THESIS

Evaluating and Predicting the Impact of Roadworks Using Mobile Phone Data

Department of Information and Computing Sciences

Faculty of Science

March 2016

Supervisor
Prof. Dr. Arno Siebes
(Utrecht University)

Second Supervisor
Dr. Marco R. Spruit
(Utrecht University)

Author
Johan Meppelink
johanmeppelink@gmail.com
Master business Informatics

External Supervisor
Wim Steenbakkers
(Mezuro)

Second External Supervisor
Menno de Pater
(Decisio)

Universiteit Utrecht mezuro DECISIO

Declaration of Authorship

The chapters 5, 6, and 8 of this thesis are largely the result of collaborative research performed by fellow graduating student Jens van Langen and myself, Johan Meppelink. The research presented in these three chapters formed a fundamental part of both of our research projects. Therefore, the research has been performed in collaboration. Jens van Langen wrote the first draft of the Data Quality (chapter 5), the Scaling chapter (chapter 6), and Appendix D. These chapters I then rewrote to better fit in with the remainder of the research. The comparison with GPS (chapter 5 section 5.5) and evaluation with road side measurement data (chapter 6 section 6.2) were written by me. Furthermore, I wrote the initial version of the Label chapter (chapter 8), which Jens van Langen adapted for his research. The other chapters and content presented in this thesis are the product of my own work.

Abstract

Roadworks affect road users all over the globe impacting society as a whole through loss of valuable time (Ministerie van Infrastructuur en Milieu, 2015; Schrank, Eisele, & Lomax, 2012). To improve our understanding of the impact roadworks on society as a whole, we need to move to a new source of information. Current techniques, such as surveys and road side measurements, require a lot of effort and resources to investigate the impact of a single roadwork (Taale, Schuurman, & Bootsma, 2002; Cáceres, Wideberg, Benitez, 2007). The costs of traditional technique imply research on the true economic impact of roadworks is only scarcely performed. Hence, we are in need of an alternative source of information if we want to learn more about the impact of roadworks. In this research we propose mobile phone data, i.e. mobility data extracted from Call Detail Records (CDRs), as a viable alternative.

We will present a method to measure the impact of roadworks using mobile phone data. Furthermore, we validate the presented method by comparing the outcomes with traditional information sources such as surveys, road side measurements, and GPS traces. The standard mobile phone data will be fine-tuned to elicit more accurate origins and destinations. After fine-tuning, we find the mobile phone data delivers results similar to the traditional sources with much greater ease and at unprecedented scale. Moreover, we show we can enrich the mobile phone data with data about crucial trip motives, e.g. home-to-work, previously only present in mobility surveys. These motives can then be used to measure the economic impact to society rather than travel time loss (Kennisinstituut voor Mobiliteitsbeleid, 2013). Where traditional techniques would require months of research to measure the impact of one roadwork; we show that mobile phone data can measure hundreds with a fraction of the time and effort.

The rich and plentiful information present in the mobile phone data will also be used to predict the impact of roadworks. Using this new found source of information, we investigate the underlying structures that result in delays and uncertainties in travel times due to roadworks. We, for one, create models that explain up to 45% of the variation in the measured impact and suggest research directions to further increase this percentage.

Acknowledgements

First of all, I would like to say thank you to everyone who contributed to this body of research through feedback, advice, data, and interviews. I would like to thank my supervisors for their constant feedback and support. Thank you prof. dr. Arno Siebes, dr. Marco Spruit, ir. Wim Steenbakkers, and MSc Menno de Pater. In particular, I would like to thank prof. dr. Arno Siebes for the regular meetings and thorough feedback on prior versions of this thesis. Wim and Menno, thank you for your help and the wonderful opportunity, allowing me to do my research at Mezuro and Decisio. Furthermore, I would like to thank everyone at Mezuro for the laughs and discussions, especially Jasper Keij. And last but not least; thank you, Jens van Langen, for the amazing collaboration!

Contents

1 Background

Roadworks are a necessary evil affecting road users all over the globe. Large investments in the road network are required to increase road capacities to support the growing number of road users (Schrank, Eisele, & Lomax, 2012). Without these investments the roads will become overly congested resulting in waste of valuable time and resources. In the US a study by Schrank et al. (2012) found the average road user already spends an unnecessary 38 hours a year in congested traffic. This is equivalent to a financial loss of up to 121 billion USDs (Schrank et al., 2012). Roadworks are performed for maintaining and expanding the road network to help alleviate congestion. However, when the roadworks are occurring the road capacity is temporarily reduced (Nagel & Schreckenberg, 1992). This by itself has a negative impact on road capacity and thus road users (Nagel & Schreckenberg, 1992; Calvert, 2010). In the Netherlands about 4.3% of all traffic jams are a direct result of roadworks (Ministerie van Infrastructuur en Milieu, 2015). This figure does not include people taking detours, driving slower, staying at home, and using other modes of transportation. By not looking at the larger picture there is a clear tendency to underestimate the impact of roadworks.

The hindrance caused by roadworks is a significant part of the true cost of roadworks (Schrank et al., 2012). Therefore, judging tenders should include factors other than price and perhaps quality to minimize the total impact roadworks have on society and the economy. A tender is a bid typically by a construction company stipulating the conditions under which they want to perform the job. The organization who wants the roadworks to be performed, judges the tenders and is tasked to choose one. From April 2016 onwards the European Union (EU) guidelines state that the tender to be chosen is the one that is the most economically advantageous, i.e. the Most Economically Advantageous Tender (MEAT) (De Koning, 2015; Crow Commercial Service, 2015). In addition to MEAT, there is another common tender method called Best Value Procurement. Best Value Procurement differs in approach, but is also judged mainly on being able to show you can meet the stated criteria such as impact on traffic flow, hindrance to the direct environment, and costs (interview with a tender manager, Appendix A). Of these criteria traffic hindrance often weighs heaviest (Taskforce Doorstroming, 2009; Crown Commercial Service, 2015). With Best Value Procurement "[t]he more confident the contractor is that we will meet the set criteria the higher the fictive discount will be" (Appendix A). In 2011, for example, the winning tender on a project of about €100 million was the one that got a fictive discount of €30.8 million because the tender involved a plan to minimize traffic hindrance (Duijnisveld, Peijs & Calvert, 2011).

Examples of initiatives to alleviate traffic hindrance are working outside peak hours, performing roadworks in stages, and paying road users to avoid driving in rush hours (Minchin, Thurn, Ellis & Lewis, 2013; Federal Highway Administration, 2003; Spitsmijden, 2009; Rijkswaterstaat, 2009b). Working at night might not have a significant impact on the cost of roadworks as increased wages are countered with an increase in productivity (Minchin, Thurn, Ellis & Lewis, 2013;

Federal Highway Administration, 2003). The latter two, i.e. performing roadworks in stages and paying road users to avoid rush hours, do result in added costs (Federal Highway Administration, 2003; Spitsmijden, 2009). Whether initiatives to alleviate traffic hindrance are worth their cost depends on how much hindrance can be avoided. For this it becomes highly important to being capable of making good predictions of traffic hindrance under different circumstances to see how traffic hindrance, and the total cost of roadworks, can be minimized.

The task of predicting traffic hindrance resulting from roadworks is nontrivial, largely because of a lack of good data about the impact of traffic hindrance during roadworks. In the Netherlands, where the MEAT procedure is already in place, the traffic hindrance associated with the research approaches in the tenders are often evaluated beforehand using simulation software (expert interview, Appendix B; Duijnisveld, Peijs & Calvert, 2011). Unfortunately, an evaluation study to investigate the impact of the roadworks is hardly ever performed (Appendix B). Hence, without these studies there is no way to test if the predictions are accurate and learn from them such that prediction can become more accurate in the future. If the predictions of the simulation tools would have been proven to be consistently accurate, or accurate enough, evaluation would be less important, but this is not the case. One study by Rijkswaterstaat, who manages the Dutch road network and is responsible for judging tenders for roadworks, included also some early predictions (Rijkswaterstaat, 2009a). The study showed the early predictions for maximum time loss and traffic jam length experienced by the road user was overestimated by 89% and 81%, respectively (Rijkswaterstaat, 2009a). Note this is also the only study by Rijkswaterstaat that to our knowledge compared the early predictions with the outcomes of the evaluation study. The observed difference is substantial and can be seen as an indication that we need to improve upon these prediction models foremost to reduce the impact of future roadworks.

2 Problem statement

2.1 Measuring the impact of roadworks

A number of large evaluation studies have been performed to investigate the impact of roadworks in the Netherlands (Taale et al., 2002; Rijkswaterstaat, 2009a). The large evaluation studies often performed by Rijkswaterstaat nearly always consisted of a combination of surveys and roadside measurements (Taale, Schuurman, & Bootsma, 2002; Rijkswaterstaat, 2009a). The reason both techniques have to be applied in the current situation is that they complement each other's weaknesses. Roadside measurements can, for example, not show what type of road user is delayed, which matters as the cost of delay is greater for some users than others (Kennisinstituut voor Mobiliteitsbeleid, 2013). However, they do provide very accurate and detailed counts on vehicles passing road section. Surveys, on the contrary, provide very rich contextual information, e.g. how many travellers take the car versus the train. However, they are often small in sample size as they are expensive. Moreover, surveys provide only a snapshot of the situation implying they will have to be repeated for each distinct roadwork (Cáceres et al., 2007). Surveys thus measure people going from A to B and roadside measurements provide information about the traffic on the road (figure 2.1).

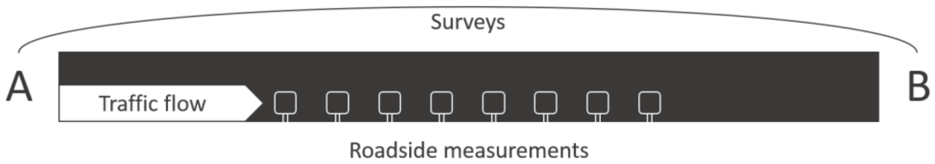


Figure 2.1, overview of existing mobility measurement techniques

The downside of using roadside measurements and, particularly, surveys is that these studies become time consuming and expensive, e.g. a large evaluation study in the Netherlands reportedly costed €400.000 (Cáceres et al., 2007; Taale, Schuurman, & Bootsma, 2002). Consequently these studies are scarcely performed, limiting our understanding of what the impact of roadworks is and how we can reduce their impact on society (Appendix B).

In this research we want to propose a novel method to measure the impact of roadworks that could drastically reduce the amount of labour required to evaluate the impact of roadworks. To do this we want to use mobile phone data. This data source contains information on how mobile phones and, assuming one phone per person, people move about**.** Using mobile phone data we can plausibly see how people travel under normal circumstances and compare this with, for example, behaviour and travel times when roadworks are on their path.

The idea to use mobile phones for traffic management have been around for some time now (Astarita & Florian, 2001). Cáceres et al. (2007), for example, already tested and showed using a simulation study that GSM data can be used to create high-quality Origin-Destination matrices (OD-matrices). One of the main advantages they see in the new technology is that the entire infrastructure for measuring the movement of phones is already in place (Cáceres et al., 2007). Another advantage, as mentioned by Astarita and Florian (2001), is the great diffusion and penetration of mobile phones.

The data available for this study covers about a fifth of the Dutch population from September and October 2015 dwarfing the 0.2% of the population surveyed by Onderzoek Verplaatsingen in Nederland (OViN), i.e. one of the largest mobility survey in the Netherlands (CBS & RWS, 2015). In contrast to OViN, the mobile phone data contains information about individuals over a longer time period. Longitudinal data is beneficial as it can provide insight into how behaviour changes before after and during an event. Large sample sizes and longitudinal data is already available with the roadside measurements. However, mobile phone data provides much more contextual information. Unlike roadside measurements, it can reveal when and where a person departs and arrives. In addition, mobile phone data is not confined to the roads and can thus measure other modes of transportation just as well (Keij, 2014). Most importantly though is that the cost profile of studying mobility using mobile phone data is completely different from that of the traditional methods. The initial effort needed to transform the raw phone data into mobile phone data might be substantial, but after the initial phase the variable cost of looking at one versus many roadworks is relatively small. Once the method is established it will become possible to cost-effectively monitor and evaluate the impact of many roadworks by using the data, i.e. Call Detail Records that mobile phone carriers are already storing for payment registration.

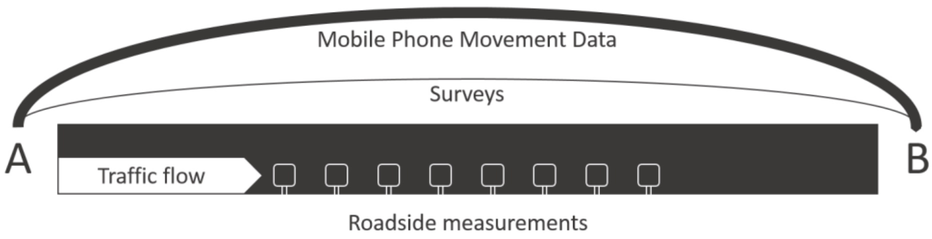


Figure 2.2, overview of new and existing mobility measurement techniques

As depicted in figure 2.2, mobile phone data provides similar information about travel behaviour as surveys, including a lot of contextual information, but at a much larger scale.

2.2 Predicting the impact of roadworks

Predicting the impact of roadworks is a nontrivial task. The study by Rijkswaterstaat (2009a) showed their early estimation for the impact of the roadworks were off by quite a margin. In literature similar unsatisfying findings are reported for predicting the impact of roadworks using simulation packages and or traffic models (Borchardt, Pesti, Sun & Ding, 2009; Calvert, 2010).

Borchardt et al. (2009) tested the accuracy of three simulation tools to estimate traffic hindrance. They used VISSIM, i.e. a widely used microscopic multi-model traffic flow simulation package, as a baseline for evaluating the other two (Borchardt et al., 2009). One simulation package showed roughly similar results whereas the other deviated by as much as 52% for one of the four test sites (Borchardt et al., 2009). Whether any of the three models predicted the impact of roadworks correctly has not been investigated. These tools also require very accurate and extensive input to correctly simulate the real world. This is input that is hard to obtain requiring and expert to provide the correct input (Borchardt et al., 2009). Furthermore, several studies report that the models are highly sensitive to the input parameters meaning small changes in the input quickly lead to bad predictions (Park, Won, & Yun, 2005; Borchardt et al., 2009; Calvert, 2010).

Most of these simulation packages, and all three of the models tested by Borchardt et al. (2009) lack fundamental information on the total impact of roadworks. The majority of these simulation packages are designed to simulate traffic flows on a microscopic scale, i.e. on the level of single vehicles and single road sections. Calvert (2010), for example, investigated the effect of roadworks on travel time over the road section where the roadworks are taking place. The purpose of his research is to improve the predictions of travel time through work zones for use in navigation software (Calvert, 2010). Nevertheless, by focussing on the microscale the predictions miss out on important information such as how many people will stay at home, who takes another mode of transportation, and what type of road user is affected and by how much. Without this information the predictions provide only incomplete information about the true impact of roadworks.

There is a logical explanation for measuring the impact of roadworks on a microscopic rather than meso- or macroscopic scale. Even though there is few information about the true impact of roadworks, there is plenty on the microscopic scale. Information on traffic jams can be a prime source of information for designing models that predicts the chance, length, and duration of traffic jams resulting from roadworks.

As the study by Schrank et al. (2012) indicated, the loss of travel time due to road congestion has a large financial impact on the economy. Roadworks, although temporarily, reduce road capacity even further leading to more congestion and traffic jams (Nagel & Schreckenberg, 1992; Calvert, 2010). Prediction on the impact of roadworks might, therefore, be most valuable when judged on economic cost rather than traffic delay. To do so, one needs to know the composition of the traffic being affected as different types of road users are associated with different values (Kennisinstituut voor Mobiliteitsbeleid, 2013). Previous evaluation studies often do not take into account cost as the focus of the research may differ, e.g.

gaining knowledge about traffic hindrance and experiences by road users. Nevertheless, some of these studies do indicate that a cost benefit analysis to see if the hindrance caused can be justified would be valuable (Rijkswaterstaat, 2008). Typical simulation packages might provide information on changes in travel time, for example, but will not be able to state the total cost. The exact valuation of the travel delay would be very difficult. Mobile phone data can potentially provide a practical solution as trip motives might have distinct trip and travel characteristics. Relations between trip and traveller characteristics could potentially be extracted from mobility surveys such as the OViN (CBS & RWS, 2015). The surveys are too small in sample to say something about the composition of road users, especially in specific circumstances such as roadworks. Nevertheless, the sample could be sufficient to draw conclusions about trip characteristics in relation to trip motives.

We thus find there is a lack of models capable of providing accurate predictions of the impact of roadworks that take into account the entire financial picture. Current models are either very hard to set up, e.g. the micro-simulation models, resulting in false predictions or include only part of the total impact, e.g. travel time near the roadworks. This might be explained by the absence of data about the total impact of roadworks, but that might be solved using mobile phone data.

3 Research questions and scope

3.1 Research questions

The proposed research is shaped by the main research question. This main research question overarches the two main problems that are highlighted in the problem statement. These problems are: the lack of a scalable method to acquire more data about the true impact of roadworks and the inaccurate and incomplete predictions of the impact of roadworks by the current standards. The main research question is split into sub research questions that are more convenient to answer and easier to interpret. Answering the main research question will be the goal of the proposed research.

Main research question

How can mobile phone data be used to improve the measurement and prediction of the impact of roadworks on highways?

The first task that needs to be executed to answer the main research question is to acquire data on the impact of roadworks on highways. For this we want to use mobile phone data. To the best of our knowledge no method exists in literature or elsewhere that tries to measure the impact of roadworks using mobile phone data. Therefore, a new method will be developed and presented in this research. The first sub research question (sub research question 1) is aimed at answering this part of the proposed research.

Sub research question 1

How can mobile phone data be used to measure the impact of roadworks on highways?

The second task is creating, testing, and evaluating a model that aims to predict the impact of roadworks on highways. A series of subtasks are generally involved in creating, testing, and evaluating a model. First, we need to figure out what attributes could plausibly influence the impact of roadworks. Sub research question 2a is concerned with this subtask. Next a model will be trained using the attributes identified to plausible influence the impact of roadworks. The idea is that the model will optimally represent the underlying structure and relations between the attributes and the impact of roadworks on highways. The impact of roadworks will be determined using our own method that is the result of answering sub research question 1. The model will be evaluated in two ways. First, the model will be evaluated by quantifying how accurate its predictions are. Next, we will compare the predictions from our model with the current state of the art. These different types of evaluations are represented in sub research questions 2b and 2c, respectively.

Sub research question 2a

What is the relation between roadwork characteristics and the impact of roadworks on highways?

Sub research question 2b

How accurately can our model predict the impact of roadworks on highways using mobile phone data?

Sub research question 2c

How does our model compare to the current state of the art in predicting the impact of roadworks on highways?

3.2 Scope

The scope of this project is confined mainly by the strengths and weaknesses of the available data. We limit the investigation on the impact of roadworks to roadworks occurring on highways or larger roads in the Netherlands. The Netherlands is chosen as the mobile phone data is only available for the Netherlands. The roadside measurements made public by the Nederlandse Databank Wegverkeersgegevens (NDW) consists primarily of measurements made on large roads and highways in the Netherlands (NDW, 2015a). Moreover, in the mobile phone data the Netherlands is split in a large set of areas. These areas are roughly 5 km to 10 km in diameter and movement within these areas cannot be detected. The data is, therefore, less suitable to observe short distance movements and better for measuring trips covering larger distances, e.g. trips on highways. Roadworks in cities, for example, might hinder many local road users, but this group is unlikely to be represented in the mobile phone data. On highways people will often travel further and are likelier to be represented in the mobile phone data. Hence the focus of this research is roadworks on highways and larger roads in the Netherlands.

The effects of roadworks can spread to people that are not naturally using the road the roadworks are occurring on as is represented in figure 3.1. When people start to migrate to detour roads this results in added traffic on these roads. The added traffic may result in delays for users that naturally use these roads, which are only indirectly related to the roadworks. Determining what detour roads are affected is a nontrivial task. For this research we will focus solely on the people that use the road where the roadworks are occurring, i.e. the people traveling from A_2 to B_2 in figure 3.1. This is most likely also the group that will be most affected by the roadworks.

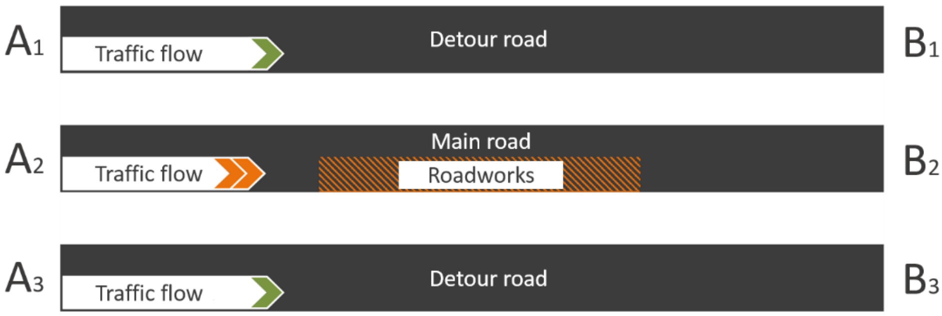


Figure 3.1, the road users influenced by the roadworks are mostly the people traveling from A_2 to B_2.

4 Research approach

The proposed research will cover two related, but distinct, solutions to the identified problems: (1) a new method to measure the impact of roadworks and (2) creating and evaluating a model to predict the impact of roadworks using mobile phone data. These problems both concern the impact of roadworks, but each will require a distinct research approach. The first is a typical design and action research type as it ultimately provides prescriptions on how to do something, i.e. how to use mobile phone data to measure the impact of roadworks (Gregor, 2006). The second leans more towards explanation and prediction research as it concerns predicting something, i.e. the impact of roadworks, by using testable propositions and testing causal relations (Gregor, 2006). The research approaches for each solution will be presented in sections 4.2 and 4.3, respectively.

Before this research can be conducted it is important to get a thorough understanding of the mobile phone data. When using the data we need to know how well it can measure how people are moving across the country. In particular, we want to know what trips are sure to be recorded and what trips are less likely to be noticed. In the section 4.1 we will go into how the quality and limitations of the data will be evaluated and a better understanding of the data used is created.

4.1 Data understanding and data quality

A vast amount of research has already been performed into investigating the use of mobile phone data to gather better insights into behaviour, social networks and mobility patterns of the masses (Daas et al., 2009; Snijkers, 2009; Ahas, Aasa, Roose, Mark, & Silm, 2008; Eagle, Pentland, & Lazer, 2009; Becker et al., 2011; Palchykov, Kaski, Kert´esz, Barab´asi, & Dunbar, 2012). Specifically, Human Mobility and Networks Lab (HUMNET) have numerous publications related to the use of mobile phone data to discover mobility patterns (Wang et al., 2012; Alexander, Jiang, Murga & González, 2015). However, there is no gold standard for creating and the mobile phone data and as such differences persists.

The most relevant description of the mobile phone data used in this study can currently be found in the works by Keij (2014) and Van Kats (2014) who did there Master theses at the organization, i.e. Mezuro, governing the data used in this study. Nevertheless, even here the data differs from the data used in this study. We have proposed and applied several techniques to improve the mobile phone data and as such the process has changed. Hence, we find it is worthwhile to again cover the process of creating the mobile phone data. This will also help to increase the replicability of this study and create a better understanding of the limitations in the data. In chapter 5 we will cover the process of creating the mobile phone data and stipulate the general data characteristics and limitations.

In chapter 5 we will also discuss the effects of the proposed improvements with a data quality study. The quality of the mobile phone data will be evaluated by comparing the origins and destinations with those acquired from accurate GPS measurements. The GPS trace is distilled to origins and destinations in similar fashion as the CDRs in the mobile phone data. The origins and destinations are thereafter compared by hand on accuracy and completeness. The results of the data quality research are reported at the end of chapter 5.

4.2 Measuring the impact of roadworks

The first section of the research will be about determining how Mobile phone data can be used to measure the impact of roadworks on highways, i.e. answering sub research question 1.

Sub research question 1

How can mobile phone data be used to measure the impact of roadworks on highways?

For this part of the research the structure has some strong parallels with the extension of the well-known Plan-Do-Check-Act cycle, or more appropriately the Plan-Do-Study-Act (PDSA) cycle (Moen & Norman, 2006). A concise explanation of the PDSA cycle is well formulated by Berwick (1996). He states "[t]he plan-do-study-act (PDSA) cycle describes, in essence, inductive learning - the growth of knowledge through making changes and then reflecting on the consequences of those changes" (Berwick, 1996, p.620).The extension is called the Model for Improvement and exists of two parts (Berwick, 1996). The first part consists of asking and answering the following three questions:

1. What are we trying to accomplish?
2. How will we know that a change is an improvement?
3. What changes can we make that will result in improvement?

In short the answers in our situation would be (1) creating a more scalable, accurate, and complete method to measure the impact of roadworks, (2) comparing the method to the current state of the art, and (3) employing and improving mobile phone data. The second part consists of doing the necessary work and is equivalent to the traditional PDSA cycle (figure 4.1).

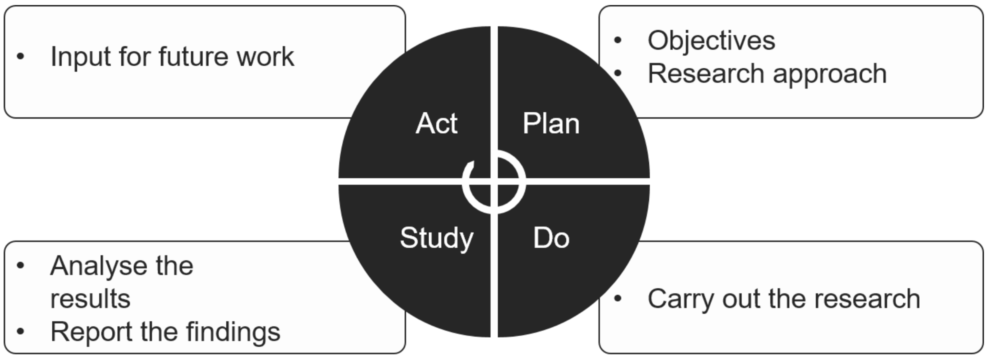


Figure 4.1, overview of the PDSA cycle, going clockwise from act to study, complete with a short description of what each phase entails.

The PDSA cycle is meant to continually revise and improve. In this study we will perform one cycle of the PDSA. We start the PDSA cycle with the Plan phase. In this phase we establish what the objectives are of the research and how to reach those objectives, i.e. in our case our research approach. Section 4.2.1 provides an overview of what needs to be taken into account when measuring the impact of roadworks. In section 4.2.2 we will describe on a conceptual level how Mobile phone data can be used to measure the impact of roadworks on highways. What will be discussed in section 4.2.2, however, is not an answer to sub research question 1, it is merely a proposal. To answer sub research question 1 we will have to validate the proposed method. By validating the proposed method we can check if we succeeded in creating a more scalable, accurate, and complete method to measure the impact of roadworks. In 4.2.3 through 4.2.7 we will go into detail on how the validation will be performed. Performing the research will be the next step and relates to the Do phase of the PDSA. After the conclusions we will provide input for future research by indicating where possible improvements can be made, i.e. the Act phase.

4.2.1 Economic impact of roadworks

The economic impact of roadworks result from a number of factors. The most profound are the economic impacts relating to travel delay and uncertainty in travel time (Kennisinstituut voor Mobiliteitsbeleid, 2013). When people arrive at work late or have to leave earlier to go to a business meeting valuable time is lost that could have been spent productively. The economic impact also depends on the motives behind the trips. When people are late for work or an important business meeting the economic costs are greater than when they are late going to a party. Furthermore, people who change their travel behaviour as a result of the roadworks do so because of the hindrance caused by the roadworks and are thus also affected. In total there are three factors at play that influence the impact of roadworks. These are: (1) changes in travel time and travel time reliability, (2) trip motives, and (3) people that change their travel behaviour as a result of the roadworks.

The economic costs of increases in travel time and travel time reliability are represented in the Value of Time (VoT) and Value of Reliability (VoR), respectively. VoT and VoR represent the Euros people would on average pay to mitigate an hour of travel time and reduce the standard deviation of the travel time, respectively. These values are available specifically for VoT and VoR relating to road users in the Netherlands (Kennisinstituut voor Mobiliteitsbeleid, 2013). And as stated earlier, the VoT and VoR depend on the trip motives and are hence specified for each motive of interest.

Three categories in particular are of interest. These are trips from and to work, business trips, and whatever is left (Kennisinstituut voor Mobiliteitsbeleid, 2013). Trip motives relate to the purpose of a trip and not to what a person does at the destination. For example, when someone leaves to go to work and travels back later that day the motives are both work. The reason for this is that the second trip is only made because a person left to work earlier that day. Trip motives rather than what happens at the destination determines the economic impact.

The figures estimated and adhered to in the Netherlands for VoT and VoR for people traveling by car are presented in table 4.1. These values are derived from Kennisinstituut voor Mobiliteitsbeleid (2013), however, they relate to 5 year old estimates and hence have to be revised to be relevant for 2015. Guidelines established by Ministerie van Verkeer en Waterstaat and Centraal Planbureau (2004), i.e. two large Dutch organizations, state the value of travel time should be increased by half of the change in wage rates. From 2010 to 2015 wages increased by an average of 11.69% (CPB, 2014). The value of time and value of reliability thus increased by 5.85%. The current and updated figures can be found in table 4.1

Table 4.1 the value of time and value of reliability are shown for trips by car for a variety of trip motives with and without brackets for the 2010 and 2015 values, respectively.

TRIP MOTIVE	VALUE OF TIME (PER HOUR)		VALUE OF RELIABILITY (PER HOUR)	
HOME TO WORK	(€ 9.25)	€ 9.79	(€ 3.75)	€ 3.97
BUSINESS	(€ 26.25)	€ 27.79	(€ 30.00)	€ 31.76
OTHER	(€ 7.50)	€ 7.94	(€ 4.75)	€ 5.03
AVERAGE	(€ 9.00)	€ 9.53	(€ 5.75)	€ 6.09

When a person decides to change his or her behaviour to avoid the roadworks he or she is also affected by the roadworks. By assuming people are rational and try to do what is best for them we can state that these people are doing something more optimal then taking the trip (expert interview, Appendix C). Nevertheless, in a normal situation, i.e. without roadworks, these people would take the trip. The economic impact experienced by these people is thus greater than without the roadworks, i.e. the baseline, and less than the impact experienced when going through the roadworks (Appendix C). We only know the spectrum and not exactly how much these people are affected. For this the 'rule of half' is developed (Appendix C; Eijgenraam, Koopmans, Tang & Verster, 2000). The rule of half states that the impact of people that change their behaviour is half of that of the people going through the roadworks (Appendix C; Eijgenraam et al., 2000).

To summarise, to measure the economic impact of roadworks we need to:

- Measure the change in travel time;
- Measure the change in travel time reliability;
- Measure how many people are usually using the road;
- Measure how many people keep traveling during the roadworks; and
- Measure how many people change their travel behaviour, e.g. stay at home or change modes of transportation.

Furthermore, we need to keep track of the motives that govern these trips as motives influence the conversion rate to Euros, i.e. the economic cost of the roadworks.

4.2.2 Method overview

Here we present the envisioned method to measure the impact of roadworks on highways using mobile phone data. The envisioned method, represented in figure 4.2, is based largely on the information described in the previous section about what makes up the impact of roadworks. The envisioned method consists of 5 steps: scale, focus, label, compare, and report. In the following paragraphs the raison d'être of each step will be explained.

Figure 4.2, the method to measure the impact of roadworks on highways using mobile phone data.

Scale

In the scale phase the goal is to upscale the sample to the population. The mobile phone data provides information about a subset of everyone in the Netherlands. The aim is to measure the impact on the entire population and thus scaling is necessary. In section 4.2.3 we will go more in depth into how a scaling method will be devised, evaluated and applied to the mobile phone data.

Focus

In the focus phase the goal is to select the people that can logically be affected by the roadwork. As discussed in the scope of this research we only want to focus on the impact the roadworks have on the people that usually take the road where the roadworks occur. Hence, we need to employ a method to assign trips to the roads people are logically using, e.g. by assuming they take the fastest route. In section 4.2.4 a concise discussion of the relevant literature and a description of the technique to assign vehicle trips to a road is given. Moreover, here in 4.2.4 we discuss how to evaluate whether the applied technique correctly determines who is logically affected.

Label

In the label phase the goal is to assign motives to the trips of the people affected. Motives, as discussed earlier, are key for calculating the economic impact and thus a fundamental part of the research. Section 4.2.5 is dedicated to how we will assign trip motives to the trips found in the mobile phone data.

Compare

In the compare phase the goal is to measure changes in travel characteristics between when a roadwork is occurring and the baseline. This depends on two factors. On the one hand, the impact of a roadwork depends on the change in travel time and standard deviation of travel time, i.e. travel time reliability. On the on the other hand, it depends on the people affected. The latter is a combination between how many people keep using the road, change their behaviour, and what the motives are behind the trips. These two factors combined suffice to calculate the true impact of a roadwork. In section 4.2.6 a detailed description is given on how to compare the situation during the roadworks with a baseline situation.

Report

The final phase is the report phase. This phase consists of calculating the economic impact of the roadwork from the data acquired in the compare phase and reporting the findings. The report phase is essential for going from the measurements and calculation performed in the compare phase to actionable information, e.g. how large was the economic impact of the roadwork and what makes up this impact.

4.2.3 Scale

In the scale phase the mobile phone data will be scaled such that it is a good representation of the entire traveling population. This step is nontrivial and ensuring the scaling is done correctly can be a tedious task. For scaling a good understanding of both the sample that is represented by the mobile phone data and the population is needed.

There is currently already a scaling method in devised and applied at Mezuro to scale the users to the inhabitants for each area, i.e. subsection of the Netherlands of which there are a total of 1259 in the mobile phone data. In Appendix E an overview and evaluation of the current scaling factor is presented. Evaluation is performed to investigate the geographic as well as demographic representativeness of the data.

In chapter 6 we will propose a new method to scale trips from mobile phone users to the traveling population. We see there is a discrepancy with scaling from users to the inhabitants per area as compared to scaling to the traveling population that we want to address. This is also a remark to other research that scale to the inhabitants per area rather than the traveling population (Toole, Colak, Sturt, Alexander, Evsukoff, & González, 2015). By scaling to the inhabitants per area one does not yet take into account demographic differences. Demographic differences are important because of two main reasons. First of all, some people are more likely to have a mobile phone than others, e.g. mobile phone penetration is lower than average for very young people (Telecompaper, 2015). Second of all, some people, e.g. children, will be more likely to be found on the highway on specific days than others. Demographic representativeness can thus be an important factor to include in the scaling method. Hence, in chapter 6 we will also present a new method to scale the mobile phone data, not to the population in general, but to the traveling population. Due to the introduction of this new scaling factor the description of the old scaling factor is less relevant for this research and hence put only in the appendix.

Our new scaling factor will be evaluated by checking whether the scaled trips from the mobile phone data comply with the actual vehicle counts on the road. Obviously, we have to adjust for the number of people per vehicle as we measure mobile phones and, assuming one phone per person, the number of people traveling. This differs from the number of vehicles as there might be more than one person per vehicle. How we translate people to vehicles and perform the final comparison will, for one, be discussed in the next section. The results will, nonetheless, be discussed at the end of chapter 6 with a description of the new scaling factor.

4.2.4 Focus

To prove mobile phone data provides insight into what happens on the road we will compare the measurements with actual roadside measurements. To evaluate whether mobile phone data can also be used to get measure vehicle counts we will compare our predictions with roadside measurement data. Roadside measurements provide accurate measurements of the number of vehicles on the road and are made publically available by NDW (NDW, 2015b). An overview of our approach to see if the data sources comply is presented in figure 4.3.

We are not the first to evaluate whether mobile phone data can be used to observe traffic characteristics on the road. Wang, Hunter, Schechtner and González (2012), for example, also derived OD-matrices from CDRs to observe traffic patterns on the road in two major cities. While they were able to evaluate the travel times with their predicted travel times, a comparison with actual traffic counts was omitted. As we do not only care about travel time, but also the number of people on the road we need to extend their analysis in this respect.

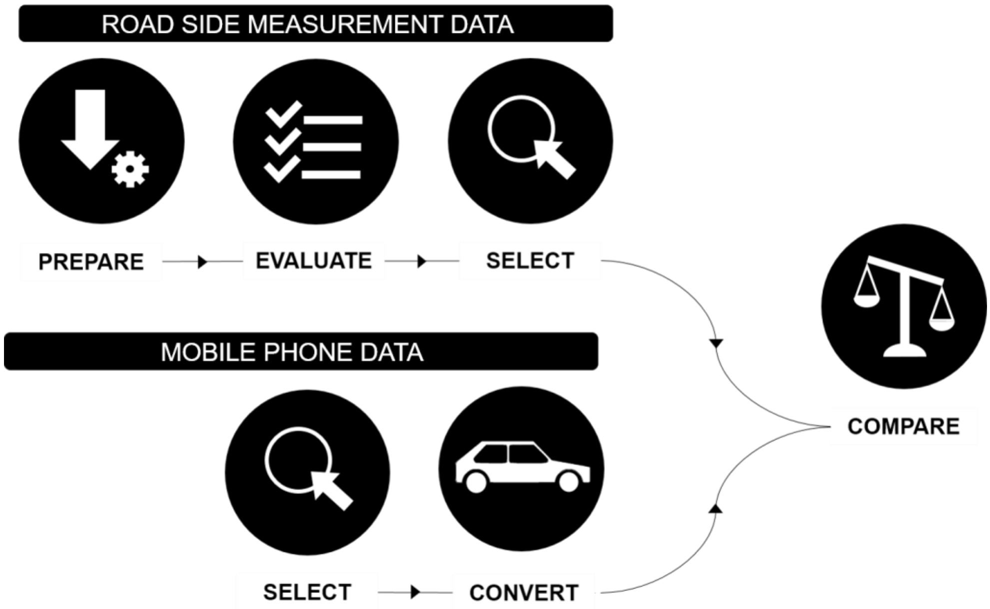


Figure 4.3, method for comparing traffic counts from roadside measurements with mobile phone data.

The steps presented in figure 4.3 are discussed in the paragraphs below.

Prepare the roadside measurement data

The first step is preparing the roadside measurement data. This step consists of downloading and converting the raw data to a workable format. Preparation, furthermore, consists of creating extra attributes and restructuring of the data for analysis.

Evaluate the roadside measurement data

The second step is evaluating the roadside measurement data. As always it is important to check the quality of the data that is used. In the best case scenario all data is perfectly accurate, but experience teaches us this hardly ever is the case.

For the final comparison we want to know (1) the accuracy of the roadside measurement sites in general, (2) when we have to discard information due to too many missing data, and (3) that the measurement site is at the stated location.

The accuracy of the roadside measurement sites is determined by comparing vehicle counts of consecutive measurement sites that are not separated by an on or off ramp. Hence, vehicles measured at the first site should also pass the second. If the measurement sites are 100% accurate they should provide the same measurements. The deviation between consecutive sites is used to quantify the accuracy of the roadside measurement sites.

Often there are some errors in measurement. This is not a big deal per se, but measurement sites providing erroneous data might indicate those measurement sites are untrustworthy. Hence we will also evaluate the relation between missing data and the accuracy of the roadside measurements using the same method described in the previous paragraph.

The final evaluation step consists of validating the location of the measurement sites. The location of the measurement sites are reported by humans and might contain discrepancies. At a conference by the NDW questions were also raised regarding the accuracy of the locations of the measurement sites. Hence, we will try to discard measurement sites that are possibly wrongly located. We aim to identify these measurement sites by checking whether they measure similar traffic counts as their neighbours. We cluster the sites on their vehicle counts and will check whether measurement sites in the same cluster are also located on the same road. If they are then they are most likely correctly located. If the measurement sites are far away from other sites in their cluster they are probably not at the stated location. This measure will be our final barrier between all measurement sites and the ones we can trust and to which the mobile phone data will be compared.

Select the roadside measurement data

The third and final step regarding the roadside measurement data is drawing a sample from all the good sites passing our evaluation tests. This sample will consists of about 100 measurement sites all located at least 5 km apart to ensure both a sufficient statistical power and independence of measurements. All sampled sites, furthermore, will have to be located on Dutch highways, as this is the scope of the research.

Select the mobile phone movement data

Here we select all trips over the road passing a selected roadside measurement site. This is done in two stages. First we identify whether a trip is performed by train or over the road using the algorithm developed by Keij (2014). Thereafter, we select all origin and destination pairs that cross the roads where the selected roadside measurement devices are located. Whether the logical route between an origin and destination crosses a specific road is determined using a route assignment algorithm.

What route assignment algorithm is applied and why will be discussed in chapter 7 (section 7.2.1). In essence what we evaluate here is whether the chosen route assignment method correctly estimates routes from the origin and destinations in the mobile phone data.

Convert the mobile phone movement data

Converting the mobile phone movement data consists of (1) scaling the trips to go from mobile phones in our sample to the traveling population and (2) translating the traveling population to vehicles on the road. The first is done by applying our own scaling method that will be presented in chapter 6. The second is done by dividing the people found on the road by the average number of people per vehicle. Compensating for people per vehicle is done, for one, by taking into account the type of day, e.g. weekdays and holidays et cetera, and trip motive. In chapter 7, i.e. the chapter about comparing vehicle counts, we will provide more information and evidence that we can use trip motives and day types to get to stable people per vehicle ratios. After scaling the sample to the traveling population and going from people to vehicles on the road we can start the comparison between the two data sources.

Compare both datasets

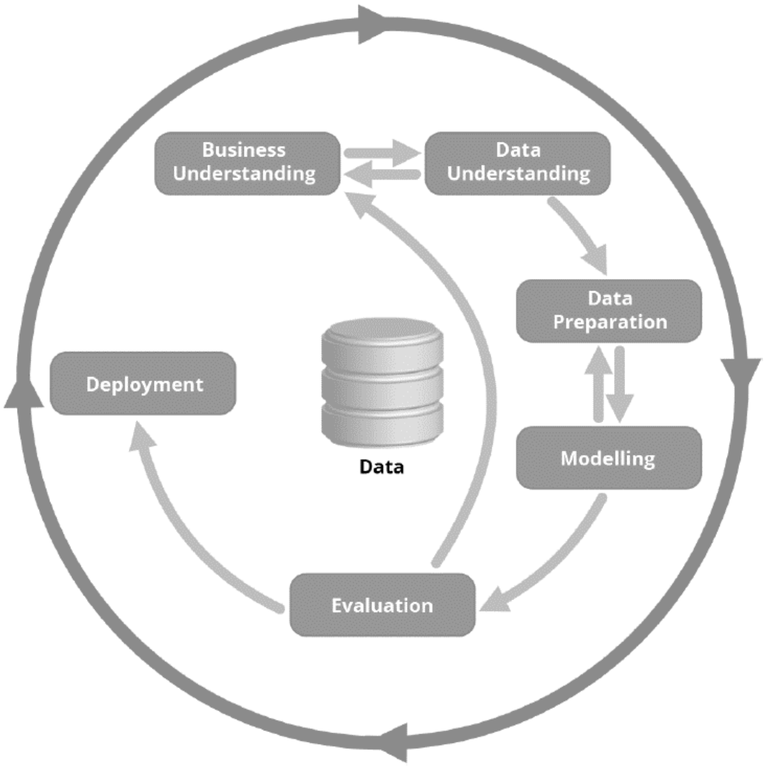
The comparison will include 100 sites spread over the Dutch road network on major highways. The comparison will include a correlation analysis, which will show if the same patterns and vehicle counts are observed from both data sources. Moreover, the comparison will also be done to see if total vehicle counts match. When the scaling is done correctly and all trips are recorded the traffic counts should be equal. Whether this is the case will also be discussed at the end of chapter 6 when we evaluate the scaling factor.

4.2.5 Label

In the mobile phone data we can see travel patterns, but not the motive of the trips. Travel patterns and trip characteristics, e.g. leaving home early in the morning, might provide enough contextual information to make an educated guess about the motive of a trip. To do so a model has to be created in which trip characteristics are linked to trip motives, i.e. the reason why a trip is taken. The three motives of interests are: home-to-work, business, and other (Kennisinstituut voor Mobiliteitsbeleid, 2013).

Trip characteristics, e.g. when does a person leave, will help to infer what the motive is of a specific trip. The idea is to use the OViN dataset to determine how trip characteristics relate to trip motive. Thereafter, we will use this information to make predictions about trip motives on the mobile phone data. A number of attributes such as start and end of a trip, leaving from or going to home, et cetera can be found in both datasets (CBS & RWS, 2015). By training a model using these common attributes on the OViN dataset will allow us to make predictions on the mobile phone dataset.

For model creation the well-known and popular Cross Industry Standard Process for Data Mining (CRISP-DM) of Wirth and Hipp (2000), depicted in figure 4.4, will be used (Marbán, Mariscal, & Segovia, 2009). CRISP-DM is a comprehensive method for data mining and knowledge discovery (Wirth & Hipp, 2000). Data mining is defined in the Oxford dictionary as "The practice of examining large pre-existing databases in order to generate new information" (Oxforddictionaries, n.d.). Data mining is represented in the steps from data understanding through modelling. The entire CRISP-DM method is devised for knowledge discovery. What we aim to do by creating a model to predict trip motives is a typical data mining and knowledge discovery type of assignment. The goal is to generate new information about how trip motives relate to trip characteristics. We want to incorporate this new knowledge in a model to predict what the motive is of a trip for trips in the mobile phone data. Hence, CRISP-DM fits our task perfectly.



Figuur 4.4, Overview of the steps and chronological relations in the CRISP-DM method.

CRISP-DM consists of six intertwining steps (Wirth & Hipp, 2000). These steps are presented in figure 4.4 and well documented by Chapman et al. (2000). The first step is business understanding. This step consists of getting a grasp of the project objectives (Chapman et al., 2000). The second step is data understanding. During data understanding the data is gathered, data quality concerns are discussed, and some first insights and hypotheses might be discovered (Chapman et al., 2000). The third step is data preparation. Here data sources are merged, variables are constructed and selected, and data cleansing is performed (Chapman et al., 2000). After this step the final dataset used to for modelling is finished. In the fourth step, i.e. modelling, various modelling techniques are applied. Because various techniques might have various data requirements there may be some back and forth between modelling and data preparation (Chapman et al., 2000). After a model is created the model is evaluated. Evaluation of the model is the fifth step in the CRISP-DM method. Evaluation is performed to see if the project objectives are satisfied (Chapman et al., 2000). If a satisfactory model is found it still has to be deployed. Deployment, i.e. the sixth and last step, consists of applying the knowledge gained during the project (Chapman et al., 2000). In our situation this step would involve applying the created model to estimate trip motives on the mobile phone data. Proving that the trip motive can be determined from trip characteristics is part of the evaluation phase in the CRISP-DM method. This will also be the validation step to show labelling is possible and can be included in the mobile phone data.

Chapter 8 will be the applied CRISP-DM method where we go over each step and present our model and results regarding trip motive prediction.

4.2.6 Compare

The compare phase is about determining the effect of roadworks on travel time, travel reliability, and travel behaviour in general of the people affected. This is a matter of putting the measurements into context. Knowing the average travel time is 20 minutes provides very little useful information without knowing that during the roadworks this increased to 28 minutes. Context thus is key. We have to compare the situation during the roadworks with a baseline. Finding a good baseline and calculating the differences between the roadwork and the baseline situation is the goal of this phase.

