd of March 2019, contains 45 working days (excluding school vacations)
December	Trips	5926900	252000	931500	4615700	279800	947900				
	Journeys	5037865	214200	791775	3923345	237830	805715	11010730	0,625	6881706	10 out of 16 working days not in vacation (from December 9th)
January	Trips	8859000	371400	1492800	7075000	441000	1504400				
	Journeys	7530150	315690	1268880	6013750	374850	1278740	16782060	0,826	13863441	19 out of 23 working days not in vacation
February	Trips	7905000	330400	1333500	6615000	392600	1393600				
	Journeys	6719250	280840	1133475	5622750	333710	1184560	15274585	0,8	12219668	16 out of 20 working days not in vacation
March	Trips	8775000	359500	1550200	7350000	439700	1596500				
	Journeys	7458750	305575	1317670	6247500	373745	1357025	17060265	0	0	0 working days not in vacation (till March 2nd)
All working days										32964815	
Average working day										732551	

Figure J.1: Calculation of BTM-journeys on working days from VRA-dashboard

When the number of journeys by bus, tram and metro has been estimated, unimodal train journeys are not yet included. To estimate the number of train journeys, NS data regarding the number of boarding and alighting passengers per station have been used (Nederlandse Spoorwegen, 2019). These values are for an average working day in 2019. It is assumed that most train trips in the MRA are directed to Amsterdam. To avoid double counts, only the number of boarding and alighting passengers on train stations in the concession area Amsterdam have been included. It has been assumed that 50% of those passengers boards and 50% alights at the station. Next, the number of passengers that access or egress the train station by bus, tram or metro has been extracted from the total number of boarding and alighting passengers. This results in the number of unimodal train journeys per station. By summing these over all train stations in Amsterdam the number of unimodal train passengers on an average working day has been obtained. The calculation thereof can be found in Table J.1. Because the number of train passengers grew by 3.9% from 2018 to 2019 (NU.nl, 2020), this factor has been applied to estimate the number of train journeys for 2018. Next to that, seasonality effects have been accounted for by assuming that the number of train journeys is 10% larger around January or February than in June or July.

Table J.1: Calculation of the number of unimodal train journeys on working days in 2019 from NS-dashboard

Train stations in Amsterdam	**Boarding and alighting passengers per average working day in 2019** [travelers/day]	**No access by BTM** [%]	**No egress by BTM** [%]	**Unimodal train journeys to and from station** [journeys/day]
Amsterdam Sloterdijk	60,472	56	58	34,469
Amsterdam Lelylaan	15,908	69	55	9,863
Amsterdam Zuid	68,699	63	65	43,967
Amsterdam RAI	4,842	89	91	4,358
Amsterdam Centraal	199,510	51	53	103,745
Amsterdam Muiderpoort	15,149	91	78	12,801
Amsterdam Amstel	36,100	65	50	20,758
Amsterdam Science Park	4,869	97	98	4,747
Duivendrecht	12,500	41	24	4,062
Amsterdam Bijlmer Arena	30,035	50	84	20,123
Amsterdam Holendrecht	4,620	77	94	3,950
Diemen	3,328	90	91	3,012
Diemen Zuid	3,535	80	77	2,775
Total				268,630

Now the total number of journeys on a working day (excluding school vacations) can be obtained by summing the BTM-journeys and train journeys for both time periods (see Table J.2). The last step is then to compare that with the number of journeys in the OD travel times data set after estimating the ranges. As can be seen in Table J.3 the estimates are valid because the number of journeys does not differ more than 15% between each data set that has been used.

Table J.2: Number of journeys on working day (excluding school vacations) based on operator data

Data set	**BTM-journeys** [journeys]	**Train journeys** [journeys]	**Total journeys** [journeys]
2018Q2	542,000	232,000	774,000
2019Q1	733,000	268,000	1,001,000

Table J.3: The number of journeys on a working day (excluding school vacations) based on dashboards data and travel time data

Data set	**Estimated number of journeys from dashboard [journeys]**	**Estimated number of journeys from travel time data [journeys]**
2018Q2	774,000	772,000
2019Q1	1,001,000	950,000

K

Validation of estimates for ranges in route data

This Appendix describes how the estimates used to replace the ranges in the route data set have been validated. The estimates should lead to a plausible number of trips per day per operator day. To check this individual, openly available operator data has been used. Because the estimates are rough and allow no distinction between the operators, the estimates are valid if the total number of trips per operator obtained from the OV-dashboard and the routes data set differ less than 20%.

To validate the estimates for the route data, the same data source has been used as for the validation of the travel times data, namely the OV-dashboard. The same trip totals per ticket type and operator have thus been used. This data is described in detail in J. However, the validation steps differ slightly. The reason for that is that the route data consists of all trips per operator per journey. This allows to distinguish between the operators while validating.

The first step is to convert the monthly trip totals on working days per operator to trip totals on working days (excluding school vacations) per operator. For this the same working day factor as for the travel times data has been applied to the monthly trip totals. Thereafter, the values per operator have been summed to obtain the total number of trips per operator on working days (excluding school vacations) that should approximate the trip totals in the before and after data set. Next, these have been divided by the number of working days for each data set.

Table K.1: Trip totals on a working day (excluding school vacations) based on dashboards data and route data for 2018Q2

Operator	**Estimated daily trips TLS data [trips]**	**Daily trips OV-dashboard [trips]**	**Absolute difference [trips]**	**Factor [-]**
CXX	101,720	104,000	2,280	0.98
EBS	23,640	26,000	2,360	0.91
GVB	560,480	508,000	-52,480	1.1

Table K.2: Trip totals on a working day (excluding school vacations) based on dashboards data and route data for 2019Q1

Operator	**Estimated daily trips TLS data [trips]**	**Daily trips OV-dashboard [trips]**	**Absolute difference [trips]**	**Factor [-]**
CXX	144,690	130,000	14,690	1.11
EBS	35,730	35,000	-730	1.02
GVB	733,780	697,000	-36,780	1.05

L
Outliers in travel times data set

This Appendix describes the removal of outliers from the travel time data set.

Table L.1 shows how much outliers are removed per data set per time period if the following criteria are applied:

- The number of journeys should be between 1 and 50
- The median travel time in that half hour block should differ 50% from the median travel time over the whole day
- The median travel time in that half hour block should differ at least 30 minutes from the median travel time over the day

Table L.1: Number of outliers in travel times data per time period before and after Noord/Zuidlijn

	2018Q2		*2019Q1*	
Time period	**Absolute number of outliers [records]**	**Relative number of outliers [%]**	**Absolute number of outliers [records]**	**Relative number of outliers [%]**
Morning	147	1.1	174	1.15
Evening	285	1.94	226	1.36
Rest of day	1940	2.26	2032	1.95
Whole day	2372	2.12	2432	1.8

M

Columns in data set after travel time data processing

This Appendix contains all columns in the travel time data set after processing.

- Part of day
- Check in cluster code
- Check in cluster name
- Check in cluster type
- Check in cluster x-coordinate
- Check in cluster y-coordinate
- Check out cluster code
- Check out cluster name
- Check out cluster type
- Check out cluster x-coordinate
- Check out cluster y-coordinate
- Average travel time before Noord/Zuidlijn
- Number of travelers before Noord/Zuidlijn
- Lowerbound number of travelers before Noord/Zuidlijn
- Estimated number of travelers before Noord/Zuidlijn
- Average travel time after Noord/Zuidlijn
- Number of travelers after Noord/Zuidlijn
- Lowerbound number of travelers after Noord/Zuidlijn
- Estimated number of travelers after Noord/Zuidlijn
- Absolute difference in travel time
- Relative difference in travel time
- Absolute difference in number of travelers
- Relative difference in number of travelers

N

Outliers in route data set

This Appendix describes the removal of outliers from the route data set.

Table N.1 shows how much outliers are removed per data set if the following criteria are applied:

- (1) The number of travelers on the route should be between 1 and 50 and in the route a transfer from train to BTM and back should take place. Such a transfer pattern is illogical because of the hierarchy in the public transport network of Amsterdam.
- (2) The number of travelers on the route should be between 1 and 50 and four or more transfers are made on that route.
- (3) The number of travelers on the route should be between 1 and 50 and three or more transfers are made on that route while a direct route exists as well. If a direct route exists on an OD-relation, a route with at least three transfers more is illogical.
- (4) The number of travelers on the route should be between 1 and 50 and the exact same route is not made in any of the other periods of the day and at least three or more transfers are made on that route.

Table N.1: Number of outliers in route data before and after Noord/Zuidlijn

	2018Q2		*2019Q1*	
Outlier removal criterion	**Absolute number of outliers [records]**	**Relative number of outliers [%]**	**Absolute number of outliers [records]**	**Relative number of outliers [%]**
1	208	0.2	373	0.2
2	4,472	3.6	12,406	5.6
3	3,972	3.2	9,972	4.4
4	9,770	7.9	23,483	10.7
Total	18,422	14.9	46,054	20.9

After outlier removal, 104,978 unique routes remain in the before data set. In the after data set 174,599 unique routes remain after removal of outliers. It must be noted that a unique route is counted for each time period it occurs in.

O

Number of OD-relations per time period per representativeness criterion for the travel times data

This Appendix presents the number of OD-relations per time period for each reliability criterion. These can be found in Table O.1.

Table O.1: The number of OD-relations with at least a certain number of travelers before and after the Noord/Zuidlijn per time period. Both directions of an OD-relation are considered separately.

Time period	**At least 51 travelers before and after Noord/Zuidlijn [OD-relations]**	**At least 12 travelers before and after Noord/Zuidlijn [OD-relations]**	**At least 4 travelers before and after Noord/Zuidlijn [OD-relations]**	**At least 1 traveler before and after Noord/Zuidlijn [OD-relations]**
Morning peak	762	1,583	2,767	3,617
Evening peak	633	1,833	3,136	4,002
Rest of day	1,804	3,503	4,161	4,545
Whole day	2,315	3,837	4,315	4,623

P

Contribution of each range to the total number of travelers on a working day

This Appendix presents the share that each range contributes to the total number of travelers on a working day. What becomes clear is that the lower bound travelers contributes most to the total number of travelers. Besides, comparing Table P.2 and Table P.4 to P.1 and Table P.3 it follows that the share of higher ranges and exact values becomes larger if the aggregation periods are longer. Additionally, it can be seen that the share of the travelers lower that fall in one of the ranges is much larger for the travel time data than for the route data. This indicates that travel demand varies much more over half hour blocks than over routes.

Table P.1: Share that the lower bound and estimated number of travelers from records of a range contributes to the total travelers on a working for the 2018Q2 travel time data. The total number of travelers is 88,300.

Range [travelers]	**Share of lower bound travelers [%]**	**Share of estimated travelers [%]**
1-50	10.3	28.6
51-100	12.0	3.3
101-150	8.0	1.5
151-200	6.0	1.0
201-250	4.1	0.5
251-300	3.0	0.3
>300	21.3	0
Total	64.7	35.3

Table P.2: Share that the lower bound and estimated number of travelers from records of a range contributes to the total travelers on a working for the 2019Q1 travel time data. The total number of travelers is 103,700.

Range [travelers]	**Share of lower bound travelers [%]**	**Share of estimated travelers [%]**
1-50	5.6	16.4
51-100	10.9	3.0
101-150	8.7	1.6
151-200	7.1	1.1
201-250	5.8	0.7
251-300	4.5	0.4
>300	34.0	0
Total	76.6	23.3

Table P.3: Share that the lower bound and estimated number of travelers from records of a range contributes to the total travelers on a working for the 2018Q2 route data. The total number of travelers is 82,200.

Range [travelers]	**Share of lower bound travelers [%]**	**Share of estimated travelers [%]**
1-50	4.8	9.7
51-100	5.7	1.6
101-150	4.9	0.9
151-200	4.0	0.6
201-250	3.4	0.4
251-300	2.7	0.3
>300	60.9	0
Total	86.4	13.5

Table P.4: Share that the lower bound and estimated number of travelers from records of a range contributes to the total travelers on a working for the 2019Q1 route data. The total number of travelers is 100,700.

Range [travelers]	**Share of lower bound travelers [%]**	**Share of estimated travelers [%]**
1-50	3.6	7.1
51-100	4.7	1.3
101-150	3.9	0.7
151-200	3.2	0.5
201-250	2.9	0.3
251-300	2.5	0.2
>300	69.0	0
Total	89.8	10.1

Q Validation results sensitivity analysis for an underestimation and overestimation of the average value for each range

This Appendix presents the journey and trip totals for an underestimation and overestimation of the average value for each range. It should be kept in mind that the totals are based on all OD-relations in the whole data set and thus not only on regional OD-relations.

Table Q.1 presents the underestimated number of daily journeys in the travel time data. Table Q.2 presents the validation results for an overestimation.

Table Q.1: Journey totals in the travel time data on a working day (excluding school vacations) for an underestimation of the average value for each range compared to the journey totals in the OV-dashboard

Data set	**Underestimated daily journeys [journeys]**	**Daily journeys OV-dashboard [journeys]**	**Absolute difference [journeys]**	**Factor [-]**
2018Q2	624,500	774,000	-149,500	0.81
2019Q1	847,400	1,001,000	-153,600	0.87

Table Q.2: Journey totals in the travel time data on a working day (excluding school vacations) for an overestimation of the average value for each range compared to the journey totals in the OV-dashboard

Data set	**Overestimated daily journeys [journeys]**	**Daily journeys OV-dashboard [journeys]**	**Absolute difference [journeys]**	**Factor [-]**
2018Q2	1,085,500	774,000	+311,500	1.40
2019Q1	1,155,600	1,001,000	+154,600	1.15

Table Q.3 and Table Q.4 present the underestimated trip totals per operator for the before and after data respectively.

Table Q.3: Trip totals in the 2018Q2 route data on a working day (excluding school vacations) for an underestimation of the average value for each range compared to the trip totals in the OV-dashboard

Operator	**Underestimated daily trips [trips]**	**Daily trips OV-dashboard [trips]**	**Absolute difference [trips]**	**Factor [-]**
CXX	82,680	104,000	-21,320	0.8
EBS	18,400	26,000	-7,600	0.71
GVB	442,080	508,000	-65,920	0.87

Table Q.4: Trip totals in the 2019Q1 route data on a working day (excluding school vacations) for an underestimation of the average value for each range compared to the trip totals in the OV-dashboard

Operator	**Underestimated daily trips [trips]**	**Daily trips OV-dashboard [trips]**	**Absolute difference [trips]**	**Factor [-]**
CXX	125,111	130,000	-4,889	0.96
EBS	29,511	35,000	-5,489	0.84
GVB	610,844	697,000	-86,156	0.88

Table Q.5 and Table Q.6 present the overestimated trip totals per operator for the before and after data respectively.

Table Q.5: Trip totals in the 2018Q2 route data on a working day (excluding school vacations) for an overestimation of the average value for each range compared to the trip totals in the OV-dashboard

Operator	**Overestimated daily trips [trips]**	**Daily trips OV-dashboard [trips]**	**Absolute difference [trips]**	**Factor [-]**
CXX	129,400	104,000	+25,400	1.24
EBS	31,120	26,000	+5,120	1.2
GVB	733,480	508,000	+225,480	1.44

Table Q.6: Trip totals in the 2019Q1 route data on a working day (excluding school vacations) for an overestimation of the average value for each range compared to the trip totals in the OV-dashboard

Operator	**Overestimated daily trips [trips]**	**Daily trips OV-dashboard [trips]**	**Absolute difference [trips]**	**Factor [-]**
CXX	173,022	130,000	+43,022	1.33
EBS	44,689	35,000	+9,689	1.28
GVB	913,200	697,000	+216,200	1.31

R

Distribution of travel time and transfer impacts over travelers and OD-relations for representativeness criterion 1

This Appendix presents the distribution of the travel time and transfer impacts for a representativeness criterion of 1. Figure R.1 and Figure R.2 present the distribution of the travel time impacts over transfers and OD-relations for a representativeness criterion of 1.

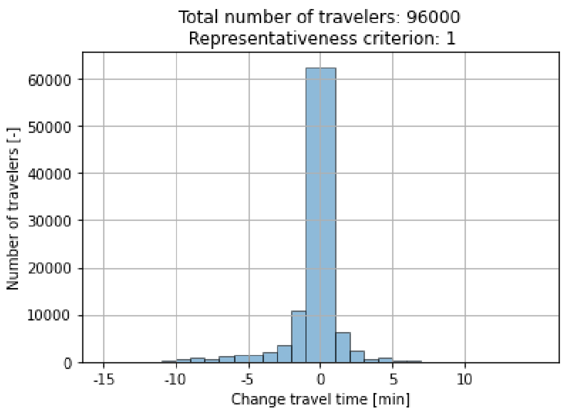


Figure R.1: Distribution of the absolute change in travel time against the number of travelers. Results are plotted for the average travel time over the whole day.

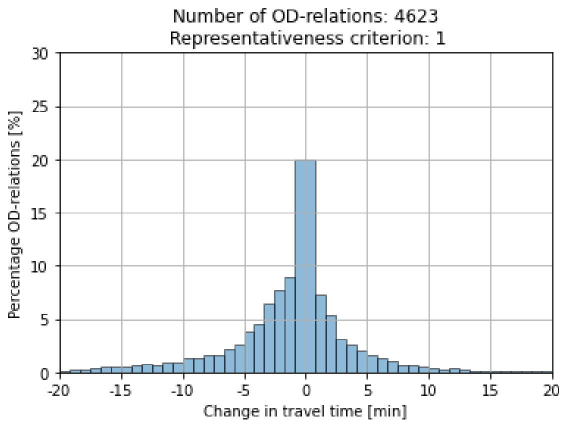


Figure R.2: Distribution of the absolute change in travel time against the number of OD-relations. Results are plotted for the average travel time over the whole day.