A good baseline is one where the only difference to the roadwork situation is the fact that it lacks the roadwork. In science this is called a controlled experiment, i.e. an experiment where only one variable is changed to measure the effect of changing that one variable. There are two distinct techniques to get a good baseline. These are making sure all external variables during the baseline match those where the roadworks take place. This implies, for one, including only moments in the baseline where the weather approximately equals that when the roadworks occur. The advantage is that the roadworks are now truly the only variable that is changed and we will approximate the controlled experiment. However, being very strict on the inclusion criteria can dramatically reduce our sample. Our data spans September and October 2015, roughly eight and a half weeks. Hence, if the roadworks occur on a Monday we have at most eight comparable days, i.e. Mondays, in our baseline. As the weather differs over these days the sample becomes smaller. Low sample sizes lead to more noise that can obstruct us in finding the true impact of the roadworks. Now one could wait and use data covering a longer period or try to correct for changing weather conditions rather than excluding these measurements. We will choose the latter and as a result we have much more data in the baseline providing a more trustworthy baseline.

To correct for variations in the baseline we will create a linear regression model. The goal of the model will be to predict what the travel time would be given the variables we cannot control for, e.g. weather and distance travelled. The difference between the actual travel time and the predicted travel time is the corrected travel time. When the corrected travel time is larger than zero, travel time is longer than average for that type of day. A negative corrected travel time implies people are traveling faster than normal.

Before establishing a good baseline we also have to select the data that will be used for analysis. Obviously we will use the mobile phone data, but within this data we can cherry pick the information fitting our needs best. We can, for example, select only the people with many data points for more accurate travel time measurements. In addition, there may be more ways in which we can cherry pick our data that will result in more accurate and trustworthy results. How to do this properly is a nontrivial task. We will use findings from all previous chapters, in particular, the one about data quality to help determine how our dataset will be composed.

In chapter 9 we will discuss first how to create the most optimal dataset from the mobile phone data to perform the analyses. Thereafter, we will create a model that helps to correct for variations in the measurements due to variances in external variables, e.g. weather.

4.2.7 Report

The report phase is the final phase in the proposed method to measure the impact of roadworks. Here metadata about the roadworks is presented along with the output from the compare phase. Moreover, the economic impact of the roadworks is calculated and reported.

Calculating the impact of the roadworks involves converting differences in travel time and travel time reliability for the road users into a monetary value, e.g. Euros. The results from this calculation are multiplied with (1) the total number of people that keep traveling over the road section where the roadworks occur and (2) half the people no longer found on the road, e.g. those who decides to stay at home. This has to be done for all trip motives because while the travel time and travel time reliability will be equal for all road users, the costs are not (Kennisinstituut voor Mobiliteitsbeleid, 2013). The total will amount to the measured economic impact of the roadwork.

Finally, a proof of concept will be included in this study. The proof of concept will serve as a practical summary of all five steps and help to show the added value of our method, i.e. the ease and accuracy at which the economic impact roadworks can be measured. Chapter 10 constitutes the proof of concept where we also discuss whether we achieved our goal of creating a more scalable, accurate, and complete method to measure the impact of roadworks. Chapter 10 will also be a prime example of the deliverable created in the report phase.

4.3 Predicting the impact of roadworks

In the proposed research we aim to find out if mobile phone data can help to improve the accuracy of predictions of the impact of roadworks on Dutch highways. For this it is necessary to investigate what characteristics influence the impact of roadworks, how accurate these characteristics help to predict the impact of roadworks, and how our prediction model compares to the current state of the art. These aims are represented in sub research questions 2a through 2c.

Sub research question 2a

What is the relation between roadwork characteristics and the impact of roadworks on highways?

Sub research question 2b

How accurately can our model predict the impact of roadworks on highways using mobile phone data?

Sub research question 2c

How does our model compare to the current state of the art in predicting the impact of roadworks on highways?

Answering each sub research question will require a unique approach. These approaches are elaborated upon in 4.2.1, 4.2.2, and 4.2.3, respectively.

4.3.1 Characteristics influencing the impact of roadworks

To predict the impact of roadworks it is important to know what may influence the impact of roadworks and how. Hence sub research question 2a is formulated. To answer this sub research question a twostep process is taken. First a list of characteristics is created. This list will consist of characteristics of the work zone, e.g. number of lanes closed, the as well as external characteristics that plausibly influence the impact of the roadworks. These characteristics include the composition of the traffic, mobility management techniques applied, and more types of characteristics that may influence the impact of the roadworks. These characteristics will be the building blocks for our model. Although ideally only the truly important characteristics are used there is no guarantee this is the case. Analysis during model creation and evaluation will help determine what characteristics are truly important. The second step is determining the structure of the relation between the characteristic and the impact of roadworks. The model fundamentally aims to describe the underlying relations. Hence, knowing their structure is key to ensure the model is able to do so. It will help when designing the model and preparing the data.

The first step consists of creating a list of all possible characteristics that may influence the impact of roadworks. These characteristics are extracted from literature and personal expertise built over the course of the study (figure 5). Proven relations between roadwork characteristics and the impact of roadworks can be extracted from literature. Moreover, relations that appear logical and have not been in literature will also be included, i.e. they are added based on personal expertise.

4.3.2 Accuracy of predicting the impact of roadworks

The first step is creating a model designed to predict the impact of roadworks. The input for this model is a combination of the characteristics that may influence the impact of roadworks and the impact of roadworks as measured using mobile phone data. The first is found by investigating literature and statistical analysis as just described. The latter, i.e. measuring the impact of roadworks, will be done using mobile phone data as described in chapter 4.1.

Gathering the data about roadworks is fundamental. Ideally information is gathered on all roadwork on Dutch highways between the start of November 2014 and the end of January 2015. Within this timeframe the mobile phone data is available for this study. One source of information is the website vananaarbeter.nl. On this website governed by Rijkswaterstaat there is a list of most roadworks in the Netherlands and some basic information such as number of lanes closed and a rough estimation of the additional travel time (Ministerie van Infrastructuur en Milieu & Rijkswaterstaat, 2015). Additional information can be found either online or via contact with Rijkswaterstaat.

Once all the data is gathered the next logical step is to create a model for predicting the impact of roadworks. Years of development in traffic engineering led to three types of macroscopic methods for modelling traffic flows: explanative deductive models, explorative inductive models, and intermediate models (Papageorgiou, 1997). The first is based purely on traffic flow theory, the second is based purely on empirical data, and the third is a mixture combining both traffic flow theory and empirical data (Papageorgiou, 1997). The most obvious direction from a data science perspective appears to be using empirical data and choosing a machine learning technique that allows the model to embody or mimic fundamental relations as identified in traffic engineering theory. Using the information from our literature study and by using visualization and experimentation we aim to get an idea of the structure of the final model. Figure 4.5 provides an early indication of how the model may appear. The overall layout is partially inspired by the model created by Calvert (2010).

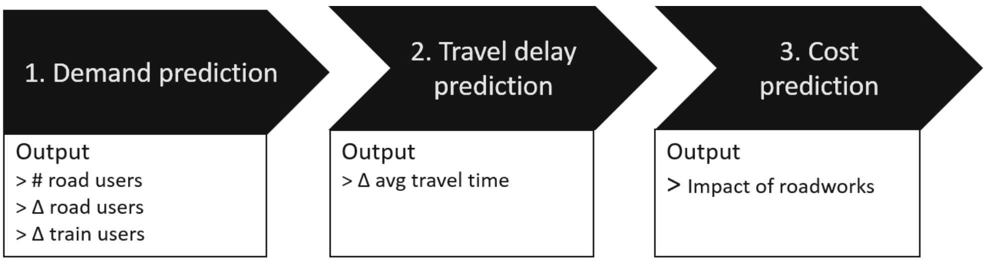


Figure 4.5, structure of a model for predicting the impact of roadworks.

Training a model is only part of the task here. The model needs to be tested and evaluated to find out its strengths and weaknesses. In chapter 11 we also evaluate our model in terms of how accurately it can predict the impact of roadworks. Where it works well and where the model is way off. In chapter 11 we will furthermore discuss what we think misses from our model that could help improve its predictive capabilities.

4.3.3 Model comparison with the state of the art

The final step in this research is comparing the model with predictions made by, for example, Rijkswaterstaat or road construction companies. This comparison will provide insight into whether the created model is any good. Without the comparison it becomes impossible to state whether predictions that deviate 10% from the measured impact of roadworks are an improvement or not.

Information about the predicted impact of roadworks is partially available online. On the website vananaarbeter.nl there is some information about the expected travel delay available (Ministerie van Infrastructuur en Milieu & Rijkswaterstaat, 2015). The range of expected travel delay, however, appears to be quite large. The provided timeframes for travel delay are 5 to 10 minutes, 10 to 30 minutes, and 30 or more minutes. 15 to 30 minutes or greater than 30 minutes. To see if this is actually the best prediction of travel delay an interview at vananaarbeter.nl will be conducted. Travel delay, nevertheless, is only part of the impact of roadworks.

The impact of roadworks is more complex than the delay in travel time. Estimations of the true impact of roadworks are not available online. Interviews at Rijkswaterstaat and road construction companies will be planned to try and extract this information.

All information related to our model and evaluations to come is presented in chapter 11

5 Data understanding and data quality

This chapter serves as an introduction and a critical review of the use of mobile phone data as a data source. In short we will discuss how the dataset at Mezuro is currently created (section 5.1). Thereafter, we discuss the proposed changes (section 5.2), general data characteristics (section 5.3) and limitations (section 5.4). Data quality is assessed in section 5.5, followed by a conclusion stating the most important findings of this chapter in section 5.6.

5.1 Creating the mobile phone data

Figure 5.1, from left to right: a cell tower with antennas for various frequencies (2G/3G/4G), a camouflaged cell tower, and a cell tower located on top of a building.

5.1.1 The physical telecom network infrastructure

The telecom network has been set up to enable mobile devices to communicate. To enable mobile devices to communicate, cells are attached to a cell tower or a high building (figure 5.1), called the cell site, and provide signal to devices within range. The area that a cell provides service to takes the shape of a two dimensional cone. The service area is controlled by the following parameters: angle, radius, direction and location (figure 5.2). The location, direction, angle and radius of all cells are registered in the cell plan. Each row in the cell plan corresponds to a single cell at a given moment in time. The cell plan thus allows us to approximate where a cell phone is located when an event is generated. When a cell tower is changed this is also added to the cell plan. Only the most current information is stored about each cell.

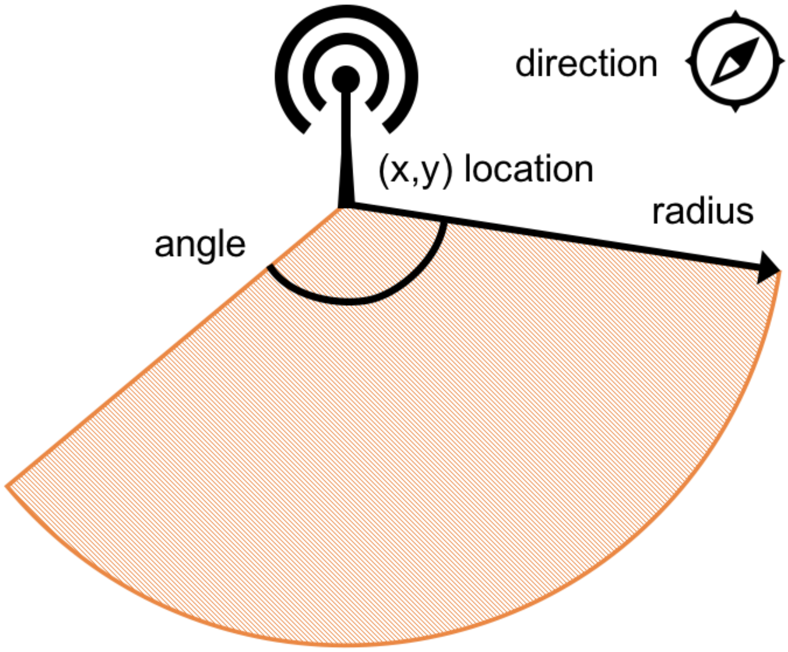


Figure 5.2, the adjustable parameters of a single cell located on a cell tower.

The second key source of information is the table with events. This table consists of records where a mobile phone has communicated with a cell tower. For each record the table has information about the unique person identifier, which is a one-way-hashed version of a person's phone number, a timestamp, event type, country code and corresponding cell tower id. These records are the Call Detail Records (CDRs) we talked about previously. Although the phone records are hashed, there is still information about the country of origin for foreigners. This helps to provide additional information about, for example, how many Germans visit the Dutch shores compared to other nationalities. There are a number of event types such as voice, SMS, and data. This information will not be used in this research, but could serve valuable when trying to infer additional information about the type of user.

The cell table provides information about the whereabouts of the cell towers and the events table shows what cell tower a person was near at what moment in time. Linked together these two data sources provide information about the whereabouts of mobile phones at specific moments in time. In the next section we will discuss how to go from these timestamped locations to origins and destinations.

5.1.2 From locations to origins and destinations

The origin and destination algorithm retrieves useful information about trips from the cells a person is connected to. The general idea is that a person is moving when he connects with a cell that the cells he has connected to at his previous location have no overlap. When a person stays in an area for 30 minutes or more without traveling that person is assigned a destination.

The algorithm in use is slightly deviates from the one just described. The above algorithm has to check if there is overlap with all previous cells since the person arrived at the previous location. This is computationally very expensive considering the enormous amount of users for which this has to be done. A much quicker way to process the data is to only check if there is overlap with the first cell touched at the previous location. The algorithm in the previous paragraph would be ideal, but is unfeasible at this moment. The current algorithm might be less accurate, but may provide good results nonetheless at a fraction of the computational time needed. The expensive algorithm will be referred to as the gold standard and the other as the current algorithm. How good both algorithms are will be evaluated in section 5.5.

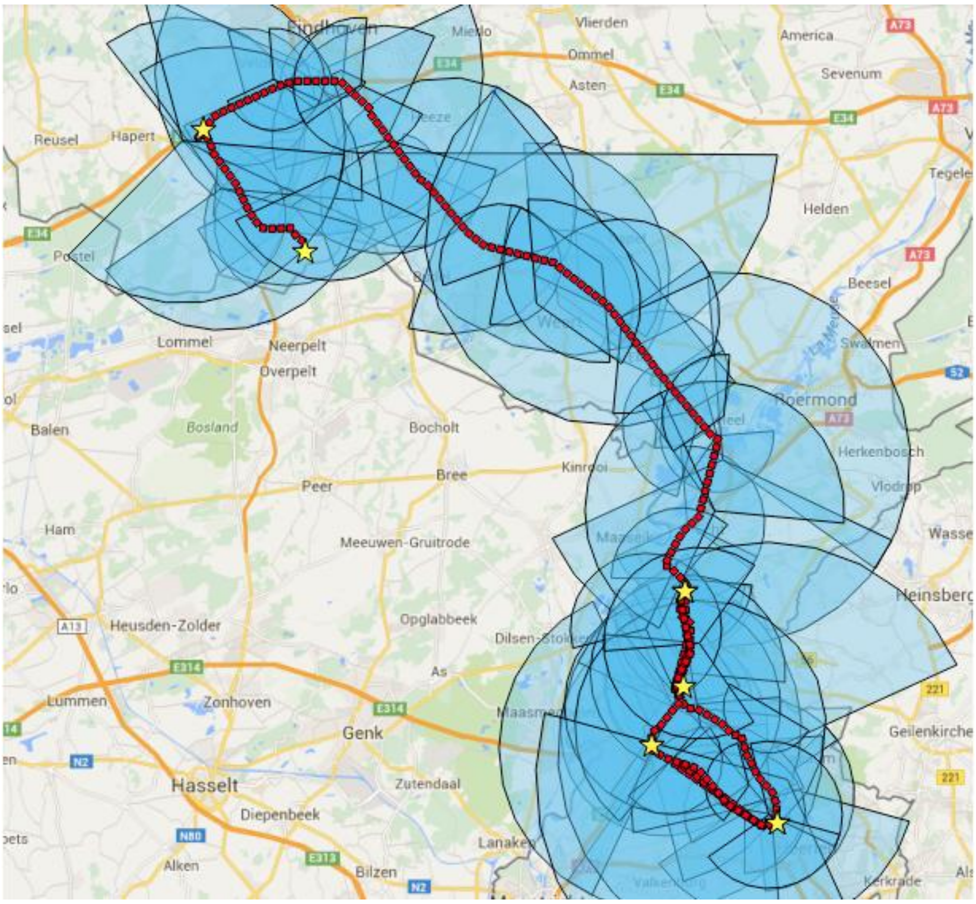


Figure 5.3, the cells connected to while traveling. The red dots is a GPS trace and the blue areas are cells that the user had events with. Destinations, i.e. where a person stays for longer period are depicted with yellow stars.

In figure 5.3 we provide an example of how a series of events at cell towers, depicted with blue in figure 5.3, leads to destinations, the yellow stars in figure 5.3.

5.2 Proposed changes

There are three main changes we want to make to the mobile phone movement data. All of them are either directly or indirectly related to producing more accurate travel times. Accurate travel times are the foundation of this research, but have been low priority for many other studies as those often are more interested in where the trip is going rather than the trip itself.

Travel time in the mobile phone data is currently defined as the difference between the last event at the origin and the first event at the destination. We have no idea whether the last event before being observed traveling was when the person was at home or already on his way. This is because it takes a while before a person leaves the area we define as the origin. For the destination the same holds. Travel distance is also measured from the centre of the origin to the centre of the destination area as depicted in figure 5.4.

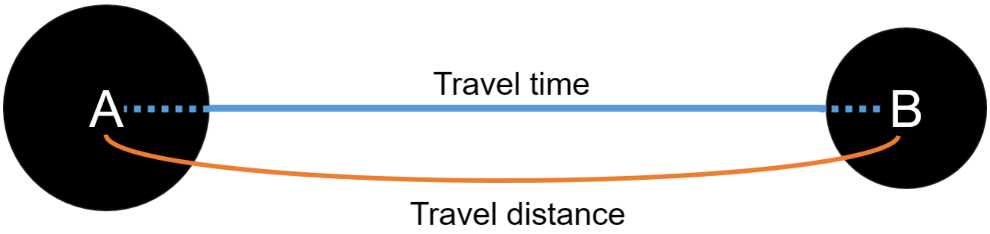


Figure 5.4, travel time versus travel distance measured in the mobile phone data.

The first change we propose is discarding events with cells that have a very large radius. When the cells have a large radius we encounter the following two issues. First of all, the cell with radii of up to 40 km cover a significant part of the country. Because we cannot see within the area below the cell we have only a very rough approximation of the person's location. Hence, trips on shorter distances will be less likely to be recorded. Also as we know less accurately when a person starts and stops moving we lose accuracy in the reported travel times. Reducing the maximum allowed cell size can thus lead to more accurate location specifications and more precise travel time recordings.

The downside to discarding events with large cell radii is losing data. To get a grasp on what we lose we analyse how many events are still available when deciding to remove cells with a radius greater than a certain value. The effects on the number of events discarded by setting a limit to the cell radius can be seen in table 5.2. Near 10 to 12.5 km there are still a 91% and 94% of events still available, respectively. Moreover, the coverage in the Netherlands is such that even without these cells with radii larger than 10 or 12.5 km the entire country is still represented in the cell plan. The introduction of the thresholds thus does not induce a geographic bias. In section 5.5 we evaluate the effect of using 10 and 12.5 km as cut-off points for the maximum allowed cell size. This is also compared to the current situation to see if the threshold put on maximum cell radius has a positive effect on data quality.

Table 5.2, percentage of events compared to the maximum allowed cell size

CELL RADIUS UP TO (KM)	PERCENTAGE OF EVENTS
2.5	60%
5	77%
7.5	86%
10	91%
12.5	94%
15	95%

The second change we propose is to allow for staying in an area for over 30 minutes up to an hour provided the person continuous to travel in the same direction. Typically a person gets a destination after 30 minutes. However, in heavy traffic the travel velocity can drop dramatically and this may result in people getting a destination while actually stuck in traffic. We think this does not have a major impact in general. With a cell of 12.5 km in radius, i.e. 25 km in diameter, it would require velocities over a 25 km stretch to be below 50 km/h, for example. Nevertheless, in it is feasible, especially with heavy roadworks occurring. Our proposed change would double the allowed duration for people that continue to travel in the same direction and thus mitigate the chance of these 'traffic jam destinations'. To test if people keep to travel in the same direction we first pick the following three destinations: the destination with a stay under an hour, the destination before that one, and the destination after that one (figure 5.5). When distance d1 plus distance d2 divided by distance d3 are smaller than the square root of two we assume the traffic jam destination is on the path from the origin to the true destination. Then the traffic jam destination will be removed and values such as travel time will be recalculated.

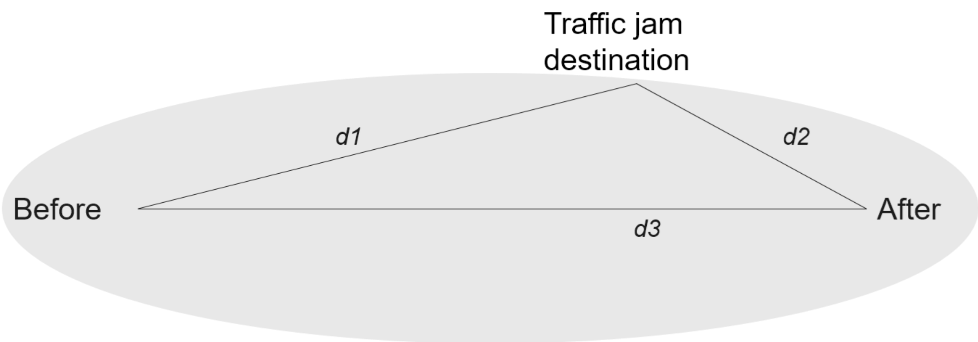


Figure 5.5, destination under an hour within the grey area are assumed to be on path from destination before that one to the destination after that one.

The third change we propose is to compensate for the number of events a person has during a day. When a person has many events the last event before leaving the origin was probably just before leaving the origin area. In contrast, when a person has for example 24 events a day, the event before leaving the origin might have been an hour before the person started traveling. The number of events per day can thus influence the measured travel time arbitrarily and we want to compensate for this.

We perform this compensation by subtracting the average time between events from the travel time. When a person, for example, has 144 events a day that comes down to 1 event every 10 minutes. Roughly between 0 and 10 minutes before the person leaves the origin area the last event was recorded at the origin. Because we do not know the exact moment we assume the event was 5 minutes, i.e. half the average time between events, before leaving the origin area. Vice versa the first event in the destination area was approximately 5 minutes after entering that area. To get a better estimation of the true travel time we will subtract the average time between events from the actual travel time for each person in the mobile phone data.

Obviously the proposed change will remove the travel time in the first and last section of the trip, i.e. the part in the blind spots at the origin and destination. To add these section we will artificially fill this time by adding the average time needed to travel in the blind spots at a given speed. On average we find cell radii of approximately 6.258 km for cells under 12.5 km (assuming we will go to smaller cell radii). Considering the trips are distributed evenly between the start and end of the blind spot, i.e. the 6.258 km, a person would travel half that distance in a blind spot. We further assume the first and last part of the trip are in urban areas, where average travel velocity (as the crow flies) is estimated at 27.5 km/h. Using this information we can calculate the average time a person travels in a blind spot that turns out to be just shy of 7 minutes, i.e. 6.258 km divided by 27.5 km/h. In total we thus have to add approximately 14 minutes to compensate for the time travelled in the blind spots.

5.3 General Data Characteristics

By analysing the general data characteristics of the mobile phone data we hope to provide an overview of the size and scope of the dataset and highlight trends that exist within the data. At the time of writing about 370 million events are generated per day by 3 million subscribers. However, the first thing that has come to our attention when analysing these numbers, is that the number of events produced on a daily basis has been increasing steadily over time (see figure 5.6). This is due to the increasing number of subscribers who are switching to 4G technology, which produces about 5 times more events than 3G technology. This trend is beneficial for the data quality, because the core of the location estimation algorithm is based on these events. The more events we have the greater the accuracy of the resulting mobile phone data.

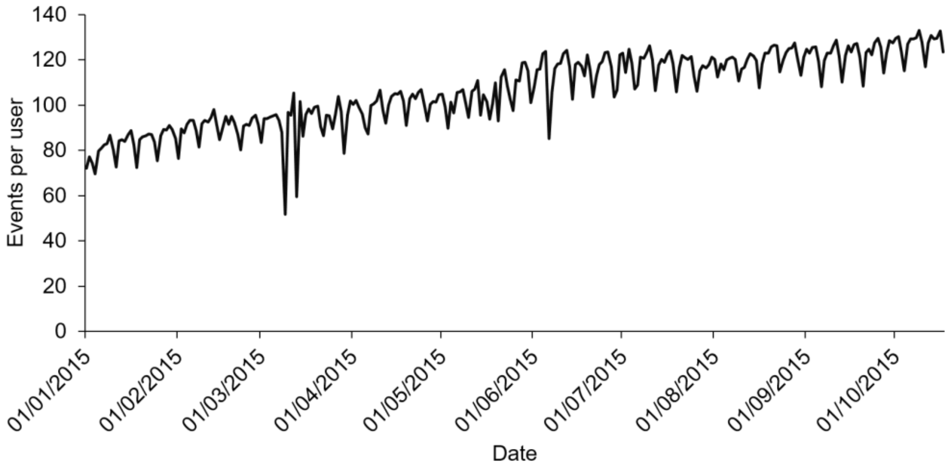


Figure 5.6, the average number of events per user in 2015.

In figure 5.6 we also see some drops in the average that require explanation. First of we must note that the exact details of when events are created are not known in detail, not even at the service provider itself as the network is mostly outsourced. However, we are aware of a positive correlation between phone usage and the number of events created distilled from experience. Thus, generally speaking, the more a phone is used the more events are created. The drops occur mostly during the weekend and especially on Sundays. We expect the drops to be caused by business users leaving their phone off during the weekend and religious people who tend to do less on Sundays in general.

5.4 Limitations

The limitations that affect the mobile phone location data can be divided into two main categories. The first category consists of limitations that are due to privacy regulations, while the second category of limitations are due to the technical specifics underlying this data source. The subsequent sections extensively discuss these limitations and elaborate upon the consequences these limitations have for this research project.

5.4.1 Limitations due to Privacy Regulations

Privacy has always been an important aspect of mobile phone location data as formalized by the Dutch Data Protection Directive (in Dutch this law is called “Wet Bescherming Persoonsgegevens”). Before the mobile phone location data can be analysed, three measures need to be taken in order to satisfy these privacy regulations.

Firstly, data at the level of the individual can only be processed at the servers that belong to the service provider. Output that is sent from the service provider to Mezuro is ridded of unique identifiers that can be used to track a device. Consequently, individuals cannot be traced.

Secondly, the phone numbers are hashed (i.e. a type of encryption) using a hash key that is changed every month, which implies that the unique key to identify a person by changes every month. Hence, the spatio-temporal trace a person leaves behind can be associated with an individual for the maximum duration of one month.

Thirdly, to prevent any unauthorised person from viewing the data of an individual person, outputs that aggregate 15 persons or less are omitted. That is because mobility patterns from mobile phone data are so unique that only four spatio-temporal records are necessary to be able to identify 95% of the users according to research by De Montjoye, Hidalgo, Verleysen, and Blondel (2013). This has implications on all instances where in the output less than 15 users are aggregated. Most notably, this has implication in for example rural areas, with very little activity, and in all areas where short time periods are selected, as these data are prone to being omitted. When these three measures are combined they ensure that it is impossible to identify individuals using this data.

5.4.2 Technical Limitations

In addition to privacy limitations there are also technical limitations. In particular, trips on shorter distances might fall below the same cell and are hence not recorded. To get a feeling of how many trips are recorded when we only include trips over a minimum distance we compared the trip counts with those of OViN, i.e. a large mobility survey in the Netherlands. In this chapter the data quality of OViN will not be discussed in depth, because this chapter is limited to a discussion about mobile phone location data. For a quality assessment of OViN we refer to chapter 8 where the OViN is discussed in greater detail as it is the main source of data used to predict trip motive.

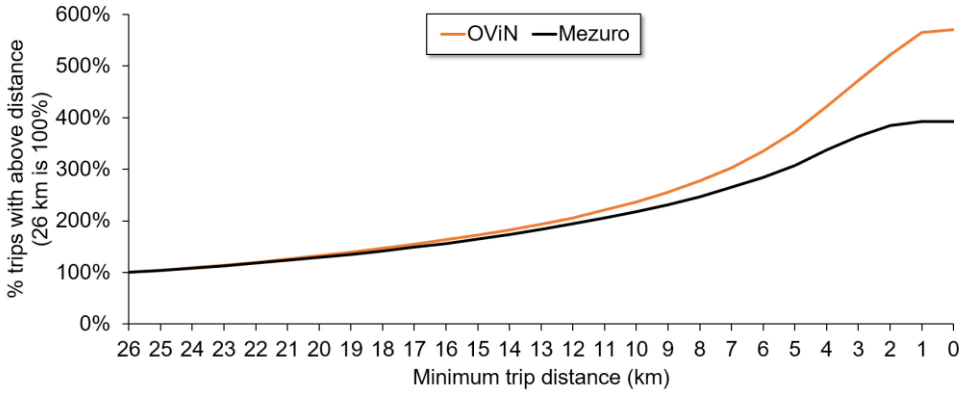


Figure 5.7, Comparison between OViN and Mezuro of the relative number of trips per travel distance. Both datasets have their number of trips set to 100% at a travel distance of 26 km.

Figure 5.7 depicts a graph that represents the number of trips per travel distance from OViN and Mezuro. Within this graph the number of trips of both datasets has been set to 100% at a travel distance of 26 km. In the OViN there would be no reason any trip would not show up in the data, thus also not a trip over 26 km. In the mobile phone data we know the largest trip within an area is 26 km. The maximum cell radii is set to 12.5 km. Because the angle of the cell may be 180 degrees this implies the maximum trip length in the blind spot of a cell would be 25 km. Why the limit is at 12.5 km will become clear in the next section.

Figure 5.7 shows that the trips on smaller distances are not reliably detected. We observe that both lines stay very close until trips smaller than approximately 10 km are included. Beyond the 10 km mark there is an increasing amount of trips absent in the mobile phone data. This is the result of trips not happening between different Mezuro areas or trips staying within the reach of the same cell towers. For this research we will not look at these shorter trips and exclude them from the data. As we look at highways, we also expect not to encounter many trips below 10 km and hence do not think this will significantly affect the outcome of the research.

5.5 Data quality evaluation study

In section 5.5.1 the results of the comparison with the GPS trace are presented. In section 5.5.2 we provide a further quality check in terms of evaluating the likelihood that realistic travel times can be recorded for each of the four algorithms. The data used in this analysis is derived from a test set with data of friends from Mezuro and consists of approximately 20 people.

5.5.1 Comparisons with GPS

In this section we provide an analysis of the data quality by comparing the mobile phone data with a GPS trace. The GPS trace is covers the entire month of February 2015. The analysis will thus cover 28 days of measurement. The GPS trace is from one the employees of Mezuro for whom the privacy regulations are lifted to do this and similar evaluation studies. In total four algorithms are compared against the GPS trace: the gold standard, the current algorithm, the current algorithms with only cells under 10 km, and the current algorithm with only cells under 12.5 km. These algorithms all provide us with information about the origins and destinations of the person of interest. We will evaluate the algorithms based on the number of trips acknowledges, the percentage of correctly predicting the origin and the percentage of correctly predicting the destination. A destination is defined similarly as in the mobile phone data, i.e. a person needs to be near stationary for half an hour or more except for traffic jam destinations as explained in 5.2. Near stationary we define in our GPS algorithm as moving less than 5 km in radius. Locations are compared based on the areas also in the mobile phone data. When a person has a destination in the GPS trace the average of the longitude and latitude at a location are mapped to these areas. We observed a total of 52 trips.

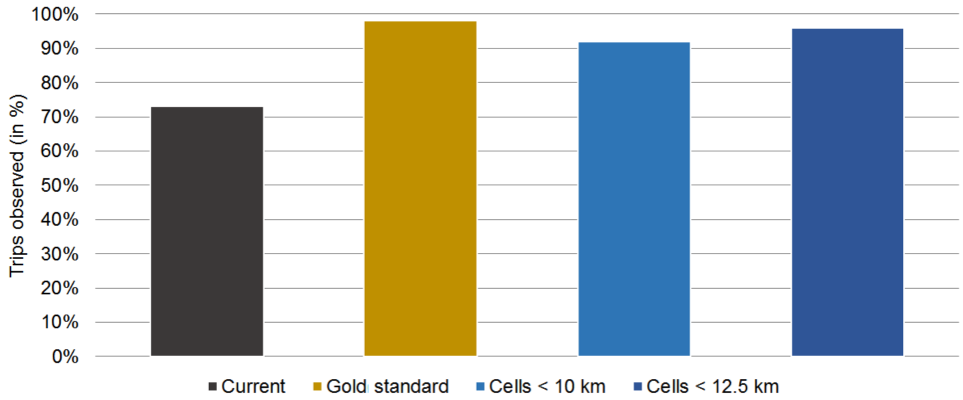


Figure 5.8, percentage of trips observed in the GPS trace that can also be found in the mobile phone data.

In figure 5.8 the percentage of trips observed for all four algorithms is depicted. The percentage of trips observed means the number of trips in the GPS data that can be traced back to the origins and destinations in the mobile phone data. When origins and or destinations are not exactly correct, but rather positioned in neighbouring areas they are still counted.

From figure 5.8 we observe that the current algorithm performs much worse than the other three, with the gold standard righteously performing best. The current algorithm with only cells under 12.5 km performs nearly as good as the gold standard. The one with only cells under 10 km does slightly worse. We expect too many events are left out resulting in worse performance. We must note that mistakes are most often made on trips with shorter distances. Long distance trips were well recorded with all algorithms.

For all trips that can be traced back from the GPS to the mobile phone data we also evaluated how often the origin was exactly correct, i.e. the same location as in the GPS trace (figure 5.9). We found all algorithms performed well. With the gold standard and cells under 12.5 km providing the best results. When the algorithms were off, they were nearly always located in a neighbouring area. The graph with the destinations is omitted as very similar results are found.

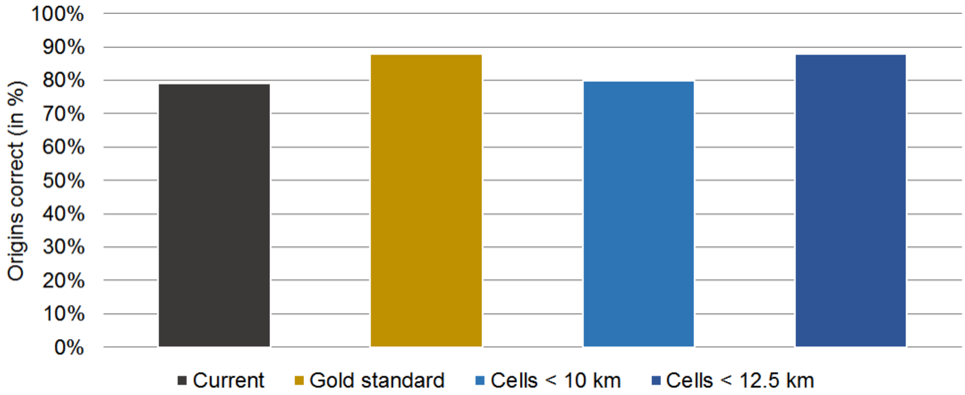


Figure 5.9, percentages of origins correct for all four algorithms.

To conclude, the current algorithm performs worse than all alternatives. The gold standard, as we expect, outperforms all algorithms. However, the idea of leaving out events with cells above a certain radius appears to be fruitful. Both algorithms presented in blue perform well. In particular, the current algorithm with only cells under 12.5 km appears to be very nearly as good as the gold standard. It manages to observe 96% of the trips found in the GPS compared to the 98% of the gold standard and just as often measures the correct origin for the observed trips.

5.5.2 Evaluation of measured travel times

The measured travel times are evaluated by measuring the percentage of unrealistic travel times. Unrealistic travel times we define here as travel times leading to a velocity, as the crow flies, greater than 145 km/h. OViN, i.e. a data source containing travel information about Dutch citizens.

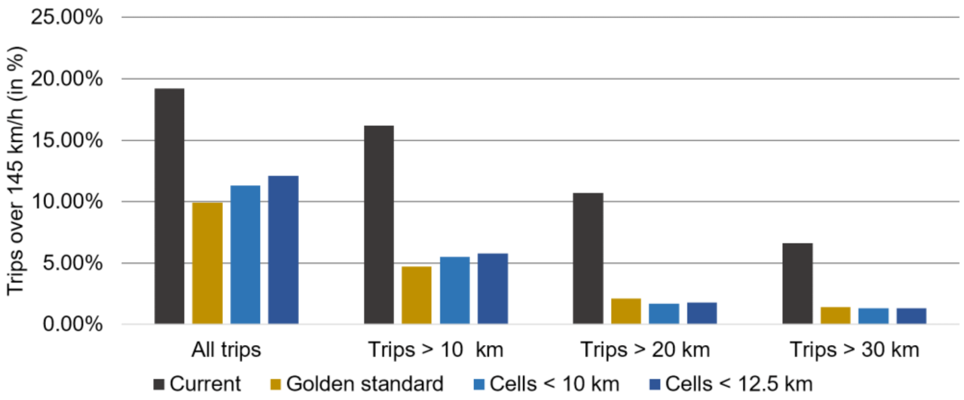


Figure 5.10, percentage of unrealistic travel times (> 145 km/h) is shown for all four algorithms and plotted against a threshold for the minimum length of a trip.

The percentage of trips that are too fast, i.e. over 145 km/h are shown in figure 5.10. There is a clear relation visible between the threshold put on the distance of the trips and the percentage of unrealistic travel times. This we would expect because the distance over which travel time is measured will become shorter relative to the distance between origin and destination (see figure 5.4). Moreover, from figure 5.10 we see the number of unrealistic travel times is much higher for the current algorithm than for the others and stays higher. Travel times are calculated by using the last event before leaving the origin and the first event at the destination. With the current algorithm cells are much larger and thus a person may have been travelling for longer before we observe them leaving the origin and vice versa after entering the destination decreasing travel times and increasing the likelihood of unrealistic travel times.

For trips with an origin or destination very close to where the roadworks occur the roadworks may be in a blind spot near the origin or destination. Delays in travel times would then not be measured. Hence, when measuring the impact of roadworks we advise to leave these trips out of the measurements related to travel time. There is, however, no reason to belief these people are not affected by the impact of the roadworks or that they are delayed much more or less than the other travellers. We will, therefore, assign the same travel delay to these people as to the rest of the travellers that are affected by the roadworks.

5.6 Conclusion

The negative effects observed by looking at descriptive statistics of the mobile phone location data are irrelevant as they occur outside the period that is analysed within the context of this research.

The positive effects observed however, are influential as the increased adoption of 4G technology by the majority of the subscribers leads to more events being produced, which in the long run leads to more accurate location estimation.

Next to that three measures are taken to satisfy privacy regulations. Firstly, data can only be processed at the mobile phone service provider. Secondly, mobile phone numbers are hashed using a hash key that changes every month. Thirdly, outputs can only be viewed in aggregate when they contain at least 15 persons. Logically, these three measures that are taken to satisfy privacy regulations cause limitations for the analysis of the data. In particular the 15 persons or more limitation is relevant for this research. The rule implies we might have to increase the time frame to ensure enough users have passed a road section to be seen. These privacy limitations are thus important to keep in mind in further analyses.

We also evaluated the impact of discarding events at cells with larger radii. By comparing with a GPS trace it showed disregarding events with cells greater than 12.5 km in radius the mobile phone data becomes more valuable. The data becomes more precise in terms of determining the correct origin and destination and becomes more complete in terms of percentage of trips observed. By including the threshold during data preparation the data quality is greatly improved.

Regarding travel times we found the measurements in general are usable. There are a number of limitations, however, that need to be kept in mind. First of all, the travel time is measured when the origin is left and the destination area is entered. Hence there are blind spots near the origin and destination where travel time is not measured. We thus advise to only take into account trips over 10 km. Furthermore, trips with unrealistic travel times, i.e. trips with an average velocity over 145 km/h, to be filtered from measurements regarding travel times. Finally, when the origin or destination of a trip is very close to the roadwork, i.e. plausibly within the blind spot, these trips and corresponding travel times should not be included. The trips can, nonetheless, be included for vehicle count analysis. When this is done we advise to assign the average delay experienced by other trips passing the roadworks to these trips.

6 Scaling

In this chapter we will discuss a new scaling method that can scale mobile phone users to the traveling population (section 6.1). Furthermore, we discuss whether the scaling factor is successful. This we do by comparing absolute vehicle counts on the road with the actual number of vehicles on the road inferred from the mobile phone data as discussed in 4.2.3 (section 6.2).

6.1 Scaling method

The new scaling factor needs to improve the current scaling factors on a number of aspects. The current scaling factor succeeds in adjusting the sample to the population by adjusting for the number of people per area (Appendix D). The current scaling factor, however, scales to the number of inhabitants per area and not to the number of people found on the road (Appendix D). This is an important distinction that has to be made as we want to measure how many people are affected by the roadworks. The scaling factor proposed here will, in essence, describe the ratio between the number of people we expect to be traveling and the number of users in the mobile phone data that we expect to observe traveling.

The chance a person is found on the road is related to the age of that person. Children under the age of 15, for example, will go to school during workdays, i.e. Monday through Friday, whereas people in the age group 40 to 45 might go drive to work. Hence, depending on the age group and the type of day the chance to see a person traveling a certain distance may change.

We make a distinction between four day types. These are workdays, Saturdays, Sundays, and workdays during the holidays. These day types provide insights into the distinct travel behaviours, with differences mainly observed for the young and old inhabitants. Further distinctions in day types did not provide additional information and would further spread our sample introducing noise. Hence we decide to only differentiate on these four. The likelihood of a person taking a trip of at least 10 km or more is represented in figure 6.1. From this figure we observe significant differences in travel behaviour based on age groups. A scaling factor to go from users to people traveling will thus have to include this information.

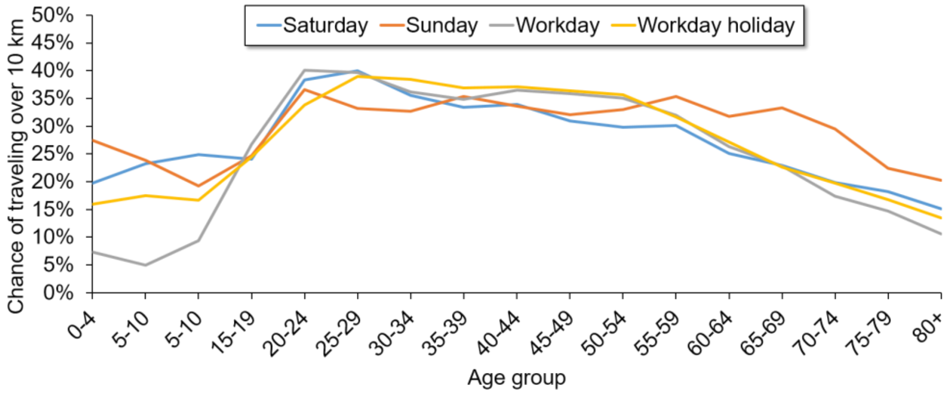


Figure 6.1, chance of observing a person making a trip over 10 km during a day depending on the age group and type of day.

In addition to differences in chance of traveling, different age groups have also a different representation in our sample. A child of 5 years-old, for example, is much less likely (4%) to possess a mobile phone than an adult aged 25 (96%) (Ofcom, 2014; Telecompaper, 2015). Therefore, we need to determine the penetration of mobile phones per age group and apply a correction on the data. This is because the 25 year old is much more likely to show up in our sample then the 5 year old.

Our scaling factor will include the above information to get a good representation of the people traveling from the mobile phones we observe traveling. The proposed scaling method is formally represented in a Process Deliverable Diagram (PDD). The PDD is shown on the next page (figure 6.2). In a PDD the processes are shown on the left and the product of the action on the right (Van de Weerd & Brinkkemper, 2008). The PDD is accompanied by two tables, one describing the processes (Appendix E, table E1) and one describing the products (Appendix E, table E2). A comprehensive fictional example of how the scaling factor is calculated can be found in Appendix F. The next paragraphs will provide a brief overview of the steps and products in the scaling method.

The number of inhabitants traveling can be calculated in three simple steps. First, multiply the number of inhabitants by the distribution of age groups across the population for that area. Second, multiply the inhabitants per age group with the chance they are traveling. Note this differs also per type of day. Third, take the sum over all age groups to get the number of inhabitants expect to travel.

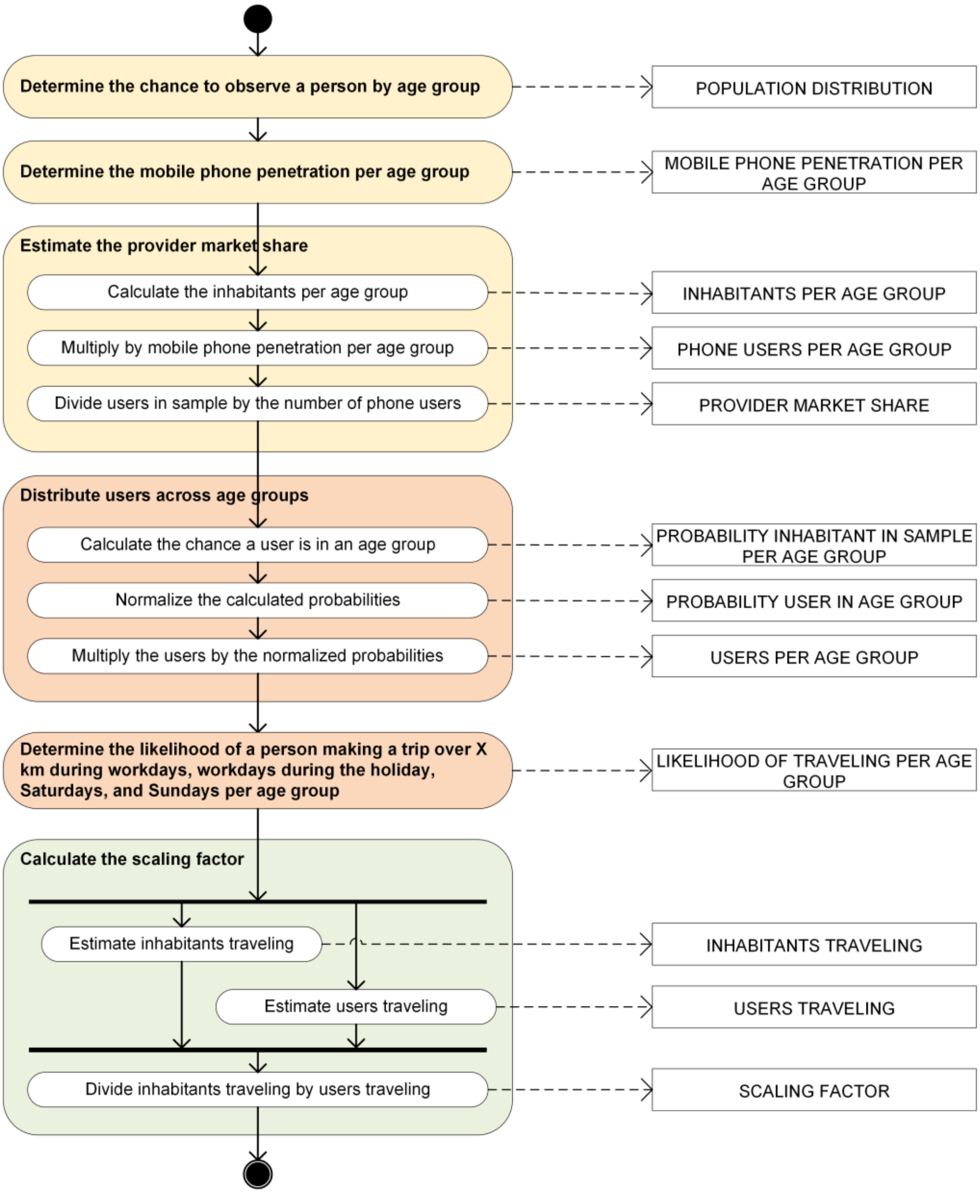


Figure 6.5, a PDD showing how the new scaling method can be calculated.

For the number of users we want to do exactly the same. However, we have information about the number of users and not the number of users by age group, which we do have for the inhabitants per area. Hence, we first need to calculate the chance of a user being in our sample. The chance of a user being in our sample depends on (1) the age distribution in that area, (2) the chance of having a mobile phone per age group, and (3) the penetration of the telecom provider in the population per age group.

The age distribution is the first step to get an estimate of the distribution of our sample over the age groups. The second is the change of having a mobile phone per age group. A child of 5 years-old, for example, is much less likely (4%) to possess a mobile phone than an adult aged 25 (96%) (Ofcom, 2014; Telecompaper, 2015). Therefore, we need to determine the penetration of mobile phones per age group and apply a correction on the data. This is because the 25 year old is much more likely to show up in our sample then the 5 year old. The third step is compensating for the penetration of the telecom provided in the population per age group. It may be that the provider targets a specific demographic resulting in a bias within the sample. By incorporating the market share of the provider per age group you can compensate for this bias with respect to the scaling factor. Unfortunately, we do not have data covering the provider market share per age group. We, therefore, make the assumption that the chance a mobile phone user is uniform across age groups in our sample. Offermans et al. (2013) got a good grasp of whether the data is representative for the population and they stated that appears to be the case. We, therefore, do not expect this to have much effect for our scaling factors. For other telecom providers with a more bias user base we do recommend to use it.

Finally, the number of traveling users is estimated from our expected number of users per age group and the chance of traveling per age group. By dividing the number of expected inhabitants traveling by the number of expected users traveling the scaling factor is determined each day for each area.

6.2 Evaluation

Here we provide the results of our comparison between absolute vehicle counts on the road and vehicles inferred from the mobile phone data. The comparison is performed on 95 distinct locations on highways in the Netherlands and spans October 2015. This is the same data we also use to test if we can observe the same patterns in traffic intensity from both the mobile phone data and the roadside measurement data. In the next chapter we go deeper into how the datasets are prepared and constructed. Hence, a description is omitted here.

To convert mobile phones to vehicles there are at least to fundamental steps. The first is converting people to vehicles, and the second is scaling the sample to the population. How the former is performed will be discussed in 7.2.2 and hence a description is omitted here. The latter, i.e. scaling our sample to the population, is done using the method described in 6.1 and is what we evaluate here.

In figure 6.5 we show the four steps and corresponding average people and vehicle counts. In blue from left to right we go from people on the road, to scaled people on the road and finally vehicles inferred from the mobile phone movement data. In yellow on the right we have the vehicles measured from roadside measurement devices. From here we can see that the scaling works well, but still produces a slight overestimation of the number of vehicles found on the road. The difference being a mere 11.7%.

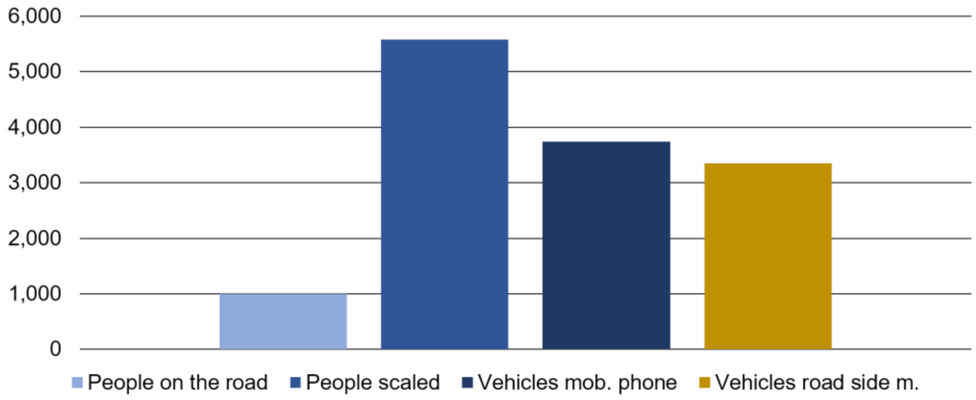


Figure 6.6, vehicles inferred from the mobile phone data to vehicles measured on the road.

As is depicted in figure 6.7 the scaling factor successfully scales the sample to the traveling population for the majority of measurement sites irrespective of the day of week. Nonetheless, there are some outliers. These outliers are also the result of the 11.7% difference we just found comparing total vehicle counts. We, for one, count 1.6 and 1.8 times as much vehicles than there are found on the road from Zwolle to the South of the Netherlands and between Amsterdam and Schiphol, respectively. On both occasions we find the roads are very close to a busy railroad. Because this makes distinguishing people from more difficult, we suspect the added vehicles found on the road are misclassified train passengers. On two other locations we infer up to 35% less vehicles than there are found on the road. One location is in Southern Limburg on the A76 between Germany and Belgium. Here it might be the case that the road is used by a lot of foreigners who are more difficult to detect, e.g.

because they leave off their phone to prevent roaming charges. The other is in the centre of Rotterdam. Here we have no explanation of why we should measure fewer vehicles. The only logical explanation here is a fault in the route assignment algorithm that assumes there is another, more optimal path, people will take.

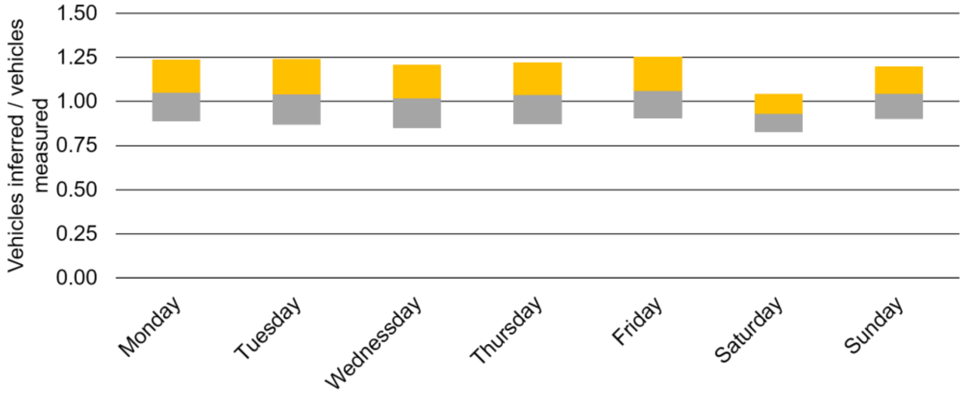


Figure 7.7, the ratio of vehicles from the mobile phone data versus those measured on the road for the middle 50% of comparison sites. The median is located where yellow meets grey.

6.3 Conclusion

In this chapter we have presented a new scaling method. The scaling factor aims to scale to the traveling rather than general population and is unique in this aspect. To do so demographic characteristics such as mobile phone use and chance of traveling are taken into account.

Furthermore, we show in 6.2 that the scaling factor helps us to get good estimates of the true traveling population. In the end that is all what the scaling factor has to do, i.e. correctly going from sample to population. Hence, we find the presented scaling factor is a success. On some occasions there is some deviation between the vehicles inferred and those measured, but this is not surprising given that we also had to distinguish train from road users, decide who travels over any specific road, and translate people to vehicles.

7 Focus

In this chapter the technique to assign people from the Origin Destination (OD) matrix to the road is evaluated. As discussed in 4.2.4, this is done by comparing roadside measurement data with the mobile phone data. The complete method describing how measurements from both sources are compared is represented in figure 7.1. This is the same figure displayed and explained in 4.2.4 and will not be further discussed here.

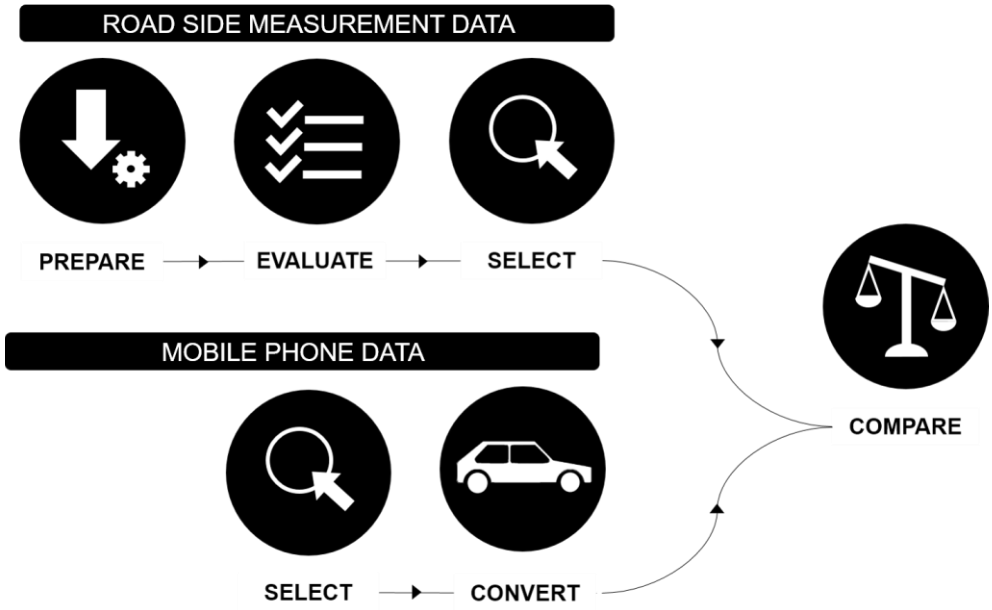


Figure 7.1, method for comparing traffic counts from roadside measurements with mobile phone data.

In section 7.1 we discuss the preparation evaluation and select phase related to the roadside measurement data. In section 7.2 the select and convert phases of the mobile phone data are discussed. After these steps we have vehicle counts for selected road sections on highways in the Netherlands from both data sources. In section 7.3 we will compare these vehicle counts to see if we can observe similar vehicle count patterns in both data sources.

7.1 Roadside measurement data

7.1.1 Prepare

The number of vehicles measured on the road is obtained via road site measurements provided as open data by Nationaal Databank Wegverkeersgegevens (NDW). NDW is a government initiative in the Netherlands that collects the measurement data from different parties such as Rijkswaterstaat (similar to the Ministry of Transportation in the USA). Most of the roadside measurements are taken using an inductive-loop measurement device placed on or in the road's surface with a self-reported accuracy upwards of 90%. An overview of the types of roadside measurement devices, the number of occurrences on the Dutch road network, and the mean self-reported accuracies of each type of device is shown in Table 7.1 (NDW, 2015).

Table 7.1, roadside measurement devices in the Dutch road network.

MEASUREMENT DEVICE	COUNTS	ACCURACY
Inductive-loop vehicle detector	21.711	99%
Automatic Number-Plate Recognition	1.484	95%
Bluetooth	1.409	Unknown
Infrared	948	100%
Floating Car Data (from navigation systems)	24	Unknown

The information represented in table 7.1 covers all measurement sites in the Netherlands, including information on the use of parking lots and gas station et cetera. Moreover, the measurements from these measurements sites contain predominantly raw data. For major roads Rijkswaterstaat cleaned the raw data by, for one, removing outliers. The algorithm processing the raw data is called Monibas, an algorithm that is proven to be highly accurate (Technical University Delft, 2006). The Monibas processed data are present in the data as a separate measurement sites. In total there are 13.693 measurement sites for which Monibas is applied, all of which employ raw data from inductive-loop vehicle detectors. Research by the Technical University Delft (2006) commissioned by Rijkswaterstaat concludes Monibas data, in particular for real-time reporting, is by far the superior algorithm among competing algorithms at the time and is also much more accurate than the raw data. Hence sites with Monibas data are preferred and are taken as a baseline for the comparisons to see how well CDRs are translatable to vehicles on the road Monibas sites.

Raw data is provided by NDW (NDW, 2015b) in xml format and converted to a manageable csv using software written in Python (Van Rossum & Drake, 1995). The csv contains information about the measurement site as a whole for 15 minute periods, e.g. the average vehicle count or average velocity. In addition, it contains information about the minimum and maximum vehicle counts and velocities as well as information on the number of trucks passing by. To keep data manageable, information about the independent lanes are aggregated to a single vehicle count. In total data size is reduced a factor 800, i.e. from approximately 80 GB to 100 MB for a day of information, while maintaining all the relevant information for this analysis.

Data for October 2015 has been downloaded and prepared using the conversion software just discussed. When downloading the data not all minutes appear to be present. Data for October 2015 was much more complete than September 2015. Hence, the data for September 2015 is left out in this analysis. All days in October has at least 1400 of the 1440 minutes present (97%).

7.1.2 Evaluate

Quality constraints will be applied to ensure only the most trustworthy and error free roadside measurements will be used to test the accuracy of the trip assignment of the mobile phone data. Research shows the roadside measurements may be reasonably accurate and the self-reported accuracy of these measurement devices is high (Nihan, Wang, & Zhang, 2002). Nevertheless, there might still be exceptions to the rule and it never hurts to double check the data quality.

For the final comparison we want to know (1) the accuracy of the roadside measurement sites in general, (2) when we have to discard information due to too many missing data, and (3) that the measurement site is at the stated location. To test this the roadside measurement sites will be evaluated on the following three aspects. These, respectively, are: (1) compliance between consecutive measurement sites, (2) missing/erroneous data, and (3) measurement site location validation.

Compliance between consecutive measurement sites

To test compliance between consecutive measurement sites we first make a selection of sites for comparison. Measurement sites are chosen such that there is no off or on ramp between two measurement sites and all vehicles passing the first measurement site also have to pass the second measurement site. In total 60 measurement sites on Dutch highways are chosen for analysis. This means there are 30 unique locations, i.e. 2 consecutive measurement sites per locations. We also consistently chose two sites in going one direction and two sites going the other direction. The points on the map in figure 7.2 show where the location of the chosen measurements sites.

Figure 7.2, locations of the consecutive roadside measurement sites (black dots).

For all sites we find a very high correlation between what consecutive measurement sites report. Pearson correlations are all near one and greater than 0.977, which confirms there is high compliance between measurement sites. Furthermore, we wanted to see if there are sites with strong biases. To do this we calculated the difference between consecutive measurement sites at each location and divided this by the average vehicle count over both sites. The measurements used are those of the average vehicle counts per 15 minutes. The bias, or deviation in average vehicles measured in percentages, are shown in figure 7.3.

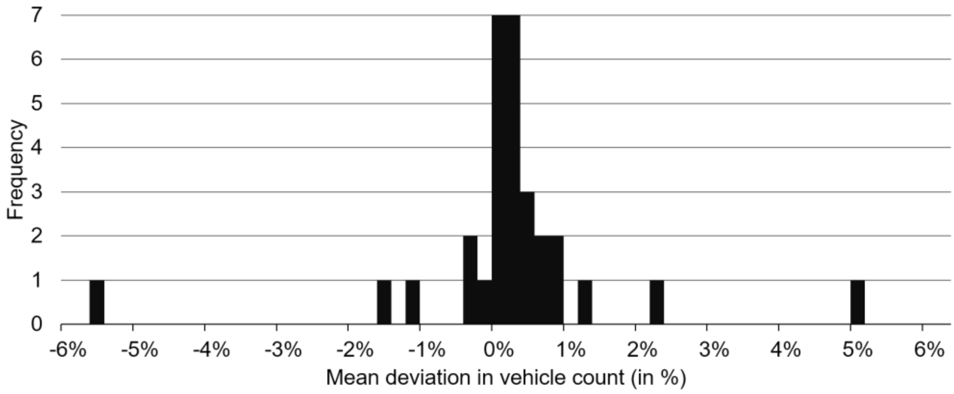


Figure 7.3, deviation in vehicle counts between consecutive measurements.

We observe from figure 7.3 that most sites have at most a very small bias. The majority of the locations show a deviation between sites between -1% and 1%. This is no cause for concern as it can be explained by people driving in between the measurement devices when switching lanes, for example. There are, however, two outliers visible that do require further investigation. One outlier can be found at 5% and one near -5.5% in figure 7.3. For these sites we are unable to detect one of the two supposedly consecutive measurement sites for both outliers with Google Street view. The measurement devices should be installed prior to the imagery on Google Street View and are typically easy to distinguish in the road. Hence we expect these measurement devices to be located elsewhere than stated by the NDW. This is also the reason that later on we will provide a measure to check whether the measurement sites are where there should be. The outliers found are discarded from further analyses in this chapter.

Missing/erroneous data

The next phase in our evaluation is investigating how the number of missing minutes relates to the quality of the measurements. The measure we employ to test data quality is the mean absolute deviation between consecutive measurement sites. The hypothesis we explore in this subsection is that data quality improves when there is less missing data. Typically we use the average of the measured vehicle counts and multiply this by 15 to get to the 15 minute vehicle count average. For example, when 6 minutes contain errors the 15 minute average is the mean vehicle count over the remaining 9 minutes. With missing data we start to lose some certainty and start estimating rather than measuring. However, by only using data without missing minutes of data we may exclude too many data points from our sample. Here we try to get a grasp of how data quality relates to having missing data to quantify when the data is still acceptable and when to discard the data.

The same sites and measurements are used here as in the previous subsection without the spotted outliers. The mean absolute deviation in consecutive measurement sites is based on 15 minute vehicle counts. Note this differs from what is presented in figure 7.3 as there we used total counts per site rather than differences per site per 15 minutes. The results are shown in figure 7.4 in relation to the maximum number of minutes with errors.

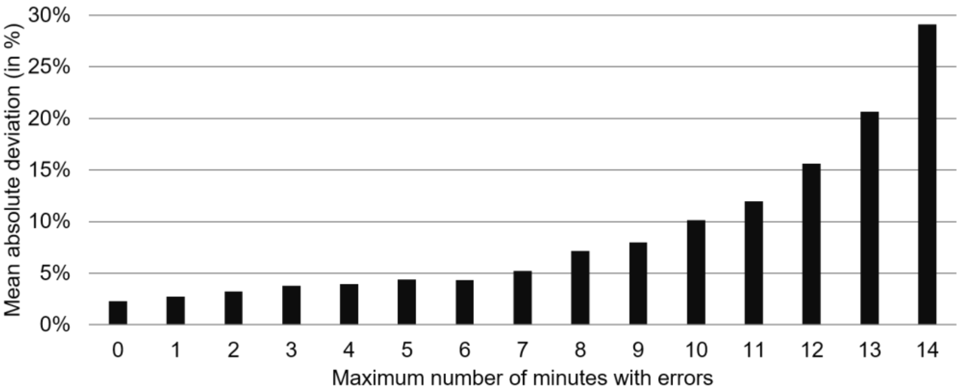


Figure 7.4, the relation between de deviation in vehicle counts and the maximum number of erroneous minutes per location.

We observe that the deviation is very small for the first part, but grows rapidly at and past 7 minutes with errors. At 7 minutes with errors we also see the mean absolute deviation growing to above 5% and resulting at a deviation of 29% at 14 erroneous minutes. For further analysis we will only include measurements made with 6 or less minutes with errors. When we use the 6 minute criteria the 15 minute vehicle counts of all measurement sites appear to follow a normal distribution shown in figure 7.5.

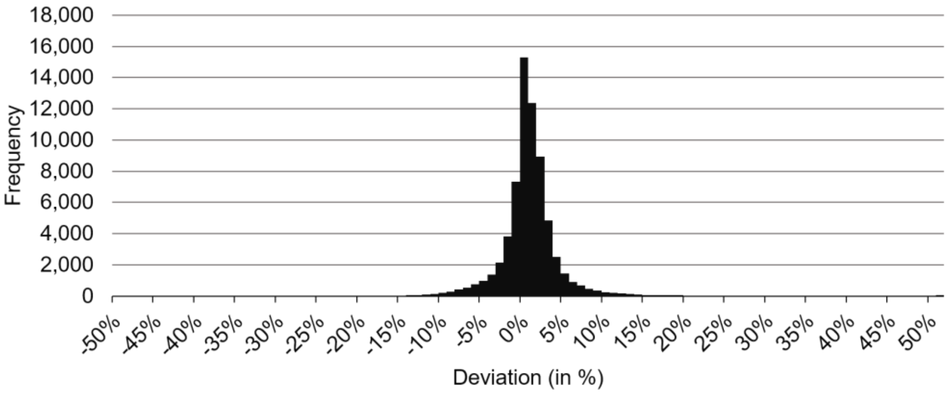


Figure 7.5, histogram of the percentual deviation in measurements per 15 minute interval for all locations grouped.

From figure 7.5 we observe that the differences between consecutive sites appear to be normally distributed with a mean at 0. The standard deviation of the shown distribution is 0.045. When the error in measurement of each site is normally distributed and equal amongst sites, the σ for each measurement site becomes approximately 0.032 following the variance sum law. This is very acceptable and provides us with a good indication of how accurate the roadside measurements are.

By excluding all measurements with more than 6 minutes we discard part of the available data. The percentage of good quarters, i.e. with six or less erroneous minutes, for all Monibas measurement sites for October 2015 is shown in figure 7.6.

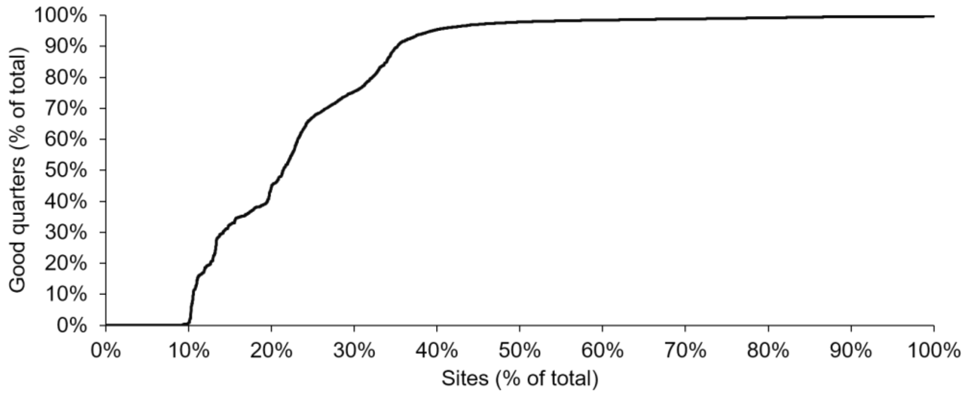


Figure 7.6, percentage of good quarters (with 6 or less erroneous minutes) plotted for all 7796 Monibas measurement sites in the database.

From figure 7.6 it becomes clear approximately 60% of the roadside measurement sites produce consistently good quarters. 30% produces good data some of the time and 10% of the sites do not produce any good quarters at all. Although it is not directly relevant here because we can cherry pick the good sites, we find the error rate of the measurement sites preposterously high. The costs per measurement site is in the range of thousands to tens of thousands of euros a year (Middleton & Parker, 2002; ITS International, 2010; Verkeersmonitoring met inductielussen, n.d.). In other word (using €10.000 as a rough approximation of the actual costs of a site) there is about €7.8 million being spend on sites that do not produce any good data and another €23.4 million on the 30% of sites that have some good quarters. The good news, however, is that there are still about 4.775 sites that have 95% or more good quarters that we can compare with the mobile phone data.

Measurement site location validation

For the final comparison we want to be sure the selected roadside measurement sites are at the correct location. For this we will opt for and present a novel automated method to decide what measurement sites are likely at the stated location and what measurement sites are not.

The method is founded on the following premise: measurement sites that are located on the same road near each other will produce more similar vehicle counts than measurement sites further apart. The idea is thus to (1) cluster measurement sites based on their vehicle counts and (2) check if sites within each cluster should be located on the same road based on the metadata about the measurement sites. If 70% of the measurement sites within one cluster belong to the same road and are headed the same direction these measurement sites are most probably where the metadata states they are. If the site belongs to a completely scattered cluster or the majority of the sites belong to one road, but this measurement site is on a different road, the measurement site is probably not where it should be.

Clustering will be performed using the K-means algorithm as described by Hartigan and Wong (1979). As stated by Hartigan and Wong (1979): "The aim of the K-means algorithm is to divide M points in N dimensions into K clusters so that the within-cluster sum of squares is minimized". For us the N dimensions the vehicles measured for specific hours for which we have measurements and the M points are unique measurement sites. Typically, one starts by randomly assigning K points and set these points as cluster centres. In the following iterations over the M points all points are assigned to the nearest cluster. Thereafter, the cluster centres will be recalculated. This process may repeat itself a number of times and finally one is left with K clusters that together incorporate all M points. We set the maximum number of iterations to 200 to ensure a good (local) optimum is found. The within-group sum of squares is the measure we try to minimize here. It is defined as the sum over all squared distances between each point and the cluster centre it belongs to. When K equals M each point is essentially a cluster centre and the within group sum of squares is 0. Selecting M clusters would be pointless. Hence, we want to select the number of clusters that provide a good trade-off between a low within group sum of squares in combination with a low K. The goal is to find this point where more clusters produce only a small reduction of the total within cluster distance. This point is called the ‘elbow point’. The plot in which the elbow point can be determined is called a scree plot. Scree plots were originally proposed by Cattell (1966) to find the number of components that should be included in Principal Component Analysis. In cluster analysis the same principles from Cattell’s (1996) research are still applied, but for determining the number of clusters rather than components.

For our analysis we first selected all measurement sites that on average have 95% or more average vehicle counts meeting the missing data criteria from the previous section. This leaves us with 4.775 measurement sites. We then discarded any hourly vehicle counts with average vehicle counts not meeting our criteria. As k-means is unable to handle missing data we than had to select all hours from October where every measurement site has good measurements. In total all this leaves us with 145 hours spread over the month October 2015 for which we have good data on all selected measurement sites. To ensure each hour gets equal importance the vehicle counts are normalized for each hour. Otherwise a 1% difference in vehicle counts during rush hour would result in a greater distance between sites than a 1% difference in vehicle counts during the quiet night time. This appears strange as we would assume two sites with, for example, 1.500 versus 1.600 vehicles are more similar than two sites with, for example, 100 versus 200 vehicles measured. Scaling will help bring vehicle counts in perspective and provide a more just measure to compare compliance between sites.

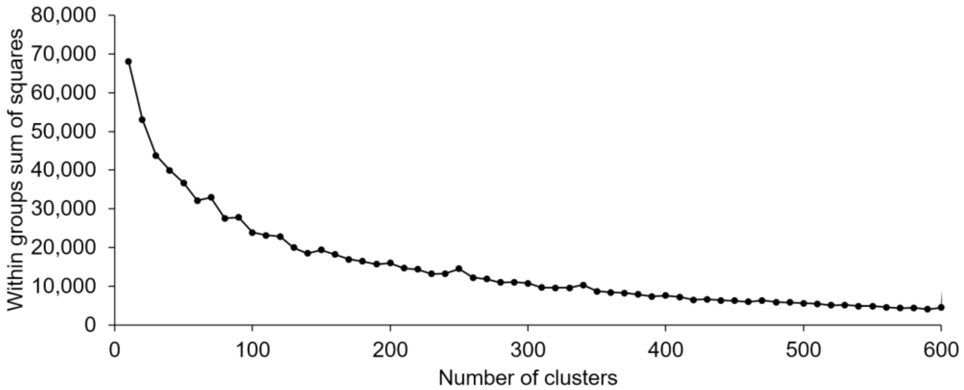


Figure 7.7, scree plot providing an indication of the number of clusters in the roadside measurement data.

Figure 7.7 shows a scree plot that helps us get a grasp of the number of clusters in our dataset. Unfortunately, figure 7.7 does not provide concluding results that drive us to state whether the true number of clusters is 70, 90, 150, or 240. At all locations the trend goes from 'strong' downward to nearly levelling out. As the true number of clusters is unknown and the scree plot is inconclusive the best we can do is taking an educated guess as to what the true number of clusters will be. The scree plot is, nonetheless, useful as it confirms that the number of true clusters is in the range of a hundred to a few hundreds. Note that selecting the correct number of clusters is more of an art than a science. On the one hand, we want the clusters to be large enough to determine whether part of the cluster is wrongly located. On the other hand, we want clusters to be small and thus specific enough such that each cluster will cover only one location.

We choose 240 as the number of clusters. At 240 clusters there are on average 14 measurement sites per cluster. From the knowledge gained by working with the roadside measurement data we expect this to be a good compromise.

Figure 7.8 shows the result of our clustering. In figure 7.8 we depicted all measurement sites on the right lanes with the colour unique for each identified cluster. Because of the large number of clusters it is rather difficult to identify the outliers from the image below. However, figure 7.8 provides us with confidence that the clustering is reasonable as we observe long strings of the same colour, i.e. cluster, on many major roads. This also shows the vehicles measured on the same roads in consecutive order often provide very similar measurements. On some occasions, nonetheless, we find a continuous stream of, for example, blue dots being separated by a single different colour point. These are plausibly the points where the measurement site either has a false location or provides false measurements.

Figure 7.8, clusters of similar measurement sites on the right lanes on the Dutch road network.

Further analysis shows that of all clusters just under half (49%) has solely measurement sites on one side of the road, i.e. exclusively left lane or right lane. When we are strict and state the measurement sites need to be (1) on the same side of the road and (2) on the same road we reduce the number of clusters to 103 (35%) and the measurement sites to 1.076 (32%). The relation between the number of completely identical locations, i.e. based on road name and side of the road, and the number of clusters is depicted in figure 7.9.

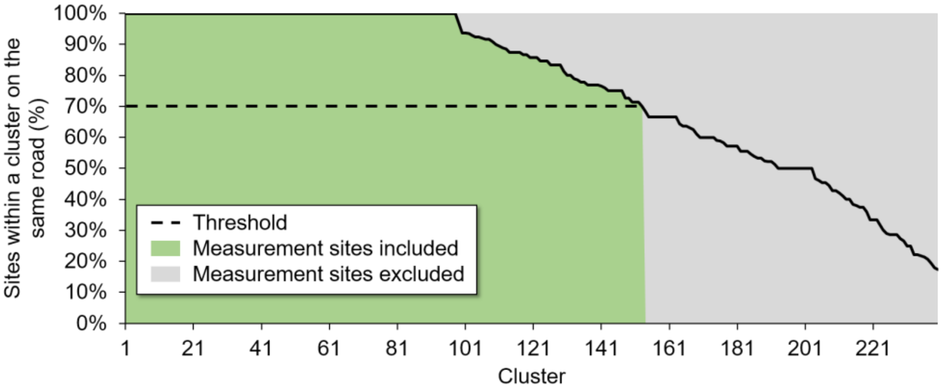


Figure 7.9, percentage of sites at the same road and on in the same direction for all 240 clusters.

The goal here, however, is not to be very strict. The goal here is to identify outliers, i.e. measurement sites that are on a different location than the other measurement sites in its cluster. We draw the line at 70% meaning that 70% of the measurement sites within one cluster need to be on the same road in the same direction. The remaining measurement sites in the clusters will be discarded as they are probably outliers. This leaves us with 1.761 good measurement sites. 140 measurement sites, although good according to the minimum number of errors criteria, are discarded as outliers. These 140 sites are plausibly wrongly located or producing trivial information.

Conclusion

We constructed two criteria that will help us to extract the most trustworthy of road side measurement sites. These are making sure the used measurements have 6 or less minutes of erroneous data and meet the clustering criteria, i.e. it belongs to the majority (70+ %) of measurement sites on the same road within a cluster. Furthermore, we found that road side measurement sites are very accurate overall, when they meet the above criteria. We may expect a standard deviation from the true vehicle counts of approximately 3.2%.

7.1.3 Select

The final step regarding roadside measurement data is selecting good measurement sites for the final comparison with the mobile phone data. To get a good sample we will (1) only select sites that we consider good, i.e. with 95% of the data meeting the 6 minutes or less of erroneous data, and (2) select sites randomly over the road network.

The step in the select phase is filtering out the measurement sites that are bad, leaving us with the good sites. Furthermore, for comparison only the sites that measure traffic staying on the highways are used. 23% of the sites measure how many people enter or leave the road at the exits. These sites we did not evaluate because no two consecutive sites measure the number of people leaving. Hence, the quality of these sites are uncertain. We expect the measurements here to be slightly worse because people might be cutting corners, for example, and consequently skipping the traffic detectors in the road (Appendix B).

The second step is selecting a subset of the good sites. We might investigate the relationship on all measurement sites, however, this will not result in a good test. The distribution of measurement sites over the road network is not uniform. Near the larger cities in the Netherlands where roads are busier, more measurement sites are positioned, e.g. between Amsterdam Rotterdam and Utrecht. Using all sites for comparison will thus result in a bias as comparisons near major cities will be more valued than those in less congested areas, which possess fewer measurement sites. To limit the bias we select measurement sites randomly such that each site on the same road have to be at least a 5 km distance from all other selected sites. In total 100 sites are selected. There are 50 sites going one direction and 50 sites going the other direction. The sample is generated in R where distances are calculated using the geosphere package (Hijmans, 2015). The locations of the selected sites will be shown later with the corresponding correlation coefficients, i.e. the results, in section 7.3, figure 7.13.

7.2 Mobile phone data

7.2.1 Select

Route assignment from OD-matrices

There is a large volume of published studies describing how to assigning vehicle trips from Origin Destination pairs (OD-pairs) (Prato, 2009; Ortúzar & Willumsen, 2011). The key assumption often made is that people are rational and take the route that minimizes their travel cost (Ortúzar & Willumsen, 2011). Travel cost can be seen as a combination of multiple factors such as travel time, distance, cost of fuel, congestion charges, et cetera (Ortúzar & Willumsen, 2011). The most important factors, explaining 60% to 80% of all route choices in practice, are travel time and distance (Ortúzar & Willumsen, 2011). Methods taking the shortest path or k shortest paths as the possible route choices between OD-pairs account for the largest group of path generation methods (Prato, 2009). At the moment the dataset available for this study uses Dijkstra's shortest time path algorithm to link OD-pairs to road sections (Dijkstra, 1959). The shortest time path is chosen by time rather than distance. This is done by taking into account the maximum speed allowed on a road section based on the information from Open Street Maps (OpenStreetMap contributors, 2014).

The shortest time path algorithm, however, has its limitations. Knowing what vehicle trips to take into account when performing post-hoc analysis of the impact of roadworks is fundamental and the shortest time path allocation method is not 100% accurate. Prato (2009) argues that the shortest time path method might not be ideal as it does not take personal preferences or abstractive measures of route attractiveness into account. By applying the shortest time path algorithm all vehicle trips from A to B are assigned to a specific road, which may be unrealistic. Nevertheless, as 60% to 80% of all route choices can be explained by a combination of time and distance there is still support for using the shortest time path (Ortúzar & Willumsen, 2011). We may find that a portion of travellers are found on other roads. Roads that are often congested, for instance, may be avoided and other roads may be used more than the shortest time path algorithm may identify. When scaling from the sample to the entire traveling population this has to be kept in mind as it may influence the scaling factor. Nevertheless, we propose to use the shortest time path algorithm for this research regardless of its imperfections. The algorithm is already in place and might still provide adequate results. As indicated earlier this research aims to improve the state of the art and encourages future research to build upon it. Implementing a newer more advanced trip assignment method is left for future research.

To get from OD-pairs to devices on a road we need to know what route, i.e. set of connected roads, a device took to get from the origin to the destination. For each OD-pair in the OD-matrix a route is pre-assigned based on the shortest time principle, i.e. the path from origin to destination that takes the least amount of time. The total amount of GSM-equipped devices assigned to a piece of road is then determined by taking the sum of all OD-pairs of which the route contains the road of interest. The digital version of the Dutch road network is obtained from Open Street Maps (OSM). From the OSM data the shortest time is calculated and the shortest time routes thus contain all road parts as supplied by OSM. This implies that for each piece of a road, up to the level of detail of OSM, devices passing that piece of road can be inferred. The shortest time route assignment is an assumption that may influence how well devices measures relates to vehicles measures on the road. In this chapter we validate, for one, whether the route assignment using the shortest path is a good assumption.

The roadside measurement sites selected in 7.1.3 are linked to the roads in OSM. This is done by assigning the measurement site to the nearest road where the road name, e.g. A2, is equal. For all Monibas measurement sites we checked by hand whether the imposed link appeared correct. We then selected the road section corresponding the roadside measurement sites selected for analysis.

Handling the minimum of 15' rule

The data from CDRs are constrained by a 'minimum of 15' rule. This rule entails that only if a minimum of 15 devices can be aggregated for an activity we get precise measurements. If less than 15 devices are measured the output will be zero, although anywhere between zero and fifteen devices are observed. The minimum of 15 rule is in place to ensure the privacy of the people owning the GSM-equipped devices as discussed in 5.4.1.

The minimum of 15 rule brings a number of unique challenges to obtain information from the CDRs. In particular, when looking at either very short time frames or investigating movements between low populated areas there is a real chance to drop below the threshold of 15. For comparison between CDRs and roadside measurements this has to be taken into consideration. Two measures taken for this analysis to decrease the chance of dropping below 15 measured devices. The first measure is to take all OD-pairs that pass a road section and group them together as one. On the one hand, there is now no way to distinguish between different OD-pairs and observe where people passing a road section originate from or go to, thus loosing detail in the information. One the other hand, a much larger sample is taken, thus the chance of dropping below the 15 threshold and loosing information in general becomes much smaller.

Figure 7.10 shows the origin (yellow) and destination (blue) of all OD-pairs passing a road section along the A2 between Amsterdam and Utrecht. Note the colour indicates the number of origin and destination links per area. It is not affected by the number of trips from each origin and arriving in each destination.

The second measure taken is to take a sufficiently large time fame. Although the roadside measurements are available on minute basis, for the CDRs a timeframes of hours are taken. This is to be sure the minimum of 15 rule is met. We tried to do timeframes of 15 minutes, but too often the minimum of 15 rule was not satisfied.

Moreover, because the timestamping is only an approximation a time frame of an hour would be more representative of when a person really drove past the road section. For the analyses in this chapter time frames of an hour are used.

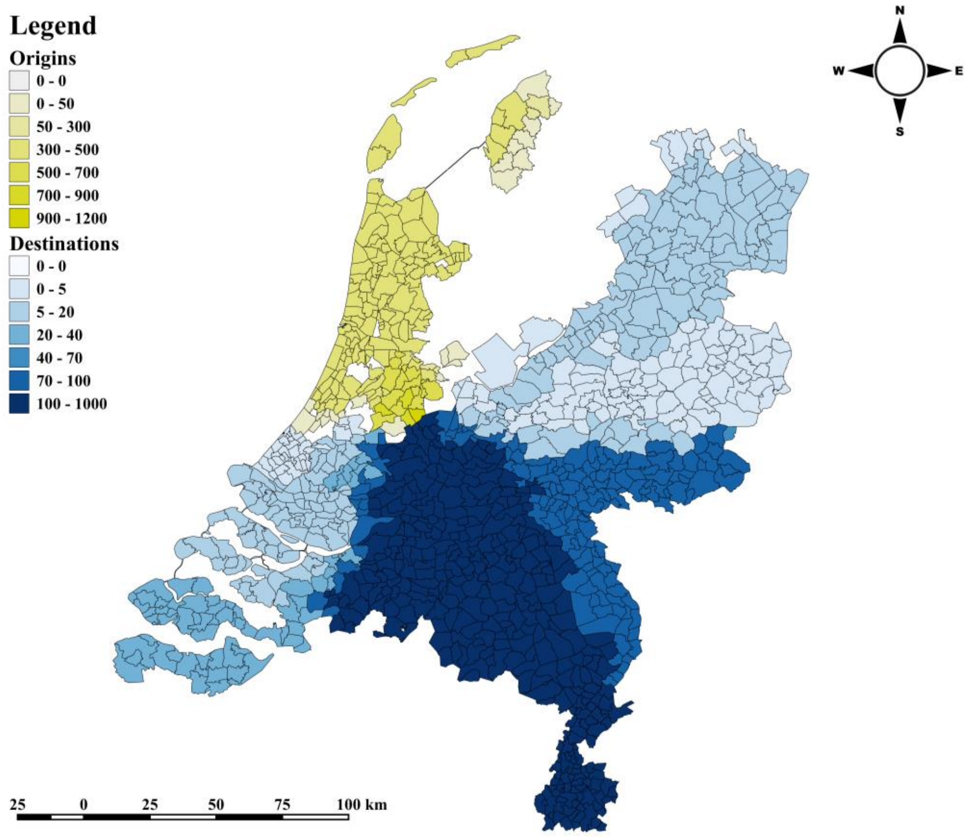


Figure 7.10, map of the origins (Yellow) and destinations (Blue) passing the A2 from Amsterdam to Utrecht.

Timestamping

When comparing the mobile phone data with the traffic counts we want to compare them based on how many vehicles are measured on the road at a certain moment in time. For the roadside measurements the timestamps are already in place. For the mobile phone data this has to be constructed as there is no information about where a person is during a trip, only that the person moved from A to B between time t0 and t1.

For timestamping we will use the middle of the trip. For example if a person leaves at 9:00 and arrives at 9:50 we assume the person crossed the road at 9:25. This is not super accurate as the road with the roadside measurement device might actually be crossed earlier or later during the trip. However, this is not a major limitation because we are never much off given we compare only hourly vehicle counts. 76% of the trips over 10 km in the Netherlands end within an hour and within 2 hours over 94% of the trips are completed (CBS, 2014a). Hence in only 24% of the cases we are able to be more than an hour off the real moment when the measurement site is crossed and in just 6% of the cases we can be off more than 2 hours. Moreover, while sometimes the middle of the trip is earlier than the actual moment of crossing the measurement device, i.e. when the road section is near the end of the trip, the opposite is just as likely. For some people the road section is close to the origin and for some close to the destination. When we make an error on one side this might is compensated by an error on the other. Overall, we thus expect that the middle of the trip will result in good estimates for the hourly vehicle counts. Hence, we will not attempt to devise a more precise timestamping method just for this comparison.

7.2.2 Convert

In essence the mobile phone data measures the movement of mobile phones and thus people rather than vehicles. A translation is thus required to go from people to vehicles on the road.

To get a good estimate of the number of people per vehicle, survey data from 5 years of OViN are used, starting at 2010 and ending at 2014. In the combined data there is information about 97.432 trips that are comparable to those in our mobile phone data. To get comparable trips in the OViN we applied some selection criteria. We discarded trips outside the Netherlands and trips under 10 km, which is the distance 'as the crow flies'. Furthermore, trips with unrealistic travel times, i.e. trips with velocities above 145 km/h, are discarded as these trips are most likely errors. OViN is chosen because it contains information about the means of transportation, e.g. train, car or bus, and for people traveling by car there is a distinction between being a passenger and a driver. This we will link to known people per vehicle ratios for the different means of transportation to get a good estimate of the chance a person is a driver. The percentage of people traveling over the road that are drivers is directly related to the number of people per vehicle and hence can be extracted from the OViN.

An overview of all means of transportation of trips greater than 10km with vehicles that can be found on highways is shown in table 7.2. The two largest groups by a margin are car (driver) and car (passenger). Here the description provides enough context to know if the person is a driver or not. The third largest group bus (public transport) accounts for nearly 5% of all trips. The chance of being a driver for this class (11%) and the class motor is extracted from literature on Dutch public transport (Otten et al., 2014). Bus (private) is assumed to be similar to bus (public transport). All other classes combined including taxi and freight truck only account for just over 1% of all trips. The assigned chances of being a driver for these classes are based on personal experience. Provided their low share amongst all trips the possible impact of mistakes due to guess work is considered to be negligible.

Table 7.2, different types of means of transportation, their prevalence on the Dutch roads, and the change of being a driver.

MEANS OF TRANSPORTATION	% OF TRIPS	CHANCE OF BEING A DRIVER
Bus (public transport)	4.93%	11%
Bus (private)	0.57%	11%
Camper	0.05%	50%
Car (driver)	43.67%	100%
Car (passenger)	17.09%	0%
Delivery van	0.52%	95%
Motor	0.44%	87%
Taxi	0.44%	50%
Freight truck	0.06%	100%
Other / not on highway	32.23%	-

We find the motive of a trip, rather than hours of the day or day of the week, can provide very stable and coherent people per vehicle ratios. In figure 7.11 the people per vehicle ratio, i.e. the inverse of the chance of being a driver, is depicted for each of the three motives, i.e. work, business and other, and hour of the day. As can be observed from figure 7.11 is that the ratio people per vehicle is stable over the majority of the day. The unstable pattern in the early morning can be attributed to a low sample size in the early hours of the day. Note that we included the motive other twice, once for workdays and once for weekends. This we did as there was a clear difference in people per vehicle between the two day types for this motive. During the weekend there are generally more people per vehicle for non-work and business related activities. The people per vehicle ratios applied are 2.02 for other weekend, 1.64 for other workdays, 1.08 for business, and 1.10 for work trips.

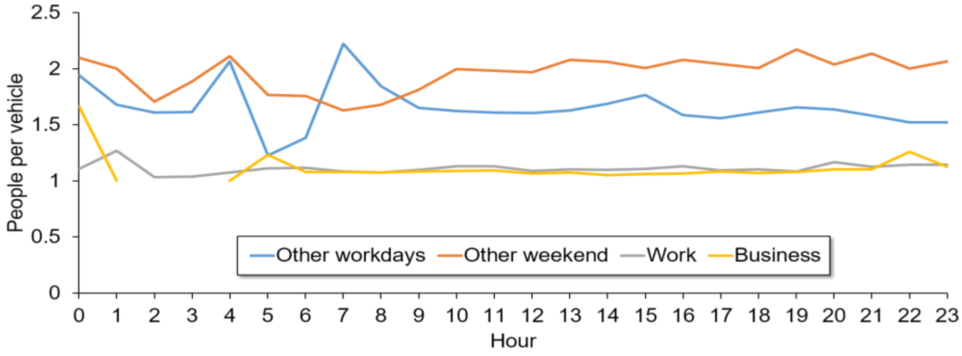


Figure 7.11, people per vehicle ratio plotted for different trip motives averaged per hour of the day for October 2015, with an additional separation for the motive other between workdays and weekend.

7.3 Comparison of the data sources

As we now finished the data gathering and preparation stages for both data sources we can now perform a sound comparison. Part of the comparison with respect to total vehicle counts has already been discussed in section 6.2 to evaluate the scaling factor. We found the absolute vehicle counts to correspond very well for the majority of measurement sites. Here, however, we will test whether patterns in traffic intensity on the road can be inferred from the mobile phone data by comparing it to the gold standard, i.e. road side measurements. If this is also the case we can say we can correctly infer vehicle counts from the mobile phone data.

The most common measure that states how well patterns occurring in one data series correspond to those in another is the Pearson correlation coefficient. If when one data series goes up by 10% the other does the same for all data points the correlation coefficient will be 1. When the opposite is true, i.e. one goes up then the other goes down, the correlation coefficient is -1, and when there is no relation it will be 0. In figure 7.12 a histogram of the Pearson correlation coefficients is shown. We can see that for the majority of the locations the Pearson correlation coefficient is above 0.9 indicating a very high correlation.

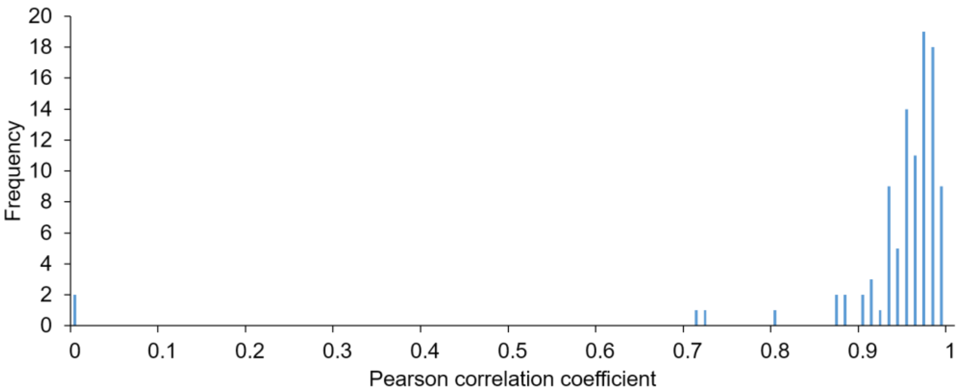


Figure 7.12, Pearson correlation coefficients for 100 selected test sites in October 2015.