Figure R.3 and Figure **??** present the distribution of the travel time impacts over transfers and OD-relations for a representativeness criterion of 1.

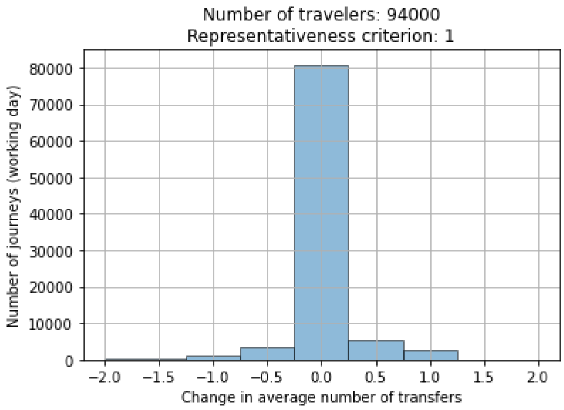


Figure R.3: Distribution of the absolute change in travel time against the number of travelers. Results are plotted for the average travel time over the whole day.

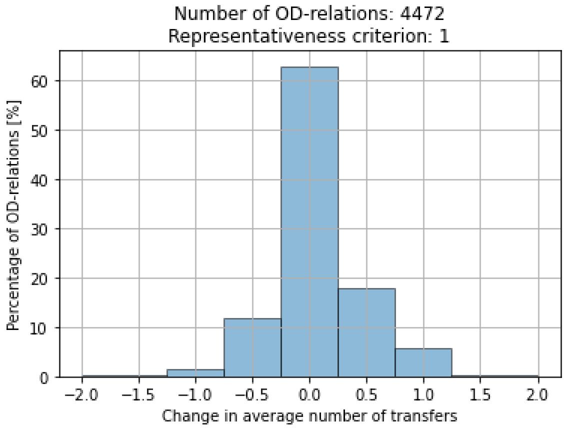


Figure R.4: Distribution of the absolute change in travel time against the number of OD-relations. Results are plotted for the average travel time over the whole day.

S

Top 20 OD-relations with the largest difference travel time change in opposite directions

This Appendix presents the top 20 OD-relations where the difference in travel time change in opposite directions is largest. These can be found in S.1.

Table S.1: OD-relations with the largest difference in change in travel time between both directions. The representativeness criterion is 51 and the time period whole day.

From/to	**From/to**	**Difference in travel time change between both directions [min]**	**Change in travel time from Amsterdam to region [min]**	**Change in travel time from region to Amsterdam [min]**
Oostelijke Eilanden en Kadijken	Edam	14.0	-0.1	-14.1
Leidseplein en Leidsestraat	Uithoorn	13.7	-9.3	4.4
Nieuwmarkt en Lastage	Zaandam Oost	13.4	-13.4	0.0
Oostelijke Eilanden en Kadijken	Ilpendam/Broek in Waterland	12.4	-3.8	-16.2
Oostelijke Eilanden en Kadijken	Zaandam Oost	12.1	1.0	-11.1
Bijlmer ArenA	Oostzaan	11.5	-10.3	1.2
Oostelijke Eilanden en Kadijken	Zaandam Midden	10.8	0.2	-10.6
Weesperbuurt en Plantage	Purmerend NS	10.7	-0.0	-10.7
Indische Buurt	Hoofddorp NS	10.3	5.0	-5.3
Bos en Lommer	Badhoevedorp	10.2	2.5	-7.7
Oostelijke Eilanden en Kadijken	Purmerend Zuid-Oost	10.1	-1.7	-11.8
Oostelijke Eilanden en Kadijken	Purmerend Noord	10.0	-0.6	-10.6
Oostelijke Eilanden en Kadijken	Volendam	9.9	-0.6	-10.5
Volendam	Westpoort	9.1	-4.8	4.4
Bijlmer-Oost	Purmerend Noord	9.0	-7.0	2.1
Nieuwmarkt en Lastage	Zaandam Midden	8.9	-12.5	-3.6
Schiphol Plaza	Noord midden	8.8	-4.5	-13.3
Oostelijke Eilanden en Kadijken	Purmerend Centrum	8.7	-5.3	-14.0
Spui	Hoofddorp NS	8.5	-5.6	-14.1
Leidseplein en Leidsestraat	Aalsmeer / Kudelstaart	8.5	10.1	1.6

T

Top 20 OD-relations with the largest relative change in travel time

This Appendix presents the OD-relations with the largest relative change in travel time before and after the Noord/Zuidlijn.

Table T.1 presents the top 20 OD-relations with the largest relative travel time decrease. Journeys with a check out in stop cluster Oostelijke Eilanden en Kadijken occur often. This can be explained by a new bus stop at the end of the IJtunnel. Furthermore, as expected mainly OD-relations with either an origin or destination around the Noord/Zuidlijn experience a large decrease in travel time. It stands out and is logical that these are mainly OD-relations for which a long part of the journey is made with the Noord/Zuidlijn.

Table T.1: Top 20 OD-relations with largest relative decrease in travel time over a whole working day. The representativeness criterion is 51 travelers.

Check in cluster	**Check out cluster**	**Change travel time [%]**
Ilpendam / Broek in Waterland	Oostelijke Eilanden en Kadijken	-51.4
Landsmeer / Den Ilp	Noord west	-46.3
Schiphol NS	Metrostation Weteringcircuit	-39.1
Schiphol NS	Pijp	-37.6
Ilpendam / Broek in Waterland	Rivierenbuurt	-37.5
Noord west	Landsmeer / Den Ilp	-36.2
Amstelveen Busstation	Noord midden	-36.1
Purmerend Centrum	Oostelijke Eilanden en Kadijken	-35.8
Metrostation Weteringcircuit	Schiphol	-34.7
Amstelveen	Noord midden	-34.2
Rivierenbuurt	Ilpendam / Broek in Waterland	-34.2
Purmurend Zuid-Oost	Oostelijke Eilanden en Kadijken	-32.9
Amstelveen Centrum	Noord midden	-32.2
Monnickendam	Rivierenbuurt	-31.9
Zuid	Ilpendam / Broek in Waterland	-31.8
Edam	Oostelijke Eilanden en Kadijken	-31.6
Rivierenbuurt	Monnickendam	-31.5
Pijp	Schiphol NS	-29.9
Hoofddorp NS	Pijp	-29.8
Landsmeer / Den Ilp	Rivierenbuurt	-29.2

Table T.2 presents the top 20 OD-relations with the largest relative increase in travel time. The results are presented for OD-relations with at least 51 travelers before and after the Noord/Zuidlijn and for a whole working day. Regarding the largest relative increase, OD-relations from Waterland to the city center stand out. Besides, stop cluster Noord midden returns a couple of times in the list. This is probably caused by the fact the that bus network in Amsterdam Noord was restructured when the Noord/Zuidlijn became operative. From the concession Amstelland-Meerlanden travel times to and from the Dam increased because the bus lines from Amstelland-Meerlanden do not penetrate the city center till Amsterdam Centraal any more. An OD-relation that can not directly be linked to the Noord/Zuidlijn is from Buitenveldert to Aalsmeer/Kudelstaart. Last, a strong decrease in number of travelers can be seen on some of the OD-relations with an increase in travel time.

Table T.2: Top 20 OD-relations with largest relative increase in travel time over a whole working day. The representativeness criterion is 51 travelers.

Check in cluster	**Check out cluster**	**Change travel time [%]**
Nieuwmarkt en Lastage	Landsmeer / Den Ilp	86.9
Landsmeer / Den Ilp	Nieuwmarkt en Lastage	70.9
Centraal Station	Landsmeer / Den Ilp	53.2
Purmerend West	Nieuwmarkt en Lastage	41.0
Nieuwmarkt en Lastage	Purmerend West	40.5
Noord midden	Landsmeer / Den Ilp	37.0
Nieuwmarkt en Lastage	Purmerend Oost	32.7
Purmerend Oost	Nieuwmarkt en Lastage	32.0
Purmerend West	Dam	31.2
Oostzaan	Noord midden	29.6
Noord midden	Oostzaan	29.2
Landsmeer / Den Ilp	Noord midden	28.4
Noord midden	Ilpendam / Broek in Waterland	27.7
Centraal Station	Purmerend West	27.0
Centraal Station	Oostzaan	25.6
Marken	Nieuwmarkt en Lastage	24.8
Noord midden	Monnickendam	22.9
Centraal Station	Purmerend Oost	21.9
Dam	Amstelveen	21.5
Buitenveldert	Aalsmeer / Kudelstaart	21.5

U

Effect of a varying representativeness criterion on the distribution of travel time impacts over OD-relations

This Appendix shows what effect selecting a less conservative representativeness criterion has on the distribution of the travel time impacts over the OD-relations. This is shown in Figure U.1.