On some occasions, however, it drops even below 0.8, which is still high in general but low compared to our other findings. To find out why these outliers are present we performed some further analyses. As it turns out there is a logical explanation of why these outliers exists. The route assignment algorithm only contains paths from the centre of an area to the centre of another area. Hence, trips cannot be assigned to roads beyond the centre of an area when that area nears the border. Hence there are no or very few origin destination combinations crossing these roads resulting in very low or non-existent vehicle counts.

In figure 7.13 we show the Pearson correlation coefficients of all 100 selected sites, note some overlap between points is possible. Below the dark blue points all the way at the left of the Netherlands are four points in total. This is where the Pearson correlation coefficients are much lower than average and vehicles are hardly detected in the mobile phone data

Figure 7.13, all 100 sites plotted with on a map with their respective colours indicating the Pearson correlation coefficient.

Points with a correlation coefficient at and below 0.8 are removed from the analyses performed in 6.2 and hence only 95 measurement sites are compared there. For the later roadwork analyses we will not investigate roadworks happening near the country border as vehicle counts may be misinterpreted.

When vehicle counts over a week are compared we get the results as shown in figure 7.14. Here we selected one of the higher correlating measurement sites and use only the first week, starting on Monday, of October 2015. The latter we do to provide a clearer graph. Note that the drop on October 6th 2015 is due to missing data and is not taken into account for the other analyses. Overall, we find the results are great and, with the exception of roads near the country border, we are confident the relatively simple least time path algorithm suffices in assigning people to the road.

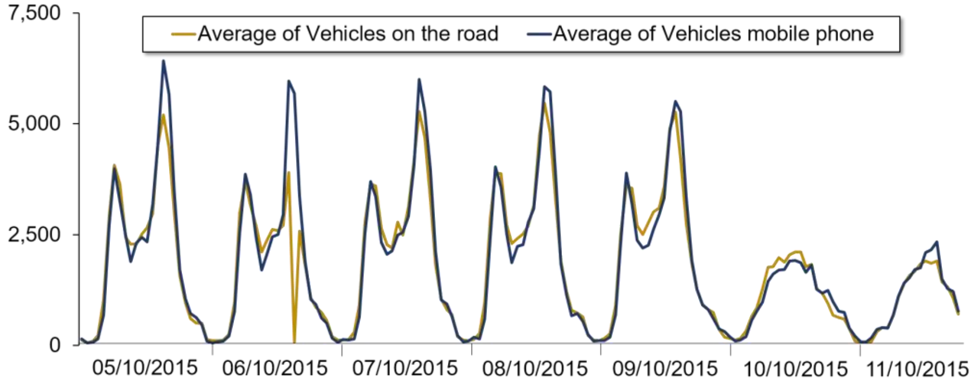


Figure 7.14, hourly averages of vehicles measured on the road and those inferred from the mobile phone data.

7.4 Conclusion

We performed an extensive analysis of roadside measurement data and explained how to use mobile phone data to go to vehicles on the road. Furthermore, we selected a 100 sites that met all criteria and performed an analysis. The analysis was performed to see whether road intensities on the road can be inferred from the mobile phone data.

The key lessons from the results in this chapter with respect to the roadside measurement data, mobile phone data, and the comparison are the following.

We found the roadside measurements are very accurate overall. The number of minutes with errors, however, do affect the compliance and thus quality of the measurements. By only including measurements with 6 or less minutes of errors the average standard deviation of measurement sites are estimated to be 3.2%, which is quite accurate. Moreover, we found that in some occasions the roadside measurements are not where they should be. Clustering was performed to check whether they are at the correct location, i.e. road section, and the results appear promising. Of the 100 selected sites we never encountered an instance where we still expected the measurement site to be wrongly located.

The mobile phone data had to be converted from phones to vehicles. This is done based on the trip motives. For different trip motives, e.g. other work and business, the people per vehicle ratio appears to provide a very stable estimate. We did, however, find that during the weekend the recreational traffic has slightly more people per vehicle than during workdays. Hence, we make a distinction for this trip motive. For the other two trip motives no correction was needed.

Overall, the results of the comparison as presented in 6.2 and 7.3 about absolute vehicle count and patterns in vehicle counts, respectively, are both very promising. However, we do have to keep in mind the shortest time path algorithm is not yet able to assign vehicles to roads near the border resulting in a few outliers. Roadworks near the country border can thus not yet be evaluated. Nevertheless, for the remainder of the country we can definitely state we are able to measure who will be affected by roadworks by using the rather simple shortest time path algorithm.

8 Label

As discussed prior, the label phase consists to determine the motive of trips. For different trip motives there are different VoT and VoR stated, i.e. the economic impact differs per motive. We want to use trip and person characteristics to predict the motive of a trip. For this a model has to be created that can determine trip motives based on a set of trip and person characteristics. To create the model the CRISP-DM method is applied as discussed in section 4.2.5. The following six sections will contain the products resulting from performing the CRISP-DM method. Ultimately, this will lead to and include adding trip motive to the mobile phone data.

8.1 Business understanding

The main objective of labelling is to be able to assign trip motives to the mobile phone data. The reason for doing this is because delay experienced by people going from and to work is more costly than delay experienced by people going shopping (Kennisinstituut voor Mobiliteitsbeleid, 2013). Cost distinctions are made for trips with the following motives: home-to-work, business, and other (Kennisinstituut voor Mobiliteitsbeleid, 2013). In OViN home-to-work trips are defined as trips to and from a work location. This can be a regular as well as temporary work location. Hence, this also includes locations for on-call employees, part-timers and voluntary workers (CBS, 2014). Business trips are defined as trips that are due to work, excluding trips that are to regular work locations. This category mainly concerns service trips, customer visits, meetings and symposiums (CBS, 2014). All trips not belonging to either of these two groups we characterize with the motive other. The goal here is to get the most accurate estimate about the motive of a trip for these three groups.

It is important to understand that we always look at travel behaviour from a population or aggregated perspective. Consider for example the following situation. What we care about is being able to estimate the composition of the travellers such that we can assign cost to the delay. There is no reason to suspect travel motive influences the delay that will be experienced. There may be some fluctuations in the delay experienced per traveller, but this is more likely to be the result of random errors in measurement than actual differences in experienced delay. Hence the aim is more geared towards predicting the distributions well rather than predicting well on individual level.

In the end the goal is to use the model to make predictions on the mobile phone data. Because of technical constructs there are a few crucial points to take into account when designing the model. First of all, for this research, the mobile phone database can only be reached through SQL-queries. This is important to keep in mind because some models are easier to convert to SQL than others. Moreover, to be of practical use in the long run the query has to run faster than it takes new data to enter the database. If it takes, for example, two days to process one day of data this is unsustainable. Ideally it takes much less than a day, e.g. one hour max, to be able to also add trip motives to historical data and allow for other processes to run.

In summary there are three critical success factors. First, the model needs to significantly improve the estimation on the classes home-to-work, business, and other compared to the baseline, i.e. the a-priori information on the underlying distributions. Second, the model needs to be applicable in an SQL type of environment. Third, trip motives should be addable to the mobile phone data within a reasonable time frame, e.g. approximately one hour.

8.2 Data understanding

To link trip and or people characteristics to trip motive it is key to have a data source in which both are present. In the Netherlands the largest source of information containing both trip characteristics and trip motives is the OViN. This is a yearly survey aimed to gain information about mobility patterns of inhabitants. The survey is performed by the Centraal Bureau Statistiek (CBS), i.e. a large organization that is tasked to gather and present statistics about the country and its inhabitants. OViN gathers information across the Netherlands and across age groups about mobility. Other data sources about mobility patterns in the Netherlands do exist. The most noteworthy is from the Mobiliteits Paneel Nederland (MPN). The key differences are that the MPN covers three days in a person's life and OViN one, but at the cost of a ten time smaller sample size. Moreover, the three days available appear to be only a slight addition. Weekly patters, for example, are still not visible with the MPN. Because of the superior sample size and the fact MPN still does not allow to see weekly patters, the OViN is preferred over the MPN. This does imply we can only distinguish daily patterns and cannot identify travel patterns that span multiple days, e.g. a person always going to work at the same location. Additional information could potentially be generated by employing both data sources. Combining these data sources would require significant effort and given the limited time available for this research this will be left for future research. Moreover, if the model created just using OViN produces satisfactory results there may be no incentive to put in the extra effort.

Results of the OViN surveys for the years 2010, 2011, 2012, 2013 and 2014 are combined to provide a more stable and robust sample. The plausibility of the results found in the OViN survey have been analysed by the CBS and both passed the tests (CBS 2010; CBS 2014b; CBS, 2014c; CBS, 2015b). This is an indication the quality of the data appears logical on a high level. Weight factors are included in the results to compensate for biases in the sample. These weight factors can compensate for biases on household, person, and trip level. Because we try to predict motives for trips the weight factors on trip levels need to be included when training our model. In total the dataset contains information on about 650 thousand trips spread over the five years.

Because of the earlier design decision to focus on trips longer than 10 km, only a subset of the results from OViN are relevant. OViN does have distance classes to state the distance of a trip, but these are not necessarily as the crow flies. In the mobile phone data only trips over 10 km are taken where the distance is measured as the crow flies. Fortunately, the OViN does include the four digit postal codes of the origin and destination. An additional database retrieved from postcodedata.nl is merged with OViN to add GPS coordinates to the origin and destination (Postcode Data, 2014). Before adding the GPS coordinates to the dataset the GPS coordinates per postal code, e.g. 1234AA, are averaged to get the average for the four digit postal code as present in OViN. In data preparation the GPS coordinates are translated to distances and by taking into account the reported travel times also the velocities are calculated.

When performing a quick scan over the velocities calculated we encounter some strange behaviour. Occasionally the velocities for trips over 10 km are much greater than can be physically possible and are incoherent with other answers. For example, one interviewee noted she walked from one side of the country to the other within ten minutes. This might have been a trip smaller than 10 km, which we would normally exclude. Unrealistic trips like these we want to keep out of our dataset. Hence, we decided trips going over 145 km/h will be excluded.

The mobile phone data is also an important dataset that needs to be understood here. On this dataset the model has to be applied. Hence, the characteristics used to predict trip motive have to be present also in this dataset. In Appendix G table G1 a list of the attributes present in the mobile phone data is presented along with a short description.

Time of departure and time of arrival are present in both datasets. However, there is an important distinction between OViN and the mobile phone data. While OViN knows the exact moment a person leaves and arrives we only have an estimation in the mobile phone data. In the mobile phone data the departure time is the first event before leaving the origin area and the departure time the first event at the destination area. Based on the average number of events of the users we estimate the true arrival and departure times as has been elaborated upon in 5.5.2. This adjusted time will then be used such that both datasets provide an indication of the true departure and arrival times. Furthermore, there is an attribute Homebased in the mobile phone data that we think can provide valuable information about the trip motive. In OViN this attribute is not present per default and hence is constructed. How the homebased attribute, among others, is added will be discussed in the following section. i.e. data preparation.

8.3 Data preparation

In this section we will discuss how additional attributes are constructed, and how the dataset is filtered to contain only useful information. Data preparation and analysis has been performed in R (R Core Team, 2014). R is a language and environment that was developed to perform statistical computation (R Core Team, 2014). R is chosen because it is freely available, has many useful extensions, and provides greater flexibility than for example SPSS or Excel.

A description is provided stating exactly how all attributes are created. The order of presentation is also the order in which the attributes are created. Some attributes are created before the dataset is reduced because of quality concerns stated in Data Understanding or because it is not relevant to our research, e.g. information outside the scope of this research. The name or names of each attribute is shown in bold and when data reduction is performed the text is italic. A description of each attribute can be found in Appendix G table G2.

The first data cleansing step is duplicates in OViN. In OViN each trip can also have sub trips. For example, a trip include going to the train station by foot, taking the train, and going to work by foot. Then the motive is assigned to all three trips. Removing 'duplicates' reduces the number of trips by approximately 9%.

Trip motive

Business trips and home-work trips area already defined in OViN. All other categories are merged into the category other.

Homebased

In OViN there is an attribute stating the goal of a trip. If this goal is going home we know the postal code of the destination is the home postal code. For all people that once stated the goal of their trip is going home we can thus infer the home location. Based on this we can decide on a person level if someone is leaving home, going home, or is traveling between locations that do not include the home postal code.

Travel distance

Travel distance is calculated by linking the postal codes of the origin and destination location with their respective longitude and latitude. The link between postal codes, longitude, and latitude are provided by an external data source (Postcode Data, 2014). Originally the external data source includes postal codes with digits and two characters, e.g. 1234AA. Longitude and latitude are averaged per four digit postal code, e.g. 1234. The longitude and latitude are than linked to the origin postal code and destination postal code. Using the R package Geosphere and specifically with the function distm the travel distance is calculated from these longitude and latitude combinations (Hijmans, 2015).

All trips under 10 km are discarded as they fall outside of the scope of this research. As the majority of trips are below this threshold the dataset is reduced by another 76%.

Departure / arrival time

Values are calculated by multiplying the start / end hour of a trip by 60 minutes and adding the start / end minute of a trip. Start / end hour and start / end minute are readily available in OViN.

First / last trip start / end

The start / end of the first trip is calculated by aggregating over the entire OViN dataset per person id and taking the start / end time of the first trip. The first and last trip start / end is added to all trips taken for each person.

Velocity

The velocity of each trip is calculated by dividing the trip's travel distance (converted from meter to km) by the difference between trip's end time and trip's start time (converted from minutes to hours).

As stated in data understanding the trips over 145 km/h are unrealistic and either the origin or destination is incorrectly reported. These trips are, therefore, also left out. This results in a data reduction of 9%. Moreover, although this may have been done earlier, the few trips that are solely abroad are subtracted. Finally, trips with no value for the attribute trip motive that we aim to predict are removed (20%). In the end this leaves us with a total of 97.423 unique trips.

Weekday

Day of week is calculated from the date attribute. Day of week is a numeric value ranging from 0 on Monday to 6 on Sunday.

Holiday

Whether it is a holiday depends on the part of the country a person lives in. In the Netherlands there are three regions: North, Middle, and South. Each region has their own holiday periods. We thus had to map the postal code from the home location, already constructed to calculate homebased, to these three regions. This we did by first going from postal codes to municipalities and mapping those to the regions defined by the government (Rijksoverheid, n.d.). After adding the region information to OViN we linked this to holiday information to know whether a trip in OViN was during a holiday period or not (Landenportal, n.d.).

8.4 Modelling

The modelling phase has two distinct phases. First, an algorithm has to be chosen. Thereafter, we need to determine how to best implement the chosen algorithm to get the best model as a result. Section 8.4.1 will discuss what type of algorithm fits our task best. Section 8.4.2 will, thereafter, focus on how to apply the algorithm, i.e. what settings to evaluate and later use, to get the best possible model.

8.4.1 Model choice

For model choice it is important to take into account the structure of the available data and our goals, which are elaborated upon in the business understanding phase. What is apparent from the data is that we have three classes to predict. The model chosen will thus have to be a classification rather than regression type of model. Furthermore, there are approximately twelve input variables and all of these are either numeric or binary (0, 1). There is one exception, i.e. the homebased variable. The homebased variable can has the value -1 when the home location is the origin, 1 if the home location is the destination and 0 when the home location is not involved in the trip. However, if needed the information can be translated by creating two binary variables. To our knowledge all popular classification algorithms allow for numeric data and thus this does not provide any noteworthy restrictions on our model choice. On the business side it is important that the model can be implemented relatively easily into SQL and that the implementation is relatively fast, e.g. runs a day of data within one hour. These constraints, as discussed earlier, are crucial for the model to be useful in practice. Moreover, these constraints significantly narrow the candidate classification algorithms.

Neural networks and support vector machines have been shown to provide solid results (Pradhan, 2013; Baesens, Van Gestel, Viaene, Stepanova, Suykens, & Vanthienen, 2003). However, implementing these algorithms into SQL is not a straightforward task. Neural networks and support vector machines also have the disadvantages of being more difficult to explain and visualize. Decision trees and probability estimation trees (PETs) are much easier to explain and convert into SQL.

The trees are basically an ordered set of if else statements that will lead to a value. In case of decision trees the values are hard label, e.g. business trip, and for PETs the values contain probabilities, e.g. business (50% chance) home work (25% chance) and other (25% chance). For predictions on groups rather than individuals soft labels, i.e. probabilities, are preferred (Niculescu-Mizil & Caruana, 2005). "[I]n data mining applications the interest is often more in the class probabilities themselves, rather than in performing a class assignment." (Hastie, Tibshirani, Friedman, 2009, p. 348). In the above example, if a 100 people would have been labelled with a hard label we would get 100 business trips. With a soft label it would be 50 business trips, 25 home work trips, and 25 of the class other. The latter will typically be a better representation of the underlying population. Hence, we prefer a PET over a standard, i.e. hard label, decision tree.

Another well-known classification algorithm is decision forest, which is basically a large collection of short decision trees. Each tree in the forest provides a prediction and based on all the predictions one final verdict is given. This could also be in the shape of a hard and soft label, i.e. probabilities of belonging to a class. Because a decision forest is in essence a collection of decision trees it is also a collection of if else statements. However, a decision forest would take much longer to make predictions. For example, a large decision tree with a depth of 20 is already large, but would per instance at most take 20 if else statements. A decision forest works well when there are hundreds up to thousands of small trees. If each tree would only consist of one if else statement it might requires orders of magnitude more computation time. Implementing a decision forest is not feasible given the long execution time.

For this study we will implement a PET. The PET is preferred over the alternatives because it is (1) easy to translate into SQL, (2) probably provides a better estimation of the distribution than a decision tree, and (3) requires little computation time at implementation.

8.4.2 Training a PET

Numerous studies have been performed to determine how to best predict class probabilities using PETs (Niculescu-Mizil & Caruana, 2005; Zadrozny & Elkan, 2001). Because the goal of decision trees, i.e. predicting hard labels with maximum accuracy, differs from probability estimation trees, i.e. estimating class probabilities, the two models have to be trained and evaluated differently (Provost & Domingos, 2000). For decision trees the standard procedure is to train the tree and perform pruning afterwards (Esposito, Malerba, Semeraro & Kay, 1997). Pruning is performed to check if leaves are worth the added complexity they bring to the model and remove them if they are not (Provos & Domigos, 2000; Zadrozny & Elkan, 2001). The idea is that complex trees model are likely to over fit the training data. Overfitting implies the model incorporates too much noise and outliers and so reduces the predictive power of the model on unseen data. Predicting well on unseen data is obviously the goal and, therefore, pruning is crucial.

For probability estimation trees, however, the story is a little different (Provos & Domigos, 2000; Zadrozny & Elkan, 2001). For getting good estimates pruning can also be a culprit, hurting results (idem). While pruning removes outliers it also tends to remove leaves if they have little predictive power because the underlying probabilities are too similar. The latter would actually help the effectiveness of probability estimation trees, because these focus on distributions and changes herein rather than the best accuracy of predicting classes (Provos 7 Domigos, 2000). An example of this can be found in figure 8.1. The bottom left two leaves do not truly add to the predictive power. In both cases the tree would predict just as many good or wrong as when these leaves would be pruned. However, for a probability estimation tree the bottom two leaves do add value. They show the chance on the bottom right leave is 50/50 and that of the left is 25/75, i.e. a noteworthy difference.

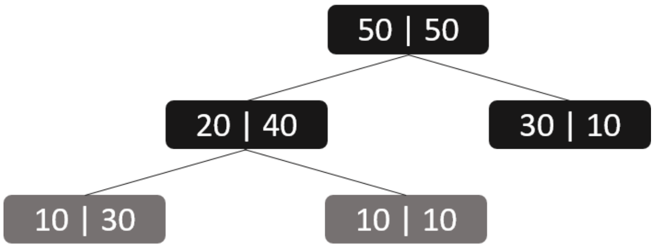


Figure 8.1, example of a probability estimation tree.

Provos and Domigos (2002) tested how well probability estimation trees would perform with a variety of settings. They did this on 25 publically available dataset commonly used for testing data mining algorithms, e.g. the Iris and Hepatitis datasets that are also built into R (idem). Pruning, for one, performed better than not pruning. They hypothesize that even though pruning removes useful information it also results in leaves that are produced from very few observations (idem). It may, for example, be the case a distribution at a leaf node is only built on five observations that all belong to one class. As a result the leaf will say all future observations belong with a 100% certainty to that class.

Laplace correction is one technique occasionally implemented to address the issue overly confident distributions at leaves with few observations. Laplace correction is applied to reduce the confidence of the leaf by adding one observation to each of the classes artificially. Consequently, instead of having five observations of one class you get six of that class and one of each other class. This will result in a more uniform distribution. When applying Laplace correction on the unpruned tree the results found were slightly better than with pruning, although not significantly so (idem).

Zadrozny and Elkan (2001), for one, point out that applying Laplace correction might not be ideal as it results in a more uniform distribution, which might not be the true underlying distribution. Rather than using Laplace correction they went for m-estimation in their study (idem). M-estimation is largely similar to the Laplace correction, but draws the distribution closer towards the a-priori distribution rather than a uniform distributions.

Equations 8.1 and 8.2 will show how class probability is calculated before and after m-estimation, respectively. In the equations p_i stands for the probability of class i, k_i stands for the number of observations of class i at the leave, and n stands for the total number of observations of the leave of interest. B_i is the apriori probability of encountering class i and m is a multiplier that determines how much smoothing is applied. The optimal value for m can be determined by using cross validation (Cussens, 1993).

$$\mathrm{p}_i = \frac{k_i}{n} \quad \text{(eq. 8.1)}$$

$$\mathrm{p}_i = \frac{k_i + m * b_i}{n+m} \quad \text{(eq. 8.2)}$$

An alternative to smoothing, e.g. by applying m-estimation, is to perform isotonic regression (Niculescu-Mizil & Caruana, 2005). The issue with isotonic regression is that it is hard to implement in our situation. Isotonic regression implies the class probability has to be continuously increasing or decreasing and that is not necessarily the case in our situation. Moreover, it is designed to be used for binary classification and extending the algorithm to a multi-class problem is non-trivial (Niculescu-Mizil & Caruana, 2005).

In addition to applying smoothing, Zadrozny and Elkan (2001) propose to perform curtailment. Curtailment is the act of ignoring leaves that have less than to be established number of observations (idem). This differs from traditional methods to reduce the complexity of a tree such as applying a threshold on the number of observations each leaf should have during creation of the tree. Curtailment occurs after a tree is created.

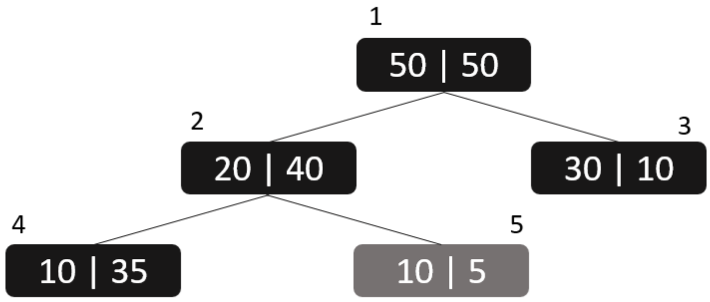


Figure 8.2, a decision tree before and after curtailment, i.e. without node 5, with at least 20 observations needed.

In figure 8.2 an example of a tree is provided where the grey leaf is a leaf that is being ignored because it does not meet the minimum number of observations criteria. When an observation would end up in a grey leaf the distribution of the parent node will be used, provided it has sufficient observations otherwise it will go to the grandparent. This process can continue all the way back to the root node. Even though Zadrozny and Elkan (2001) show promising results when applying curtailment we show here that curtailment can result in biases in prediction. Imagine each leaf should have at least 20 observations to provide sound estimations of the distribution. Curtailment will than result in the black tree as shown in figure 8.2 where leaf 5 is left out because the lack of observations. Whenever an observation ends up in that leaf during prediction the distribution of node 2, i.e. its parent node, will be assigned. Now imagine 60 new observations entering node 2 and we want to predict the distribution. If the tree is correct this would result in 20 yesses and 40 no's, i.e. using the left class as yes and the right class as no. In the tree after curtailment about forty-five observations go to node 4, which results in 10 yesses, and fifteen go to node 2 as node 5 does not meet the criteria, which results in a further 5 yesses. In total we get 15 rather than the expected and correct 20 yesses. This example shows curtailment can result in artificial biases and, therefore, we abstain from using it. The general idea of setting a minimum to the number of observations that should be in a leaf appears useful nonetheless. For this we propose to use minleaf, which is an old technique that only allows the tree to grow further leaves if at least a certain number of observations will end up in each leaf. Setting minleaf at 20 in our example would have resulted in a tree without nodes 4 and 5 which implies the 20/40 distribution at leaf 2 would be assigned to new observations, which is correct. Cross validation, as with curtailment, will be used to determine what the threshold on minleaf should be to deliver the best results.

8.5 Evaluation

Evaluation is about finding out how good our model performs under different circumstances when trained using a variety of settings. Finally, this will help us to answer questions such as “Does the model add to what we already know?”, “How accurate is the model in different circumstances?”, and “What settings will result in the best model?”.

This section is divided in four subsections. In 8.5.1 the method to correctly evaluate the model is described. In 8.5.2 the evaluation results are presented with analysis following in the subsequent section, i.e. section 8.5.3. A description of the best model including information about what attributes add much to the predictive power of the model can be found in section 8.5.4.

8.5.1 Evaluation method

When evaluating the created models it is key to have a good measure of how good a model is. Moreover, the measure will have to fit the task the model is designed for. A popular quote states: “Everybody is a genius. But if you judge a fish by its ability to climb a tree, it will live its whole life believing that it is stupid.” The same goes for model evaluation. The goal of our model is to provide accurate estimations of the trip motives of people driving past a certain road section. We do not necessarily care about each individual. Hence, the goodness measure should be about predicting distributions accurately and not individual observations.

As a goodness measure we will use the Chi-square test. The Chi-square test evaluates whether a distribution, in our case the estimated distribution, might represent the actual distribution, i.e. the distribution in the test set. When the distribution do not significantly differ, with a confidence level of .95 (α is .05), we assume the prediction is correct and otherwise that the prediction is false. The Chi-square test has two assumptions that have to be satisfied for the test to be meaningful. These are:

- Independence of observations, i.e. each observation needs to be unique and unrelated to any other observation (Field, Miles & Field, 2012).
- Each class should have at least five observations. When there are less than five observations it is generally assumed the test has too little statistical power (Field, Miles & Field, 2012). Hence, with less than five observations per class the test does not provide conclusive evidence.

A few attributes selected for our model go beyond trip level. For example, the first trip a person takes during a day. Hence, not every observation is completely independent. To ensure independence observations all people present in the test data will have their trips removed from the training data. By doing this independence is assured. The minimum of five observations rule depends greatly on how the test data is constructed from the total dataset, which occurs randomly. Hence, we will check each time whether there are enough observations per class with each Chi-square test that will be performed. By doing so both assumptions will be met and the results from the Chi-square tests will be meaningful.

In addition to the Chi-square test we will also compare estimates of the resulting average VoT for subsets of the population compared to the country average. In the end the VoT and VoR are the values that will convert loss in travel time and travel time reliability into monetary values. Being able to predict these well is one of the main reasons to do this trip motive prediction and it is thus important to also judge the models in this respect.

Once a goodness measure has been established the next step is to determine in which situation and with which settings the model performs well. For us it is important that the model is able to predict trip motive for people driving over a particular road. The model thus does not only have to perform well on a country level, but also when zoomed in to province and even municipality level. There are thus three unique situations in which the model has to perform well that differ along the axis of granularity. In terms of settings there are two axis. As established in 8.4.2 there are two settings, i.e. adjustable parameters, that will determine how well our final model will perform. These are the weights set for m-estimation, i.e. the number of dummy variables taken into account, and the value for minleaf, i.e. the minimum number of observations needed per leaf to continue growing the tree. Figure 8.3 provides an overview of all axes along which potential models will be evaluated.

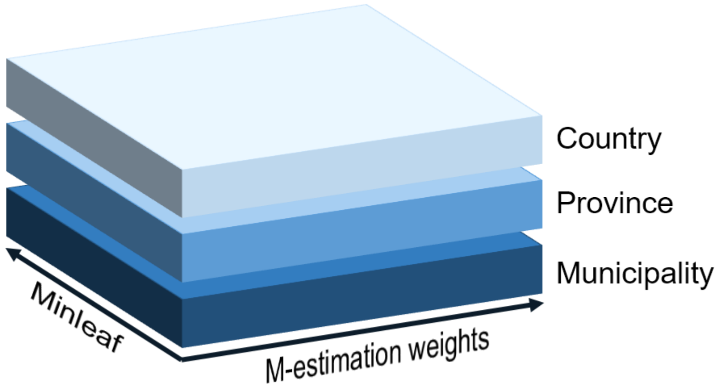


Figure 8.3, axes of evaluation. Potential models will be tested on three levels of granularity with a variety of settings for minleaf and m-estimation weights.

Evaluation will be performed by training the model on a subset of the data and testing the model on the remainder of the data. Note that the train data in this case has the trips for people removed that occur also in the test data to ensure independence of observations. Chi-square tests will be performed to see how well the model can predict the distribution of trip motives in the test data. To test whether the models perform significantly better than the baseline, i.e. an estimation using the a-priori distribution, the McNemar-test will be performed. The McNemar-test tests if there are differences between two groups based on one dichotomous, i.e. yes or no, dependent variable. Where the dependent paired t-test evaluates continuous variables, the McNemar-test evaluates dependent dichotomous variables. In our case the variables are dependent, for example, how well the model predicts trip motives for people going to Amsterdam versus the baseline model. The outcomes are dichotomous as the distribution is either equal or it is not. McNemar-test thus appears to be a good fit for what we aim to do. Dietterich (1996) also indicated the McNemar-test is one of the most promising statistical tests to test whether one classification algorithm outperforms another. All assumptions for the McNemar-test are also satisfied given our variables are truly dichotomous, i.e. there is no overlap between classes. The McNemar-test will thus provide good insight into how well our model performs and what the added benefit is of employing our model. An alpha of .05 will be used to test for significance.

8.5.2 Evaluation results

In total we evaluated each level, e.g. country, with fourteen unique values for minleaf and m-estimation resulting in a total of 84 combinations.

On country level 20 folds are made and evaluated. The results are shown in table 8.1. The a-priori distribution was indistinguishable from the true distribution, i.e. with an alpha over .05. Hence, the model has no chance of performing better, only worse.

Table 8.1, results on country level. Showing how often the model's predictions are indistinguishable from the true distribution of trip motives. Post minleaf 300 no changes are found and hence the results are omitted.

	M-ESTIMATION					
MINLEAF	**0**	**10**	**20**	**30**	**40**	**50**
0	80%	0%	0%	0%	0%	0%
50	100%	90%	85%	75%	55%	35%
100	100%	100%	95%	90%	85%	80%
150	100%	100%	100%	95%	95%	90%
200	100%	100%	100%	100%	100%	95%
300 +	100%	100%	100%	100%	100%	100%

In the Netherlands there are a total of 12 provinces. Per combination of minleaf and m-estimation settings we created a model on 11 of the 12 provinces to test how good predictions are on trips going to the other province. The results are shown in table 8.2. For reference, the a-priori distribution, which was 100% accurate on country level, is only 50% accurate on province level. On country level the best models predict 8 out of 12 correct versus 6 out of 12 for the baseline, i.e. a-priori, model. Although we can already see improvements by applying the model it is not significant. The p-value produced by the McNemar-test is .47.

Table 8.2, results on province level. Showing how often the model's predictions are indistinguishable from the true distribution of trip motives.

	M-ESTIMATION					
MINLEAF	**0**	**10**	**20**	**30**	**40**	**50**
0	50%	25%	33%	42%	50%	58%
50	67%	58%	67%	67%	67%	67%
100	67%	67%	67%	67%	58%	67%
150	67%	67%	67%	67%	67%	67%
200	58%	58%	67%	67%	67%	67%
300	67%	67%	67%	67%	67%	67%
400	67%	67%	67%	67%	67%	67%
500	58%	67%	67%	67%	67%	67%
750	50%	58%	58%	58%	67%	67%
1000	58%	58%	58%	58%	58%	67%
1250	58%	58%	58%	67%	67%	67%
1500	50%	50%	50%	58%	58%	58%
1750	50%	50%	50%	50%	58%	58%
2000	42%	42%	50%	50%	50%	58%

Although there are many municipalities in the Netherlands, we are unable to test how well the model performs on each one. The assumptions of the Chi-square distribution are the culprit here. When there are less than 5 observations the test is unreliable. We, therefore, chose to only evaluate the 30 most frequent visited municipalities in our dataset. The a-priori distribution was correct in 16 of the 30 cases (53% accurate) while the best model(s) predicted the distribution well in 22 of the 30 cases (73%) as is shown in table 8.3. The McNemar-test produces a p-value of .04, which is below the .05 criterion we use for significance. On this lowest level our model thus predict the underlying trip motives to the top 30 visited municipalities significantly better than the a-priori distribution.

Table 8.3, results on municipality level. Showing how often the model's predictions are indistinguishable from the true distribution of trip motives.

	M-ESTIMATION					
MINLEAF	**0**	**10**	**20**	**30**	**40**	**50**
0	60%	60%	57%	60%	63%	63%
50	63%	63%	60%	63%	63%	60%
100	70%	67%	67%	67%	67%	70%
150	73%	67%	67%	67%	67%	67%
200	67%	67%	67%	67%	67%	67%
300	67%	67%	67%	67%	67%	67%
400	67%	67%	67%	67%	67%	67%
500	67%	67%	67%	67%	67%	67%
750	67%	67%	67%	67%	67%	67%
1000	70%	67%	67%	67%	67%	67%
1250	67%	67%	67%	67%	67%	67%
1500	67%	67%	67%	67%	67%	67%
1750	67%	67%	67%	67%	67%	67%
2000	60%	60%	57%	60%	63%	63%

8.5.3 Evaluation analysis

The key finding here is that the model helps to stay accurate when we zoom in to specific areas. On country and province level we did not find any significant differences in performance though the model is never worse than the a-priori. On municipality level we encounter the first significant differences with our model providing the better distributions.

In terms of settings we find that m-estimation does not provide any benefit to training our PET. More specifically, the best models are often created where m-estimation is not performed. Changes in the minleaf constraint does make a large difference in the accuracy of the created models. We observe that a too low value for minleaf leads to bad results, which may relate to the model overfitting the data. On the other end of the spectrum we see that a large minleaf also leads to bad results when drilling down from country to smaller areas. This might be because a large minleaf results in a very general model that is unable to capture necessary relations.

In terms of finding the best model, in our opinion, there is a clear winner. On each level the model with a minleaf of 150 and no m-estimation produces the best results. The model that will be used for implementation is will be the one trained on the entire dataset using these settings.

8.5.4 Model description

The model used for implementation is a PET trained on the entire dataset with minleaf set at 150. The final model should perform at least as good and probably better as more data is used for training. In addition to more data, there are also no biases in the data, as is the case when leaving out specific provinces and municipalities.

Note that the dataset contains only trips over 10 km in distance, which is the distance as the crow flies. The model is, therefore, only applicable for these trips. It might be used to classify shorter trips, but for those trips the accuracy of the model has not been determined.

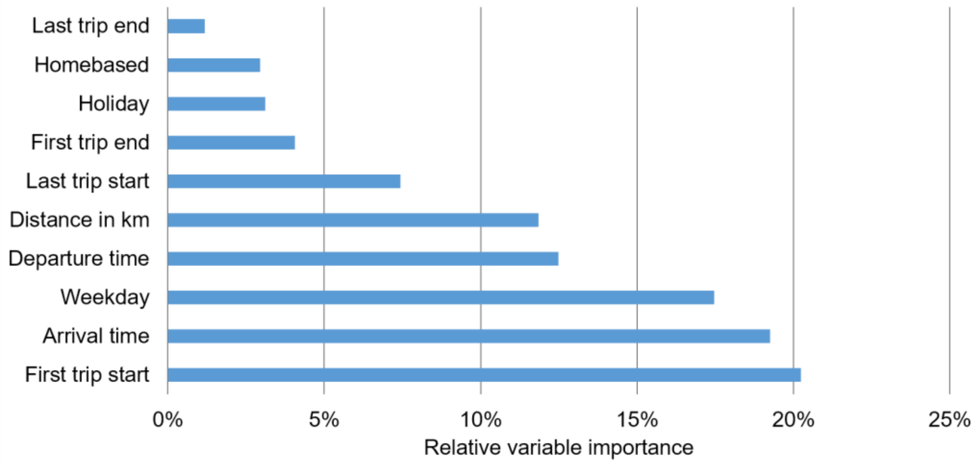


Figure 8.5, variable importance scaled to 100% for the final tree created.

In terms of variable importance we find the first trip start, arrival time and weekday (see figure 8.5). Runner ups are departure time and travel distance. Together these five attributes make up 81% of the 'variable importance'. Variable importance is a rather vague measure, nonetheless. Variable importance is measured by observing how many times a variable is used in to make a split in the tree and how often it was a surrogate split, i.e. second or third choice. To calculate the variable importance we used the function varImp from the caret package (Kuhn, 2016). In figure 8.4 the variable importance is scaled to 100%.

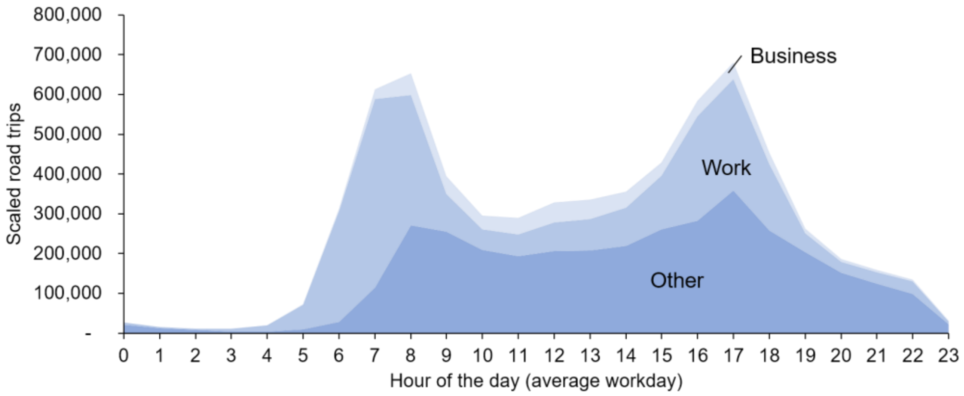
8.6 Implementation

Now a good model has been created we will implement the model on the mobile phone data. Implementing the model is a non-trivial task. We need to make sure the attributes the model is trained on match the ones in the mobile phone data. Furthermore, we need to translate the model from R into SQL. The clearest way to show the implementation is successful is comparing the acquired distributions to the OViN. This we do in table 8.4. From table 8.4 we can see the distributions are very similar for trips over the road. Nevertheless, there are some deviations. More specifically, we have slightly more business trips and trips in the category other. Overall, we do not have any reason to doubt the implementation. The comparison in table 8.4 is also comparing five years of OViN with September and October of 2015, for which we do not have the OViN. Changes over time, e.g. the economic crisis, might also explain why there is a slight decrease in the number of work trips.

Table 8.4, distributions of trip motives over all trips for the OViN and mobile phone data.

MOTIVE	OVIN	MOBILE PHONE DATA
Other	60.0%	62.5%
Work	35.6%	31.6%
Business	4.4%	6.0%

For a typical workday the distribution is shown in figure 8.6. We see the distribution appears logical with large peaks of work during the rush hours and more recreational traffic towards the end of the day. Furthermore, business trips occur mostly during the middle of the day as is also the case in figure 8.6. Overall, we thus find these results encouraging and are confident the model adds value.



Figuur 8.6, trip distributions for people traveling by car for an average workday in September and October 2015.

8.7 Conclusion

There are a number of key findings in this chapter about trip motive prediction. In terms of business understanding we made clear the trip distribution rather than individual trips are more important for this research. Furthermore, we find that the OViN provides enough resources to build and evaluate a model. Survey data provided a clear added value to the mobile phone data by matching characteristics found in the survey to similar characteristics found in the mobile phone data.

We provided a comprehensive analysis of how to correctly train a PET. PETs are the models of choice for this research because the tree structure matches closely to the case when statements present in the SQL language. This allowed us to implement the model created in R into a SQL environment.

A number of models were created and tested against the a-priori distribution using the Chi-square test for distributions. These models were built on three levels of granularity: country level, province level, and municipality level. We show that the model gains more value once we start to zoom in to smaller areas. Here the a-priori distribution increasingly starts to misrepresent the actual trip distributions to these areas. When we get to a municipality level the model starts to become significantly more accurate in predicting the trip distributions. Furthermore, we looked at a variety of settings for minleaf and m-estimation. The latter turned out to be of little added value. Tweaking the former, however, had a large impact on the model accuracies. When minleaf was too small the model over-fitted, and when it was too large it was unable to incorporate fundamental relations. With a minleaf of 150 the best resulting model was created.

In terms of variables that help determine trip motives we find the most important once are arrival time of the trip, departure time of the first trip during a day and the day of week. When performing similar studies we advise to take at least these into account.

We also provided indications that the implementation was successful. This we did by comparing the outcomes with (1) the OViN on which the model was trained and (2) our common sense. In both cases the resulting values from the models did not deviate much

9 Compare

In this chapter we will focus on putting the measured values at roadworks into context. These values consist of travel time, standard deviation of travel time, and people affected. At roadworks we expect these values to be affected, e.g. we expect an increase in travel time due to larger traffic hindrance on the road. To quantify how much these values are affected we need to (1) create a dataset that measures good travel times as well as accurate counts of the affected people and (2) create a proper baseline.

In section 9.1 we discuss how to construct a dataset from the mobile phone data that measures good travel times and accurate counts of the people affected. In section 9.2 we go into detail on how a baseline is constructed to compare the roadwork situation with the non-roadworks. The findings will be summarized in section 9.3.

9.1 Dataset construction

9.1.1 Mobile phone data

Creating a dataset that works well on all fronts is a non-trivial task that requires deep understanding of the mobile phone data. We know the counts of people on the roads to be accurate in both absolute counts as shown in section 6.3 and over different hours of the day as shown in section 6.3. In this section we will thus mainly focus on getting accurate travel times from mobile phone data.

There are two prime reasons why travel times would be unreliable in the mobile phone data. These are:

- Being in a blind spot, i.e. under a cell tower near the start or end of a trip.
- Having few events, i.e. a low density of data points from which travel times are calculated.

As discussed in the data quality chapter (chapter 5) the travel times at the start and end of a trip are not measured, but estimated based on the number of events a person has. We can thus not measure true differences in travel time at the start and end of a trip as a result of, for instance, roadworks. For roadwork analysis the trips ending and starting near the roadworks have to be discarded as the delay would not be measureable. We decided to exclude all trips with an origin or destination within 12.5 km of the roadwork's location to get more accurate travel times. Furthermore, the more events a person has the more accurate travel times are measured. On average a person has about 140 events a day nowadays (see figure 5.6). To prevent losing too much of our sample we stated people needed to have on average at least 100 events per day, i.e. once every 14.4 minutes. This reduces our margin of error while maintaining enough trips to stay above the minimum of 15 rule. With these two measures in place we can create a dataset from the mobile phone data with accurate travel times.

Now we have two datasets each with their own strengths. One dataset for accurate travel times that removes origins and destinations at blind spots and uses only people with relatively frequent events. And one dataset that contains everyone and is shown in chapter 6 and 7 to provide accurate counts of people on the road. Note that both these datasets are again only containing trips over 10 km long, which is similar to what we did in the majority of this research. These two distinct datasets will be merged into one such to create a 'best of both worlds' dataset.

In table 9.1 all relevant attributes from the mobile phone data are presented along with the original data source. The dataset with accurate travel times is dubbed the Travel time data and the dataset with accurate count data is named the Count data. They are merged on the first three attributes, i.e. the date hour and road id.

Table 9.1, origins of the attributes to create a 'best of both worlds' dataset from the mobile phone data.

ATTRIBUTE	DATA SOURCE
Date	Both
Hour	Both
Road id	Both
Users on the road	Travel time data
Users on the road scaled	Count data
Users on the train scaled	Count data
Urbanity (average)	Travel time data
Travel time for road users (average)	Travel time data
Distance travelled for road users (average)	Travel time data
Travel time for road users (std)	Travel time data
Users on the road scaled with motive Other	Count data
Users on the road scaled with motive Work	Count data
Users on the road scaled with motive Business	Count data

9.1.2 External sources

The attribute road id in table 9.1 is not naturally in the mobile phone data. This attribute originates from the shortest path algorithm. Trips are selected based on whether they cross the road based on the shortest path algorithm. In general one would know the road id where the roadworks occur. In our case, however, we only have descriptive information about the location of the roadworks. We have a road name, e.g. A2, a direction, e.g. Right, and a number corresponding to a sign by the road, e.g. 110.2. Because we have over a thousand unique roadworks in our sample the process of linking a roadwork to a road id had to be automated. This we did by first extracting geolocations from shapefiles about road signs and linking these to a road directory to extract information about corresponding road names and driving directions (Nationaal register, 2016). Thereafter, we used the information to get the road id from the Open Street Map database, which is linked to our mobile phone data, by picking the roads closest to the longitude and latitude from the road sign where the road names match. Only roads with 15 meters from the road sign were taken into considerations. Finally, we linked the descriptive information from our roadworks data to the road sign data. A roadworks starting on the A2 in direction R at 110.2 km will then, for example, be mapped to the road sign on the A2 in direction R with number or close to the number 110.2.

The information about roadworks originate from Systeem Planningen en Informatie Nederland (SPIN) and cover September and October 2015. SPIN is the system that is used by contractors that add data and Rijkswaterstaat who communicates the information to the road users (Rijkswaterstaat, n.d). Note that when we say unique roadworks we mean uninterrupted work at a specific road under fixed circumstances. When people are working on a road and they close one lane the first two hours and close a second lane the two hours after that we count two roadworks happening. One with the setting on one lane closure and one with two lanes closed.

As we found in chapter 7, we are unable to correctly measure people traveling over a road when the road is located near the Dutch border or shores. Hence, we will have to remove the roadworks happening in those locations as we know we cannot accurately measure their impact. In figure 9.1 we provide an overview of all the roadworks present in our database. In figure 9.1 we, furthermore, make a distinction between included roadworks and excluded roadworks, which are in an area near the shore.

In addition to information about roadworks, we also want information about weather at the roadworks. Weather can affect road capacity and influence travel behaviour in general (Cools, Moons & Wets, 2010). Hence, knowing what weather it is at the roadworks can provide valuable information. We also want to correct for the effects of weather variations to establish a purer baseline, which we will do in section 9.2. The information is thus crucial. Weather information is acquired from the Koninklijk Nederlands Meteorologisch Instituut (KNMI) and consists of hourly weather information from a number of weather stations (KNMI, 2015). We have selected eight weather stations across the Netherlands to extrapolate local weather information near the roadworks. The locations of these weather stations are also depicted in figure 9.1. Local weather information near the roadworks is calculated using a distance metric. Based on the premise that the closer the roadworks are to a weather station the more probable it is the weather is like that at the weather station. We calculated the inverse of all weather stations to any roadwork and used this as weights to estimate local weather information. For example, if a roadwork is 10 km to a weather station where it rained and 20 km to a weather station without rain, we state it rained for two thirds of the hour at the roadworks.