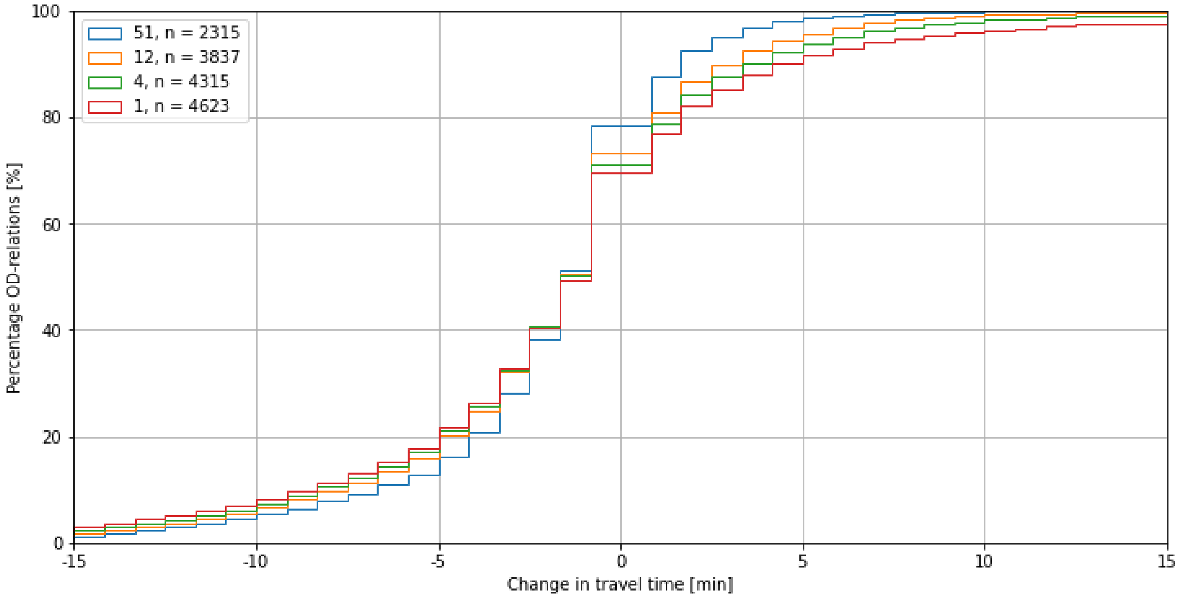


Figure U.1: Cumulative distribution function of the absolute change in travel time against the percentage of OD-relations for a varying representativeness criterion. Results are plotted for the average travel time over the whole day.

V

Visualization of the absolute change in travel time over all OD-relations

In this Appendix the change in travel time per OD-relation is visualized. In Figure V.2 the change in travel time from the region to Amsterdam is shown over the whole day. In Figure V.1 the opposite direction is presented. For both these figures the representativeness criterion is 12. The stop clusters have been grouped geographically.

In Figure V.3 the change in travel time from Amsterdam to the region is shown for representativeness criterion 1. It becomes clear that this results in some changes in travel time that cannot be attributed to the Noord/Zuidlijn. In Figure V.4 the change in travel time is visualized for representativeness criterion 51. As can be seen this way a large amount of OD-relations is excluded from the analysis.

Figure V.1: The absolute change in travel time for OD-relations from the region to Amsterdam over the whole day. The representativeness criterion is 12.

Figure V.2: The absolute change in travel time for OD-relations from Amsterdam to the region over the whole day. The representativeness criterion is 12.


Check out cluster
Check in cluster
15
10
5
0
−5
−10
−15

Figure V.4: The absolute change in travel time for OD-relations from Amsterdam to the region over the whole day. The representativeness criterion is 51.

W

Effect of different time periods on the number of travelers per working day

This Appendix provides insights in the growth in number of travelers as a result of the different aggregation periods of the smart card data. The smart card data before the Noord/Zuidlijn has been aggregated between the 16th of June and the 21th of July, 2018. The period after the Noord/Zuidlijn is from the 9th of December, 2018 to the 2nd of March, 2019.

Figure W.1 shows that the number of travelers per working day increases by a factor around 1.17 between the winter and summer months. Only ticket types that are present in interoperable smart card data have been included.

	Amsterdam		Region		Number of working days
	Abonnementen	Saldo reizen	Abonnementen	Saldo reizen	
dec-17	4278800	5390400	985300	1218800	15
jan-18	6707000	7845000	1723700	1618600	22
feb-18	6575000	7230000	1715400	1561600	20
Total	46849600		8823400		
Per working day	821923		154796		
jun-18	6037000	6750000	1454500	1375000	21
jul-18	2896900	4817500	879500	1005100	15
Total	25215500		4714100		
Per working day	700431		130947		
Factor	1,17		1,18		

Figure W.1: The number of travelers per area ('Amsterdam' and 'Region'), ticket type ('Abonnementen' and 'Saldo reizen') and period (December 2017 - February 2018 and June 2018 - July 2018) and the factor change between these periods.

X

Share of trips made using single-use and paper tickets

Figure X.1 presents the share of single-use and paper tickets in 2018 in the VRA. The share differs per concession area and modality.

Overview trip totals and share of single-use and paper tickets				
2018				
Concession			total trips (* 1.000)	*share CT/paper tickets*
Amsterdam		19,0%	252.220	47.922
	Tram	*26,7%*	*112.387*	*30.007*
	Metro	*13,5%*	*90.345*	*12.197*
	Bus	*11,2%*	*49.487*	*5.543*
Amstelland Meerland		9,2%	35.031	3.223
Waterland		7,2%	10.592	763
Zaanstreek		10,3%	6.152	634
Total		17,2%	303.994	52.366
		weighted average		

Figure X.1: The share of trips made using single-use and paper tickets over all trips made in the VRA.

Y

Code travel time data processing

```
#!/usr/bin/env python
# coding: utf-8

# In[3]:



# --- Import packages ---
import pandas as pd
import scipy
import scipy.stats as ss
import math
import numpy as np
from varname import nameof


# In[ ]:



# --- Variables ---
data_2018Q2 = dataset before Noord/Zuidlijn
data_2019Q1 = dataset after Noord/Zuidlijn
data_2018Q2.name = '2018Q2'
data_2019Q1.name = '2019Q1'

cluster_info = dataset cluster info
cluster_info['name_type'] = cluster_info['name']+ ' ' + cluster_info['type']

# --- Half hour blocks per time period ---
morning = ['7.00-7.30', '7.30-8.00', '8.00-8.30', '8.30-9.00']
evening = ['16.00-16.30', '16.30-17.00', '17.00-17.30', '17.30-18.00']
rest_day = ['0.00-0.30', '0.30-1.00', '1.00-1.30', '1.30-2.00', '2.00-2.30',
    ↪ '2.30-3.00', '3.00-3.30', '3.30-4.00', '4.00-4.30', '4.30-5.00', '5.00-5.30',
    ↪ '5.30-6.00', '6.00-6.30', '6.30-7.00', '9.00-9.30', '9.30-10.00',
    ↪ '10.00-10.30', '10.30-11.00', '11.00-11.30', '11.30-12.00', '12.00-12.30',
    ↪ '12.30-13.00', '13.00-13.30', '13.30-14.00', '14.00-14.30', '14.30-15.00',
    ↪ '15.00-15.30', '15.30-16.00', '18.00-18.30', '18.30-19.00', '19.00-19.30',
    ↪ '19.30-20.00', '20.00-20.30', '20.30-21.00', '21.00-21.30', '21.30-22.00',
    ↪ '22.00-22.30', '22.30-23.00', '23.00-23.30', '23.30-0.00']
whole_day = ['0.00-0.30', '0.30-1.00', '1.00-1.30', '1.30-2.00', '2.00-2.30',
    ↪ '2.30-3.00', '3.00-3.30', '3.30-4.00', '4.00-4.30', '4.30-5.00', '5.00-5.30',
```

```
    ↪ '5.30-6.00', '6.00-6.30', '6.30-7.00', '7.00-7.30', '7.30-8.00', '8.00-8.30',
    ↪ '8.30-9.00', '9.00-9.30', '9.30-10.00', '10.00-10.30', '10.30-11.00',
    ↪ '11.00-11.30', '11.30-12.00', '12.00-12.30', '12.30-13.00', '13.00-13.30',
    ↪ '13.30-14.00', '14.00-14.30', '14.30-15.00', '15.00-15.30', '15.30-16.00',
    ↪ '16.00-16.30', '16.30-17.00', '17.00-17.30', '17.30-18.00', '18.00-18.30',
    ↪ '18.30-19.00', '19.00-19.30', '19.30-20.00', '20.00-20.30', '20.30-21.00',
    ↪ '21.00-21.30', '21.30-22.00', '22.00-22.30', '22.30-23.00', '23.00-23.30',
    ↪ '23.30-0.00']

# --- Add column 'dagdeel' to dataframe based on values for column 'tijdsperiode' ---
data_2018Q2['dagdeel'] = np.where(data_2018Q2['tijdsperiode'].isin(morning), '
    ↪ ochtendspits',
                                np.where(data_2018Q2['tijdsperiode'].isin(evening), '
                                    ↪ avondspits', 'restdag'))

data_2019Q1['dagdeel'] = np.where(data_2019Q1['tijdsperiode'].isin(morning), '
    ↪ ochtendspits',
                                np.where(data_2019Q1['tijdsperiode'].isin(evening), '
                                    ↪ avondspits', 'restdag'))

# --- Selection of region clusters and Amsterdam clusters ---
all_clusters = ['0', '1', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
    ↪  '2', '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '3', '30',
    ↪ '31', '32', '33', '34', '35', '36', '37', '38', '39', '40', '41', '42', '43',
    ↪ '44', '45', '46', '47', '48', '49', '5', '50', '51', '52', '53', '54', '55',
    ↪ '56', '57', '58', '59', '6', '60', '61', '62', '63', '64', '65', '66', '67',
    ↪ '68', '69', '7', '70', '71', '72', '73', '74', '75', '76', '77', '78', '79',
    ↪ '8', '80', '81', '82', '83', '84', '85', '86', '87', '88', '9', 'ah', 'alm', '
    ↪ almm', 'almp', 'amf', 'amr', 'asa', 'asb', 'asd', 'asdh', 'asdl', 'asdm', '
    ↪ asdz', 'ass', 'assp', 'bkl', 'brn', 'bv', 'dmn', 'dmnz', 'dv', 'dvd', 'hfd', '
    ↪ hil', 'hlm', 'hn', 'ht', 'hvs', 'hwzb', 'ledn', 'lls', 'ndb', 'pmr', 'rai', '
    ↪ rm', 'rtd', 'shl', 'ssh', 'tb', 'ut', 'utg', 'vl', 'wp', 'zd', 'zl']
cluster_region = region clusters
cluster_amsterdam = amsterdam clusters

# --- Number of working days in each data set ---
ndays_2018Q2 = 25
ndays_2019Q1 = 45


# In[ ]:


# --- Select day ---
def select_day(dataset, day=['werkdag (buiten schoolvakantie)']):
    dataset = dataset[dataset.loc[:,'dag'].isin(day)].copy()
    return dataset

# --- Select certain check in clusters ---
def select_timeperiod(dataset, sel_tijdsperiode):
    dataset = dataset[dataset.loc[:,'tijdsperiode'].isin(sel_tijdsperiode)].copy()
    return dataset

# --- Select certain check in clusters ---
def select_checkin(dataset, clusters):
    dataset=dataset[dataset.loc[:,'cki_zone'].isin(clusters)].copy()
```

```
    return dataset

# --- Select certain check out clusters ---
def select_checkout(dataset, clusters):
    dataset=dataset[dataset.loc[:,'cko_zone'].isin(clusters)].copy()
    return dataset

# --- Select regional OD ---
def select_regional_OD(dataset, amsterdam, region):
    dataset=dataset = dataset[(((dataset.loc[:, 'cki_zone'].isin(amsterdam)) & (
        ↪ dataset.loc[:, 'cko_zone'].isin(region)))
                              | ((dataset.loc[:, 'cki_zone'].isin(region)) & (dataset.
                                  ↪ loc[:, 'cko_zone'].isin(amsterdam))))].copy()
    return dataset

# --- Fill in ranges ---
def fill_in_ranges(dataset, rng_est=[13, 65, 120, 175, 225, 275], rng_bins=['1 - 50',
    ↪  '51 - 100', '101 - 150', '151 - 200', '201 - 250', '251 - 300']):
        dataset = dataset.copy()

# --- If (number of travelers = 1-50 & mean = median) then number of travelers = 1
        dataset.loc[(dataset['aantal_reizen'] == '1 - 50') & (dataset['
            ↪ mediaan_reistijd_seconden'] == dataset['gemiddelde_reistijd_seconden'])
            ↪ , 'aantal_reizen'] = 1

# --- Replace other ranges by estimates ---
        dataset.loc[:,'aantal_reizen'].replace(to_replace=rng_bins, value=rng_est,
            ↪ inplace=True)
        dataset.loc[:, 'aantal_reizen'] = pd.to_numeric(dataset.loc[:, 'aantal_reizen
            ↪ '])
        return dataset


# In[ ]:



# --- Determine number lower bound and estimated number of travelers per half hour
    ↪ block ---
def exact_estimated(dataset, rng_est=[13, 65, 120, 175, 225, 275], lb=[1, 51, 101,
    ↪ 151, 201, 251]):

    total = dataset['aantal_reizen'].sum()

    # --- Split ranges in exact (lb) and estimated (rng_est - lb) part ---
    exact_part = (#len(dataset.loc[(dataset['aantal_reizen'] >= lb[0]) & (dataset['
        ↪ aantal_reizen'] <= (lb[0]+49) )]) * lb[0] +
                  len(dataset.loc[(dataset['aantal_reizen'] >= lb[1]) & (dataset['
                      ↪ aantal_reizen'] <= (lb[1]+49) )]) * lb[1] +
                  len(dataset.loc[(dataset['aantal_reizen'] >= lb[2]) & (dataset['
                      ↪ aantal_reizen'] <= (lb[2]+49) )]) * lb[2] +
                  len(dataset.loc[(dataset['aantal_reizen'] >= lb[3]) & (dataset['
                      ↪ aantal_reizen'] <= (lb[3]+49) )]) * lb[3] +
                  len(dataset.loc[(dataset['aantal_reizen'] >= lb[4]) & (dataset['
                      ↪ aantal_reizen'] <= (lb[4]+49) )]) * lb[4] +
                  len(dataset.loc[(dataset['aantal_reizen'] >= lb[5]) & (dataset['
                      ↪ aantal_reizen'] <= (lb[5]+49) )]) * lb[5])
```

```
    estimated_part = (#len(dataset.loc[(dataset['aantal_reizen'] >= (lb[0]+1)) & (
        dataset['aantal_reizen'] <= lb[0]+49 )]) * (rng_est[0] - lb[0]) +
                len(dataset.loc[(dataset['aantal_reizen'] >= lb[1]) & (dataset['
                    aantal_reizen'] <= lb[1]+49 )]) * (rng_est[1] - lb[1]) +
                len(dataset.loc[(dataset['aantal_reizen'] >= lb[2]) & (dataset['
                    aantal_reizen'] <= lb[2]+49 )]) * (rng_est[2] - lb[2]) +
                len(dataset.loc[(dataset['aantal_reizen'] >= lb[3]) & (dataset['
                    aantal_reizen'] <= lb[3]+49 )]) * (rng_est[3] - lb[3]) +
                len(dataset.loc[(dataset['aantal_reizen'] >= lb[4]) & (dataset['
                    aantal_reizen'] <= lb[4]+49 )]) * (rng_est[4] - lb[4]) +
                len(dataset.loc[(dataset['aantal_reizen'] >= lb[5]) & (dataset['
                    aantal_reizen'] <= lb[5]+49 )]) * (rng_est[5] - lb[5]))

    # --- If median =/= mean, then at least 3 travelers exactly known ---
    exact_part += (len(dataset.loc[(dataset['aantal_reizen'] >= lb[0]) & (dataset['
        aantal_reizen'] <= (lb[0]+49) ) & (dataset['mediaan_reistijd_seconden'] !=
        dataset['gemiddelde_reistijd_seconden'])]) * 3)
    estimated_part += (len(dataset.loc[(dataset['aantal_reizen'] >= lb[0]) & (dataset
        ['aantal_reizen'] <= (lb[0]+49) ) & (dataset['mediaan_reistijd_seconden']
        != dataset['gemiddelde_reistijd_seconden'])]) * (rng_est[0] - 3))

    # --- If median = mean, then only 1 traveler exactly known ---
    exact_part += (len(dataset.loc[(dataset['aantal_reizen'] >= lb[0]) & (dataset['
        aantal_reizen'] <= (lb[0]+49) ) & (dataset['mediaan_reistijd_seconden'] ==
        dataset['gemiddelde_reistijd_seconden'])]) * lb[0])

    exact_counts = exact_part + dataset.loc[(dataset['aantal_reizen'] > 300)].loc[:,'
        aantal_reizen'].sum()
    return exact_counts, estimated_part, total


# In[4]:


# --- Calculate average and median travel time including outliers ---
def daily_tt(dataset, checkin, checkout, day=['werkdag (buiten schoolvakantie)'],
    rng_est=[13, 65, 120, 175, 225, 275], rng_bins=['1 - 50', '51 - 100', '101 -
    150', '151 - 200', '201 - 250', '251 - 300']):
    dataset = select_day(dataset, day)
    dataset = select_checkin(dataset, checkin)
    dataset = select_checkout(dataset, checkout)

    daily_tt_list = []
    for idx_i, i in enumerate(checkin):
        subset_ci_dataset = dataset.loc[(dataset['cki_zone']==i)].copy()
        for idx_j, j in enumerate(checkout):
            median_tt_list = []
            temp_dataset = subset_ci_dataset.loc[(subset_ci_dataset['cko_zone']==j)].
                copy()

            if temp_dataset.empty == True:
                daily_tt_list.append([i, j, np.nan, np.nan])
                pass