Figure 9.1, locations of the weather stations and roadworks that are integrated with the mobile phone data for further analysis.

9.2 Creating the baseline

9.2.1 Defining the baseline

The baseline for a situation where a roadwork is present is one at the same location at similar moments in time where no roadworks are present. When the roadworks occur on road A on a Monday between 6 AM and 7 AM then the baseline will be the data from people driving over road A on all other Mondays between 6 AM and 7 AM. That is the case provided there are no other roadworks on the same road on Mondays between 6 AM and 7 AM.

Our data spans September and October 2015, but unfortunately there are some special cases in this time period. In the Netherlands as it happens to be there are holidays in different parts of the country during the last two weeks of October. There is reason to belief that travel behaviour could be influenced by the holidays we cannot use data from that time period. The roadworks and their respective baselines can thus only span from the first of September to the 17th of October.

As can be seen from figure 9.1 there are many roadworks occurring throughout the country, sometimes also on the same road. When creating a baseline for roadwork X we do not want delays of roadwork Y to end up in our baseline. When roadwork X and roadwork Y are located on the same road in the same direction within 15 km of each other we will not include the data when roadworks Y are present in the baseline for roadworks X. For example, if roadwork X occurs in week 1 and roadwork Y in week 2, then we will only consider data from week 3 up to week 7 for our baseline. The threshold of 15 km rather than the entire road is specified because some roads are very long. Being on the same road thus does not imply they influence each other, e.g. there will be no effect from roadworks on the A2 near Maastricht on roadworks on the A2 near Amsterdam (217 km apart). By using a threshold of 15 km we inherently assume people on the same road will not be affected by roadworks further away, but might be by roadworks within this range. For now we assume the 15 km threshold is correct and leave finding a better threshold for future research.

Furthermore, we want to have at least four data points in our sample. In total we can have up to six data points, i.e. one for every seven weeks in our sample minus one for when roadworks are present. If we would compare a roadwork against one measurement in the baseline this would give an unreliable image. A small baseline would allow for large effects of noise, e.g. delays caused by traffic accidents. The larger the baseline the better it will become. In table 9.2 we show the average number of baseline points per hour for all roadworks in our dataset. With our threshold of four points in the baseline we discard 222 roadworks and keep 638. Note when we started we had just over a thousand roadworks and thus already lost quite a few. This is because those roadworks had no baseline data all together and were discarded altogether.

Table 9.2, average number of data points in our baseline rounded to integers and the corresponding number of roadworks for which we have that many baseline points.

# BASELINE POINTS	1	2	3	4	5	6
# ROADWORKS	61	50	111	215	292	131

9.2.2 Correcting travel time

The travel time we measure is based on the time it takes a person to go from A to B. While trips always have an origin and destination these may change over time resulting in changes also in travel time. To get a fair comparison of travel times we want to correct for these factors that do affect travel time, but are due to external factors rather than variations due to the roadworks. For the corrections we use only data with 500 to 1500 vehicles per lane. The bottom limit is to ensure we have a proper sample size and the upper limit is to ensure we do not start to model the effects of road being congested quicker under certain circumstances. Rain, for example, can result in a road capacity reduction of 4 to 30% (Stern, Shah, Goodwin, & Pisano, 2003; Unrau & Andrey, 2006). For this we do not want to correct because it directly relates to the delay experienced at roadworks, which also reduce road capacity leading to more congestion. Moreover, we discard all information at roads when roadworks are present when building a model to correct for external factors influencing the travel times.

In total we identify four crucial external factors that may influence travel time. These are:

- Average travel distance, i.e. the distance travelled in km over the road
  (ranging from 21 to 215 km)
- Rain duration, i.e. ratio of the rain duration to the entire hour
  (ranging from 0 to 1)
- Wind max, i.e. the maximum wind speed in km/h
  (ranging from 2 km/h to 184 km/h)
- Night, i.e. whether it is between 7 am and 20 pm
  (0 for day and 1 for night)

Travel distance is obvious, the further one travels the longer the trip will take. The other three are have a less direct impact on travel time. When it rains and when there are strong winds people may be more cautious and drive slower. During the night time people might be in a hurry and roads are less congested which may result in higher velocities. In figure 9.2 we show how these attributes are related. In figure 9.2 the baseline is the travel time where there is no rain and wind max is below its 33% quantile. The baseline, but only where it rains, is the blue rain line depicted in figure 9.2. Together with the wind line where wind max is at or above its 66% quantile the velocities are consistently lower than the baseline. All values are obtained by taking the average per 5 km. For example, the average travel time at the baseline between 50 km and 55 km is shown at 50 km (x-axis) and 55.8 minutes (y-axis).

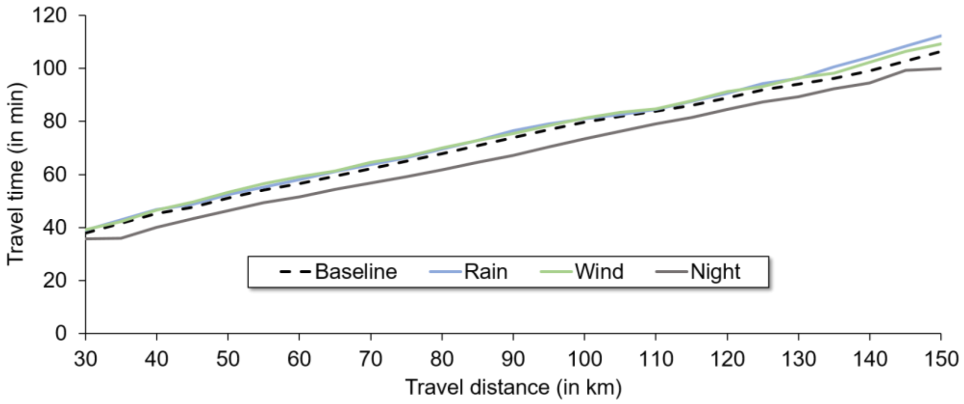


Figure 9.2, average travel time during the baseline, rain, wind and night scenario.

Wind and rain have very similar effects on travel time. Both are consistently just above the baseline with some larger effects near the end. During the night time scenario, travel times appear to be a few minutes lower. There does not appear to be any change in effect with respect to distance travelled.

Provided the clear linear relations shown in figure 9.2 we opt for a Linear Model (LM) to predict travel times. We want to predict travel time based on the four factors identified. The precise formulation of the model we will create is shown in equation 9.1. In equation 9.1 t stands for travel time and d for travel distance. W stands for a weight that will result from the LM. These weights are why we create a linear model as they describe the relation between the attribute, e.g. distance, and travel time. The attributes rain wind and night correspond to the other three factors above. Note wind speed is measured in km/h, rain duration is always between 0 and 1 and night is a Boolean with 1 standing for night. Finally, we see an error term in equation 9.1. This error term is introduced to allow for errors. Note the error term in a linear model assumes the error is normally distributed, which we will have to test together with other assumptions to see if the resulting LM produces meaningful results. The beta coefficient is there to allow for an intercept. This we do because people might be spending some time in the car, e.g. in front of traffic lights, independent of rain or distance covered. The beta term, i.e. intercept, will incorporate this time spent. We also added an additional term for night as from figure 9.2 we can see from extrapolating the line at night that the intercept is much lower than during the baseline situation.

$$t = d * (\omega_1 + \omega_2 * rain + \ \omega_3 * wind) + night + \beta + \varepsilon \quad \text{(Eq. 9.1)}$$

In equation 9.1 we multiplied distance with rain duration and wind. This we do because the relation between any of these attributes and the difference in travel time from the baseline we belief is dependent on the travel distance. Moreover, when we look closer at the above equation we see that the relations seem logical once we interpret the meaning of the weights. The first weight, i.e. ω_1, for example is equivalent to one over the baseline travel velocity ($V_{baseline}$) (eq. 9.2). The right side of this equation originates directly from the definition of velocity from classic physics. The inverse of each weight that our LM will produce thus says something about velocity. The average velocity, for the first weight, and the change in velocity in given circumstances for the other three, e.g. rain.

$$t_{baseline} = w_1 * d = d/V_{baseline} \quad \text{(Eq. 9.2)}$$

For a LM to be meaningful we need to check whether the assumptions on which the model is built are met (Field et al., 2009). These are the following assumptions:

- Normality of the residuals, i.e. the errors around the fitted line have to follow a normal distribution (Field et al., 2012).
- Homoscedasticity of the error variance, i.e. whether the residuals are distributed similarly across the spectrum of predicted values (Field et al., 2012).
- Independence of the residuals, i.e. there should be no relation between one measurement and another (Field et al., 2012).
- No multicollinearity, i.e. the attributes should have no correlation with others in the model (Field et al., 2012).

We will test whether these assumptions are valid and apply corrections when necessary. Because we have great amounts of data most statistical tests will either fail or be significant even when only miniscule patterns in the data are observed. Hence, we will test these assumptions not with statistical tests, but by visualizing and interpreting the underlying data. For the normality of residuals we will provide a QQ-plot. To test for homoscedasticity we will plot the residuals over the predicted values. Finally, independence of residuals is evaluated by a lag plot. A lag plot shows what the next residual will be given the previous one. If there is no relation the plot would result in an unstructured cloud of residuals that is normally distributed along both axes. The only test that we can easily perform is the VIF test for multicollinearity. The VIF tests evaluates multicollinearity. Ideally the values are 1, but VIF scores below 10 are generally sufficient to say the no multicollinearity assumption is met (Field et al., 2012).

During the assumption tests we did not find any worrisome outcomes. The created linear model met all assumptions implying the results are fully valid and interpretable (see Appendix H figure H1, figure H2, and figure H3). We did find a slight positive skew in the residuals, but nothing that would have a large influence on the created model. The created model is presented in table 9.3.

Table 9.3 our LM estimation for predicting travel time based on non-roadwork related factors. The model has an R^2 of 0.85.

	WEIGHT	STD ERROR	T VALUE	PR(>|T|)
Intercept	24.3200	0.0394	617.14	<2e-16
Night intercept	-7.4950	0.0419	-179.06	<2e-16
Distance	0.5386	0.0006	965.06	<2e-16
Distance * rain	0.0608	0.0007	87.57	<2e-16
Distance * wind	0.0001	0.0000	19.81	<2e-16

Note the travel time is denoted in minutes. The intercept of 24 thus denotes people spend on average 24 minutes on the road in addition to the time spend traveling at a certain velocity. This number drops by 7.4 minutes during the night time. Discarding the intercept and night intercept we see that every km travelled relates to an increase in travel time of 0.539 minutes. When we convert this into velocity we find every km travelled is travelled at about 111 km/h. When it's raining every km travelled is done so in 0.539 plus 0. 061 minutes. The velocity during an hour of rain will thus be on average 100 km/h, lower than during the baseline situation. During the night the intercept is lower the total average travel time could still be lower, even though the velocity while covering distance is lower than the baseline. Overall, we are not put off by these results. The average velocity found for traveling of 111 km/h is very reasonable as maximum speeds on highways are limited at 100, 120 or 130 km/h in the Netherlands. Furthermore, we see decreases in average velocities during rain and strong winds, which confirm what we hypothesized. Although the effect of wind appears minimal. All relations as shown in table 9.3 are significant. P values are found far below the standard .05 used to test for significance indicating all relations found are very likely to persist when we would measure the entire population rather than just our sample.

Finally, to correct for travel time we will subtract the predicted travel time, i.e. prediction on our complete dataset using the model we just created, from the measured travel time. By doing so we remove the differences in measured travel time that can be explained by our model and thus the external factors mentioned above. When evaluated over the entire dataset we find an $R^2$of 0.85 between the predicted travel time and the true travel time. The remaining 0.15 or 15% of the variance in travel times cannot be explained by the model. This we assume might be explained by, for example, congested roads and people having to take detours when roads are closed. The corrected travel times are those that will be used in the remainder of this research. This includes measuring as well as predicting the impact of roadworks.

9.2.3 Correcting travel time reliability

In addition to correcting for travel time in general we also want to correct the travel time reliability. Travel time reliability is denoted in the standard deviation of the measured travel times. Standard deviation of travel time in our dataset is literally the standard deviation of the measured travel times during an hourly period on a specific road. When we have a large average travel distance there may also be large variations in trip length resulting in greater differences in travel time. To investigate this hypothesis we make a similar plot to the one found in the previous section. Here again we use the same attributes and create the baseline similarly as for figure 9.2. The relations between standard deviation of travel time, distance travelled, rain duration, wind, and night are depicted in figure 9.3.

From figure 9.3 we observe there is indeed an increase in the standard deviation of travel times when distances become larger. Furthermore, travel times appear to be much more consistent, i.e. have a lower standard deviation of travel time, during the night. Rain and wind have only a minor impact here, but appear to introduce more unreliability when the average distance grows. Night time does the opposite, here the travel times become more reliable at larger distances in comparison with the baseline.

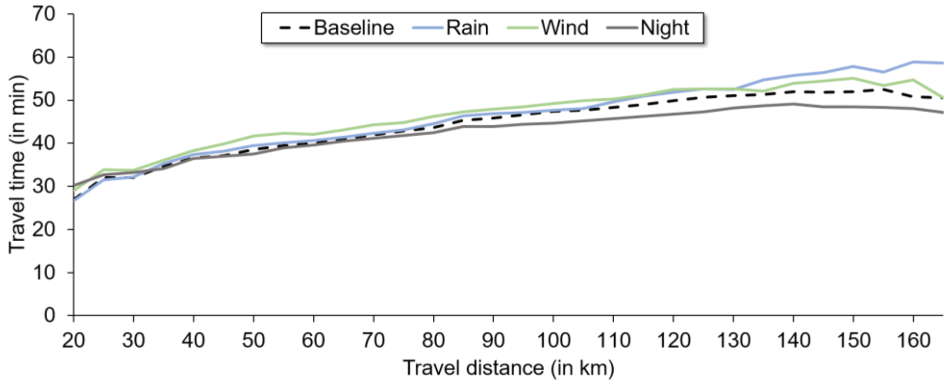


Figure 9.3, average travel time standard deviation during the baseline, rain, wind and night scenario.

Most importantly, however, we observe the standard deviations start to decrease, at least for the baseline, for average distances greater than 150 km. As we want to use a LM to correct for travel time reliability the lines should, in reality, be straight. If we now try to draw a straight line describing the baseline based on distance, we might overestimate the standard deviation distances over 150 km, and most likely underestimate those at shorter distances. We thus require some further investigation to see whether the observed behaviour is structural or accidental before we continue to fit a LM.

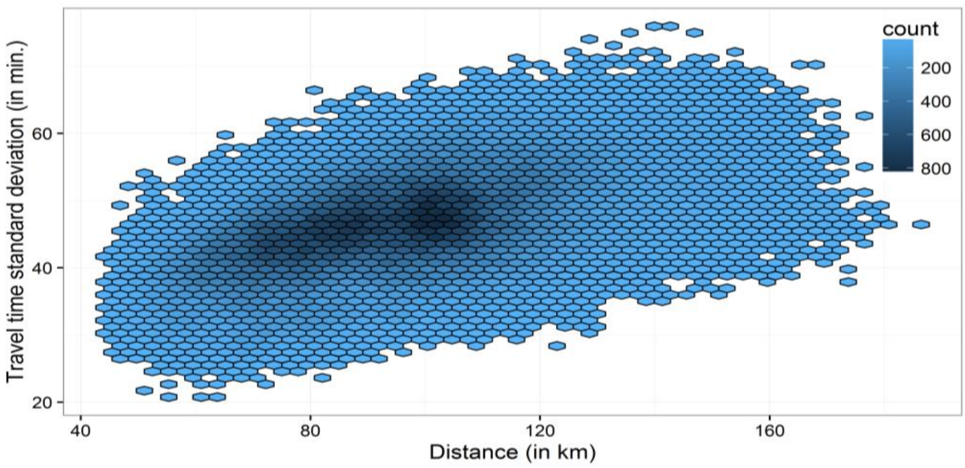


Figure 9.4, travel time reliability (standard deviation in minutes) for all observations plotted against distance travelled using the hexbin and ggplot2 packages (Car et al., 2015; Wickham, 2009).

Further analysis of the data behind the graph in figure 9.3 shows us the dip observed post 150 km might be due to outliers (see figure 9.4). In figure 9.4 the darker the colour the more observations there are behind each point. We can clearly see the majority of our observations, the darker areas, are between 60 and 120 km.

At 150 km we have less than 2% of our data left. The few observations all the way at the bottom right of the graph are most likely outliers. These are the ones that pull down the aggregated mean in figure 9.3. Looking at figure 9.4 we are thus not convinced the drop in travel time reliability is anything other than randomness. Over the remainder of the spectrum, i.e. below 150 km, we do find the relation is linear and we would not be surprised if the true relation is actually linear. Hence, we will continue to fit a LM to our data. Furthermore, we will leave out data with distances greater than 160 km. The distance between these points and the fitted line would be relatively large as we expect a large standard deviation of travel times while the observations at the end of the distance spectrum show low values. Hence, these few outliers may have a disproportionate effect on the coefficients estimated by the linear model that can deteriorate the accuracy of the model on the majority of the data.

From figure 9.3 we infer the following structure (equation 9.3) underlying the relation between travel time reliability (r), travel distance (d), night, wind and rain. There is a clear relation between distance and reliability. We further see a small increase in (un)reliability when it rains, which grows with when the distance increases. Although we see a changing impact of night depending on distance, we find high VIF scores when this relation is included. We hence, only included night as an intercept like variable. The beta coefficient again stands for the intercept. Even when people do not travel, or travel below 10 km, we could still imagine some differences in travel times. The error term at the end is there to allow for the normally distributed error that is assumed in a LM.

$r = d * (\omega_1 + \omega_2 * rain + \omega_3 * wind) + night + \beta + \varepsilon$ (Eq. 9.3)

When we train our model we again find there is autocorrelation between residuals, i.e. the previous residual can explain part of the variance in the following residual. This finding is confirmed by a Pearson correlation coefficient of 0.41 between consecutive residuals. Because the created LM might be biased as a result of the correlation between consecutive residuals, we resort to the Cochrane-Orcutt estimation (Cochrane & Orcutt, 1949). Not satisfying the autocorrelation assumption implies the error at the previous moment in time relates to the error observed next. This might be the case when, for example, people on one road are consistently faster than average. When this is the case the errors measured at t-1 will be related to that at t-0 for that measurement site, and hence the errors are related. Cochrane & Orcutt (1949) developed a procedure that estimates the error at t-1 and uses it as an added dimension to the standard LM. Now the weights in the LM can be set irrespective of the error at t-1 as a value, or an estimate to be precise, of the error at t-1 is included. Because the next prediction already incorporates the information of the residual at t-1 the effect of autocorrelation decreases or disappears completely. When the information at t-1 is already included it cannot explain why the estimate at t0 would be wrong. Overall this would lead to less biased and thus trustworthy outcomes. We do lose one observation, however, as there is no observation at t-1 for the first record in our dataset, which is trivial provided our 200.009 observations. In theory one can continue to apply the Cochrane-Orcutt procedure until independence of residuals is achieved (Cochrane & Orcutt, 1949).

The resulting model resulting from Cochrane-Orcutt estimation does not necessarily find a global optimum, but rather is guaranteed a local optimum (Dufour, Gaudry & Liem, 1980). This results from having to model the lagged residuals in addition to the normal residuals which adds another dimension to the problem. Dufour et al. (1980) do mention, however, that with typical data and a unimodal distribution of the residuals, which we have, it is very rare not to find the global optimum. Irrespective of finding the best weights, i.e. at a global optimum, we are still able to evaluate the model and, for example, calculate how well the model can explain variations found standard deviations in travel time.

After one iteration, Cochrane-Orcutt estimation helped us to find a linear model that meets all required assumptions (see Appendix H, figure H4, figure H5, figure H6 and table H2). The weight coefficients and tests showing whether the relations, either positive or negative, are significant. We use the common .05 criteria for alpha to test for significance here. We find that all proposed relations, with p-values below $2 * 10^{-16}$, are significant within our model (table 9.4). The fact that adverse weather negatively affects travel times confirms what is found in literature (Tu, 2008). The effect of night and distance is to our best knowledge new information. Given the extremely low p values we are confident the found relations are representative of the effects on travel time reliability in general.

Table 9.4, our LM post Cochrane-Orcutt estimation for predicting the standard deviation (in minutes) of travel times. The model has an R^2 of 0.37.

	WEIGHT	**STD ERROR**	**T VALUE**	**PR(>|T|)**
Intercept	29.1339	0.0526	554.04	<2e-16
Intercept night	-9.0660	0.0320	-283.13	<2e-16
Distance	0.1725	0.0007	248.79	<2e-16
Distance * rain	0.0417	0.0007	62.80	<2e-16
Distance * wind	0.0003	0.0000	52.04	<2e-16

Although the model can explain a decent amount of variance in the complete dataset, i.e. 37%, the majority of the variation still requires some explaining.

9.2.4 Correcting people on the road

Finally, we will correct the number of people found traveling for variations in weather. In particular, maximum temperature, maximum wind speeds and rain can influence traffic intensities (Cools et al., 2010). Cools et al. (2010) found increases in traffic intensities for high temperatures and decreases during strong winds and rain. Unfortunately, they did not provide clear measures that help us to correct for the number of people on the road. We will thus have to construct our own correction mechanism.

The correction is necessary to prevent external effects distorting our image about the impact of roadworks. For example, when we find 5% less people on the road during roadworks we have to decide whether these people would normally make the trip. If they would, then we could assign them with half the impact experienced by the people who keep traveling, following the rule of half. If the people would normally not use the road, because they would not when it rains, then we should not include them in any further impact analysis.

As temperature, wind and rain depend on time of day, e.g. temperatures are always lower at night, we will compare daily rather than hourly traffic counts. Moreover, the weather information will also be aggregated to daily levels. Furthermore, we expect weather will influence traffic intensities differently at different moments. During workdays people will still go to work, even when it rains. During the weekend, however, people might decide not to go to the beach when it rains. Hence, we will allow our model to create different correction factors for weekend and not weekend. To create our training data we will note per road per day the maximum temperatures, maximum wind speeds, and average rain duration. Furthermore, we will divide the total vehicle counts per road per day by the average vehicle counts on for that road and day of the week. This will ensure we model ratios, e.g. on a rainy day there are 5% less people on the road than average, rather than absolute counts. Ratios provide more information and are more widely applicable than absolute numbers. A decrease in traffic flow of 5% could be explaining variation throughout the country whereas a decrease of 500 vehicles on one road, is specific only to that one road. This is also the reason why we do not use the models developed by Cools et al. (2010) as a decrease of X vehicles on a highway in Belgium cannot be applied to any highway in the Netherlands at any moment in time, e.g. day and night.

After constructing our dataset we find a few outliers are present as shown in table 9.5. The lowest 1% and top 1% show extreme values compared to the other 98%. When we dive into the raw data we find these outliers only exist on a few roads and they all occur during the weekend. Some large events may attract a lot of people causing peaks in traffic intensities on some roads and keeping people away from others. Because we want to model structural rather than occasional changes in travel behaviour we remove the top and bottom 1% of relative counts from our data.

Table 9.5, quantiles of the relative people counts in our sample.

Quantile	**0%**	**1%**	**10%**	**25%**	**50%**	**75%**	**90%**	**99%**	**100%**
Rel. counts	0.77	0.90	0.95	0.97	1.00	1.03	1.05	1.14	1.68

When we train the model described above, however, we find high vif scores for wind and temperature both in and outside of the weekend. This indicates the relations might be independent of whether it is weekend or not. Therefore, we decided to not make distinctions for weekend and not weekend for these variables. The final model is formulated in equation 9.4 where rc stands for the relative counts.

$$rc = \omega_1 * rain_{weekend} + \omega_2 * rain_{not\ weekend} + \omega_2 * wind + \omega_4 * temperature + \beta + \varepsilon \qquad \text{(Eq. 9.4)}$$

The trained model is shown in table 9.6. We checked and meet all the required assumptions for a LM. The plots and VIF scores that show the assumptions are met, along with a brief description and analysis of the findings, is presented in Appendix H (figure H9, figure H10, figure H11, figure H12 and table H3). We see all relations are significant with an alpha of .05. Wind, however, only has a very small impact of relative counts of people on the road. We can also infer from our model that rain during the weekend has a much larger effect than rain during weekdays.

Table 9.6, our LM estimation for predicting the change in relative counts of people traveling due to weather variations. The model has an R^2 of 0.09.

	WEIGHT	STD. ERROR	T-VALUE	PR(>|T|)
Intercept	1.0950	0.0014	768.43	<2e-16
Wind	-0.0002	0.0000	-24.36	<2e-16
Temperature	-0.0044	0.0001	-56.52	<2e-16
Rain weekend	-0.2532	0.0082	-30.83	<2e-16
Rain not weekend	-0.0560	0.0023	-24.66	<2e-16

We find, unlike Coots et al. (2010), that temperature also has a relatively large effect on vehicle counts. Note our data only covers September and October 2015 and are not necessarily descriptive of behaviour during the other months of the year. During summer, for example, people might be triggered more by weather, and specifically by nice summer days, to participate more in outdoor activities and go to the beach, for instance. Our findings thus do not dispute those by Coots et al. (2010) in general as the data only covers September and October 2015. What it does do, however, is indicate travel behaviour will plausibly be affected differently by weather in different times of the year.

Interestingly, we find that the created model is only able to explain a small part of the variation (9%) in the relative number of people traveling on the road. Although a small group might change travel behaviour due to weather, most people appear to be driven by other motives.

Although the model only explains a small portion of the variance, it still explains some. Hence, we will use this model to correct the number of people traveling. This, as stated earlier, helps to get a purer estimation of how many people would normally travel during roadworks.

9.3 Conclusion

We first created a dataset from the mobile phone data that provides good vehicle counts and is able to measure accurate travel times. The latter we achieved by taking quality users, i.e. with at least 100 data points a day, and removing trips with an origin or destination near the roadworks. Thereafter, we added information about roadworks and weather by linking descriptions about road positions to roads in our database and extrapolating weather information using GPS coordinates.

From this dataset we created a baseline. The most important characteristics of the baseline are that the baseline of a roadwork hour has to be on the same road on the same weekday hour combination where no other roadworks are present within 15 km on the same road. The last two weeks of October are excluded because they contain a holiday that could otherwise distort the measurements in the baseline. Furthermore, we only include hours of roadworks in future analyses where we have at least 4 hours in the baseline to ensure the baseline is stable.

Finally, we created models to predict travel time, travel time reliability and traffic flow using external factor such as weather fluctuations, nigh and day time, weekend, and travel distance. We do this such that we can predict travel times using factor unrelated to the impact of roadworks. By looking at the difference between what we predict and what we measure we can calculate the unexplained variation in travel time, travel time reliability, and traffic flow. This unexplained variation could then be used when we compare travel times, for example, during the baseline with those during the roadwork. The variation could then be attributed to the roadwork rather than, for example, rain.

In short we carefully craft our dataset from the mobile phone data and correct for variations caused by external factors such that we can accurately evaluate the impact of roadworks.

10 Report

In this chapter we will provide a case to discuss in greater detail the steps needed to measure the impact of roadworks. We will go over the individual steps for one case in section 10.1. In addition, we will present the impact of all roadworks that happened on highways in the Netherlands during October and September 2015 in section 10.2. One of our main motivations for wanting to measure the impact of roadworks is to drastically reduce the time and effort required to get this information. A conventional study might take months to complete (Taale et al., 2002). When we manage to calculate the impact of 638 roadworks here, we show how and that mobile phone data can be used to improve the measurement of the impact of roadworks on highways. Note, while we had over a thousand roadworks before, we only have 638 for which we also have data available. This could be the result of not reaching the minimum of 15 rule or having roadworks that are repeated over the entire two months, which means we cannot establish a baseline.

10.1 Case

The roadwork we selected for our case is on the A2 near Maastricht airport headed South. Meta-data about the roadworks is presented in table 10.1. We chose the roadwork because of the long duration of the roadworks and because large hinder was expected. Rijkswaterstaat estimated 10 to 30 minutes of delay for this particular roadwork.

Table 10.1, meta-data about the roadwork used for the case study.

QUESTION	AWNSER
Where were the roadworks?	A2 near Maastricht airport headed South
When did the roadworks start?	Friday, September 5, 2015, 21h
When did the roadworks end?	Monday, September 8, 2015, 2h
What happened?	Both lanes were closed, traffic was diverted

In terms of data present in our dataset, we have information on all 53 hours of when the roadworks are present. Furthermore, we have 256 hours of baseline information. We lack to hours during the baseline at 5 o'clock Sunday morning, one hour at 3 o'clock on Sunday, and another hour on Saturday at 4 in the morning. This is because we do not always have 15 people in our sample, and due to privacy regulation we cannot use this data. For all other moments in time, i.e. for each hour per weekday, we have 5 hours in our baseline.

As stated in chapter 9 we will compare the measurements during roadworks with those at similar moments in time. The measurements during the roadworks at 12 o'clock on Saturday will be compared to those at 12 o'clock during the baseline. Travel times, standard deviation in travel times, and people counts are corrected for external variables such as weather (see chapter 9).

The first step in our analysis, though the order is trivial, is analysing the number of people that decided not to travel (table 10.2). This is the number of people, corrected for changes in behaviour due to weather and such (see chapter 9), who decide not to take the trip. On average we find 173 people per hour decided not to make the trip. Over the entire 53 hours the roadworks are present the number adds up to 9.169 people for which the rule of half should be applied.

Note, this does not include people who decided to take a detour, but does include people who change mode of transportation. We cannot make a distinction between people making a detour and people who keep using the road where the roadworks are present. In our case the answer is straightforward, everyone made a detour as the road is closed. The rule of half will not be applied to people taking a detour. Although some might argue taking a detour is also finding a 'more optimal solution than keep traveling over the road', which suggests the rule of half should be applied (Appendix C). However, this is not always an option, and our case is a prime example.

Table 10.2, the corrected hourly counts of people on the road during the baseline and roadwork. The differences are shown under Delta.

MOTIVE	BASELINE	ROADWORK	DELTA
Other	784.09	627.52	-156.57
Work	75.84	60.64	-15.20
Business	5.22	4.05	-1.17

In addition to the people for whom the rule of half should be applied there are also 692 people per hour, on average, that continue to travel by car when the roadworks are present. In total this accumulates to 36.687 people affected by the roadworks. These people who are directly affected by the roadworks will be assigned the full impact.

During the roadworks we find higher travel times, on average, and greater deviation in travel times (see table 10.3). Here, as discussed in section 9.2.2, we have corrected for variation due to variations in distance travelled and weather.

Table 10.3, corrected travel times and travel time standard deviations per person during the baseline and roadwork. The differences are shown under Delta.

	BASELINE	ROADWORK	DELTA
Individual travel time (in min.)	-3.41	-0.88	2.53
Individual travel time std (in min.)	8.52	10.42	1.90

We can clearly see an increase in travel time and decrease in travel time reliability. The differences though, are less dramatic than the 10 to 30 minutes estimated by Rijkswaterstaat, which we know from our roadwork data. The impact presented by them, however, is predicted before measures are taken to help reduce the impact. The traffic alleviating measures put in place could explain both the lower counts of people on the roads and the small increase in average travel time. We do find a few minutes extra delay when we would not correct for weather fluctuations. However, this would be due to the added rain and not due to the roadworks. It rained about 9.8% of the time during the roadworks while it only rained 1.3% of the time during the baseline. By correcting for this we measure the delay caused by the roadwork, rather than the rain as discussed in chapter 9.

The differences in travel times and those in people found on the road as presented in tables 10.2 and 10.3, respectively, are not the numbers we will use to estimate the total impact. We showed them to (1) provide insight into the data and (2) better explain how we do measure the true impact. The difference between what we actually do and what we just discussed is that we look on hour level rather than roadwork level. We compare the people counts and travel time information at 11 o'clock on Sunday when the roadwork is present with those on an average Sunday 11 o'clock during the baseline. For each hour we then calculate the amount of people for whom to apply the rule of half, the number of people affected per motive, the increase in travel time in hours, and the increase in travel time reliability in hours.

The reason we operate on an hour rather than roadwork level is because the impact of roadworks in time dependent. The situation during rush hour can be completely different from that during the night, for example. It may be that during the night there is hardly any impact while during rush hour there is a lot. During rush hour there may also be more business and work related trips, which have a large economic impact. Hence, we prefer to look on hour rather than roadwork level. Ideally, we would get this information on even lower level, but this is not possible due to privacy imposed limitations.

In figure 10.1 we show the average costs per hour of the day for the roadwork we investigate in this case. Note the values are averaged. The data from Friday through Monday are all averaged to get the values shown in figure 10.1, showing every hour of every day produced an unreadable graph.

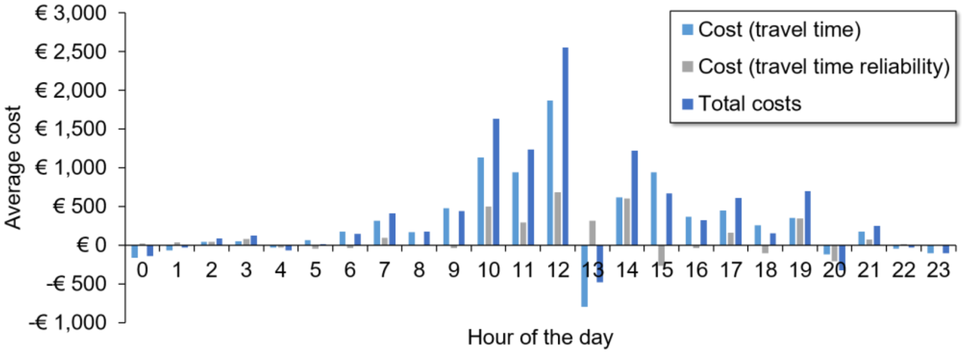


Figure 10.1, average cost for a specific hour of the day during which the roadwork was present. The costs due to travel time, travel time reliability and total costs are presented.

What we observe from figure 10.1 are two things. First the impact is mostly positive and much larger when there are more people on the road. The nightly hours hardly had any effect on travel time or travel time reliability as there is no clear positive or negative trend during these hours. During the day, with exception of 1 PM, travel time and the standard deviation in travel time are mostly positive. This implies the roadworks cause most delays during busy periods of the day, which confirms our expectations. At 1 PM there is some strange behaviour as travel times apparently drop. This might be because there were traffic jams, for example, during the baseline which did not exist during the roadworks. There is thus clearly some randomness in our data that can obscure the true impact of roadworks. Nevertheless, we are able to measure the overall trends. Furthermore, the traffic alleviating measures put in place could also be responsible for the lack of delays here. What we measure is not only the impact of the roadwork itself, but everything related to the roadwork. This includes measures to divert and reduce traffic intensities.

The complete impact of the roadwork sums to €18.676, with €4.870 being the result of less reliable travel times and €13.806 due to losses in travel time. Note the conversion from losses travel time and travel time reliability is performed using the VoT and VoR, respectively, that were presented and discussed in section 4.2.1.

10.2 Evaluation of all roadworks

To evaluate the impact of each roadwork we go through the same procedure as during the case above. For all 639 roadworks in our dataset we calculated the economic impact due to a decrease in travel time reliability and an increase in travel time. Together these form the complete economic impact of a roadwork. All relevant costs for each roadwork are presented in figure 10.2.

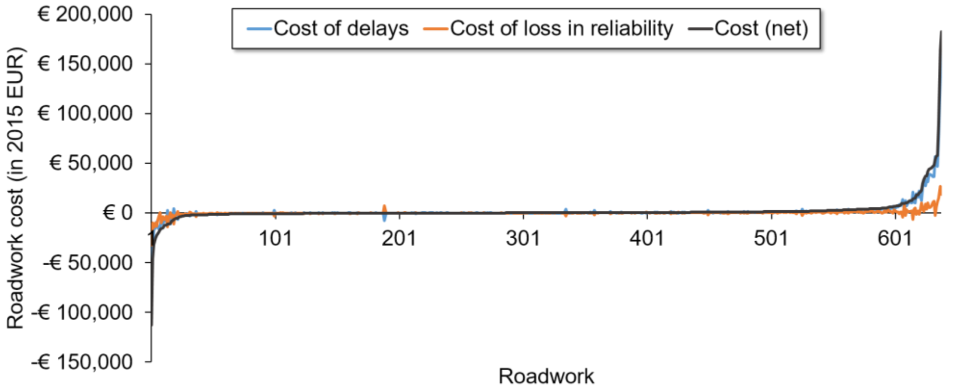


Figure 10.2, economic impact of all roadworks during September and October 2015, for which we can establish a baseline. Total costs are presented together with the costs due to increasing travel time and decreasing travel time reliability.

From figure 10.2 we observe the majority of roadworks have no net impact whatsoever. There are even a few roadworks during which travel times actually decreased and travel times became more reliable. This would be possible, for example, when there are traffic jams during the baseline and none during the roadworks. The roadworks that showed large negative costs occurred during the Tuesday and Wednesday morning rush hours. Plausibly there would thus be roadworks during the baseline and traffic alleviating initiatives could have prevented traffic jams from occurring. When we only look at the delays and decreases in travel time reliability this results in a completely different image (figure 10.3).

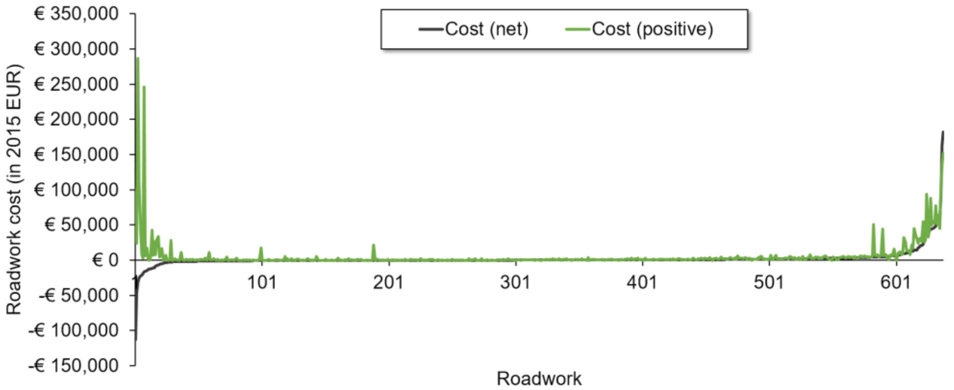


Figure 10.3, the overall costs (Costs net) measured during the roadworks and those discarding people traveling faster and more reliably during roadworks (Costs positive).

In figure 10.3 we see that the roadworks that have a negative net impact can still also cause delays. Similar to in our case we saw some hours might differ from the trend. In the case we found a negative impact at 1 o'clock in the afternoon while the remainder of the day travel times increased. For the roadworks with a net negative impact the same might be the case. There are apparently some hours which do negatively affect the population, but also some hours that were better than the baseline. Some randomness thus always exists. What we also observe from figure 10.3 is that the roadworks that appeared irrelevant in figure 10.2, i.e. in the middle of the graph, remain irrelevant. There is no apparent countering effect where some roadworks have both a large negative and positive effect on the people on the road.

When we aim to predict the impact of roadworks, starting in the next chapter, it will be fruitful to first detect what roadworks are interesting and what roadworks are not. As we just found, the majority of the roadworks are uninteresting from a societal perspective as they have very little impact on people. Nevertheless, knowing when a roadwork has little impact is important. When we can make a distinction between what roadworks are likely to have an impact and what roadworks do not this could be extremely valuable information. When roadworks are unlikely to have any impact we do not have to put a lot of effort in minimizing the little impact that is left and can focus on the bigger ones.

When we look at roadworks, especially the larger ones, it would be unfair to look only at the total impact. The total impact depends on the impact experienced over all hours during the roadwork. While some hours might have a profound impact some other might not. Determining whether a roadwork might have an impact thus has to be viewed within time frames. When we look at the impact of each hour where roadworks are present, neglecting the times where roadworks were better than the baseline, we get to figure 10.4. Here we see that 82% of the hours we measure an impact below a thousand Euros, over half of which are even less than 100 Euros. The other 18% ranges from 100 to 25.000 euro an hour. These are the hours we should focus on. The biggest benefit from a road user perspective would be achieved when we can create a shift towards fewer roadworks in expensive hours and more in expensive hours.

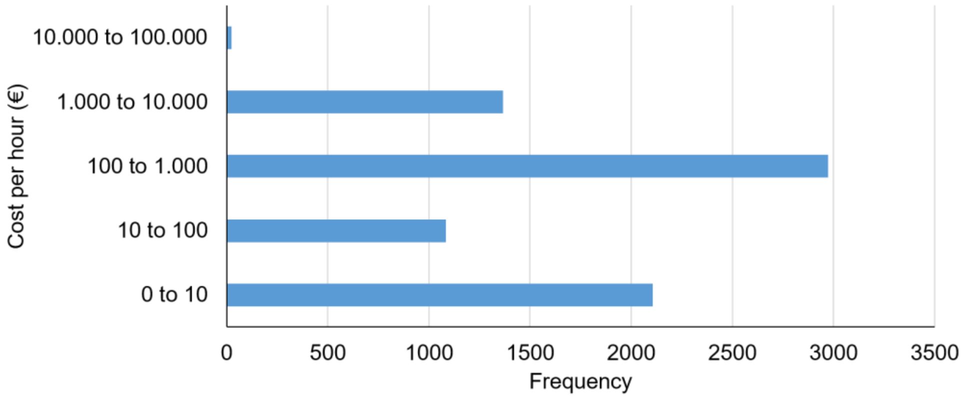


Figure 10.4, cost of roadworks per hour. In total there are 7.552 hours that haven been evaluated.

As promised we also calculated the total impact of all roadworks during September and October 2015. We found the net impact to be €1.109.548. The impact where we discarded any positive effects the roadworks came to a total of €3.543.513. This is the impact calculated for 638 of the 1.448 roadworks that actually happened during September and October 2015. Over half of the roadworks failed to be in the criteria we established to ensure we get a proper baseline. Furthermore, some roadworks were also in the last two weeks of October, which we discarded all together as these weeks were during holidays. The impact of recurring roadworks requires a larger baseline than was available for this study, but could, nonetheless, be calculated in exactly the same fashion.

The values we found are relatively low compared what experts expect. When we read a tender got a fictive discounts of up to €30.8 million because of a good traffic hindrance plan this appears out of proportion (Duijnisveld et al., 2011). Why would we somehow prefer a roadwork with a good traffic hindrance plan over one that is €30.8 million cheaper given the total impact of 638 roadworks, consisting of a total of 7.552 hours, is just over a thirtieth of this? It may be that the impact we measure is low because of the good traffic management we have in the Netherlands. However, there comes a point when the means surpass the goal and we start to spend more than we could initially save. We cannot provide a definite answer whether we overvalue traffic hindrance alleviating measures, but we at least find evidence suggesting it is worth investigating.

10.3 Conclusion

In this chapter we presented a case to show have the impact of a roadwork can be calculated. We also calculated and evaluated the impact of all roadworks on highways in the Netherlands during September and October 2015, provided we could establish a good baseline for the roadwork. Doing the same using conventional surveys and road side measurement devices would have taken an enormous amount of time and effort. Mobile phone data can thus much faster and with a lot less effort measure the impact of future roadworks. Knowing this we can start to also evaluate whether certain traffic alleviating measures help to reduce the impact of roadworks and how much. The latter will be important because we do not want to spend more on alleviating traffic hindrance than the impact we would otherwise experience.

Furthermore, we find there is a large difference between the net impact and positive impact, i.e. with only hours where the roadwork resulted in delays or less reliable travel times. The 638 roadworks evaluated had a €1.109.548 net impact and €3.543.513 positive impact. We, furthermore, find this number, despite it includes only part of all roadworks, to be low in comparison to the value assigned to traffic hindrance alleviating plans ranging into the tens of millions (Duijnisveld et al., 2011). We will note this under potential future research. We cannot state definitively the valuations are too high as our measurements are about the impact including the measures to reduce the impact. Hence, what we measure is not purely the impact due to a lane being closed, for example, but the roadwork in total, including the traffic alleviating measures taken.

11 Predicting the impact of roadworks

In this chapter we will predict the impact of roadworks. Roadworks are a complex phenomenon with many different moving parts. In order to better understand the parts involved and the fundamental relations behind the impact of roadworks we start with a literature study. In this literature study we will focus on (1) how people and related traffic intensities change in the presence of roadworks and (2) why roadworks cause delays. The important concepts and relations are discussed in section 11.1 finishing with a list of key attributes to take into consideration. In section 11.2 we will describe the data available for his study regarding roadworks, e.g. number of lanes closed, and briefly cover how some attributes are constructed. In section 11.3 we create our model to predict the impact of a roadwork. Note we will do this on hour level rather than roadwork level as the impact of roadworks can differ over time, e.g. rush hour versus night time (see chapter 10). Thereafter, we will evaluate the results from our model in section 11.4, also in comparison with the current 'state of the art'. In section 11.5 we discuss whether or not we would implement our models or not. In section 11.6 we will provide a concise summary of the main results of this chapter.

The overall structure of this chapter represents the steps in the CRISP-DM. We start with business understanding in section 11.1 followed by data understanding and preparation in 11.2. Then we create and evaluate our model in sections 11.3 and 11.4, respectively, and discuss the usefulness of implementing our model in section 11.5. Subchapter headings are also related to the names of the corresponding steps in CRISP-DM wherever possible.

11.1 Business understanding

11.1.1 Changing behaviour

We find the following variables can influence how people change behaviour due to roadworks. For each variable we will provide a brief overview of the relevant literature and how the variable can influence travel behaviour. The variables are:

- Transport demand management
- Accessibility to public transport
- Availability of alternative routes

Transport demand management

Transport demand management refers to initiatives that adjust traffic demand to reduce travel times and improve travel time reliability. Transport demand management can range from telling people to take a detour via matrix signs above the roads to paying people to stay out of rush hour traffic or charging a congestion tax (Knockaert, Tseng, Verhoef, & Rouwendal, 2012; Parry, 2002).

Transport demand management can have a strong impact on how people travel (Rijkswaterstaat, 2007). Rijkswaterstaat (2007) indicate that in extreme cases a reduction of 40% in traffic demand can be observed. Taale et al. (2002) reported similar, though slightly less optimistic, findings. Charging a congestion tax that varies, for example, by hour would be an efficient mechanism to reduce peak loads on the road network (Parry, 2002). Charging a tax specifically when roadworks are occurring, however, would be very harsh on the people affected. This would imply the people who are negatively affected by the roadworks would get an additional burden. The opposite of charging people more to drive in the rush hour traffic is to subsidize people to avoid driving in the rush hour traffic. In the Netherlands there have been trails to test the successfulness of this approach called Spitsmijden. The effects of Spitsmijden in the Netherlands appear promising with up to half of the participants deciding to take the monetary reward and only travel outside the rush hour period (Spitsmijden, 2009). The subsidy needed to have people travel at later times is also less than the VoT (Spitsmijden, 2009). People traveling outside rush hour will have more reliable travel times and will spend less time on the road. Avoiding rush hour thus has a benefit besides the added subsidy. Initiatives to better spread the load on the road can help to reduce the impact of roadworks by decreasing traffic demand at the work zone and by doing so help traffic to flow smoothly.

Accessibility to public transport

When people have better access to public transport this can serve as an alternative to experience the hinder at roadworks. When people live in a city with good public transportation the switch from car to train might be easy. Without good access to public transport this switch will be harder. When there are large roadworks we thus expect to see less people on the road when there is a good alternative, e.g. public transport. Accessibility of public transport could thus influence how many people will switch to alternative modes of transportation.

As a measure for accessibility to public transportation we will use urbanity. Urbanity has a large influence on the percentage of people taking the train as we can infer from OViN (figure 11.1). Urbanity in the Netherlands is defined in five categories from not urban to very urban which relates to the number of addresses per square kilometre (CBS, n.d.). Assuming the percentage of trips taken by train relates to the accessibility of public transport, we can link urbanity to accessibility of public transport. Hence, urbanity will also be incorporated in our model to predict the number of users found on the road during roadworks.