            else:
```

```
            # --- Replace ranges by estimates and convert to numeric ---
                for idx_h, h in enumerate(rng_bins):
                    temp_dataset.loc[(temp_dataset['aantal_reizen'] == '1 - 50') & (
                        ↪ temp_dataset['mediaan_reistijd_seconden'] == temp_dataset['
                        ↪ gemiddelde_reistijd_seconden']), 'aantal_reizen'] = 1
                    temp_dataset.loc[:,'aantal_reizen'].replace(to_replace=rng_bins[
                        ↪ idx_h], value=rng_est[idx_h], inplace=True)
                temp_dataset.loc[:, 'aantal_reizen'] = pd.to_numeric(temp_dataset.loc
                    ↪ [:, 'aantal_reizen'])

                for k in range(len(temp_dataset)):
                    median_tt_list.append(temp_dataset.iloc[k, 5] * [temp_dataset.iloc[
                        ↪ k,6]])
                flat_median_tt_list = [item for sublist in median_tt_list for item in
                    ↪ sublist]

                temp_dataset['aantal * gem_reistijd'] = temp_dataset['aantal_reizen']
                    ↪ * temp_dataset['gemiddelde_reistijd_seconden']
                daily_tt_list.append([i, j, np.median(sorted(flat_median_tt_list)),
                    ↪ round(temp_dataset['aantal * gem_reistijd'].sum()/temp_dataset['
                    ↪ aantal_reizen'].sum(),1)])
    df_tt = pd.DataFrame(daily_tt_list, columns=['cki_zone', 'cko_zone', 'median
        ↪ travel time over day', 'avg travel time over day'])

    return df_tt, daily_tt_list


# In[5]:


# --- Calculate average travel time and total # travelers for a selected day and
    ↪ period for a selection of checkin and checkout clusters. Add the number of
    ↪ exact and estimated travelers. ---

def tt_matrix(dataset, sel_timeperiod, amsterdam=cluster_amsterdam, region=
    ↪ cluster_region, outlier_abs=1800, outlier_rel=0.5, day=['werkdag (buiten
    ↪ schoolvakantie)'], rng_est=[13, 65, 120, 175, 225, 275], rng_bins=['1 - 50',
    ↪ '51 - 100', '101 - 150', '151 - 200', '201 - 250', '251 - 300']):
    name = str(dataset.name)

# --- Subset data for type of day and regional OD-relations ---
    dataset = select_day(dataset, day)
    dataset = select_regional_OD(dataset, amsterdam=amsterdam, region=region)
    whole_day_dataset = dataset
    dataset = select_timeperiod(dataset, sel_timeperiod)

# --- Fill in ranges ---
    dataset = fill_in_ranges(dataset, rng_est=rng_est, rng_bins=rng_bins)

# --- Add daily median travel time to data set to enable removing outliers ---
    HB_median_avg_tt = []
    for i in amsterdam:
        for j in region:
            HB_median_avg_tt.append(daily_tt(whole_day_dataset, checkin=[i], checkout
                ↪ =[j], rng_est=rng_est, rng_bins=rng_bins)[1][0])
```

```
            HB_median_avg_tt.append(daily_tt(whole_day_dataset, checkin=[j], checkout
                ↪ =[i], rng_est=rng_est, rng_bins=rng_bins)[1][0])
    HB_median_avg_tt = pd.DataFrame(HB_median_avg_tt, columns=['cki_zone', 'cko_zone
        ↪ ', 'median travel time over day', 'avg travel time over day'])

    dataset = pd.merge(dataset, HB_median_avg_tt, left_on=['cki_zone', 'cko_zone'],
        ↪ right_on=['cki_zone', 'cko_zone'], how='outer')

# --- Remove outliers ---

# --- Drop rows in which the avg travel time meets the outliers criterium: ---
# 1-50 travelers in half hour block &
# (1-outliers_rel) * median travel time over day > median travel time in half hour
    ↪ block OR median travel time in half hour block > (1+outliers_rel) * median
    ↪ travel time over day &
# |median travel time in half hour block - median travel time over day| >
    ↪ outliers_abs
    dataset = dataset[~((dataset['aantal_reizen'] <= 50) &
                    (((1-outlier_rel) * dataset['median travel time over day'] <
                        ↪ dataset['mediaan_reistijd_seconden']) | ((1+outlier_rel) *
                        ↪ dataset['median travel time over day'] > dataset['
                        ↪ mediaan_reistijd_seconden'])) &
                    (np.abs(dataset['mediaan_reistijd_seconden'] - dataset['median
                        ↪ travel time over day']) > outlier_abs))]

# --- Calculate average travel time, total, lower bound and estimated number of
    ↪ travelers per OD-relation ---
# --- From amsterdam to region ---
    agg_tt = []
    for idx_i, i in enumerate(amsterdam):
        subset_amsterdam = dataset.loc[(dataset['cki_zone']==i)].copy()
        for idx_j, j in enumerate(region):
            subset_amsterdam_region = subset_amsterdam.loc[(subset_amsterdam['cko_zone
                ↪ ']==j)].copy()

            if subset_amsterdam_region.empty == True:
                pass
            else:
                exact_counts, estimated_part, total = exact_estimated(
                    ↪ subset_amsterdam_region, rng_est=rng_est, lb=[1, 51, 101, 151,
                    ↪ 201, 251])
                subset_amsterdam_region['aantal * gem_reistijd'] =
                    ↪ subset_amsterdam_region['aantal_reizen'] *
                    ↪ subset_amsterdam_region['gemiddelde_reistijd_seconden']
                agg_tt.append([i, j, round(subset_amsterdam_region['aantal *
                    ↪ gem_reistijd'].sum()/subset_amsterdam_region['aantal_reizen'].
                    ↪ sum(),1), subset_amsterdam_region['aantal_reizen'].sum(),
                    ↪ exact_counts, estimated_part])

# --- From region to amsterdam ---
    for idx_i, i in enumerate(region):
        subset_region = dataset.loc[(dataset['cki_zone']==i)].copy()
        for idx_j, j in enumerate(amsterdam):
            subset_region_amsterdam = subset_region.loc[(subset_region['cko_zone']==j)
                ↪ ].copy()
```

```
            if subset_region_amsterdam.empty == True:
                pass
            else:
                exact_counts, estimated_part, total = exact_estimated(
                    ↪ subset_region_amsterdam, rng_est=rng_est, lb=[1, 51, 101, 151,
                    ↪ 201, 251])
                np.seterr(all= 'ignore')
                subset_region_amsterdam['aantal * gem_reistijd'] =
                    ↪ subset_region_amsterdam['aantal_reizen'] *
                    ↪ subset_region_amsterdam['gemiddelde_reistijd_seconden']
                agg_tt.append([i, j, round(subset_region_amsterdam['aantal *
                    ↪ gem_reistijd'].sum()/subset_region_amsterdam['aantal_reizen'].
                    ↪ sum(),1), subset_region_amsterdam['aantal_reizen'].sum(),
                    ↪ exact_counts, estimated_part])

    HB_tt_table = pd.DataFrame(agg_tt, columns=['cki_zone', 'cko_zone', 'avg travel
        ↪ time','total # travelers', 'lowerbound travelers', 'estimated travelers'])

# --- Add column that can be used to select representativene OD-relations ---
    HB_tt_table['tt based on lowerbound travelers'] = HB_tt_table['lowerbound
        ↪ travelers']
    HB_tt_table['tt based on estimated travelers'] = HB_tt_table['estimated travelers
        ↪ ']

    HB_tt_matrix = pd.pivot_table(HB_tt_table, index='cki_zone', columns='cko_zone',
        ↪ values='avg travel time')
    return HB_tt_matrix, HB_tt_table, num_outliers, perc_outliers


# In[ ]:


# --- Merge before and after table ---
def table_after_before(dataset_before, dataset_after, sel_timeperiod, amsterdam,
    ↪ region, rng_est):
    HB_tt_matrix_before, HB_tt_table_before, num_outliers_before,
        ↪ perc_outliers_before = tt_matrix(dataset_before, sel_timeperiod, amsterdam,
        ↪  region, rng_est=rng_est)
    before_total = HB_tt_table_before.iloc[:,0:8]

    HB_tt_matrix_after, HB_tt_table_after, num_outliers_after, perc_outliers_after =
        ↪ tt_matrix(dataset_after, sel_timeperiod, amsterdam, region, rng_est=rng_est
        ↪ )
    after_total = HB_tt_table_after.iloc[:,0:8]

# --- Rename columns ---
    table = pd.merge(before_total, after_total, left_on=['cki_zone', 'cko_zone'],
        ↪ right_on=['cki_zone', 'cko_zone'], how='outer')
    table.columns = ['Check in cluster', 'Check out cluster', 'Travel time before', '
        ↪ Travelers before', 'Travelers before lowerbound', 'Travelers before
        ↪ estimated', 'Representativeness tt before lowerbound', 'Representativeness
        ↪ tt before estimated', 'Travel time after', 'Travelers after', 'Travelers
        ↪ after lowerbound', 'Travelers after estimated', 'Representativeness tt
        ↪ after lowerbound', 'Representativeness tt after estimated']

# --- Convert number of travelers to number of travelers per day ---
```

```
    table['Travelers before'] = table['Travelers before']/ndays_2018Q2
    table['Travelers before lowerbound'] = table['Travelers before lowerbound']/
        ↪ ndays_2018Q2
    table['Travelers before estimated'] = table['Travelers before estimated']/
        ↪ ndays_2018Q2
    table['Travelers after'] = table['Travelers after']/ndays_2019Q1
    table['Travelers after lowerbound'] = table['Travelers after lowerbound']/
        ↪ ndays_2019Q1
    table['Travelers after estimated'] = table['Travelers after estimated']/
        ↪ ndays_2019Q1

# --- Calculate absolute and relative change in travel time ---
    table['Change travel time absolute'] = table['Travel time after'].subtract(table
        ↪ ['Travel time before'])
    table['Change travel time relative'] = (table['Change travel time absolute']/
        ↪ table['Travel time before']).multiply(100).round(1)

# --- Calculate absolute and relative change in travelers ---
    table['Change travelers absolute'] = table['Travelers after'].subtract(table['
        ↪ Travelers before'])
    table['Change travelers relative'] = (table['Change travelers absolute']/table['
        ↪ Travelers before']).multiply(100).round(1)

    return table


# In[ ]:
```

Z

Code route data processing

```
#!/usr/bin/env python
# coding: utf-8

# In[ ]:


# --- Import packages ---
import pandas as pd
import scipy
import scipy.stats as ss
import math
import numpy as np
from varname import nameof


# In[ ]:


# --- Variables ---
data_2018Q2 = dataset before Noord/Zuidlijn
data_2019Q1 = dataset after Noord/Zuidlijn
data_2018Q2.name = '2018Q2'
data_2019Q1.name = '2019Q1'

cluster_info = dataset cluster info
cluster_info['name_type'] = cluster_info['name']+ ' ' + cluster_info['type']

# --- Half hour blocks per time period ---
morning = ['7.00-7.30', '7.30-8.00', '8.00-8.30', '8.30-9.00']
evening = ['16.00-16.30', '16.30-17.00', '17.00-17.30', '17.30-18.00']
rest_day = ['0.00-0.30', '0.30-1.00', '1.00-1.30', '1.30-2.00', '2.00-2.30',
    ↪ '2.30-3.00', '3.00-3.30', '3.30-4.00', '4.00-4.30', '4.30-5.00', '5.00-5.30',
    ↪ '5.30-6.00', '6.00-6.30', '6.30-7.00', '9.00-9.30', '9.30-10.00',
    ↪ '10.00-10.30', '10.30-11.00', '11.00-11.30', '11.30-12.00', '12.00-12.30',
    ↪ '12.30-13.00', '13.00-13.30', '13.30-14.00', '14.00-14.30', '14.30-15.00',
    ↪ '15.00-15.30', '15.30-16.00', '18.00-18.30', '18.30-19.00', '19.00-19.30',
    ↪ '19.30-20.00', '20.00-20.30', '20.30-21.00', '21.00-21.30', '21.30-22.00',
    ↪ '22.00-22.30', '22.30-23.00', '23.00-23.30', '23.30-0.00']
whole_day = ['0.00-0.30', '0.30-1.00', '1.00-1.30', '1.30-2.00', '2.00-2.30',
    ↪ '2.30-3.00', '3.00-3.30', '3.30-4.00', '4.00-4.30', '4.30-5.00', '5.00-5.30',
```

```
    ↪ '5.30-6.00', '6.00-6.30', '6.30-7.00', '7.00-7.30', '7.30-8.00', '8.00-8.30',
    ↪ '8.30-9.00', '9.00-9.30', '9.30-10.00', '10.00-10.30', '10.30-11.00',
    ↪ '11.00-11.30', '11.30-12.00', '12.00-12.30', '12.30-13.00', '13.00-13.30',
    ↪ '13.30-14.00', '14.00-14.30', '14.30-15.00', '15.00-15.30', '15.30-16.00',
    ↪ '16.00-16.30', '16.30-17.00', '17.00-17.30', '17.30-18.00', '18.00-18.30',
    ↪ '18.30-19.00', '19.00-19.30', '19.30-20.00', '20.00-20.30', '20.30-21.00',
    ↪ '21.00-21.30', '21.30-22.00', '22.00-22.30', '22.30-23.00', '23.00-23.30',
    ↪ '23.30-0.00']

# --- Add column 'dagdeel' to dataframe based on values for column 'tijdsperiode' ---
data_2018Q2['dagdeel'] = np.where(data_2018Q2['tijdsperiode'].isin(morning), '
    ↪ ochtendspits',
                                  np.where(data_2018Q2['tijdsperiode'].isin(evening), '
                                      ↪ avondspits', 'restdag'))

data_2019Q1['dagdeel'] = np.where(data_2019Q1['tijdsperiode'].isin(morning), '
    ↪ ochtendspits',
                                  np.where(data_2019Q1['tijdsperiode'].isin(evening), '
                                      ↪ avondspits', 'restdag'))

# --- Selection of region clusters and Amsterdam clusters ---
all_clusters = ['0', '1', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19',
    ↪  '2', '20', '21', '22', '23', '24', '25', '26', '27', '28', '29', '3', '30',
    ↪ '31', '32', '33', '34', '35', '36', '37', '38', '39', '40', '41', '42', '43',
    ↪ '44', '45', '46', '47', '48', '49', '5', '50', '51', '52', '53', '54', '55',
    ↪ '56', '57', '58', '59', '6', '60', '61', '62', '63', '64', '65', '66', '67',
    ↪ '68', '69', '7', '70', '71', '72', '73', '74', '75', '76', '77', '78', '79',
    ↪ '8', '80', '81', '82', '83', '84', '85', '86', '87', '88', '9', 'ah', 'alm', '
    ↪ almm', 'almp', 'amf', 'amr', 'asa', 'asb', 'asd', 'asdh', 'asdl', 'asdm', '
    ↪ asdz', 'ass', 'assp', 'bkl', 'brn', 'bv', 'dmn', 'dmnz', 'dv', 'dvd', 'hfd', '
    ↪ hil', 'hlm', 'hn', 'ht', 'hvs', 'hwzb', 'ledn', 'lls', 'ndb', 'pmr', 'rai', '
    ↪ rm', 'rtd', 'shl', 'ssh', 'tb', 'ut', 'utg', 'vl', 'wp', 'zd', 'zl']
cluster_region = region clusters
cluster_amsterdam = amsterdam clusters

# --- Number of working days in each data set ---
ndays_2018Q2 = 25
ndays_2019Q1 = 45


# In[ ]:



# --- Select day and time period ---
def select_timeperiod(dataset, sel_tijdsperiode=['Werkdag (buiten schoolvakanties)
    ↪ avondspits',
      'Werkdag (buiten schoolvakanties) ochtendspits',
      'Werkdag (buiten schoolvakanties) restdag']):
    dataset = dataset[dataset.loc[:,'tijdsperiode'].isin(sel_tijdsperiode)].copy()
    return dataset

# --- Select regional OD-relations ---
def sub_HB_relations(dataset, amsterdam=cluster_amsterdam, region=cluster_region):
    dataset = select_timeperiod(dataset)
    dataset = dataset[((dataset.loc[:, 'check_in_halte_groep'].isin(amsterdam)) & (
        ↪ dataset.loc[:, 'check_uit_halte_groep'].isin(region)))
```

```
                             | ((dataset.loc[:, 'check_in_halte_groep'].isin(region)) &
                                  ↪  (dataset.loc[:, 'check_uit_halte_groep'].isin(
                                  ↪ amsterdam)))].copy()
    return dataset

# --- Fill in ranges ---
def fill_in_ranges(dataset, rng_est, rng_bins=['1 - 50', '51 - 100', '101 - 150',
    ↪ '151 - 200', '201 - 250', '251 - 300'], lb = [1, 51, 101, 151, 201, 251]):
    dataset = dataset.copy()
    dataset = select_timeperiod(dataset)
    for h in range(len(rng_bins)):
        dataset.loc[:,'aantal_reizen_op_deze_route'].replace(to_replace=rng_bins[h],
            ↪ value=rng_est[h], inplace=True)
    dataset.loc[:,'aantal_reizen_op_deze_route'] = dataset.loc[:, '
        ↪ aantal_reizen_op_deze_route'].apply(pd.to_numeric)
    dataset['lowerbound'] = dataset['aantal_reizen_op_deze_route']
    for idx, i in enumerate(rng_est):
        dataset.loc[(dataset.loc[:, 'aantal_reizen_op_deze_route'] == rng_est[idx]), '
            ↪ lowerbound'] = lb[idx]
        dataset.loc[(dataset.loc[:, 'aantal_reizen_op_deze_route'] == rng_est[idx]), '
            ↪ estimated'] = rng_est[idx] - lb[idx]
    return dataset

# --- Add column with the number of transfers ---
def add_number_transfer_route(dataset):
    dataset.loc[(dataset.loc[:,'rit_5_check_uit_halte_groep'].notna()), 'Number of
        ↪ transfers in route'] = 4
    dataset.loc[(dataset.loc[:,'rit_4_check_uit_halte_groep'].notna()) & (dataset.loc
        ↪ [:,'rit_5_check_uit_halte_groep'].isna()), 'Number of transfers in route']
        ↪ = 3
    dataset.loc[(dataset.loc[:,'rit_3_check_uit_halte_groep'].notna()) & (dataset.loc
        ↪ [:,'rit_4_check_uit_halte_groep'].isna()), 'Number of transfers in route']
        ↪ = 2
    dataset.loc[(dataset.loc[:,'rit_2_check_uit_halte_groep'].notna()) & (dataset.loc
        ↪ [:,'rit_3_check_uit_halte_groep'].isna()), 'Number of transfers in route']
        ↪ = 1
    dataset.loc[(dataset.loc[:,'rit_1_check_uit_halte_groep'].notna()) & (dataset.loc