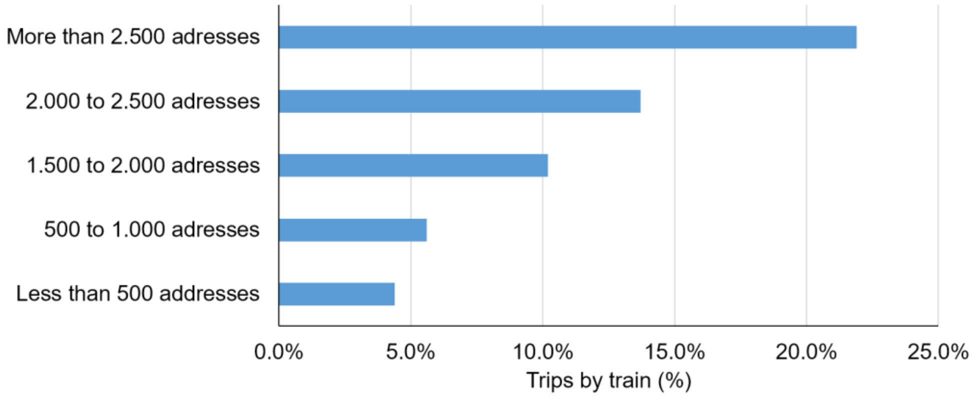


Figure 11.1, percentage of trips by train based for each of the 5 levels of urbanity, calculated using our OViN dataset from chapter 8.

Availability of alternative routes

The availability of alternative routes could also be an important variable. When people have no alternatives they are forced to continue to travel over the road where the roadworks are occurring. Hence there are less possibilities to better spread the load on the road network leading to more severe congestion and delays. A lack of alternatives imply people will have to travel of a specific route possibly where roadworks are present. The smaller the road network, i.e. the fewer the alternatives, the less the road network is capable of spreading the load of traffic. Although we do not make a distinction between people on the road and those taking an alternative route this information is still valuable. When people can take detours to avoid congestion this might reduce the overall travel time.

In general, we think, the greater the distance travelled the more alternative routes are possible. For example, there are less roads leading from Amsterdam to Utrecht than there are leading from Amsterdam to Paris. While people going to Utrecht might have to go over the road with roadworks, the people going to Paris might pick an alternative route. Distance can thus be seen as an indirect measure for the availability of alternative routes.

11.1.2 Causing delays

In this section we focus on literature that aims to explain causes for traffic delays. The simple answer is that a road user can never be faster than the road allows. This implies when all lanes are covered with vehicles you can at most travel as fast as the fastest person in front. When there are many people this can hinder average travel velocities as the fastest people are slowed down, thus resulting in delays. Moreover, the road can only allow a certain amount of traffic before a critical point is reached where traffic jams start to occur. This critical point is the road's capacity. Road capacity is also where the majority of literature focusses. Traffic jams are widely regarded as a major concern because this is where the largest delays are being experienced. The literature presented in this section will thus mainly focus on variables that are found to influence road capacity. The variables that will be discussed are:

- Lanes open / closed
- Changing road cross section
- Speed restrictions
- Lane width reduction
- Use of hard shoulder
- Ramps
- Driver familiarity
- Heavy Goods Vehicles (HGV)
- Day and night fluctuations
- Weather fluctuations

Lanes open / closed

Road capacity depends heavily on the amount of open lanes. General figures about road capacity are, therefore, often expressed in terms of vehicles per hour per lane (HCM, 2000; Goemans, Daamen, & Heikoop, 2011). Nevertheless, lower road capacity in itself does not result in delays. When road capacity is reduced the chance to exceed this threshold increases making traffic jams more probable. The number of lanes open is thus a vital variable in terms of estimating road capacity and resulting delays.

Changing road cross section

When a road is closed this will result in a change in the road's cross section. People will have to shift lanes causing additional chaos on the road. Research has shown that the weaving of traffic can lead to a reduction in capacity (Lertworawanich & Elefteriadou, 2003). Two separate studies found lane closure on average reduces the capacity of the remaining open lanes by approximately 5% (Heaslip et al., 2008; Ober-Sudermeier & Zackor, 2001). What lanes are closed appears to be relevant information as well. Rijkswaterstaat, for one, uses different reference values for road capacity depending on the cross section of the road (Goemans et al., 2011). The more traffic has to deviate its normal route the more the capacity appears to decrease (Goemans et al., 2011). Closing the right most lane, for example, is found to have a greater negative effect as there are more people who travel on the right lane than there are on the other lanes (Goemans et al., 2011). The study by Ober-Sudermeier and Zackor (2001), for example, found that crossing over all the way to the opposite carriageway results in a further reduction in capacity ranging from 5% to 10% for long term roadworks. In our model it would thus be fruitful to keep in mind the position of the lane that is closed.

Speed restrictions

Speed restrictions will also affect travel time. Lower travel speed will implicitly result in longer travel times as the same distance is still to be travelled. Speed is also found to be an important factor when estimating road capacity. According to the well-known Highway Capacity Manual (HCM) (2000) a reduction of 10 km/h will typically also reduce road capacity by 50 to 150 vehicles per hour per lane. Benekohal et al. (2003) go one step further and created an equation stating the relation between speed (U) in miles per hour and vehicles per lane per hour (q). Equation 11.1 shows the found relation (Benekohal et al., 2003).

$$q = 145.68 * U^{0.6875} \qquad \text{(Eq. 11.1)}$$

Speed restriction can thus decrease travel times because (1) people will have to drive slower and (2) lower speeds results lower road capacity and thus an increased chance of congestion.

Lane width reduction

Fitzpatrick, Carlson, Brewer and Wooldridge (2001) found lane width can explain up to 25% of the variability in average traffic speed. When lane width increases so does traffic speed (Firzpatrick et al., 2001). Plausibly, lane width affects the perception of safety of road users. When the lanes are smaller people will become more cautious and as a result drive slower. Aarts and Schaagen (2006) indicate narrow lanes do increase the change of being in an accident. This result is found even when taking into account lower speeds typically imply fewer accidents (Aarts & Schaagen, 2006).

The negative effects of lane width reduction, however, appear to be more profound when the lane width becomes smaller than 12 ft., i.e. approximately 3.6 meter (Benekohal et al., 2003). Below 12 ft. the reduction in speed due to lane width reduction is also nonlinear (idem). Benekohal et al. (2003) measured a reduction in speed of 1.9 mph at 11 ft. and 6.6 mph at 10 ft. compared to the baseline speed at 12 ft. Thus, the tighter the road the greater the reduction of speed resulting from a decrease in lane width.

Use of hard shoulder

A hard shoulder implies there is little room for error for the drivers. A hard shoulder is, for example, the presence of a concrete wall on the road used to separate the closed from the open lanes. Similarly to lane width reduction, the use of a hard shoulder may impose caution on the drivers resulting in a decrease in traffic speed. Although there are good indications the hard shoulder may affect road capacity there is hardly any conclusive research on whether the impact is significant and how large the impact is (Benekohal et al., 2003; Calvert, 2010).

Ramps

A ramp is a where a vehicle can enter or leave, for example, a highway. When people enter or leave the highway via a ramp they have to change lanes. This has a negative effect on the capacity of a carriageway (HCM, 2000). Heaslip et al. (2008) finds the incoming traffic can greatly reduce lane capacity. The impact of an onramp, i.e. where traffic enters the carriageway, depends on the amount of incoming traffic (Heaslip et al., 2008). The more traffic enters the road via the ramp the greater the impact (Heaslip et al., 2008). When a ramp is closed this can also affect people who would normally leave there. These people will now have to take a detour resulting in greater travel times. Ramps thus have a double effect, although we expect the impact to be small.

Driver familiarity

To date a number of studies have suggested that driver familiarity affects how effectively the road and local detour roads are used (Berkum & Huerne, 2014; Heaslip et al., 2007; HCM, 2000). Berkum and Huerne (2014) argue that road users that are familiar with the area might find it easier to find detours and circumvent the roadwork. They are people who are unfamiliar with the road are more likely to stay to the assigned route and major roads (Berkum & Huerne, 2014). The HCM (2000) states driver familiarity is a known influence. Plausibly due to a good way to measure driver familiarity, only a distinction between weekday and weekend days is made. The HCM (2000) suggests a 15% reduced capacity should be used for weekend days as in the weekend people are more likely to travel over unfamiliar roads.

Heaslip et al. (2007) performed a study where roadside measurements are combined with video footage to identify how driver characteristics relate to road capacity. In total four driver characteristics are identified and estimated in their study. There are driver familiarity, adaptability, aggressiveness, and accommodation. They found the four factors could result in a 37.5% increase or 26% decrease in road capacity in the most optimal and least optimal configuration, respectively (Heaslip et al., 2007). To get to these estimates they monitored video footage of two highways in the USA over a one day period. Although the results hint there is a difference it is hard to use aspects such as adaptability, aggressiveness and accommodation in a model as these are hard to estimate and for that reason will not be incorporated in this study. Familiarity is easier as it can be linked to, for example, the percentage of work traffic. People going to work often do so regularly and are most likely familiar with the road. Hence, percentage work traffic will be used as a measure for familiarity in this study.

Heavy Goods Vehicles (HGV)

The effect HGVs have on road capacity are significant and well documented (HCM, 2000; Calvert, 2010). In particular, the percentage of HGVs to the total of vehicles on a road is often taken into account when predicting road capacity and or the impact of roadworks (HCM, 2000; Calvert, 2010; In de Vegte & Van Toorenburg, 2014; Benekohal et al., 2003). There are a number of logical explanations that would explain why HGV result in a decrease of capacity. Research shows traffic jams are the result of people driving at different speeds resulting in fluctuations that can create instability and result in a traffic jam (Sugiyama et al., 2008). HGVs typically drive at a slower speed and when they have to change lanes this requires others vehicles to either brake or speed up. Because HGVs cause additional fluctuations they logically increase the chance of having a traffic jam, which is equivalent to stating the capacity of the road decreases when more HGVs are present.

Heaslip et al. (2008) used simulations to test a total of 243 unique scenarios. One of the key variables they investigated was the effect of HGVs on road capacity. Their results show the amount of HGVs have little impact when they are less than 10% of the total traffic. Between 10% and 20% a significant drop in capacity is observed of approximately 6% (Heaslip et al., 2008).

Day and night fluctuations

Day and night fluctuations have a double effect on the road capacity. The absence of natural light can harm visibility in general. On the one hand, a decrease in visibility people can result in drivers being more cautious and drive slower. On the other hand, less visibility gives people less time to react, which increases the chance on accidents and thus also reduces road capacity.

Research finds the poorer light conditions in the night can reduce capacity by 4% (Heaslip, 2008). Other research further confirms the significance of the impact of light fluctuations (Al-Kaisy & Hall, 2000). Al-Kaisy and Hall (2000) did a study to quantify the capacity reduction due to darkness. They found the impact of darkness on capacity ranged from a reduction of 3.25% to a reduction of 7.5% in capacity depending on the work site (Al-Kaisy & Hall, 2000). The lack of visibility due to increased darkness as is present during the night time can thus have detrimental effects on road capacity.

The effect of day night fluctuations, however, do not end here. Wanvik (2009) analysed data on traffic accidents on Dutch roads from 1987 to 2006 and found the chance of accidents on unlit roads to be approximately twice as high as on well-lit roads (Wanvik, 2009). Effects are less severe in bad weather conditions, e.g. fog or snow, as the chance on having an accident is already increased due to bad visibility (Wanvik, 2009). Although Wanvik (2009) did not make a direct relation with the effect of daytime versus night time it seems highly probable from his results that night time induced darkness can also have an effect on the number of road accidents. Accidents by themselves have a large negative effect on road capacity. This is because lanes might be unusable till the vehicle is removed and a further reduction of speed might be required for safety reasons.

Weather fluctuations

Weather can have a significant impact on both traffic demand and road capacity (Cools et al., 2010). Precipitation, for example, is found to decrease road capacity, with capacity reductions ranging from 4% to 30% (Stern, Shah, Goodwin, & Pisano, 2003; Unrau & Andrey, 2006). In chapter 9 where we used weather to correct for travel time and travel time reliability we only used data where vehicle counts are far below road capacity. We thus did adjust for people driving slower and having less reliable travel times during rains and strong winds, but did not for the reduction in road capacity. For road capacity weather fluctuations are thus still interesting to keep in mind.

11.1.3 Overview

Here we present a quick overview of all variables discussed in the previous two sections. We find some are related specifically to how many people we will find on the road while others are more related to the road capacity (table 11.1). Weather fluctuations are the exception as they both directly affect road capacity as well as traffic demand, as we found in chapter 9 (Cools et al., 2010).

Table 11.1, of roadwork characteristics and their subcategories.

KEY CHARACTERISTIC	VARIABLE
Traffic demand	Transport demand management
	Accessibility to public transport
	Availability of alternative routes
Road capacity	Lanes open / closed
	Changing road cross section
	Speed restrictions
	Lane width reduction
	Use of hard shoulder
	Ramps
	Driver familiarity
	Heavy Goods Vehicles (HGV)
	Day and night fluctuations
Both	Weather fluctuations

11.2 Data understanding & preparation

11.2.1 Data understanding

When we create our model we have to keep in mind that more vehicles on the road generally lead to more delay and especially when road capacity is exceeded.

There is a whole range of variables that influence how many people per lane a road can handle before traffic jams occur. All of them have distinct reasons for the capacity to decrease, e.g. fear of getting into an accident. Most of them, however, have a very similar structure in terms of their effect. They all tend to decrease road capacity by a certain percentage. The variables and their effects, both quantitatively and in word, are summarized in table 11.2.

Table 11.2, description of the variables used for modelling and their effect on road capacity

VARIABLE	EFFECT ON ROAD CAPACITY	SOURCE
Lanes open / closed	Multiplier	Capacity is measured per lane
Changing road cross sect.	-5% to -10% closing right versus left lane	(Goemans et al., 2011)
Lane width reduction	non-linear decreasing (approx. quadratic) with smaller lanes	(Benekohal et al., 2003)
Speed restrictions	non-linear decreasing (approx. quadratic) with greater restrictions	(Benekohal et al., 2003)
Driver familiarity	Increases with more familiar drivers	(Heaslip et al., 2007)
Day and night	Decrease at night	(Wanvik, 2009)
Weather fluctuations	-4% to -30% during bad weather	(Stern, Shah, Goodwin, & Pisano, 2003; Unrau & Andrey, 2006)

In terms of people on the road we find fewer effects. The main effects being access to public transport and availability of alternative routes, both of which we hypothesized, and transport demand management. The latter is a proven method that consists of putting measures in place to reduce or better direct traffic (Knockaert, Tseng, Verhoef, & Rouwendal, 2012; Parry, 2002; Rijkswaterstaat 2007).

For modelling we have to look at both traffic demand and road capacity as these are the key components leading to delays on the road. Traffic demand is needed to see if it is (1) busy on the road, which leads to slower speeds, and (2) whether the road's capacity is exceeded, which leads to traffic jams and thus delays.

11.2.2 Data preparation

In chapter 9 and 10 we already provided an extensive explanation of the data that will be used. A few things, however, we did not yet discuss. This relates solely to the data available about roadworks and in specific about the information we have on roadwork level.

In table 11.3 we present all attributes also present in table 11.1 including whether we have the necessary information present in our data. Note the information originates from the SPIN system that is used by many large organizations and is both up to date and extensive. The majority of the variables we found relevant for the study is available in this system. Nevertheless, we still needed to add some from external sources or create them from other data. In table 11.3 we also indicate the source of the data and whether they had to be constructed from other data. In case of the latter we also provide a brief description of how we accomplished this.

Table 11.3, roadwork related variables and their sources used in the model to predict the impact of roadworks.

VARIABLE	SOURCE	CONSTRUCTED
Transport demand management	SPIN	Yes
Accessibility to public transport	CBS	Yes
Availability of alternative routes	Mobile phone data	Yes
Lanes open / closed	SPIN	Yes
Changing road cross section	SPIN	Yes
Speed restrictions	SPIN	
Lane width reduction		
Use of hard shoulder		
Ramps		
Driver familiarity	Mobile phone data	Yes
Heavy Goods Vehicles (HGV)		
Day and night fluctuations		Yes
Weather fluctuations	KNMI	

We lack information about HGV, hard shoulders, and ramps. This data we do not have available at the moment. Their effect on the impact of roadworks can thus not be established here. We will suggest this would be done in future research.

Transport demand management is indirectly present in the SPIN data. In the SPIN data there is an attribute called traffic hindrance category, which relates to how much impact is expected for the roadwork. When roadworks are expected to have a large impact, measures to reduce traffic intensities are often in place. This we will use to determine, though indirectly, whether transport demand management is used for the roadwork. In SPIN the hindrance classes are E, D, C, and B. In theory A represents the highest impact, but this is not present in our data. E is the lowest. Definitions can be found in the document by Taskforce Doorstroming (2009).

Accessibility of public transport is inferred from the level of urbanity as discussed in section 11.1.1. We link the urbanity level of each destination to each trip in the mobile phone data where 1 stands for low urbanity and 5 for high urbanity. Urbanity is used as a measure for accessibility to public transport.

Availability of alternative routes is expressed in average distance travelled over the road. The further a person travels, the larger the greater the chance on alternative routes (see section 11.1.1).

In SPIN there is information about what the signs above the highway indicate. For example, lane 1, lane 2 and lane 3 are open and lane 4 is closed. Based on this we can infer the number of lanes and the number of lanes closed. Moreover, on some occasions there are no signs above the road. As we then lack vital information we cannot use these roadworks when creating and evaluating our model. In addition to the number of lanes closed we also note whether the right most lane is closed (0) or the left most lane is closed (1).

Driver familiarity we relate to the percentage of work travel. Home to work travel is consistent and often repetitive. The greater the percentage of work travel in amongst all trips we reckon the greater the chance of drivers being familiar with the road. We find, as we will later discuss, that the square root of the percentage of work travel is most valuable in the modelling stages. Driver familiarity we will therefore measure as the square root of the ratio work trips over all trips.

Day and night fluctuations are fixed with night being before 7 AM and post 8 PM.

11.3 Modelling

There are two main aspects of importance here, i.e. traffic flow and road capacity. The interplay between these two will help us to determine the impact of roadworks. For our model to predict the impact of roadworks these two distinct factors and there interplay has to be described.

We will start out by investigating how traffic flow changes during roadworks and will, thereafter, focus on how travel times and travel times reliability are affected by changes in road capacity. We will make a special case out of roadworks where the entire road is closed. There the number of people still driving remains relevant, although the road capacity does not. Hence, we choose to start by investigating changes in traffic intensities as this will be input for all models to come.

The model to predict traffic flow will be presented in section 11.3.1. Then we continue with roadworks where the complete road is closed in section 11.3.2 and we will finish off with a model to predict the impact during roadworks where lanes are still open.

Note we will not create a model to measure the impact of roadworks where no roads are closed. We would expect and find no large influences at these roadworks. On average we see an increase in travel time of 0.9 minutes with a standard deviation of 3.7 minutes. With travel time reliability we measure increases of 0.2 minutes with a standard deviation of 5.5 minutes. These values, as we will shortly seem are trivial compared to the impact of the other roadworks. In total we discard investigating the impact of 7% of roadworks by not including the ones without lane closures.

11.3.1 Traffic flow model

As stated in 11.2.1 the key attributes to focus on are accessibility to public transport, availability of alternative routes, and transport demand management.

To test the impact we first need to establish a baseline to which we will compare the vehicle counts during roadworks. To do so we apply the same procedures as with establishing the baseline when measuring the impact of roadworks, i.e. pick similar moments in time where no roadworks are present. We also only use the corrected vehicle count information here, just like we did when measuring the impact of roadworks. Once we got vehicle counts during the roadwork hours and their corresponding baselines we divided the vehicle counts during each roadwork by the average count in its baseline. Due to the randomness in data with small sample sizes we focus only on hours where there are at least 1.000 people (scaled) crossing the road during that hour. By doings so we hope to reduce outliers without losing too much valuable information. Roadworks with only a few people can never have a very large impact and thus is the information at those moments not that crucial. Vice versa, when there are many people it will become more important to provide accurate measurements.

After selecting and preparing our data we visualized the relation between the ratio vehicles roadworks over vehicles baseline with respect to availability of alternative routes and accessibility to public transport. For neither of these two relations we could observe any trend. The data points are all very close to one with a few outliers that did not seem to be related to the number of people left on the road during roadworks. The two created graphs are shown in figure 11.2.

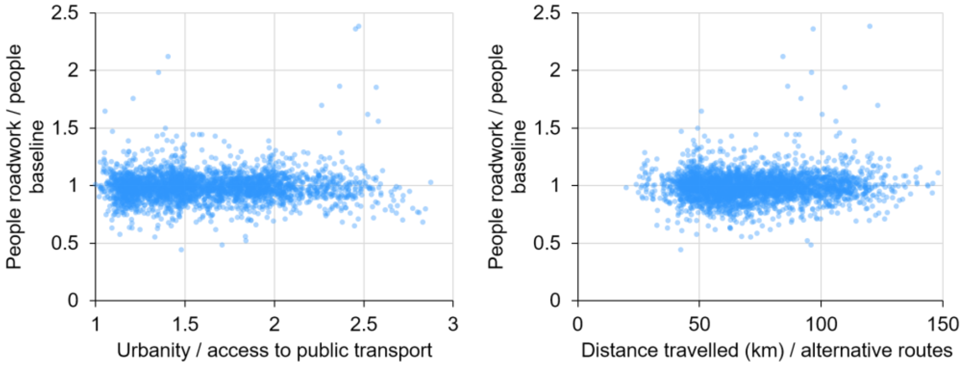


Figure 11.2, ratio people roadworks / people baseline compared to access to public transport and availability of alternative routes.

Transport demand management, however, showed a clear trend. We can clearly see there are fewer people on the road when the roadworks are classified as a B or C, i.e. roadworks with a large predicted impact. The average values per group and the group sizes are displayed in table 11.4. While during the majority of the roadworks we see no noteworthy change, we do at the roadworks of classes B and C. At these large roadworks, which is what B and C stands for, we see average numbers of people on the road drop by an average of 12%. Most likely there were some measures put in place to reduce traffic intensities at these moments, as we would expect from literature.

Table 11.4, mean ratio people roadworks / people baseline and the number of times a roadwork hour with a specific class is in our sample.

CLASS	**B**	**C**	**D**	**E**	**BLANK**
Mean	0.870	0.898	1.024	1.009	0.992
N	26	28	330	113	2847

To test whether these findings are significant we performed two two-sampled t-tests. Once to see whether the roadworks in groups B and C significantly differed from the rest and once to see whether roadworks in group B differed from those in group C. For the first test we found a significant difference in the change in average vehicle counts between roadworks in class B and C (Mean = 0.884, SD = 0.128) and the rest (Mean = 0.996, SD = 0.137) with a p-value of $4.4 * 10^{-8}$, which is way below .05 used for significance. Our second test showed there is no difference between roadworks in class B (Mean = 0.870, SD = 0.095) and those in class C (Mean = 0.898, SD = 0. 154). We found a p-value of .43, which is over our alpha of .05, and generally quite high.

We will use our new found information when predicting the impact of roadworks. We now know there is unlikely any change in travel behaviour, in terms of people switching to the train or staying at home, unless the roadworks are marked with a B or C. When roadworks belong to classes B or C we now estimate 12% are staying away from the road while 88% keep on driving.

11.3.2 Road closure model

Here we will create a model to predict the impact of a roadwork during an hour when the roads are completely closed.

As the road is completely closed we do not directly have to investigate the effect of road capacity changing variables. Instead we will focus on:

- Availability of alternative routes,
- Familiarity, and
- People affected.

These are all variables that could influence how well people can still get from A to B without the road they normally travel on. With many alternatives available it might be less worrisome when a road is closed. Furthermore, when people are familiar with the surrounding they might be more capable of finding alternative routes. Finally, when there are many people being affected, the surrounding road network could be over flooded with the extra demand resulting in delays. When only few people are on the road this should be less of an issue. Hence, we hypothesize these three relations might influence the impact of road closure.

The total impact results from the costs due to increasing travel times and part from the cost of decreasing travel time reliability. As the above described relations might differ for each subpart of the total impact we will create a model and investigate the relations for both.

We will first discuss the relations with respect to increasing travel times and, thereafter, do the same for decreasing travel time reliability. Combining both to predict the impact of roadworks with road closures and evaluating the results will be done in 11.4.1 in the section about evaluation.

Travel times at road closures

We will start by visualizing the relation between travel time and the three variables mentioned above. When we visualize the relations, later model them, we use the information from the baseline unless otherwise specified. This we do because the future is unpredictable and we want our final model(s) to be able to predict the impact with values that can be gathered beforehand. We do know some aspects might change, e.g. number of people on the road, and we try to correct for this whenever we think it is fruitful, hence our traffic flow model.

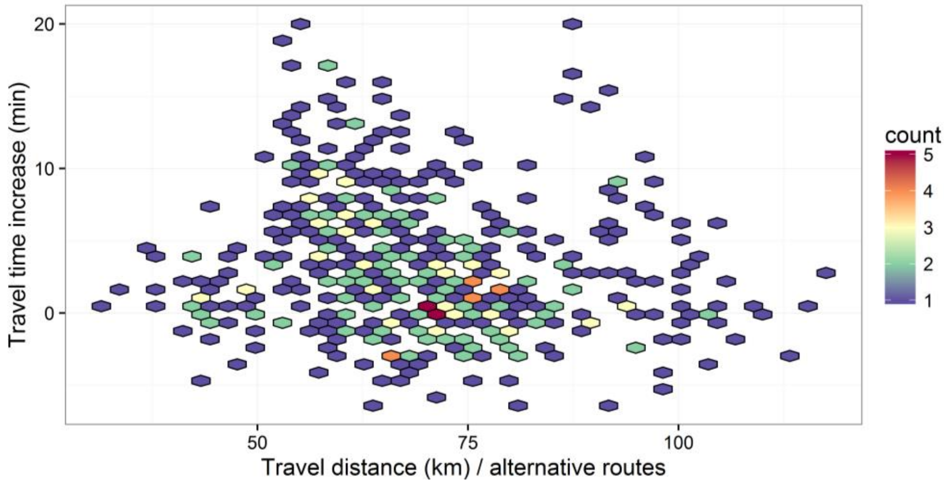


Figure 11.3, travel time increase plotted against the average travel distance during the baseline, which we use as a measure for the number of alternative routes.

In terms of alternative routes or, conversely, travel distance we see no apparent relation with increases in travel time (see figure 11.3). We also find a correlation of -0.16 between the two variables indicating there may be a small decrease in travel time at greater distances, although the effect appears to be marginal.

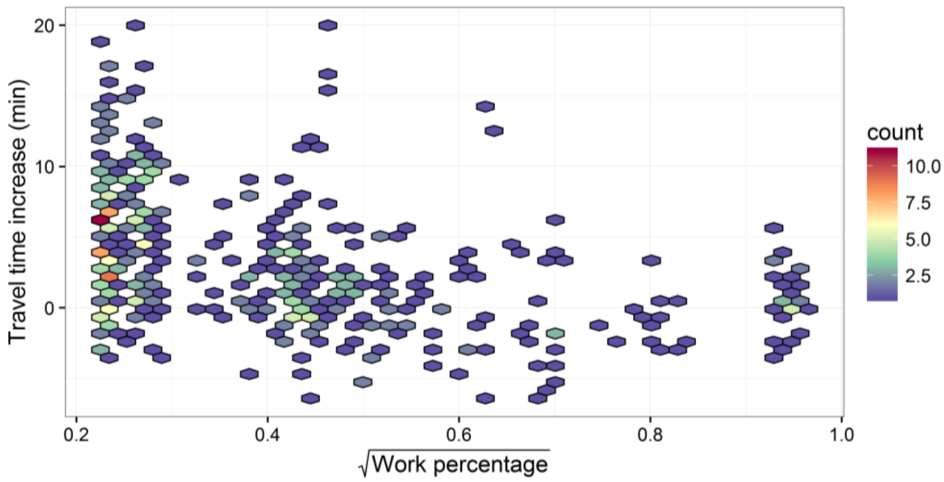


Figure 11.4, travel time increase plotted driver familiarity, which we measure as the percentage of work traffic.

When we look at figure 11.4 we can see a clear downward trend in the increase of travel time when the percentage of people with work motives increases. We used the square root of the work percentage here as the non-adjusted relation did not appear linear. When driver familiarity is low the increase in travel times appear substantial. Travel time increases of between 4 to 8 minutes are common. When we look at the other end of the spectrum we see little to no increase in travel times.

Work percentage also relates to the week versus weekend differences. As large roadworks often occur during the weekend we though this might explain the above found relation. We, therefore, also investigated whether weekend rather than work percentage can explain the increase in travel times. What we found was that the square root of work percentage can better explain the variation in travel times. While the square root of work percentage can explain 17% (Pearson correlation of -0.41) the attribute weekend could only explain 5% (Pearson correlation of 0.22). As weekend and the square root of work percentage are correlated (Pearson correlation of -0.36) we chose to leave weekend out of the equation. We did not include weekend in our model as it might result in multicollinearity.

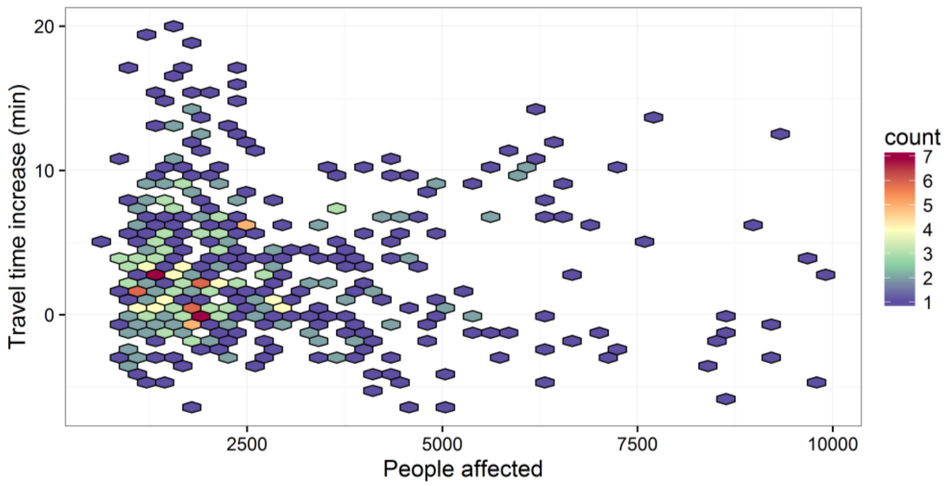


Figure 11.5, travel time increase plotted against the corrected number of people affected. The number of people affected (during the baseline) are reduced by 12% in accordance with our findings in 11.3.1.

There appears to be no apparent relation between the travel time increase and the number of people on the road during road closures (figure 11.5). We do see that the average travel time increase is quite substantial, e.g. about 4 minutes. Nevertheless, this average does not appear to be affected by the number of people.

The only variable that shows a clear relation with travel time increase is the square root of the percentage of work trips, i.e. driver familiarity. Because we found high correlation between consecutive residuals with the standard LM we applied Cochrane-Orcutt estimation (Cochrane & Orcutt, 1949). After one iteration we find the model presented in table 11.5. The final model met all necessary assumptions (Appendix I, figures I1, I2, and I3). Note we did not have to calculate VIF scores as we only have one variable. As it turns out we meet all assumptions and the model shown in table 11.5 is fully interpretable and trustworthy.

Table 11.5, model for predicting the increase in travel time during road closures. The model can explain 6% of the variance in our data.

	WEIGHT	STD ERROR	T VALUE	PR(>|T|)
Intercept	2.454	0.222	20.76	<2E-16
Driver familiarity	-2.257	0.416	-12.61	6.03E-08

As expected we a significant effect for driver familiarity. When drivers are more familiar with the road, the lower the increase in travel time during road closures. When a road is closed we expect delays of 2.5 to 0.2 minutes depending on driver familiarity, which theoretically ranges from 0 to 1. The created model can explain a total of 6% of the variations in travel time increase during road closures. Our model can thus only explain a minor part of the variation in the data.

Travel time reliability at road closures

For travel time reliability we will investigate the relations like we did for travel time.

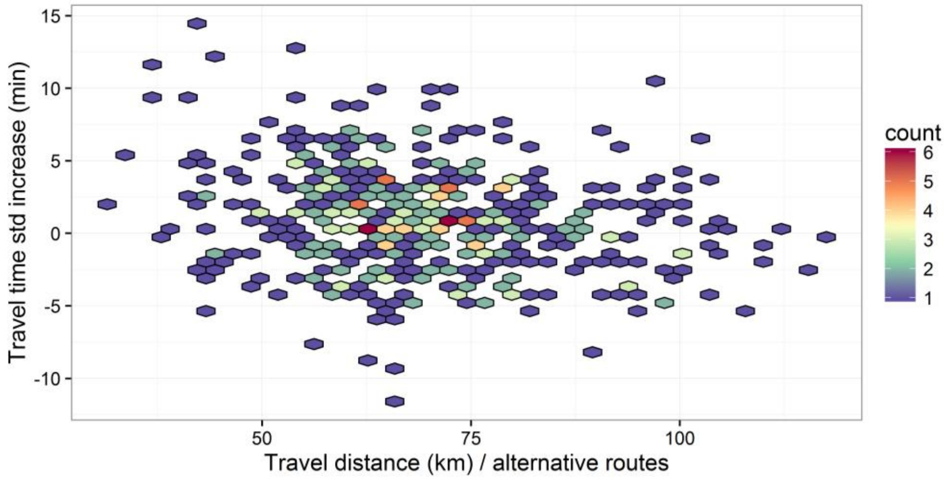


Figure 11.6, increase in the standard deviation of travel time plotted against the average travel distance during the baseline, which we use as a measure for the number of alternative routes.

It is hard to see a clear upward or downward pattern in figure 11.6. On visuals alone, we do not expect any noteworthy relation. We do find there is a correlation of 0.21 between travel distance and the increase in the standard deviation of travel times, but this is still only a very weak correlation. Moreover, when we look at figure 11.6 the points on the left appear higher than those on the right, which does not match the found positive correlation. We are, therefore, not confident that travel distance truly has an effect on the increase in travel time standard deviation during roadworks.

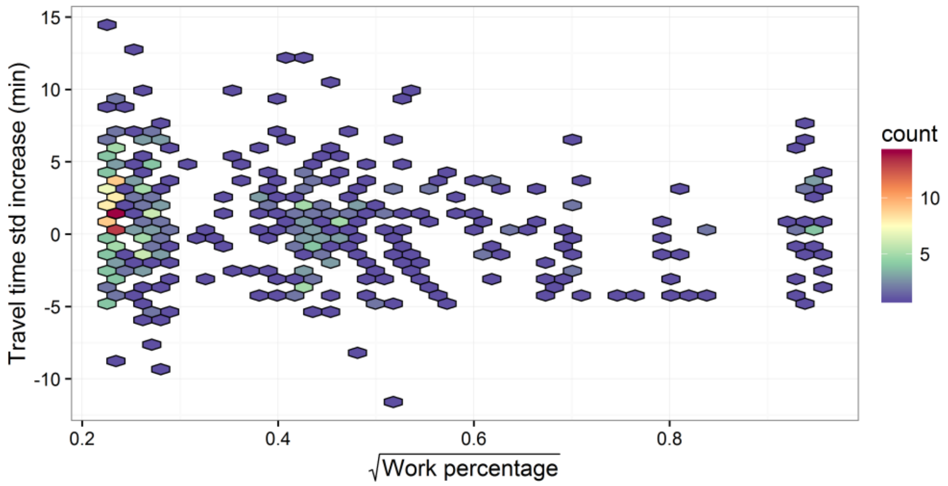


Figure 11.7, increase in the standard deviation of travel time plotted against square root of driver familiarity, which we measure as the percentage of work traffic.

Unlike the clear relation we saw when comparing familiarity with travel time, there appears to be no relation here (figure 11.7). A Pearson correlation of -0.12 confirms what we see.

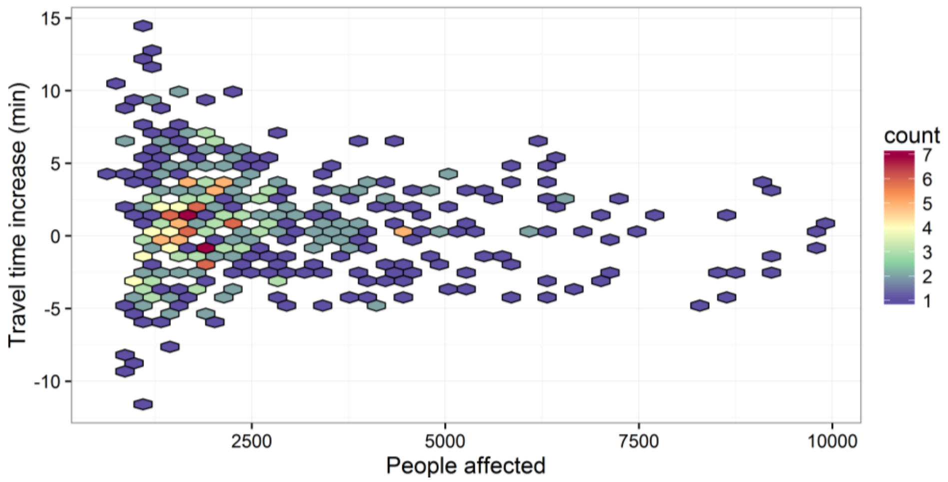


Figure 11.8, increase in the standard deviation of travel time plotted against the corrected number of people affected.

The number of people affected also has no clear effect on the decrease of travel time reliability (Pearson correlation of -0.09). From figure 11.8 we do see the largest deviations are found when there are few people on the road. Furthermore, we observe a large deviation around the mean increase in travel time standard deviation that we cannot explain.

As we find none of our variables can explain any meaningful amount of the variation in the data we choose to simply take the mean travel time deviation as our best guess. The mean increase in travel time standard deviation of 0.9 minutes will be used as the only measure to estimate the increase in travel time reliability.

11.3.3 Lane closure model

Here we will evaluate the impact of roadworks where at least one lane is closed. In total 4.204 of the 8.039 hours in our dataset belong to type of roadwork. Similarly to section 11.3.2, we will first discuss the effect of roadworks with lane closure on travel time and, thereafter, their effect on travel time reliability.

For both the analyses we will perform we investigate the impact in combination with a series of attributes. Unlike roadworks where the road is completely closed, however, here the attributes related to road capacity and travel velocities over the road are also important. In particular, the number of vehicles per lane we expect to be a crucial indicator of both travel time loss and loss in travel time reliability. This we expect because the number of vehicles per lane relates to whether road capacity is exceeded and traffic jams, i.e. large delays, are a given.

Note when we refer to vehicles per lane we use the vehicle counts during the baseline, adjust using our traffic flow model (see 11.3.1), and divide by the number of open lanes during the roadwork.

In the following model both the attributes relating to lane closures and those relating to traffic demand are all taken into consideration. Traffic demand management and familiarity with the road can also help people find alternatives and spread the load on the road. The vehicles per lane we measure might be an overestimation of what is actually on the road provided the increase in people taking detours. Finding out whether a road reached its capacity is thus not as straightforward as with conventional road side measurement devices that can measure traffic flow directly where a roadwork occurs. Traffic demand related attributes can perhaps help explain some variation in what we expect versus what we measure. Attributes related to road capacity will also be used to help explain why sometimes road capacity is exceeded and delays are measured and why no delays are measured during other times.

The data used to create the following models is equivalent to the data used through section 11.3 with a few additions. It consists of data on hours during which roadworks are present, with at least one lane closed and no full closure. Each row of data also contains information from the baseline situation that will be used for prediction. A special case here is the number of vehicles during the baseline that corrected using our traffic flow model (see 11.3.1). Using this predicted number of vehicles we also calculate the vehicles per lane by dividing the number of vehicles by the number of lanes open. We, furthermore, included an attributed to quantify how many more vehicles per lane are present during the roadworks that during the baseline. Other attributes that have been constructed are percentage of roads open / closed. With this data we will try to explain changes in travel time and travel time reliability during roadworks with lane closures.

Travel times at lane closures

The first thing we want to investigate is the fundamental relation between travel time increase and vehicles per lane. Near and beyond 2000 vehicles per lane we would expect significant increases in travel (HCM, 2000). The data regarding the relation between travel time increase and vehicles per lane is shown in figure 11.9.

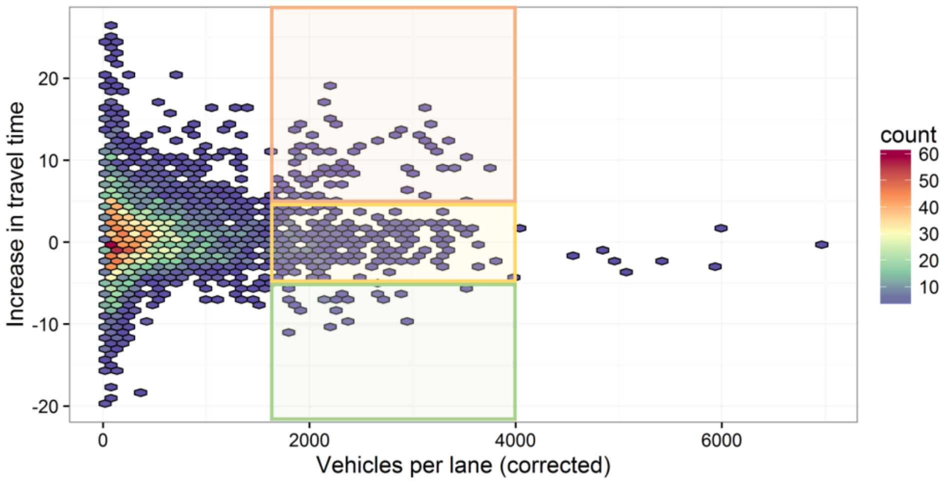


Figure 11.9, the increase of travel time plotted against the vehicles per lane. The red, yellow, and green boxes indicate three areas we are particularly interested in.

First of all, we have to explain why we have such high numbers on the vehicles per lane axis. Typically it is not possible to exceed road capacity when using road side measurement devices. However, we can exceed road capacity as our way of measuring and hence what we measure slightly differs from the definition of vehicles per lane per hour. We are able to go beyond road capacity as we do not take into account detours and due to the procedure with which we timestamp people crossing a road. When people queue in traffic they might not cross the road section in reality, but the middle of their trip might still 'cross' the road. Imagine, for example, a person leaves at 7 o'clock and arrives at 9 o'clock while another leaves and arrives at 8 o'clock. When both travel over the same road we say both went over that road between 8 and 9 o'clock. We thus do not take into account whether the first person was stuck in traffic before or after crossing the road which can also slightly distort what we see in the data.

Second of all, we have to explain why we are interested in the three highlighted regions. When we look at the left part of figure 11.9 we see an approximately normally distribution of increase in travel time over the vehicles per lane dimension. This would imply the deviations from the mean could be noise and does not imply a structural relation. That said, there are some points at around 1.000 vehicles per lane that look as if there might be something going on that resulted in the increasing travel time. At 1.000 vehicles per lane most points are between -5 and 5 minutes of travel time increase and a small cloud of observations has a much greater increase in travel time that might indicate they might differ from the group in some aspects. Nevertheless, we want to focus on the group with slightly higher vehicle per lane counts. From about 1.500 vehicles per lane to 4.000 vehicles per lane the distribution appears to be non-normal. The group of observations with no increase (yellow region) is relatively small compared to the number of times large travel time increases are measured (red region). Occasionally, we even find some points where travel times decrease (green region). This area of the graph we, therefore, find worthwhile to further investigate. Our data also focusses mostly on what influences road capacity and is thus most suited to investigate behaviour where vehicle per lane values are higher.

The group over 4.000 vehicles per lane is excluded because we find the values to be out of the ordinary. After further investigation we find three of the four roadworks where this is the case are located near large roads that could have coped with the added traffic demand (see figure 11.10). For the other roadwork we could not explain the high vehicle per lane value, but suspect someone wrongly inputted the information about the roadwork as it says one lane is open and the maximum road width is 0 meters. We, therefore, removed all other measurements regarding this roadwork as well.

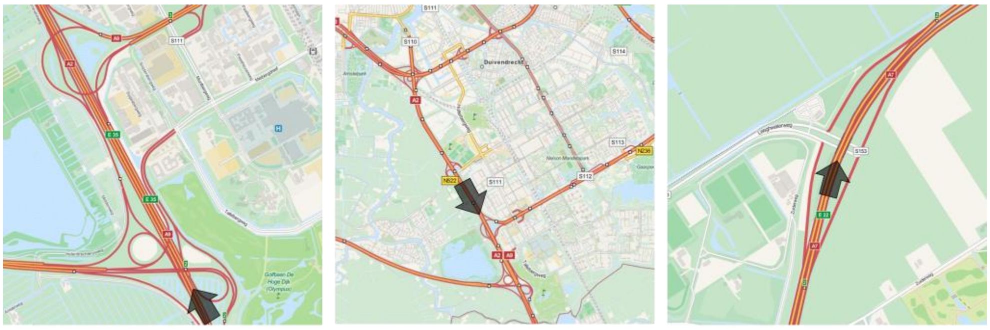


Figure 11.10, three roadworks with high vehicle per lane values that did not saw an increase in travel time, probably due to potential alternative roads nearby that could cope with the traffic flow.

As we saw in our outlier analysis, the presence of alternative routes ensured travel times did not increase even when vehicles per lane values were extraordinarily high. To see if the presence of alternative routes could explain why sometimes roadworks at high vehicle per lane counts resulted in increases in travel we created figure 11.11. In figure 11.11 the increase in travel time is plotted against the average travel distance, which we use as a measure for the number of alternative routes.

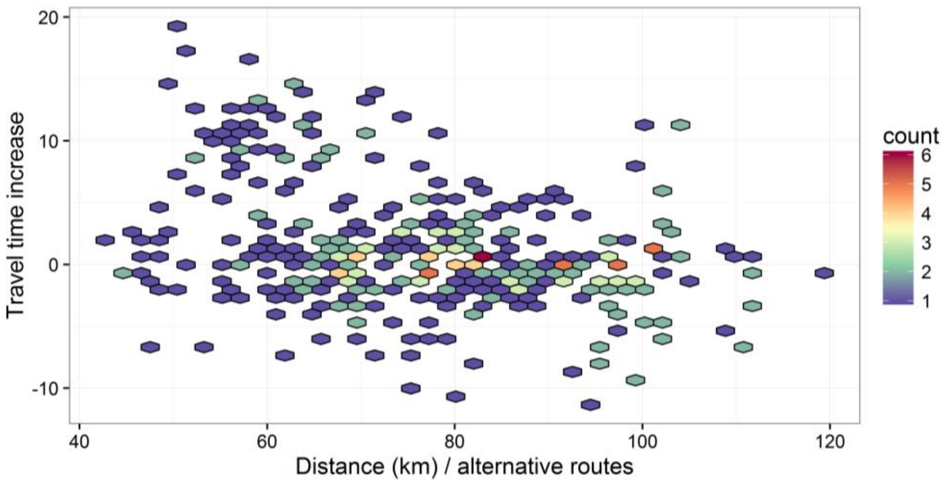


Figure 11.11, the increase in travel time is plotted against the average travel distance.

Figure 11.11 shows us the hours where roadworks caused delays are located mainly when average distances are relatively short. Nevertheless, even at shorter distances some roadworks do cause delays and others do not. This could be because the effect is only present when more people are having to be rerouted, for example. The effect of a single variable could also differ depending on others. To gain insight in the interplay between variables that could explain why sometimes delays are measured while at other times they are not, we will construct a decision tree. A decision tree is valuable because it can look at combinations of factors to elicit the complex structures underlying the reason why some roadworks do and some roadworks do not result in longer travel times. Furthermore, decision trees have a clear structure that is both easy to interpret and easy to depict.