        ↪ [:,'rit_2_check_uit_halte_groep'].isna()), 'Number of transfers in route']
        ↪ = 0
    return dataset

# --- Count number of total, lowerbound and estimated number of transfers per route
    ↪ ---
def count_number_transfer_passengers(dataset):
    dataset['Total transfers'] = dataset['aantal_reizen_op_deze_route'] * dataset['
        ↪ Number of transfers in route']
    dataset['Total transfers lowerbound'] = dataset['lowerbound'] * dataset['Number
        ↪ of transfers in route']
    dataset['Total transfers estimated'] = dataset['estimated'] * dataset['Number of
        ↪ transfers in route']
    return dataset

# --- Remove outliers ---
def remove_outliers(dataset,replace_1_50):

# --- Remove routes with modal split: train - BTM - train ---
```

```
    dataset = dataset.loc[~((dataset.loc[:,'rit_1_modaliteit'] == 'Trein') & (dataset
        ↪ .loc[:,'rit_2_modaliteit'] != 'Trein') & (dataset.loc[:,'rit_3_modaliteit']
        ↪  == 'Trein') & (dataset.loc[:,'aantal_reizen_op_deze_route'] == '1 - 50'))
        ↪ ,:]
    dataset = dataset.loc[~((dataset.loc[:,'rit_2_modaliteit'] == 'Trein') & (dataset
        ↪ .loc[:,'rit_3_modaliteit'] != 'Trein') & (dataset.loc[:,'rit_4_modaliteit']
        ↪  == 'Trein') & (dataset.loc[:,'aantal_reizen_op_deze_route'] == '1 - 50'))
        ↪ ,:]
    dataset = dataset.loc[~((dataset.loc[:,'rit_3_modaliteit'] == 'Trein') & (dataset
        ↪ .loc[:,'rit_4_modaliteit'] != 'Trein') & (dataset.loc[:,'rit_5_modaliteit']
        ↪  == 'Trein') & (dataset.loc[:,'aantal_reizen_op_deze_route'] == '1 - 50'))
        ↪ ,:]

# --- Remove 4 transfers or more and 1-50 travelers ---
    dataset = add_number_transfer_route(dataset)
    dataset = dataset.loc[~((dataset.loc[:, 'Number of transfers in route'] == 4) & (
        ↪ dataset.loc[:, 'aantal_reizen_op_deze_route'] == replace_1_50))]

# --- Remove if a direct route is possible for OD-relation and 3 transfers or more
    ↪ and 1-50 travelers ---
    df_direct_OD = dataset[dataset.loc[:,'Number of transfers in route'] == 0].loc[:,
        ↪  ['check_in_halte_groep', 'check_uit_halte_groep']].drop_duplicates()
    df_direct_OD['Direct exists'] = 'Yes'

    dataset = dataset.merge(df_direct_OD, left_on=['check_in_halte_groep', '
        ↪ check_uit_halte_groep'], right_on=['check_in_halte_groep', '
        ↪ check_uit_halte_groep'], how='left')

    dataset = dataset.loc[~((dataset.loc[:,'Direct exists'] == 'Yes') & (dataset.loc
        ↪ [:,'Number of transfers in route'] >= 3) & (dataset.loc[:, '
        ↪ aantal_reizen_op_deze_route'] == replace_1_50))]

# --- Remove routes that only occur in one time period and more than 3 transfers and
    ↪ 1-50 travelers ---
    dataset = dataset.loc[~((dataset.duplicated(subset=['kwartaal', '
        ↪ check_in_halte_groep', 'check_uit_halte_groep',
       'rit_1_check_in_halte_groep',
       'rit_1_check_uit_halte_groep', 'rit_1_modaliteit', 'rit_1_vervoerder',
       'rit_2_check_in_halte_groep', 'rit_2_check_uit_halte_groep',
       'rit_2_modaliteit', 'rit_2_vervoerder', 'rit_3_check_in_halte_groep',
       'rit_3_check_uit_halte_groep', 'rit_3_modaliteit', 'rit_3_vervoerder',
       'rit_4_check_in_halte_groep', 'rit_4_check_uit_halte_groep',
       'rit_4_modaliteit', 'rit_4_vervoerder', 'rit_5_check_in_halte_groep',
       'rit_5_check_uit_halte_groep', 'rit_5_modaliteit', 'rit_5_vervoerder'], keep=
           ↪ False) == False) & (dataset.loc[:,'aantal_reizen_op_deze_route'] ==
           ↪ replace_1_50) & (dataset.loc[:,'Number of transfers in route'] >= 3)),:]

    return dataset


# In[ ]:


def f(x):
    d = {}
    d['Total lowerbound'] = x['lowerbound'].sum()
```

```
    d['Total estimated'] = x['estimated'].sum()
    d['Avg transfer'] = x['Total transfers'].sum()/x['aantal_reizen_op_deze_route'].
        ↪ sum()
    return pd.Series(d, index=['Total lowerbound', 'Total estimated', 'Avg transfer
        ↪ '])

def avg_transfer_OD(dataset, rng_est=[3, 65, 120, 175, 225, 275]):
    dataset = select_timeperiod(dataset)
    dataset = sub_HB_relations(dataset)
    dataset = remove_outliers(dataset, '1 - 50')
    dataset = fill_in_ranges(dataset, rng_est=rng_est)
    dataset = add_number_transfer_route(dataset)
    dataset = count_number_transfer_passengers(dataset)
    avg_transfer = dataset.groupby(by=['check_in_halte_groep', 'check_uit_halte_groep
        ↪ ']).apply(f)

    return avg_transfer

def comp_avg_transfer_OD(dataset_before, dataset_after, rng_est=[3, 65, 120, 175,
    ↪ 225, 275]):
    avg_transfer_2018Q2 = avg_transfer_OD(dataset_before, rng_est)
    avg_transfer_2019Q1 = avg_transfer_OD(dataset_after, rng_est)

    comp_avg_transfer = pd.concat([avg_transfer_2018Q2, avg_transfer_2019Q1], axis=1)
    comp_avg_transfer.columns = ['Total lowerbound travelers 2018Q2', 'Total
        ↪ estimated travelers 2018Q2', 'Avg transfer 2018Q2', 'Total lowerbound
        ↪ travelers 2019Q1', 'Total estimated travelers 2019Q1', 'Avg transfer 2019Q1
        ↪ ']
    comp_avg_transfer['absolute difference'] = comp_avg_transfer['Avg transfer 2019Q1
        ↪ '] - comp_avg_transfer['Avg transfer 2018Q2']

# --- Compute average number of travelers ---
    comp_avg_transfer['avg travelers'] = ((comp_avg_transfer['Total lowerbound
        ↪ travelers 2018Q2'] + comp_avg_transfer['Total estimated travelers 2018Q2'])
        ↪ /ndays_2018Q2 + (comp_avg_transfer['Total lowerbound travelers 2019Q1'] +
        ↪ comp_avg_transfer['Total estimated travelers 2019Q1'])/ndays_2019Q1)/2
    comp_avg_transfer = comp_avg_transfer.reset_index()

# --- Add check in and check out information ---
    comp_avg_transfer = comp_avg_transfer.merge(cluster_info.iloc[:,0:8], left_on='
        ↪ check_in_halte_groep', right_on='code', how='left').rename(columns={'x':'
        ↪ cki_x', 'y':'cki_y', 'name':'cki_name', 'type':'cki_type', 'area':'cki_area
        ↪ ', 'concession':'cki_concession', 'analysis_area':'cki_analysis_area', '
        ↪ geographical area':'cki_geographical_area'})
    comp_avg_transfer = comp_avg_transfer.merge(cluster_info.iloc[:,0:8], left_on='
        ↪ check_uit_halte_groep', right_on='code', how='left').rename(columns={'x':'
        ↪ cko_x', 'y':'cko_y', 'name':'cko_name', 'type':'cko_type', 'area':'cko_area
        ↪ ', 'concession':'cko_concession', 'analysis_area':'cko_analysis_area', '
        ↪ geographical area':'cko_geographical_area'})


    return comp_avg_transfer
```