The decision tree will be built using the rpart package in R that builds a decision tree using an algorithm mostly similar to the one described by Breiman et al. (1984; Therneau, Atkinson & Ripley, 2014). Post pruning will be performed to remove the splits from the tree that model noise rather than the true underlying structure. In our case cost complexity pruning is applied. The idea of cost complexity pruning is that splits that add little value at the cost of relatively large complexity, i.e. size, should be removed. Cost complexity pruning asks “What would be the increase in error per leaf if we remove this split?” The tree with the lowest error per leaf is the one chosen.

To create our decision tree we first divided the relevant data in three classes. These are Increase, Neutral, and Decrease, and correspond to the data beneath the red, yellow, and green areas in figure 11.9. Increase stands for an increase in travel time of at least 5 minutes. Decrease implies a decrease in travel time of at least 5 minutes, and Neutral is everything in between. We allowed the tree to choose from all the attributes available in our data, including traffic hindrance class, weather information, day of the week, weekend, and night. The resulting pruned tree is shown below in figure 11.12.

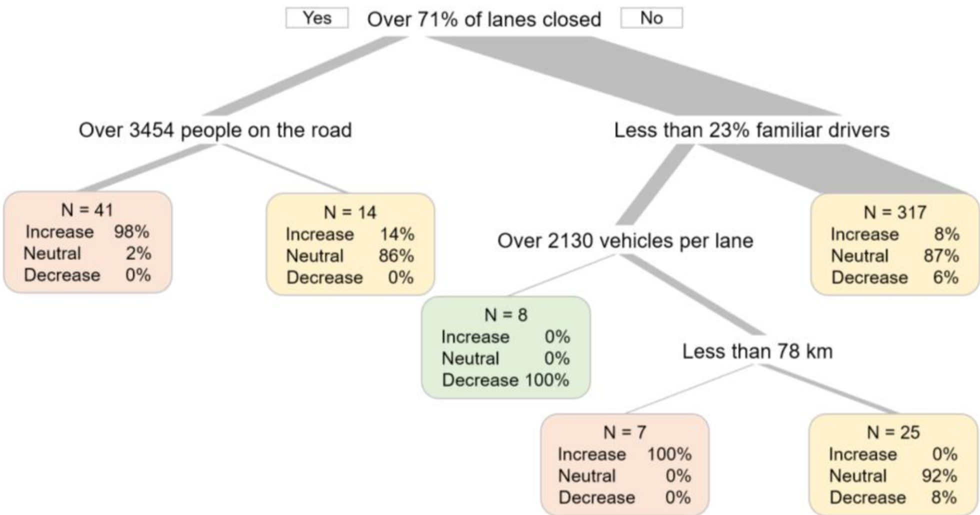


Figure 11.12, decision tree showing when travel times increased (Red), decreased (Green) or stayed neutral (Yellow) when we expect between 1.500 and 4.000 vehicles on the road.

The relations found in figure 11.12 give us several useful, though sometimes predictable, insights in the data. The percentage of lanes closed appears to have most explanatory power in terms of why some roadworks have a large impact. With over 71% of lanes closed we see the impact is often large. In particular, when there are over 3454 people on the road, i.e. when it is busy, there is a large increase in travel times. Otherwise the increase in travel time occurs only occasionally (14% of the measurements). When we have less than 71% of the lanes closed we see the majority of the measurements have little or no impact. In particular, when drivers are familiar with the road we almost never encounter any large increases or decreases in the data. Note driver familiarity is measured here as the square root of the percentage of people with the motive 'work' and ranges from 0 to 1, i.e. 0% to 100%. When drivers are unfamiliar with the road (under 21%) we get some strange findings, though this can also be due to the limited number of observations left. We see decreases in travel times when there are over 2130 vehicles per lane, which goes against all the literature we previously discussed. We, therefore, think this is the result of a tree overfitting the data. The final split relates to distance. As we already saw in figure 11.9, at shorter distances we see large increases in travel times.

At the evaluation we will use a slightly modified version of the model above to predict how much travel times will increase. We remove the split "Over 2130 vehicles per lane" and replace it with the split below about distance. When distances are below 78 km we will predict a large increase in travel times and else we will assign them to the class Neutral. The observations in the Decrease class all fall into the leaf currently at the bottom right (Neutral). The change we make to the tree here is made because we know the relation is highly unlikely to be structural and expect overfitting. When predicting travel time increase we need to do so in minutes rather than classes. We will use a conversion factor for each class to do so based on the class averages. Observations are given an expected increase in travel time of -0.1 minutes and 9.7 minutes for the classes Neutral and Increase, respectively. These are the average travel time increases for the respective classes. Note we will never predict a Decrease and hence also do not require a conversion for this class. All data that is not included in our model, i.e. below 1.500 and above 4.000 vehicles per lane, are assigned an expected increase in travel time of 0.9 minutes.

Travel time reliability

Besides travel time, lane closure can also affect travel time reliability. Similarly to what we did before we start by looking at the fundamental relation between vehicles per lane and the increase in the standard deviation of travel times, i.e. travel time reliability (see figure 11.13).

In figure 11.13 there is no useful relation visible between vehicles per lane and the increase in travel time reliability. The only relation present is the decrease of the bandwidth on which standard deviations are found. The fewer vehicles per lane the larger the variation in the measured increase in travel time reliability is. This relation, however, can be explained by the fact that a lower sample size often results in a higher standard deviation. Moreover, there are more measurements at the low end of the vehicles per lane spectrum making extreme values more likely. The relations found thus appear to be random and not related to vehicles per lane.

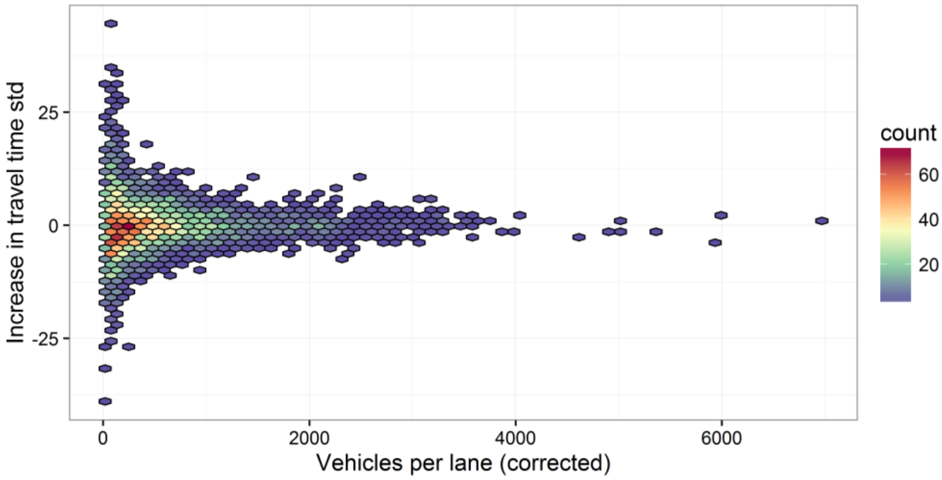


Figure 11.13, increase in travel time standard deviation (std) plotted against the expected number of vehicles per lane, which is corrected for transport demand management and lane closures.

To find out if there is a hidden pattern in the data that might explain why travel time standard deviations are high, neutral, or low, we again made a decision tree. Standard deviations over 5 minutes we label ‘high’, those below 5 minutes ‘low’, and the rest we labelled ‘neutral’. Pruning on this occasion was not necessary. Of all the attributes we allowed the model to split on none could provide any information. Travel time reliability at lane closures is something we cannot predict. Our best guess is thus the average increase in travel time standard deviation, which is 0.2 minutes.

11.4 Evaluation

Here we will test how well our models perform in terms of measuring the impact of a roadwork. There are two scenarios that we have investigated. The first could be applied to predict the impact of road closure, and the second to predict the impact of a roadwork with only part of the road closed.

From our traffic flow model (11.3.1) we learned a decrease of 12% in people found traveling, most probably, due to transport demand management when roadworks are of classes B and C (see 11.3.1). For roadworks labelled as one of those classes we will apply the rule of half for the missing 12%, which is 13.6% of the people we expect on the road. We do not make a distinction between different motives as we did not find any reason to investigate this. We did not see any change greater than 0.5% in the percentage of work trips when comparing the baseline versus the roadwork values for any of the roadwork classes. Hence, we will leave investigating the relation with respect to different motives for future research and assume the 12% decrease is the same amongst all classes.

In section 11.4.1 we will evaluate our model regarding road closure and in section 11.4.2 our model regarding roadworks with lane closure.

11.4.1 Road closure model

We will evaluate the road closure model with respect to hourly vehicle counts and with respect to entire roadworks. On both occasions the data used during training is also the data used for evaluation. The results might thus be overly positive provided we do not correct for overfitting here. We do, however, belief overfitting will not have a large influence here given the simplicity of our road closure model. The road closure model is the sum of both the model that predicts traffic flow, the model for travel time increase during road closure and the model for travel time reliability during road closure. Each of the models is very basic with at most one variable to split on or use in a linear model. The model regarding travel time reliability is even a simple average over all observations. Hence, overfitting would hardly be possible.

We will evaluate the model in two parts. First we discuss how well it can predict costs in time frames of an hour, and, thereafter, we will discuss how well it can predict the total costs of a roadwork.

Hour level

What we already learned from the R^2 of 0.04 is that our model regarding the increase of travel time can hardly explain any variation in the increase in travel times during road closures. Furthermore we are unable to find a significant relation between the chosen variables and travel time reliability. Hence, we will use the mean increase in travel time reliability as our best guess. On hourly basis we thus know our road closure model is unable to successfully explain much of either travel time increase or decrease in travel time reliability. In figure 11.14 we show a scatter plot displaying the relation between the predicted costs and expected costs.

To get to the predicted costs we calculated the expected number of work, business, and other trips by multiplying the predicted number of people on the road (traffic flow model) with the percentage of trips per motive during the baseline. Then we multiplied these values with the respected costs for travel time increase (in hours) and the increase in the standard deviation in travel time (in hours) and the predicted values for both accordingly. The predicted values for both are calculated with our models relating to travel time increase and decrease in travel time reliability. Whenever we found the rule of half should be applied, i.e. for roadworks of class B or C, we added an additional 6.8% to the total costs.

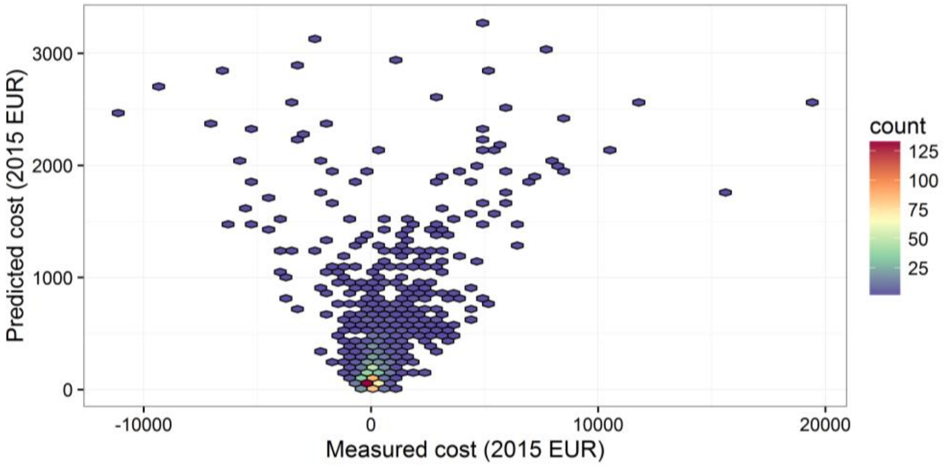


Figure 11.14, measured versus predicted cost per hour for our road closure model.

In figure 11.14 we can see there is only a slight relation between what we predict and what is measured. A correlation of 0.29 does suggest there is somewhat of a correlation. Furthermore, the predicted values appear to be conservative in comparison with the measured values. The range of predicted values goes from 0 to just over 3.000 while the measured cost range from below -10.000 to nearly 20.000 euro. The fact that our model can only explain 8% of the costs of a roadwork does not necessarily imply it is bad.

We will have to put it into context with other prediction models to determine this. Unfortunately, we do not have the information available to do so. The only information available is about the hindrance class and expected delay already in the data. These predictions, however, are generally a gross overestimation of the increase in travel times. Heavy roadworks are often assigned delays of 10 to 30 minutes and occasionally even greater than 30 minutes. These are values we almost never encounter in our measurements. Nevertheless, as it is our only benchmark we will try to get some insight into how well their predictions are when with respect to road closures. To do so, we first convert the classes “below 10 minutes delay”, “10 to 30 minutes delay”, and “over 30 minutes delay” to 0, 1, and 2. We then performed a Spearman correlation to see if higher delay classes also result in higher costs. The answer is rather surprising. We found a Spearman correlation of 0.29, which is equivalent to the correlation of 0.29 we found with our model. These classes, however, are predicted on roadwork rather than hour level so perhaps they will perform better when we evaluate them on that level.

In addition travel delays the roadwork data also has a prediction about the total hindrance. These are the classes E through B that we discussed earlier. The largest hindrance is expected at class A, but roadworks with this class are not present in our data. We perform the same procedure as with expected delay, i.e. E is converted 0, C is 1 and B gets a 2, to again perform a Spearman correlation. The correlation indicated a negative relation (-0.13). We must note, however, only 254 of the 1236 hours with complete road closure have a traffic hindrance class and the comparison is thus not completely fair. Furthermore, based on the traffic hindrance class traffic hindrance alleviating measures will be employed. The prediction is made prior to this. The situation for which the prediction is made thus differs from the measured situation. The actions to prevent the traffic hindrance might cause this result.

Overall, the created model for predicting the impact during hours at which a road is closed perform OK. The model is often wrong, but so is the current state of the art. There is simply a lot of variation that is hard to explain. Future research will have to dive deeper into the problem to improve these predictions.

Roadwork level

Here we will do the same as what we just did, but aggregated to the level of unique roadworks rather than loose hours.

On roadwork level we immediately find much better results. We find our model can explain 46% of the variation in measured costs between roadworks. In figure 11.15 the relation between predicted and measured costs is depicted. We see that we can successfully determine most roadworks have no impact and can also predict when roadworks do. There are some wrong predictions, nonetheless, where we expected an impact and found a negative impact.

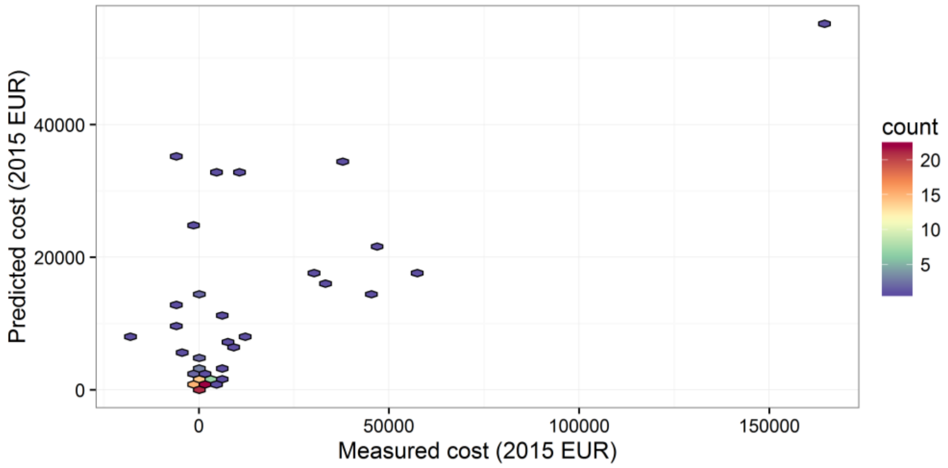


Figure 11.15, measured versus predicted cost for all roadworks where the entire road was closed.

That we find much better results on this level indicates our models do embody the underlying relations resulting the impact of roadworks with road closures. The variation we could not explain on hour level appears to level out when we aggregate over multiple hours to get to roadwork level. This suggests the previous found variation is most likely noise. Moreover, it shows our model is not much influenced by the noise.

We again compare the model with the expected delay and traffic hindrance classes in our roadwork data as we did in the previous section. On roadwork level we get Spearman correlations of 0.35 and -0.03, respectively. If we would apply a Pearson correlation over the expected delay classes, ranging from 0 to 2 for less than 10 minutes delay up to greater than 30 minutes delay, we get to a correlation of 0.53 (R^2 of 0.29). On roadwork level our model appears to strongly outperform the current 'state of the art' though we have to take these results with a grain of salt. We did use the same data for training and evaluation, and the traffic hindrance classes are calculated prior to the traffic alleviating measures, which skews results. Nevertheless, these findings do suggest our model is better than we expected after the relatively poor explanatory power on hour level.

11.4.2 Lane closure model

Similarly to the evaluation of our model on road closure we will evaluate the model for roadworks with lane closure in two parts. First we discuss how well it can predict costs in time frames of an hour, and, thereafter, we will discuss how well it can predict the total costs of a roadwork.

Hour level

On hour level we find our model performs worse than on roadworks with road closures. There is much more variation visible (figure 11.16) and we can only explain 2% of the measured cost. The indications about expected delay and traffic hindrance, however, do not perform any better with Spearman correlations of 0.04 and 0.05, respectively. There is simply a lot of unexplained variance that no-one yet appears to be able to explain.

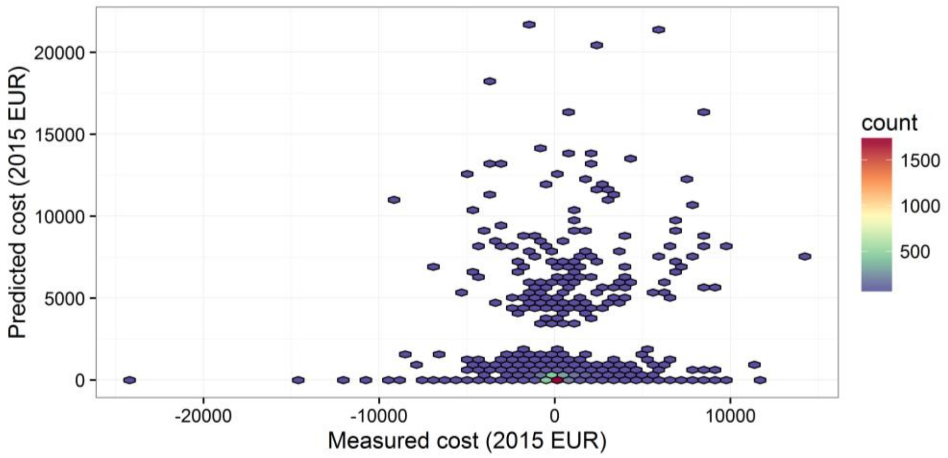


Figure 11.16, measured versus predicted cost for all hours where a lane was closed.

Roadwork level

On roadwork level we see the same pattern as with the road closure model. The noise appears to level out and the variation in the data becomes easier to explain, although the problem maintains non trivial (figure 11.16). Our model can explain 8% of the variation on roadwork level.

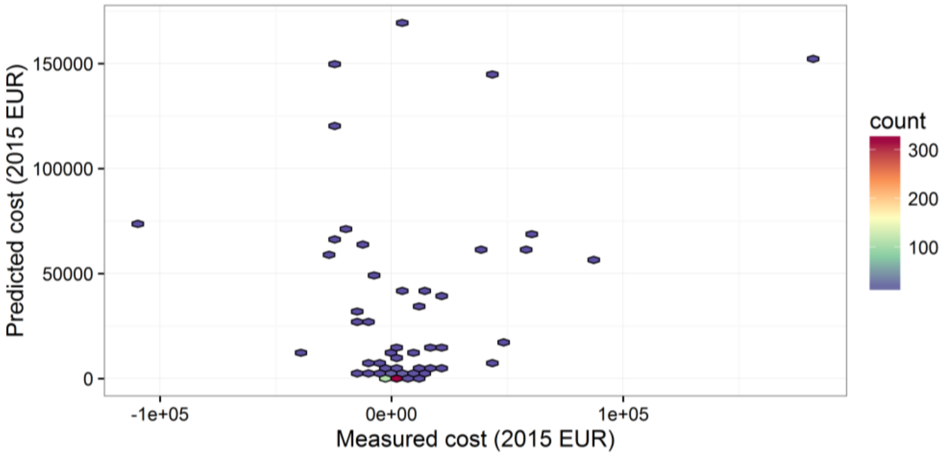


Figure 11.17, measured versus predicted cost for all roadworks where a lane was closed.

On roadwork level also the predictions in the roadwork data about estimated delay and total traffic hindrance start to correlate better with the measured costs. We find Spearman correlations of 0.08 and 0.11, respectively. Our model also performs better than the predictions in the SPIN system, from which we got our roadwork data. Although we again must state the comparison is unfair as we (1) trained our models on the evaluation data, which leads to more optimistic results, and (2) the predictions in the SPIN system do not necessarily anticipate traffic alleviating measures. Furthermore, the predictions in the SPIN system have to look further ahead in the future while our baseline consists of weeks directly surrounding the roadwork.

11.5 Implementation

Here we will discuss what mobile phone data can offer with respect to predicting the impact of roadworks in terms of added benefit over conventional data.

The main advantage, in our opinion, is that we measure travel time regardless of whether people continue to driver over the road affected or take detours. By doing so mobile phone data can provide a much more realistic estimation of the expected travel times. We see travel times increase at heavy roadworks by approximately 10 minutes at the high end of the spectrum whereas the predictions in SPIN are often much higher. This probably results from them focusing primarily on what happens exactly at the roadworks whereas mobile phone data can also provide information about people taking detours whether over larger roads or local roads.

Another benefit of using mobile phone data is the scale and ease at which the data can be gathered. Although this relates more to measuring the impact of roadworks, it also helps when predicting. The more data available the higher the chance you will find the underlying patterns beneath the noise. We found our models for road closures and lane closures to work well, particularly, when we aggregate the predictions to roadwork level, which removes the noise. This is a level the models have not been trained on, and still they are able to find structure in the noise.

We should not, however, overstate the value of the models as there is still a lot of variation left unexplained. In particular, the impact for roadworks with lane closures is very hard to predict.

The model that added most value was the one regarding complete road closure. Here the final model, though perhaps over-fitted despite its simplicity, could explain 45% of the measured costs. This is nearly twice as much as the classifications about expected delay currently in the SPIN, i.e. the system where information about roadworks in the Netherlands is stored. We think the situation with complete road closures is particularly interesting in combination with mobile phone data. Unlike traditional techniques, e.g. using roadside measurements, it is easy to track whoever usually uses the road and keep track of travel times using mobile phone data. Mobile phone data can also provide more insight in the origins and destinations of the people affected which helps when you want to predict the availability of detour roads. We did not take this into account, but future research might. There is still a lot to explain regarding what causes the impact of roadworks.

The model regarding road closure we advise people to start using as we found strong indications the model provides better estimations than the current 'state of the art'. The model regarding lane closure we do not recommend. Here it would be valuable to have more details about the roads and what specifically is going on there to improve predictions. Predictions results were bad and we reckon an expert might do better here.

11.6 Conclusion

The main thing we learned in this chapter is that predicting the impact of roadworks is non-trivial. There is a lot of variation in the data and even when we look at the impact from many angles.

We started with a literature study where we found a number of key attributes relating to the impact of roadworks. Most of the literature focussed on what reduces the road capacity and hence increases the chance on traffic jams. Many of the attributes, such as lane width reduction and rain, were never highlighted as important in relation to the impact of roadworks. These effects relate mostly to when road capacity limits are reached, which we found is not very common. The relations that are found to reduce capacity by a few percent have only a minor influence. This influence was too small to be noticed by any model that we created. What we did find to be of greater importance are (1) transport demand management, (2) distance travelled which relates to the number of alternative routes available, and (3) driver familiarity which relates to being able to find alternative routes as well as increase capacity (Berkum & Huerne, 2014; Heaslip et al., 2007). A number of key attributes such as the percentage of HGV are missing in our dataset. Future research could include these to find out if they can explain the unexplained variance in the measured costs of roadworks.

On hour level the costs are extremely difficult to predict. On this level one accident during the baseline or any event really that influences travel behaviour or road characteristics can induce variation. If there is enough data, however, it is possible to look through the noise and find underlying structures when enough data is available. On hour level neither of our models, i.e. the one for complete road closures and the one for roadworks with lane closures, could explain much if any of the variation in the measured costs. When we aggregated the hours per roadwork we found we could predict up to 45% of the variation in the measured costs for our road closure model.

In total we created three model: a traffic flow model, a road closure model, and a lane closure model. The first predicts the decrease of people on the road. Note we already corrected for external factors such as weather (see chapter 9). The effects that are left we are trying to estimate here. In our traffic flow model we found roadworks with classes B and C, which stand for high impact roadworks, saw a decrease in vehicles on the road by an average of 12% (Taskforce Doorstroming, 2009). This is most likely the result of traffic alleviating measures being applied. Other factors were unimportant in comparison. When roadworks were of lower classes we did not find a decrease in traffic flow. With respect to the other models we only found relations explaining increases in travel time. Travel time reliability could not be explained by any of the many variables in our data. Driver familiarity, percentage of lanes closed and vehicles per lane were crucial in the lane closure model while the road closure model only looked at driver familiarity and a standard average delay.

The mobile phone data shows it is a valuable source of. The main advantages are being able to track people no matter how they travel and the size that helps to look through the noise.

12 Conclusion

In this research we set out to find a new rich and scalable source of information to find out how roadworks affect mobility. Traditional techniques such as surveys and roadside measurements are both non-scalable small in sample and require a lot of effort to elicit mobility patterns. We suggest mobile phone data as a rich and scalable source of information to do the same. Where traditional techniques would require months of research to measure the impact of one roadwork; we showed that mobile phone data can measure hundreds with a fraction of the time and effort.

Mobile phone data is a rich and scalable source of information that can be used to measure patterns in mobility. This we showed by correctly scaling our sample to the traveling population and assigning trips to the road. Comparisons with a hundred road side measurement devices on various highways across the country have been made. The results clearly show mobile phone data can be used to get accurate vehicle counts on an hourly basis. Pearson correlations of 0.9 and upwards are found for 90% of the comparison sites. In addition to measuring traffic counts, mobile phone data can also be enriched. For example, as the impact of roadworks differs per motive this is crucial information to have when evaluating their impact. We, therefore, created a probability estimation tree using trip characteristics such as departure time and home location that help explain why a person travels. By carefully selecting and constructing attributes that are available in both the mobile phone data and the survey data we could transpose our model to the mobile phone data. The created model helps us to better predict the composition of vehicles traveling over a road by motive and thus get a better image of the impact of roadworks.

To measure the impact of roadworks we carefully craft our dataset from the mobile phone data. We, furthermore, correct for variations caused by external factors such as weather to ensure we can accurately evaluate the impact of roadworks. After selecting a proper baseline free of externalities we calculated the impact of 638 roadworks on Dutch highways in October and September 2015. We found a net impact of €1.109.548. This is excluding the roadworks that happened week after week on the same road for which we could not establish a baseline, which was just over half of all roadworks.

Knowing what the impact was is one aspect of this study; predicting the impact is the other. After constructing a list from literature of key roadwork characteristics that might influence roadworks we created three models. One to predict how traffic flow in absolute counts changes due to roadworks. Here transport demand management turns out to be key. One model is created for predicting the increases in travel times and decrease in travel time reliability due to road closure. Here we find driver familiarity is most important. Compared to the current state of the art we find strong indications our basic model could outperform the current 'state of the art'. Our final model regards roadworks with lane closures. After extensive modelling we found the well-known structures underlying the impact of these roadworks: high percentage of lanes closed, busy roads, and unfamiliar drivers.

We will round of this conclusion with a clear answer to our main research question:

How can mobile phone data be used to improve the measurement and prediction of the impact of roadworks on highways?

To improve how we measure the impact of roadworks we should use mobile phone data. In this research we devised and validated our proposed method describing how to use mobile phone data specifically for this task (see 4.2.2). The method consists of the following five steps: (1) accurately scale to the traveling population, (2) select the people who are affected by the roadwork, (3) determine their motive, (4) compare the roadworks with a proper baseline corrected for external factors, and (5) report the findings in a meaningful way. Using this method mobile phone data can measure the impact of roadworks fast, at high quality and low costs. The information is far richer than that from road side measurement of which the only advantage is that they provide basic information quickly. The information is quicker available than with surveys, and at much larger scale. Next to this the information can be produced at low cost with the infrastructure already in place and nearly no difference in effort between measuring the impact of ten versus a hundred roadworks (see figure 12.1). Mobile phone data can improve the measurement of the impact of roadworks on highways on every meaningful front.

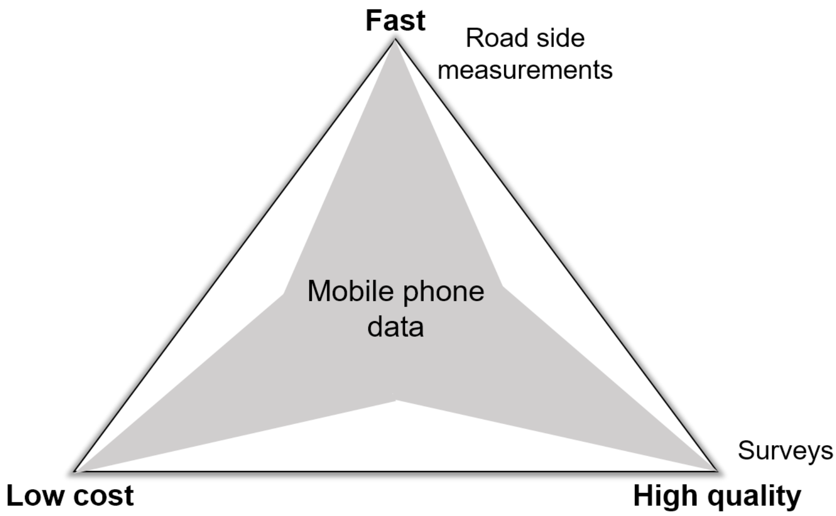


Figure 12.1, mobile phone data is able to combine high quality, quick results, at low costs. Surveys and road side measurements can only do one each.

In terms of predicting the impact of roadworks the main strength of the mobile phone data is in the abundance of information. The abundance of information helps to see through the noise and discover the hidden underlying structures. Predicting the impact of roadworks remains a nontrivial task. Nevertheless, we are confident this new source of information will allow others to discover more about the hidden effects resulting in the impact of roadworks.

13 Discussion

In this chapter we will reflect on the limitations, provide a brief overview of how the research can advance the work of others in the field and provide input for future research.

13.1 Limitations

Most of the limitations and side notes to our findings have been elaborately discussed throughout the research. There are two, however, we find deserve special attention as they are in the foundations of this research.

Semantics

When we measure the impact of a roadwork we measure, foremost, the increase in travel times and travel time reliability of the people logically traveling over the road where the roadwork occurs.

Good traffic management can successfully reduce the impact of roadwork. When traffic management is properly applied we measure the impact of closing, for example, a lane in addition to the effect of traffic management. We cannot distinguish between the two. The impact of roadworks in this research is the impact that results from both.

Because we measure both effects it is difficult to distinguish the effect of each, e.g. what would the impact be without the traffic alleviating measures. To research this actual experiments should be set up with some roadworks having these measures and some roadworks without these measures.

Detours

Traditionally the rule of half is applied to everyone who decides to change travel behaviour (Eigenraam et al., 2000). Taking a detour is changing travel behaviour as well. The motivation for this argument in line with the rule of half is that these people also found something more optimal than traveling over the road where the roadworks are occurring (expert interview, Appendix C).

We are currently unable to distinguish between travellers that stay on the logical route and those taking detours. Hence, we also do not apply the rule of half to the group taking a detour. In part that is because we simply cannot make the distinction. Mostly though, we think the rule of half should not be applied to people taking a detour. Take, for example, the case where a road is completely closed during roadworks. Everyone will have to take a detour. With the rule of half the impact would thus be half as well. Furthermore, from a personal stand point I do not see the difference between a 5 minute delay stuck in traffic or a 5 minute delay because I try to avoid a roadwork, both are a direct effect of the roadwork and both cost me just as much delay. If anything taking a detour should be weight more heavily considering the potential extra mileage to the vehicle.

Not being able to notice who takes a detour is a current limitation. With respect to the impact of roadworks, as we just argued, we do not see this as an issue.

13.2 Scientific relevance

This research can provide input for research in the field of human mobility research using mobile phone data and traffic management.

Human mobility research

The body of literature regarding mobility research using mobile phone data. Numerous researchers have found the benefit of measuring mobility with mobile phones, e.g. unobtrusive and large scale (Daas et al., 2009; Snijkers, 2009; Ahas et al., 2008; Eagle et al., 2009; Becker et al., 2011; Palchykovet al., 2012; Wang et al., 2012).

In this research we have showed how to increase the accuracy by, for one, reducing the maximum allowed cell sizes (chapter 5). We, furthermore, validated we can elicit more accurate results when we apply the threshold in combination with the described method to elicit origins and destinations from Call Detail Records (CDRs). In addition, we proposed a new scaling method that scales to the traveling rather than general population (chapter 6) in collaboration with Van Langen (2016). To our knowledge this research also contains the first evidence we can correctly scale to the population and use mobile phone data to infer accurate vehicle counts on major highways. We also provided a prime example of how surveys can enrich mobile phone data (chapter 8).

The results from this research such as the proposed and validated scaling factor could directly be implemented in other research to get, for example, a good representation of the traveling population rather than just the sample.

Traffic management

The largest gains are for the field of traffic management. We proposed a new source of rich and plentiful information that can drive research for years to come. More importantly though, we showed throughout this research that mobile phone data can be used to measure travel behaviour and thus changes in travel behaviour on major roads.

Similar to what we did with roadworks could be done with any type of event in the road network. One could measure, for example, the effects of traffic alleviating measures, e.g. a congestion tax, on country level with relatively little effort. Conventional techniques, e.g. road side measurements, can also provide some information. The advantage of mobile phone data is the additional information that can be linked to the traffic counts, e.g. where people come from, where they go, and now also their motive of traveling. This additional information could prove valuable in explaining the observed travel behaviour.

The main benefit to traffic management has been showing and further validating that mobile phone data can be used to measure travel behaviour on the roads. We are not the first to show this, but we did provide more confirmation that mobile phone data can be used for research in this field (Wang et al., 2012; Toole et al., 2015).

13.3 Future research

Over the course of this research we encountered a number of points that we regard as worthy for future research.

Benefit of a more advanced route assignment method

In chapter 7 we already extensively discussed the advantage and disadvantages of different route assignment methods. We applied a fairly basic one, i.e. shortest time path algorithm, and got more than decent results. Nevertheless, more advanced methods are available that have been shown to produce even better results (Prato, 2009). Hence, it may be worthwhile to implement one of these more advanced methods and compare the results with the baseline we established. The better we can detect what roads people are using, the better we can focus on the people influenced by events on a specific road.

Predicting the impact of roadworks with more information

Key information could still be added to the roadwork information to improve the predictions regarding the impact of roadworks. Information about Heavy Goods Vehicles (HGVs), for example, was lacking, but is found to be an important factor according to literature (Heaslip et al., 2008). In our brief outlier analysis in chapter 11 we also saw the supporting infrastructure in close proximity around roadworks might explain why some large roadworks have almost no impact.

With more information, more can hopefully be explained regarding how roadworks impact society.

Overestimating the value of traffic management

Tenders are won with discounts for good traffic hindrance plans of about €30.8 million, on tenders of €100 million (Duijnisveld et al., 2011). This is a significant sum of money to reduce the impact of roadwork. Although we cannot say how much the impact would be without traffic alleviating measures; we do know the net impact of 7.552 hours of roadworks during September and October 2015 is just over a thirtieth of the fictive discount.

As discussed, we cannot say with certainty traffic hindrance plans are overvalued. Nevertheless, given the large divide between the €30.8 million discount and the €1.1 million impact over 7.552 hours we guess it is worth an investigation.

Bibliography

Aarts, L., & Van Schagen, I. (2006). Driving speed and the risk of road crashes: A review. *Accident Analysis & Prevention, 38*(2), 215-224.

Ahas, R., Aasa, A., Roose, A., Mark, U., & Silm, S. (2008). Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. *Tourism Management, 29* (3), 469–486.

Alexander, L., Jiang, S., Murga, M., & González, M. C. (2015). Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies, 58*, 240-250.

Al-Kaisy, A. F., & Hall, F. L. (2000). Effect of darkness on the capacity of long-term freeway reconstruction zones. Transportation Research Circular E-C018, 164-175.

Astarita, V., & Florian, M. (2001). The use of mobile phones in traffic management and control. In *Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE* (pp. 10-15). IEEE.

Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the operational research society, 54*(6), 627-635.

Becker, R. A., C´aceres, R., Hanson, K., Loh, J. M., Urbanek, S., Varshavsky, A., & Volinsky, C. (2011). A tale of one city: Using cellular network data for urban planning. *IEEE Pervasive Computing*, 10 (4), 18–26.

Benekohal, R. F., Kaja-Mohideen, A.-Z., & Chitturi, M. V. (2003).Evaluation of constructionwork zone operational issues: Capacity, queue, and delay (Tech. Rep.).Illinois Trans-portation Research Center.

Berwick, D. M. (1996). A primer on leading the improvement of systems. *BMJ: British Medical Journal*, *312*(7031), 619-622.

Bolla, R., Davoli, F., & Giordano, A. (2000, November). Estimating road traffic parameters from mobile communications. In *Proceedings 7th World Congress on ITS, Turin, Italy*.

Borchardt, D.W., Pesti, G., Sun, D., & Ding, L. (2009). *Capacity and road user cost analysis of selected freeway work zones in Texas* (No. FHWA/TX-09/0-5619-1). Texas Transportation Institute, Texas A&M University System.

Breiman L., Friedman J. H., Olshen R. A., and Stone, C. J. (1984) *Classification and Regression Trees*. Wadsworth.

Cáceres, N., Wideberg, J. P., & Benitez, F. G. (2007). Deriving origin destination data from a mobile phone network. *Intelligent Transport Systems, IET*, 1(1), 15-26.

Calvert, S. C. (2010). *A-Priori Travel Time Predictor for Long Term Roadworks on Motorways.* Doctoral dissertation, TU Delft, Delft University of Technology.

Carr, D., Lewin-Koh, N., Maechler, M., & Sarkar, D. (2015). *hexbin: Hexagonal binning routines* [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=hexbin(R package version 1.27.1)

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate behavioral research, 1*(2), 245-276.

CBS. (n.d.). Stedelijkheid (van een gebied). Retrieved October 5, 2015, from http://www.cbs.nl/nl-NL/menu/methoden/begrippen/default.htm?ConceptID=658

Centraal Bureau Statistiek (CBS). (2011). Onderzoek verplaatsingen in Nederland 2010. Retrieved from http://www.cbs.nl/NR/rdonlyres/D1D6A7E6-A1C9-4329-9908-70EA47E1D8F6/0/2010ovinonderzoeksbeschrijvingart.pdf

Centraal Bureau Statistiek (CBS). (2014b, August 26). *Bevolking; kerncijfers*. Retrieved June 30, 2015, from http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=37296ned&D1=a&D2=0,10,20,30,40,50,60,(l-1),l&HD=130605-0924&HDR=G1&STB=T

Centraal Bureau Statistiek (CBS). (2014c). *Plausibiliteitsanalyse gereviseerde OVIN bestanden 2011 en 2012*. Retrieved from http://www.cbs.nl/NR/rdonlyres/215BB84A-2869-4C1E-A673-9C0FAA62C315/0/2014ovinplausibiliteitsanalyse.pdf

Centraal Bureau voor de Statistiek (CBS). (2014a). *Onderzoek Verplaatsingen in Nederland 2014 - OViN 2014. DANS*. http://dx.doi.org/10.17026/dans-x95-5p7y. (Retrieved: 2015-10-02)

Centraal Bureau voor de Statistiek (CBS). (2015a). *Begrippen; Stedelijkheid van een gebied.* Retrieved from: http://www.cbs.nl/nl-NL/menu/methoden/begrippen/default.htm?ConceptID=658

Centraal Plan Bureau (CPB) (2014). *Macro Economische Verkenning 2015.* Retrieved from https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2014/09/16/macro-economische-verkenningen-2015/mev-2015.pdf

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). Crisp-dm 1.0 step-by-step data mining guide.

Cochrane, D., & Orcutt, G. H. (1949). Application of least squares regression to relationships containing auto-correlated error terms. *Journal of the American Statistical Association*, *44*(245), 32-61.

Cools, M., Moons, E., & Wets, G. (2010). Assessing the impact of weather on traffic intensity. *Weather, Climate, and Society*, *2*(1), 60-68.

Crown Commercial Services (2015) *A Brief Guide to the EU Public Contracts Directive (2014).* Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/449295/A_Brief_Guide_to_the_EU_Public_Contract_Directive_2014_-_July_2015.pdf

Daas, P., Roos, M., de Blois, C., Hoekstra, R., ten Bosch, O., & Ma, Y. (2011). New data sources for statistics: experiences at statistics Netherlands.

De Koning, J. (2015, March 26). *Nog één jaar te gaan*. Retrieved from Cobouw: http://www.cobouw.nl/nieuws/aanbesteden/2015/03/27/nog-een-jaar-te-gaan

De Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific reports, 3.*

Dietterich, T. G. (1996). Statistical tests for comparing supervised classification learning algorithms. In *Oregon State University Technical Report* (Vol. 1, pp. 1-24).

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische mathematik*, *1*(1), 269-271.

Dufour, J. M., Gaudry, M. J., & Liem, T. C. (1980). The Cochrane-Orcutt procedure numerical examples of multiple admissible minima. *Economics Letters*, *6*(1), 43-48.

Duijnisveld, M. A. G., Peijs, L. & Calvert, S. C. (2012). Sturen op verkeershinder op het tracé Schiphol-Amsterdam-Almere. *CROW Infradagen. Thema 1: Aanbesteding en contractvormen, 1-12*.

Eijgenraam, C.J.J., Koopmans, C.C., Tang, P.J.G., Verster, A.C.P. (2000). *Evaluatie van Infrastructuurprojecten; Leidraad voor Kosten-Batenanalyse.* Retrieved from http://www.mkba-informatie.nl/index.php/download_file/force/108/195/

Esposito, F., Malerba, D., Semeraro, G., & Kay, J. A. (1997). A comparative analysis of methods for pruning decision trees. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *19*(5), 476-491.

Federal Highway Administration (2003). *Full Road Closure for Work Zone Operations - A Cross-Cutting Study.* Retrieved from http://www.ops.fhwa.dot.gov/wz/construction/full_rd_closures.htm

Federal Highway Administration. (2003). *Full Road Closure for Work Zone Operations - A Cross-Cutting Study.* Retrieved from http://www.ops.fhwa.dot.gov/wz/construction/full_rd_closures.htm

Ferri, C., Flach, P. A., & Hernández-Orallo, J. (2003). Improving the AUC of probabilistic estimation trees. *Machine Learning: ECML 2003* (pp. 121-132). Springer Berlin Heidelberg.

Field, A., Miles, J. & Field, Z. (2012). *Discovering statistics using R*. London Thousand Oaks, Calif: Sage.

Fitzpatrick, K., Carlson, P., Brewer, M., & Wooldridge, M. (2001). Design factors that affect driver speed on suburban streets. *Transportation Research Record: Journal of the Transportation Research Board*, (1751), 18-25.

Geerts, J. (2014). *Ophogingsmethodiek van devices naar personen.* (Unpublished internal document)

Goemans, J.W., Daamen, W., & Heikoop, H. (2011, November). Handboek Capaciteitswaarden Infrastructuur Autosnelwegen (CIA) Volledig Vernieuwd. In *Nationaal Verkeerskundecongress, Nieuwegijn, Nederland*.

Gregor, S. (2006). The nature of theory in information systems. MIS quarterly, 611-642.

Hartigan, J. A. & Wong, M. A. (1979). *A K-means clustering algorithm*. Applied Statistics 28, 100–108.

HCM (2000). Highway Capacity Manual. Transportation Research Board, National Research Council, Washington, D.C., U.S.A.

Hijmans, R. J. (2015). *geosphere: Spherical trigonometry* [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=geosphere(R package version1.5-1)

In de Vegte, N.K., & Van Toorenburg, J. (2014, November 12). *Invloed van invoegstroom en wegvaklengte op capaciteit.* Paper presented at Verkeerskundecongres 2014, Utrecht.

Iqbal, M. S., Choudhury, C. F., Wang, P., & González, M. C. (2014). Development of origin–destination matrices using mobile phone call data. Transportation Research Part C: Emerging Technologies, 40, 63-74.

ITS International. (2010). Cost-effective alternatives to traditional loops. Retrieved September 21, 2015, from http://www.itsinternational.com/sections/cost-benefit-analysis/features/cost-effective-alternatives-to-traditional-loops/

Keij, J. (2014). *Smart phone counting: Location-Based Applications using mobile phone location data* (Unpublished master's thesis). Delft University of Technology, Delft, The Netherlands.

Kennisinstituut voor Mobiliteitsbeleid. (2013). *The social value of shorter and more reliable travel times*. Den Haag, Ministerie van Infrastructuur en Milieu: Publisher. Retrieved from http://www.kimnet.nl/sites/kimnet.nl/files/the-social-value-of-shorter-and-more-reliable-travel-times_reprint.pdf

KNMI. (2015, December 3). Uurgegevens van het weer in Nederland. Retrieved December 4, 2015, from https://www.knmi.nl/nederland-nu/klimatologie/uurgegevens

Knockaert, J., Tseng, Y. Y., Verhoef, E. T., & Rouwendal, J. (2012). The Spitsmijden experiment: A reward to battle congestion. *Transport Policy, 24,* 260-272.

Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *International Joint Conference on Artificial Intelligence, 14*(12), 1137–1143. http://doi.org/10.1067/mod.2000.109031

Kuhn, M. (2016). caret: Classication and regression training [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=caret(R package version 6.0-64)

Landenportal. (n.d.). Schoolvakanties in 2015-2016 per regio. Retrieved March 21, 2016, from http://www.landenportal.nl/informatie/schoolvakanties-2015-2016.htm

Lertworawanich, P., & Elefteriadou, L. (2003). A methodology for estimating capacity at ramp weaves based on gap acceptance and linear optimization. *Transportation Research Part B: Methodological, 37*(5), 459-483.

Marb´an, O., Mariscal, G., & Segovia, J. (2009). A data mining & knowledge discovery process ´ model. *Data Mining and Knowledge Discovery in Real Life Applications. IN-TECH* , 2009 , 8.

Maze, T., Schrock, S., & Kamyab, a. (2000). Capacity of freeway work zone lane closures. *Mid-Continent Transportation Symposium 2000 Proceedings*, 178–183. Retrieved from http://ntl.bts.gov/lib/9000/9500/9563/Maze.pdf

Middleton, D., & Parker, R. (2002). *Vehicle detector evaluation* (No. FHWA/TX-03/2119-1,). Retrieved from: http://d2dtl5nnlpfr0r.cloudfront.net/tti.tamu.edu/documents/2119-1.pdf

Minchin Jr, R. E., Thurn, S. B., Ellis Jr, R. D., & Lewis, D. W. (2013). Using contractor bid amounts to estimate the impact of night construction on cost for transportation construction. *Journal of Construction Engineering and Management, 139(8*), 1055-1062.

Ministerie van Infrastructuur en Milieu & Rijkswaterstaat (2015, May 22). Werkzaamheden op wegnummer. Van A Naar Beter. Retrieved from http://www.vananaarbeter.nl/werkzaamheden/wegwerkzaamheden

Ministerie van Infrastructuur en Milieu. (2015). *Publieksrapportage Rijkswegennet*. Retrieved from rijksoverheid: http://www.rijksoverheid.nl/documenten-en-publicaties/rapporten/2015/02/12/publieksrapportage-rijkswegennet.html

Ministerie van Verkeer en Waterstaat & Centraal Planbureau (2004). *Directe Effecten Infrastructuurprojecten: Aanvulling op de Leidraad OEI.* Retrieved from http://www.mkba-informatie.nl/index.php/download_file/force/109/279/

Moen, R., & Norman, C. (2006). Evolution of the PDCA Cycle. *Society*, 1–11.

Nagel, K., & Schreckenberg, M. (1992). *A cellular automaton model for freeway traffic*. Journal de physique I, 2(12), 2221-2229.

Nationaal Databank Wegverkeersgegevens (NDW). (2015a). 201411281300_MST_NDW01_MT_662.xml. Retrieved from http://www.ndw.nu/documenten/nl

Nationaal Databank Wegverkeersgegevens (NDW). (2015b). Datalevering. Retrieved from http://www.ndw.nu/pagina/nl/103/datalevering/

Nationaal Georegister. (2016, January 1). Retrieved January 10, 2015, from http://www.nationaalgeoregister.nl/geonetwork/srv/dut/search#|a9b7026e-0a81-4813-93bd-ba49e6f28502

NDW. (2015). *MST - Dit bestand geeft aan waar de NDW meetlocaties zich bevinden.* Nationaal Databank Wegverkeersgegevens. Retrieved from http://www.ndw.nu/documenten/

Niculescu-Mizil, A., & Caruana, R. (2005, August). Predicting good probabilities with supervised learning. *In Proceedings of the 22nd international conference on Machine learning (pp. 625-632). ACM.*

Nihan, N. L., Wang, Y., & Zhang, X. (2002). *Evaluation of dual-loop data accuracy using video ground truth data* (No. WA-RD 535.1,). Transportation Northwest, Department of Civil Engineering, University of Washington.

Ofcom. (2014, October). *Children and parents: Media use and attitudes report 2014* (Tech. Rep.). Southwark, London.

Offermans, M., Priem, A., & Tennekes, M. (2013). Rapportage project impact ict; mobiele telefonie. *Programma Impact ICT, 1* (9).

OpenStreetMap contributors. (2014) Planet dump [Data file from n.a.]. Retrieved from http://planet.openstreetmap.org

Ortúzar, J. D. D., & Willumsen, L. G. (2011). *Modelling Transport*. *Modelling Transport*. http://doi.org/10.1002/9781119993308

Otten, M.B.J., ‘t Hoen, M.J.J., Den Boer, L.C., (2014). STREAM personenvervoer 2014 VERSIE 1.1. Retreived from: http://www.ce.nl/?go=home.downloadPub&id=1478&file=CE_Delft_4787_STREAM_personenvervoer_2014_1.1_DEF.pdf

Oxford dictionaries. (n.d.). In Oxford dictionaries, Retrieved September 24, 2015, from http://www.oxforddictionaries.com/definition/english/data-mining

Palchykov, V., Kaski, K., Kert´esz, J., Barab´asi, A.-L., & Dunbar, R. I. (2012). Sex differences in intimate relationships. *Scientific reports, 2. [A REFERENCE BY JENS – MIGHT BE A MISTAKE LOOKING AT THE TITLE]*

Park, B., Won, J., & Yun, I (2005). *Application of Microscopic Simulation Model Calibration and Validation Procedure: A Case Study of Coordinated Actuated Signal System*. Department of Civil Engineering, University of Virginia

Parry, I. W. H. (2002). Comparing the efficiency of alternative policies for reducing traffic congestion. *Journal of public economics, 85*(3), 333-362.

Pradhan, B. (2013). A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Computers & Geosciences*, *51*, 350-365.

Prato, C. G. (2009). Route choice modeling: past, present and future research directions. *Journal of Choice Modelling*, *2*(1), 65-100.

Provost, F., & Domingos, P. (2000). Well-trained pets: Improving probability estimation trees.

Rijksoverheid. (n.d.). Welke regio's zijn er voor de spreiding van de schoolvakanties? Retrieved October 16, 2015, from https://www.rijksoverheid.nl/onderwerpen/schoolvakanties/vraag-en-antwoord/welke-regio-s-zijn-er-voor-de-spreiding-van-de-schoolvakanties

Rijkswaterstaat. (2008). *Evaluatie verkeershinderbeperkende maatregelen Moerdijkbrug*. Rijkswaterstaat IJsselmeergebied afdeling WVV: Author. Retrieved from http://epomm.eu/maxeva/uploads/Evaluatie_Moerdijkbrug.pdf

Rijkswaterstaat. (2009a). *Verkeerskundige evaluatie werkzaamheden Ketelbrug A6*. Unpublished manuscript.

Rijkswaterstaat. (2009b). *Evaluatie Mobiliteitsbeïnvloedende maatregelen A6 Hollandse Brug*. Rijkswaterstaat IJsselmeergebied afdeling WVV: Author. Retrieved from http://epomm.eu/maxeva/uploads/EvaluatieA6HollandsebrugHoofdrapport.pdf

Rijkswaterstaat. (n.d.). 05 Systeem Planning en Informatie Nederland (SPIN). Retrieved October 17, 2015, from https://www.rijkswaterstaat.nl/zakelijk/verkeersmanagement/wegverkeer/spin.aspx

Schrank, D., Eisele, B., & Lomax, T. (2012). TTI’s 2012 urban mobility report.*Texas A&M Transportation Institute. The Texas A&M University System*.

Snijkers, G. (2009). Getting data for (business) statistics: What’s new? What’s next. In *European conference for new techniques and technologies for statistics (ntts)* (pp. 18–20).

Sokovic, M., Pavletic, D., & Pipan, K. K. (2010). Quality improvement methodologies–PDCA cycle, RADAR matrix, DMAIC and DFSS. *Journal of Achievements in Materials and Manufacturing Engineering*, *43*(1), 476-483.

Spitsmijden. (2009). *The effects of rewards in Spitsmijden 2: How can drivers be persuaded to avoid peak periods.* Retrieved from http://www.spitsmijden.nl/resultaten/rresultaten2a/R apport%20The%20effects%20of%20rewards%20in%20Spitsmijden%202%20okt09.pdf

Stern, A. D., Shah, V., Goodwin, L., & Pisano, P. (2003, January). Analysis of weather impacts on traffic flow in metropolitan Washington DC. In *Proceedings of the 19th International Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology.*

Sugiyama, Y., Fukui, M., Kikuchi, M., Hasebe, K., Nakayama, A., Nishinari, K., Tadaki, S., & Yukawa, S. (2008). Traffic jams without bottlenecks experimental evidence for the physical mechanism of the formation of a jam. *New Journal of Physics,* 10 (3), 033001.

Taskforce Doorstroming. (2009). *Handreiking Emvi Deel 1*. Retrieved (June 18th, 2015) from http://www.halftime.nl/wp-content/uploads/2014/01/HT_Li_MH01-HandreikingEMVI.pdf

Technische Universiteit Delft, Faculteit Civiele Techniek en Geowetenschappen, Afdeling Transport & Planning. (2006). Evaluatie en Analyse van Reisinformatie. Retrieved from http://rijkswaterstaat.nl/rws/corporate/Wegen%20naar%20de%20toekomst/floating%20car%20data%20-%20evaluatie%20en%20analyse%20van%20reisinformatie.pdf

Telecompaper. (2015, Q3). *Smartphone penetration Netherlands 2015 q3* (Tech. Rep.). Southwark, London.

Therneau, T., Atkinson, B., & Ripley, B. (2014). rpart: *Recursive partitioning and regres-sion trees* [Computer software manual]. Retrieved from http://CRAN.R-project.org/package=rpart(R package version 4.1-8)

Toole, J. L., Colak, S., Sturt, B., Alexander, L. P., Evsukoff, A., & González, M. C. (2015). The path most traveled: Travel demand estimation using big data resources. *Transportation Research Part C: Emerging Technologies, 58*, 162-177.

Tu, H. (2008). *Monitoring travel time reliability on freeways*. TU Delft, Delft University of Technology.

Unrau, D., & Andrey, J. (2006). Driver response to rainfall on urban expressways. *Transportation Research Record: Journal of the Transportation Research Board,* (1980), 24-30.

Van de Weerd, I., & Brinkkemper, S. (2008). Meta-modeling for situational analysis and design methods. *Handbook of research on modern systems analysis and design technologies and applications, 35.*

Van Kats, J. (2014). *Supporting Incident Management with Population Information Derived from Telecom Data* (Unpublished master's thesis). Utrecht University, Utrecht, The Netherlands.

Van Langen, J.J. (2016). The attractiveness of cities (Unpublished master's thesis). Utrecht University, the Netherlands.

Van Rossum, G., & Drake Jr, F. L. (1995). *Python reference manual*. Amsterdam: Centrum voor Wiskunde en Informatica.

Verkeersmonitoring met inductielussen. (n.d.). Retrieved September 21, 2015, from http://verkeer.wikia.com/wiki/Verkeersmonitoring_met_inductielussen#Bronnen

Wang, P., Hunter, T., Bayen, A. M., Schechtner, K., & González, M. C. (2012). Understanding road usage patterns in urban areas. *Scientific reports, 2*.

Wanvik, P. O. (2009). Effects of road lighting: An analysis based on Dutch accident statistics 1987–2006. *Accident Analysis & Prevention*, *41*(1), 123–128. http://doi.org/10.1016/j.aap.2008.10.003

Wickham, H. (2009). ggplot2: *Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from http://ggplot2.org

Wirth, R., & Hipp, J. (2000). Crisp-dm: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (pp. 29–39).

Appendix A: Interview with Mark van Dord

Table 0.1, meta-data about the interview with Mark van Dord

INTERVIEWEE	MARK VAN DORD
POSITION	Tender manager at BAM Infra Project Management
DATE	7th of November, 2015
LOCATION	NA

ME

Before we start I would like to thank you for your time. On your LinkedIn I see that you are a Tendermanager at BAM Infra Project Management. Is it correct if I say that you have a lot of knowledge about how new projects are acquired regarding, for example, roadworks?

INTERVIEWEE

Yes that is correct. At BAM Infra we have roadwork projects. However, there are also a lot of other large infrastructure projects that we and I are involved in. Performing road maintenance and extending the current road network is only part of what BAM Infra does.

ME

You have told me before that the EMVI is one of the procedures being used to judge tenders and determine who gets to do the job. Is it only the EMVI that is being used?

INTERVIEWEE

What I now often encounter is the Best Value Procurement, which we often refer to as best value. Different from the EMVI is that the company rather than Rijkswaterstaat for example take the lead. With best value there is always an expert from the company involved in the meetings and conversations with the client. We have to show that we have good knowledge about the project and that we know how to complete the project successfully. Similarly to the EMVI we will also be judged on the quality of our approach. We have to present documents to show that past results and or current calculations show we will meet or surpass the criteria. These criterias are goals such as minimizing the impact on traffic flow, hindrance to the direct environment, social acceptance and costs. The more confident the contractor is that we will meet the set criteria the higher the fictive discount will be. When you score well on these aspects this can reduce your fictive price by up to 75%.

ME

What do you mean by a reduction of the fictive price?

INTERVIEWEE

We have to say how much money we need to do the project. However, in the past people might have gone for the cheapest option that is not always the best option. Because quality aspects are also important we have the best value. If you can show you deliver good quality, meaning that traffic hindrance et cetera is reduced to a minimum they are willing to choose you over a cheaper competitor. The more they value your approach the lower is the fictive price that they will use to compare companies. The one with the best value gets picked.

ME

Thank you. For my research I am mostly interested in what the economic impact of roadworks is. I use mobile phone data to see what the travel time is and also what the added uncertainty in travel time is for people traveling along a road where roadworks are occurring. How do you typically assess what the impact is on people?

INTERVIEWEE

For this we have our own models. When we have to show that our approach is the best, models are typically used to calculate the impact of the design choices. By using models we can more easily communicate that we tried different options and prove that the choices made lead to the best results. On a side note, when you want to reduce the impact on travellers it would be ideal if you can tell them where to go. By telling travellers where to go and what roads to use and avoid you can optimally spread the traffic loads and minimize the chance of traffic jams and delays in general. Alternatively, you can try to make sure the impact on site is as little as possible, but this can be costly and is not always feasible. A colleague of mine might also be interested to do an interview once if you want?

ME

Great! I would first like to have some more results such that I can more confidently say what I can and cannot do, but it would be nice to also talk to someone else. Thank you again for your time and input! We will be in touch.

Appendix B: Interview with Henk Taale [Dutch]

Table A1, meta-data about the interview with Henk Taale

INTERVIEWEE	**DR. IR. HENK TAALE**
POSITION	Assistant Professor at TU Delft, Coordinator at Traffic Quest
DATE	28th of July, 2015
LOCATION	Rijkswaterstaat Lange kleiweg 34 Rijswijk, The Netherlands

ME

Hoi, bedankt dat u vandaag wat tijd heeft voor een gesprek. Mijn naam is Johan Meppelink. Ik ben student aan de Universiteit Utrecht, en doe mijn afstudeeronderzoek bij Mezuro, een bedrijf die verplaatsingsdata heeft van Vodafone klanten, en Decisio, een consultancy die veel doet met kosten baten analyses voor onder andere mobiliteitsvraagstukken. Voor mijn afstuderen ben ik aan het onderzoeken hoe verplaatsingsdata van telefoons gebruikt kan worden om de impact van wegwerkzaamheden te evalueren en te voorspellen. Heeft dit ook aansluiting bij wat u hier doet bij Rijkswaterstaat?

INTERVIEWEE

Ja. Zelf werk ik bij zowel de TU Delft als hier bij Rijkswaterstaat. Ik richt me vooral op verkeersmanagement en de effecten daarvan. Daar hoort ook verkeersmanagement bij WIU bij. Vanuit Rijkswaterstaat maak ik ook deel uit van TrafficQuest een samenwerkingsverband tussen RWS, TNO en de TU Delft op het gebied van verkeersmanagement. Hier werken we aan heel veel vraagstukken. Bijvoorbeeld wat er gebeurd met de verkeersinformatie op de matrix borden, weguitbreidingen, en ook wat er gebeurt bij wegwerkzaamheden. Maar wat is het nou precies dat jij aan het onderzoeken bent?

ME

Voor vandaag had ik nog een korte presentatie gemaakt. Zou ik die er even kort bij kunnen pakken? Dan kan ik beter uitleggen waar mijn onderzoek over gaat. Met de telefoondata wil ik onderzoeken hoe de reizen van mensen veranderen wanneer er wegwerkzaamheden zijn. Met de tellussen en andere tel systemen in de weg kun je een goed beeld krijgen van de impact op een specifiek punt, maar blijft het lastig om een totaal beeld te krijgen. Sommige mensen zullen thuisblijven, omrijden, of een ander vervoersmiddel kiezen omdat er wegwerkzaamheden zijn. Ik wil kijken of de telefoondata hier inzichten in kan geven en zo een beeld krijgen van de maatschappelijke kosten van wegwerkzaamheden. Hiervoor heb ik eerst gekeken of de telefoondata overeenkomt met de data op de wegen. De lussen op de weg, ook al zijn deze niet perfect, gebruik ik als een gold standard. Volgens het NDW hebben de lussen een zeer hoge accuracy (99% showing on the slide). Ondanks dat niet alle lussen altijd informatie geven lijken ze inderdaad vrij accuraat.

INTERVIEWEE

Wat wordt er bedoeld met 99% accuracy? Er is een verschil tussen betrouwbaarheid, dat de lus technisch goed werkt, en nauwkeurigheid, dat de lus de voertuigen goed telt. Voor beide worden bepaalde normen vastgehouden aan welke de apparaten moeten voldoen. De lussen zijn uiteraard niet 100% nauwkeurig. Je kunt hebben dat mensen lussen op de weg bijvoorbeeld missen wanneer een auto tussen de lussen door rijdt of in de bocht een stukje afsnijdt.

ME

Voor mijn onderzoek wil ik ook graag een beeld krijgen van wat er op het moment gebeurt om de impact van wegwerkzaamheden te voorspellen en evalueren. Op vanAnaarBeter staan wel aangegeven extra reistijden, maar deze lijken heel grof. Weet u hoe dit in de praktijk nu ongeveer in z’n werk gaat?

INTERVIEWEE

Ja. Meestal wordt er bij een aanbesteding een inschatting gedaan van de hinder die wordt veroorzaakt door de wegwerkzaamheden. Verschillende inzending/offertes worden vergeleken aan de hand van een model. Op deze manier worden alle inzending gelijk beoordeeld. Evaluatie oftewel controle op de daadwerkelijke impact van de wegwerkzaamheden tijdens of na de werkzaamheden gebeurt niet of sporadisch. Er zijn wel onderzoeken geweest vanuit onder andere Rijkswaterstaat. Zelf heb ik aan een onderzoek meegewerkt. Dit onderzoek was ongeveer een jaar of tien geleden uitgevoerd, deze zal ik je doorsturen. Er zijn hiernaast ook nog enkele kleinere maar vergelijkbare onderzoeken geweest naar de impact van wegwerkzaamheden.

INTERVIEWEE

Hoe ga jij de impact van de wegwerkzaamheden berekenen? Hoe is die vertaalslag naar euro’s?

ME

Hiervoor hoop ik onder andere wat expertise bij Decisio te kunnen lenen, dit is een van de twee bedrijven waar ik mijn afstudeerstage doe. Zij doen veel op het gebied van kosten baten analyses, ook voor mobiliteit vraagstukken. Hoe ik de kosten van thuisblijven en het kiezen voor de trein in plaats van de auto ga berekenen weet ik nog niet zeker. Voor vervoer met de auto denk ik dat de vertaalslag te maken is door te kijken naar de extra reistijd ten gevolge van wegwerkzaamheden en deze te vermenigvuldigen met een standaard waarde. Volgens mij had ik ergens getallen gezien waarin de kosten van files op vergelijkbare manier waren berekend.

INTERVIEWEE

Er zijn online ook documenten hiervoor van Rijkswaterstaat of het

Kennisinstituut voor Mobiliteitsbeleid. Hierin staat ook aangegeven welke kosten worden aangehouden voor reistijdverlies. Dit is ook onderverdeeld in woon-werk verkeer, zakelijk, vrachtvervoer, et cetera. Deze kun jij wellicht ook aanhouden.

ME

Wie of waarvoor zou het onderzoek dat ik nu uitvoer waarde kunnen hebben?

INTERVIEWEE

Wellicht dat het voor Rijkswaterstaat of kennis instellingen interessant is. Ook met betrekking tot optimalisatie van werken, dus wanneer moeten welke werken gepland worden et cetera. Voor Rijkswaterstaat is dit interessant voor de volgende zaken:

- Optimalisatie van de planning van WIU. Meer inzicht in de herkomsten, bestemmingen en gedrag van reizigers kan helpen om de planning beter te maken.
- Evaluatie van maatregelen in het algemeen. Net als bij WIU kan inzicht in reispatronen en de veranderingen daarin helpen bij de evaluatie van maatregelen
- Verkeersmodellen. Deze data kan waardevol zijn om invoer van verkeersmodellen beter te maken, in het bijzonder de herkomst-bestemmingsmatrix.

Appendix C: Interview with Niels Hoefsloot [Dutch]

Table 0.1, meta-data about the interview with drs. Niels Hoefsloot

Interviewee	Drs. Niels C. hoefsloot
POSITION	Partner at Decisio
Date	27th of October, 2015
Location	Decisio BV
	Valkenburgerstraat 212
	Amsterdam, The Netherlands

[INTRO skipped]

ME

Van je collega Menno de Pater, mijn tweede begeleider, begreep ik dat jij veel weet over Maatschappelijke Kosten Baten Analyses. Hij gaf aan dat jij op het moment cursusleider van de cursus MKBA’s bent bij Decisio. Voor mijn onderzoek bekijk ik naar de impact van wegwerkzaamheden aan de hand van de telefoon data van Mezuro. Volgens mij zijn jullie daar mee bekent. Hierin kan ik zien wat de reistijd is van mensen die van A naar B gaan en die heb ik toegekend aan een bepaalde weg. Wat ik wil doen is de economische kosten berekenen van wegwerkzaamheden. Deze hangen af van verlies in reistijd voor verschillende reismotieven en ook van mensen die opeens met de trein gaan reizen om de wegwerkzaamheden te mijden of thuisblijven. Van de mensen die blijven rijden kan ik de reistijden overzetten naar kosten met standaard waardes hiervoor. Voor thuisblijvers et cetera heb ik geen idee hoe ik de kosten zou moeten toekennen. De vraag is hoe kan ik de economische kosten het beste berekenen?

INTERVIEWEE

Als je de motieven hebt dan kun je de standard waardes hiervoor inderdaad makkelijk toepassen. Dit moet je gewoon doen. Vertraging heeft een negatieve economische impact en dit kun je op de koop toenemen. Omrijders leggen wel meer kilometers af. Dit levert ook slijtage op en daarmee kosten. Deze variabele autokosten zou je eigenlijk moeten meenemen. Naast verlies in reistijd worden de kosten hierdoor hoger voor deze reizigers.

ME

Omrijders kan ik heel moeilijk meenemen. In de data zie ik alleen wie er van A naar B gaan. Deze ken ik dan toe aan de weg aan de hand van de kortste route qua tijd.

INTERVIEWEE

Dus je hebt mensen die blijven rijden en je hebt mensen die omrijden. Daarvan weet ik niet of je daar iets mee kan, met die extra kilometers. Hier heb je wel de tijd van dus eigenlijk. Mensen gaan nog steeds van A naar B. Dan heb je mensen die de trein nemen. Weet je dat ook?

ME

Ja, ik weet hoeveel mensen er met de auto gaan en hoeveel mensen er met de trein gaan. En dus ook de verschillen daartussen. Ik kan ook zien of iemand normaal met de auto gaat en overstapt naar de trein en vice versa. Ik heb de hoeveelheid treinreizigers en de hoeveelheid treinreizigers.

INTERVIEWEE

Ja, wat dan eigenlijk een soort slimmigheid is om de zeggen van hoe waardeert men dan. Mensen die iets anders gaan doen dan dat ze eerst deden die zouden ook gewoon kunnen blijven verplaatsing kunnen maken. Alleen dan hebben ze die vertraging. Kiezen ze ervoor iets anders te gaan doen dan doen ze dat om hun eigen schaden te minimaliseren. Eigenlijk vaak bij dit soort dingen als je het niet precies weet zeggen we dan daar hebben wij daar de rule of half voor. Zeggen we van ja de mensen maken al de verplaatsing dus dat zouden ze ook gewoon doen als we de gewone reistijd hebben. Nu is er die vertraging en doen ze het niet meer dus hoe erg vinden ze dat nou. Ze weten wel iets anders te vinden. Hoe erg vinden ze dat dan is eigenlijk de vertraging keer de helft. Dat is eigenlijk heel simpel van als ze het nut van de vertraging nog steeds groter zou zijn dan de vertraging zelf dan zouden ze de vertraging blijven maken. Op het moment dat ze dat niet meer doen, dus ook het moment dat ze iets anders doen, dan doen ze iets wat dus blijkbaar voor hun een betere uitkomst is dan die vertraging meemaken.

ME

Dus dan krijgen ze een halve vertraging.

INTERVIEWEE

Dan reken je gewoon de halve vertraging toe ja. Dat is de rule of half. En dan ja en je kan ook nog als je iets over die omrijkosten zou weten zou je dat ook nog mee kunnen nemen, maar ik denk dat dat heel lastig wordt. Dus dan heb je eigenlijk alle reizen, de mensen die blijven, en die iets anders gaan doen. Of dat nou omrijden of de trein pakken of thuisblijven is. Dat is je rule of half. Rule of half is een benadering voor allemaal. Dus dat kan je op die manier gewoon waarderen. Dat is een geaccepteerde methode in kosten baten analyses.

ME

Dat is een goede om te weten.

INTERVIEWEE

Dan kun je dezelfde value of time toepassen en de vertragingsfactor die je uit de telefoon data haalt.

ME

Dat is dan de vertraging die de autoreizigers oplopen. Ik weet wel eventueel wat voor vertraging de treinreizigers oplopen. Kan ik die niet gebruiken?

INTERVIEWEE

Het lastige is dat het sterk afhangt van persoonlijke voorkeur en ook hoe ver iemand van het station af woont. Voor de ene persoon is de trein wel een goed alternatief en voor de andere niet. De ene persoon zal misschien bij een vertraging van een minuut al de trein pakken en de ander pas bij tien. Daarom hebben we de rule of half omdat het gemiddeld dan allemaal wel goed zou komen. Als iemand kiest voor het comfort van de trein omdat hij dan ook een kopje koffie kan drinken en de krant kan lezen dan neemt hij misschien de extra reistijd op de koop toe. De rule of half gaat niet zo zeer over tijd maar meer over voorkeuren van mensen. De feitelijke kosten en reistijd in de trein is dus eigenlijk niet zo relevant. Het gaat erom dat je eigenlijk normaal gesproken de voorkeur hebt om met de auto de verplaatsing te maken en het daarom ook doet. Op het moment dat de verplaatsing langer duurt doe je het niet, maar de enige reden dat je een andere verplaatsing maakt is omdat die langer duurt. Het kan zelfs zijn dat mensen zeggen dat ze normaalgesproken het tijdsvoordeel van de auto sterker wegen dan het comfort van de trein. Op het moment dat die vertraging er is vind ik dat misschien niet meer zo. Misschien zou hij met de auto alsnog sneller zijn, maar geeft dus nu niet meer de voorkeur aan de weg. Dus het is een negatief effect. Uiteindelijk heeft het meer voorkeuren te maken. Met de rule of half gaat dit dus wel goed. Het is volgens mij ook de enige manier om dit goed te benaderen.

ME

Super. Bedankt voor de uitleg. Geldt dit dan ook voor mensen die thuisblijven?

INTERVIEWEE

Ja. Die mensen zouden anders ook gewoon de reis hebben gemaakt. Ze zouden waarschijnlijk wel vaker thuis kunnen blijven, maar dan doen ze dat niet. Het feit dat ze op werk collega's en andere kunnen spreken dan vinden ze dat zo goed dat ze de reistijd en kosten op de koop toe nemen. Op het moment dat die reistijd langer wordt dan is dit dus niet meer zo. Dan heeft thuisblijven de voorkeur. Rule of half geldt voor alle reizigers die er normaal wel zijn en door de werkzaamheden er niet meer zijn.

ME

Het gaat om wat mensen willen.

INTERVIEWEE

Het gaat om die voorkeuren. De individuele voorkeuren.

ME

En dan blijft de zakelijke rijder een andere waardering krijgen?

INTERVIEWEE

Ja, je past dus wel per motief de value of time toe. Voor de mensen die blijven rijden is dat economische schade. Voor de mensen die er niet meer zijn doe je dat door de helft.

ME

Dat is mooi. Moet ik alleen nog bepalen wanneer iemand normaal een trip zou hebben gemaakt. Die zie ik nu nog niet. Maar dit is iets wat ik zelf moet gaan bepalen.

[netwerk effecten overgeslagen - vanaf 12:30]

ME

Is er nog iets dat ik moet meenemen? Iets wat standaard voorkomt bij een kosten baten analyse of kanttekeningen die genoemd moeten worden?

INTERVIEWEE

Sowieso zitten er hier een hoop kanttekeningen bij. Ook die standaard kerngetallen voor de value of time. Dat is op basis van echt grootschalig onderzoek van wat mensen het waardvinden om reistijd te besparen of als dingen wat kosten. Dat kan per individu en rit anders zijn. Mensen met hoge inkomens bijvoorbeeld zullen waarschijnlijk meer geld overhebben om tijd te verminderen. Dit is iets wat nog mee kan spelen. Ook als je naar verschillende gebieden kijkt. En er zitten altijd bandbreedtes omheen bij dit soort inschattingen die je uiteindelijk uitrekent.

ME

Mijn berekeningen over motieven en dergelijke zijn ook niet 100% accuraat. Die bandbreedtes krijg ik inderdaad sowieso.

INTERVIEWEE

Precies. Wat hier nog wel speelt is dat niet alleen tijd maar juist ook betrouwbaarheid van reistijden voor mensen heel belangrijk is voor hun om keuzes te maken. Dat wordt wel beter met de realtime informatie maar normaal gesproken weet je niet exact hoe lang jou rit duurt als je in de auto stapt, en zeker in de spits want dan is er altijd kans op file. Je gaat er buiten de spits waarschijnlijk vanuit dat er geen file is. Dus dan is dat redelijk betrouwbaar, maar bijvoorbeeld bij wegwerkzaamheden is die zekerheid er niet. Die betrouwbaarheid is ook wel wat waard. Daarvoor is ook die KiM publicatie over value of time. Daarin staan ook gegevens over de waarde van betrouwbaarheid van de reistijd. Volgens mij kun jij dat ook uit je data halen. De variatie van reistijd. Die zegt iets over de betrouwbaarheid van reistijden. Misschien kun je variatie wat moeilijker voorspellen, maar van gedane wegwerkzaamheden is dit prima te berekenen.

ME

Dat is goede! Ik wist dat de reistijdbetrouwbaarheid van belang is, maar had er nog niet bij stil gestaan dat ik die misschien kon halen uit de variatie van reistijden. Variatie lijkt mij inderdaad een goede maatstaaf voor de betrouwbaarheid.

INTERVIEWEE

Zeker. Mensen is dat ook gevraagd hoeveel heeft u ervoor over als de rit de ene keer 20 minuten en de andere keer 40 minuten ten opzichte van 25 en 35 minuten. Hoe minder variatie hoe beter. Daar kun je zeker nog naar kijken. Er zijn ook studies die aantonen dat betrouwbaarheid soms net zo belangrijk of zelfs zwaarder kan wegen dan reistijd verlies.

ME

Dat kan ik me voorstellen. Als je ergens op tijd moet zijn dan neem je rekening met de betrouwbaarheid. Wanneer je denkt dat je een half uur in de file kan staan dan ga je een half uur eerder weg. Wat er soms voor zorgt dat je een half uur te vroeg ergens aankomt. Dit is dan geen vertraging, maar wel hinderlijk.

INTERVIEWEE

Ja inderdaad. Te laat komen vinden mensen vaak erger dan te vroeg komen. Hier kun je zeker naar gaan kijken. Met de data die je hebt kun je die volgens mij gewoon berekenen. Dit kan een hele goede aanvulling zijn.

ME

Dat denk ik ook. Dit kan nog wel wat toevoegen.

INTERVIEWEE

Dan heb je eigenlijk alle economische kosten. Nou ja. Dat omrijden, dat blijft een lastige.

ME

Of mensen omrijden. Dat kan ik eigenlijk niet zien met de telefoon data. Ik heb wel de data van de tellussen in de weg. Wellicht dat ik daar wel iets uit kan halen.

INTERVIEWEE

Op het moment dat mensen gaan omrijden heb je altijd een onderschatting van mijn inschatting voor de impact van wegwerkzaamheden. Dit omdat omrijden extra kost en dit niet wordt meegenomen. Verder over kosten baten analyses kun je zeggen dat het meestal geen exacte wetenschap is. Met de data die je hebt heb je wel je wel meer detail dan wij meestal vooraf hebben. Wat wij dan vaak missen is die variatie van reistijd. Vaak wordt er met een verkeersmodel iets berekent en die geeft een gemiddelde reistijd op een werkdag in de toekomst. Misschien berekent het model die wel, maar geeft deze niet. Vaak hebben wij dus alleen een gemiddelde en dan werken wij met een opslag. Hier kun jij hem gewoon berekenen, dus dat zou ik dan ook doen.

ME

Top! Ik denk dat ik heel veel informatie heb nu. Bedankt.

Appendix D: The current (old) scaling method

In order to go from users to the population the data should adequately represent the population. In other words, to go from the sample to the entire population a multiplication factor should be calculated and applied. Within the context of this research the entire population is equal to all trips over 10 km within the Netherlands. To do this the following three steps are taken.

First, a baseline population is determined to compensate for people being abroad, on a holiday or on a business trip. The procedure they use is to select the week where the maximum number of unique devices are detected in March or November. During these two months the least number of people are away from home, according to tourism statistics, making it the best option available to set a baseline (Geerts, 2014).

Second, a correction is applied for local inconsistencies in the use and possession of mobile phones. To do this Mezuro calculates a multiplication factor based only on the geographic dimension of the data. This means a multiplication factor is calculated by determining the penetration of mobile phone subscribers for each Mezuro area. To put it more simply, they calculate the number of active users per Mezuro area and divide that by the number of inhabitants per Mezuro area. Arguably, this ratio provides a good indication to correct for some inconsistencies in the use and possession of mobile phones.

Finally, a correction is applied for people who are on holiday within their country of origin. This is based on their determined place of residence and activities during a month. The scaling factor per area is the result of the number of inhabitants per area divided by the number of active users.

With a view on using this method to go from the sample to the population within the context of this research, we will continue our analysis by looking at the geographic and demographic dimensions of the data in the following two subsections.

Geographic Representativeness

Here we analyse what the size is of the inconsistencies within the geographical dimension of the Mezuro data are present. To determine how well the Mezuro data is represented spatially, the first thing we do is analyse the multiplication factor as it is currently being used. The multiplication factor is applied to correct for the variation in geographical spread of the number of subscribers per Mezuro area and is calculated each day. The multiplication factor is defined as the ratio of the active users compared to the traveling population in that area, which basically represents the penetration of subscribers within a certain area.

Figure D1 (see next page) depicts the average multiplication factor per area. What can be noticed is that the multiplication factor varies a lot nationally. The categories are chosen in such a way that they each contain 1/6th of the total number of Mezuro areas. Therefore, we conclude that two third of the Mezuro areas have an average multiplication factor between 4.29 and 7.43, during the course of the first nine months of 2015. Meaning the penetration of subscribers within two third of the

Mezuro areas lies between 13,3% (100% / 7.43) and 23,8% (100% / 4.29) of the traveling population. Note that this differs from the market share of the provider because we divide by the number of inhabitants and not the number of mobile phones in the area.

Because the multiplication factor is calculated per day it is possible to see how it changes over time. We express this by calculating the standard deviation of the multiplication factor. However, when we look at the standard deviation on its own it does not tell how many people are affected. To take that into account we correct for the number of active users in that area by dividing the standard deviation by the average multiplication factor. These results are depicted in figure D2. The classes here are also chosen in such a way that they an equal number of Mezuro areas are contained within each class. Most interestingly, the highest standard deviations form a characteristic pattern that highly resembles the "Bible belt". The Bible belt is a nickname for a collection of highly religious areas within The Netherlands. Further research confirms that when isolating this effect, by just looking at the standard deviations of the multiplication factor on Sundays, the effects are even more prevalent. This behaviour is most likely explained by religious people turning off their mobile phones on Sunday and go to church.

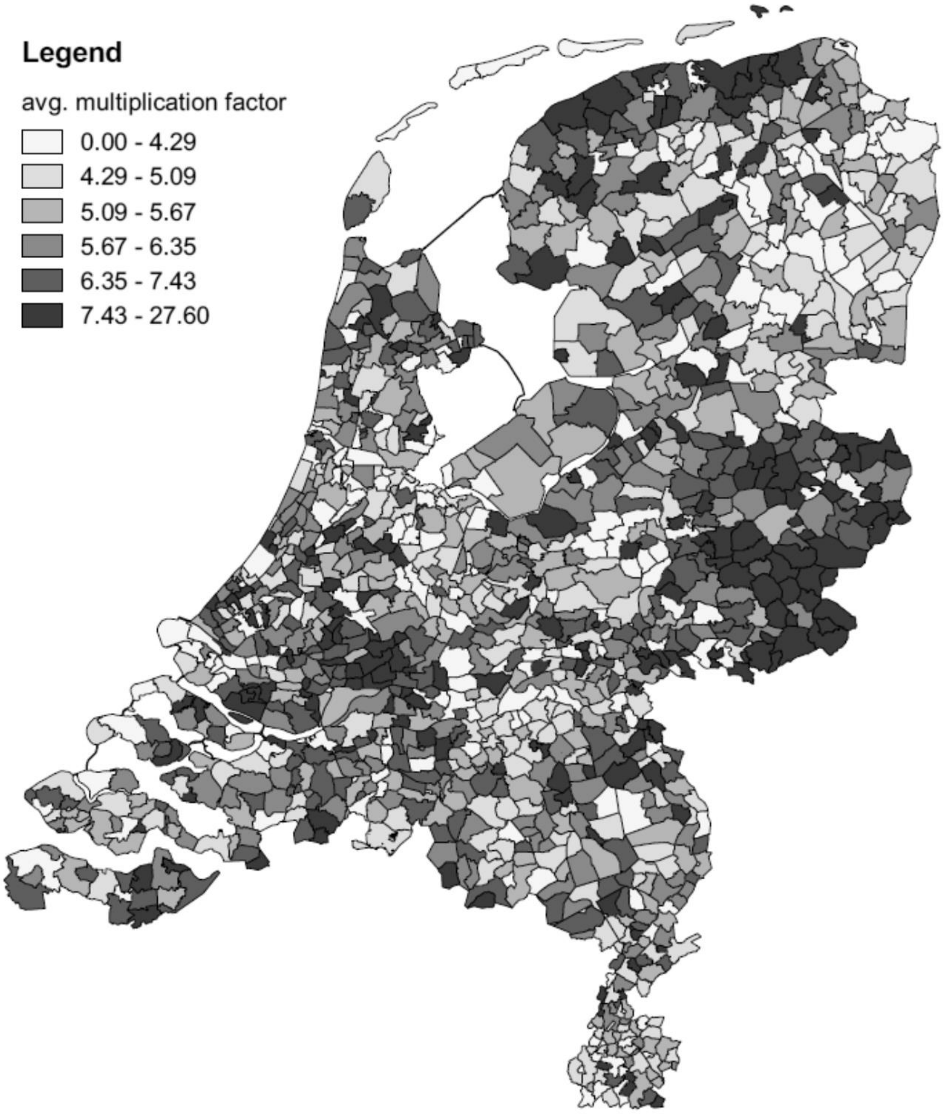


Figure D1, Average scaling factor per Mezuro area over the first nine months of 2015.

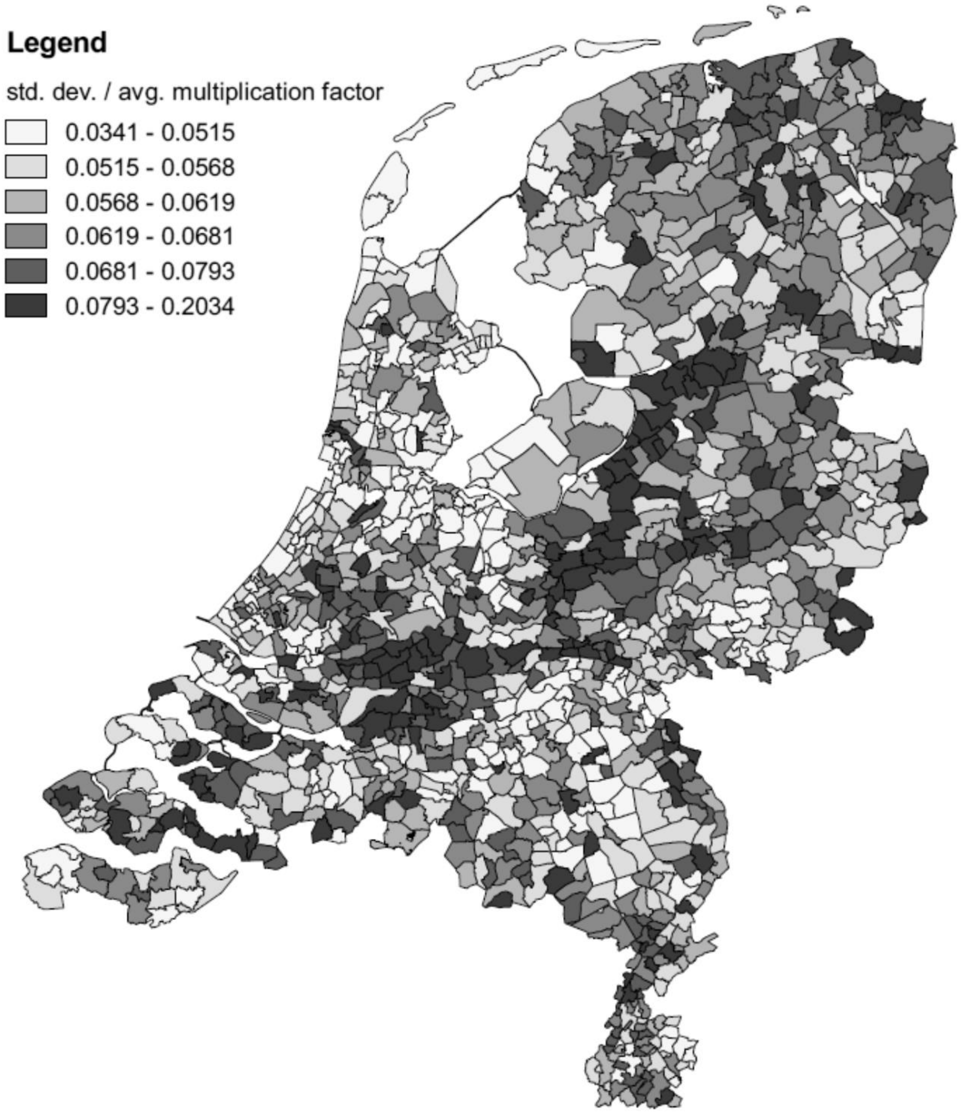


Figure D2, Geographical variation of the scaling factor corrected for the traveling population.

Demographic Representativeness

The dataset of Mezuro should be an unbiased representation of the mobility patterns of the entire traveling population of the Netherlands.

Offermans et al. (2013) did a study to evaluate, for one, the demographic representativeness of the user in the mobile phone data in this study. Offermans et al. (2013) compared the demographic details of subscribers such as age and gender, with demographic details from the municipal population register (in Dutch: Gemeentelijk Basis Administratie). Figure D3 shows the penetration of the sample in the Dutch population. They did, however, only have the part of users that do not have a pre-paid or business contract so the exact distribution might differ. Furthermore, they note that young people might be underrepresented in the data they evaluated because these young people might use phones that are contracted on their parent's name. The overall demographic representativeness over the Dutch population appears to be good (Offermans et al., 2013).

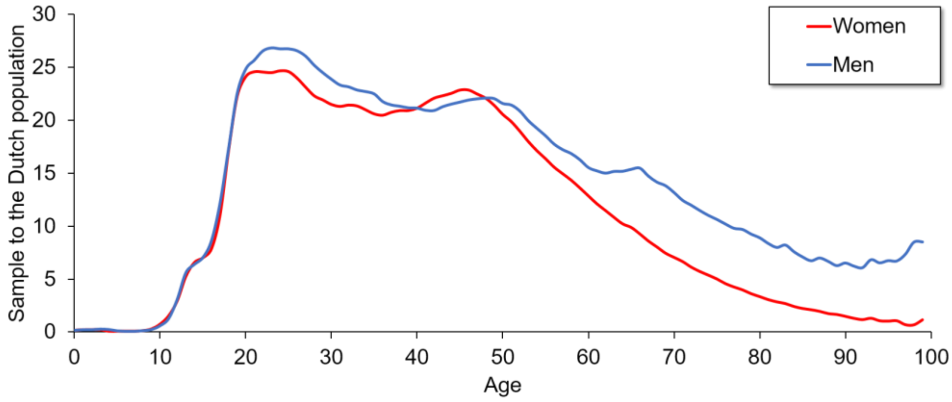


Figure D3, the penetration of subscribers per age group per gender (Offermans et al., 2013).

We, however, are less interested in how the data relates to the population in general and are more interested in how the data relates to the traveling population. Children might be less likely to be affected by roadworks, for example, as they are usually not on the road. We want the scaling factor to take into account not only the chance the relation between users in the sample to the population, but also the relation between users of different age groups in relation to the people found traveling.

Appendix E: PDD tables for the new scaling factor

The activity table and concept table describe the activities and deliverables, respectively, of the new scaling method presented in the PDD in figure 6.5.

Table D1, activity table belonging to the PDD shown in figure 6.5.

ACTIVITY	SUB-ACTIVITY	DESCRIPTION
DETERMINE THE CHANCE TO OBSERVE A PERSON BY AGE GROUP		Divide the number of people per age group by the total number of inhabitants in that area.
DETERMINE THE MOBILE PHONE PENETRATION PER AGE GROUP		Retrieve data concerning the mobile phone penetration per age group from (a) trusted source(s).
ESTIMATE THE PROVIDER MARKET SHARE	Calculate the inhabitants per age group	Multiply the total number of inhabitants by the POPULATION DISTRIBUTION. Note, the total number of inhabitants used here is adjusted for people being abroad, on a holiday or on a business trip (Geerts, 2014).
	Multiply by mobile phone penetration per age group	Multiply the INHABITANTS PER AGE GROUP by the MOBILE PHONE PENETRATION PER AGE GROUP.
	Divide users in sample by the number of phone users	Divide the number of users in the area by the number of phone users in the area, i.e. the sum of the PHONE USERS PER AGE GROUP.
DISTRIBUTE USERS ACROSS AGE GROUPS	Calculate the chance a user is in an age group	Multiply the POPULATION DISTRIBUTION by the MOBILE PHONE PENETRATION PER AGE GROUP and by the PROVIDER MARKET SHARE.

	Normalize the calculated probabilities	Divide the PROBABILITY INHABITANT IN SAMPLE PER AGE GROUP by the sum of the PROBABILITY INHABITANT IN SAMPLE PER AGE GROUP.
	Multiply the users by the normalized probabilities	Multiply the users in the area by the chance of observing a user in a certain age group, i.e. the PROBABILITY USER IN AGE GROUP.
DETERMINE THE LIKELIHOOD OF A PERSON MAKING A TRIP OVER X KM DURING WORKDAYS, WORKDAYS DURING THE HOLIDAY, SATURDAYS, AND SUNDAYS PER AGE GROUP		Gather information about the chance that a person of a certain age groups makes a trip longer than X kilometres on a day. We use OViN to determine this and take the differences in weekday, weekend and holiday separately into account.
CALCULATE THE SCALING FACTOR	Estimate inhabitants traveling	Multiply the INHABITANTS PER AGE GROUP by the LIKELIHOOD OF TRAVELING PER AGE GROUP.
	Estimate the users traveling	Multiply the USERS PER AGE GROUP by the LIKELIHOOD OF TRAVELING PER AGE GROUP.
	Divide inhabitants traveling by users traveling	Divide the sum of the INHABITANTS TRAVELING by the sum of the USERS TRAVELING.

Table D2, concept table belonging to the PDD shown in figure 6.5.

CONCEPT	DESCRIPTION
POPULATION DISTRIBUTION	The probability that an inhabitant belongs to a certain age group.

MOBILE PHONE PENETRATION PER AGE GROUP	The probability that a Dutch citizen of a certain age group possesses a mobile phone.
INHABITANTS PER AGE GROUP	The absolute number of inhabitants per age group.
PHONE USERS PER AGE GROUP	The absolute number of inhabitants that possess a mobile phone per age group.
PROVIDER MARKET SHARE	The market share of the network provider. Hence, in this case the market share is equal for all age groups.
PROBABILITY INHABITANT IN SAMPLE PER AGE GROUP	The probability that an inhabitant is in our sample per age group.
PROBABILITY USER IN AGE GROUP	The probability that a user in our sample is in a certain age group.
USERS PER AGE GROUP	The absolute number of users per age group.
LIKELIHOOD OF TRAVELING PER AGE GROUP	The probabilities that a person of a certain age group makes a trip that is longer than X kilometre on a specific day of the week.
INHABITANTS TRAVELING	The number of INHABITANTS PER AGE GROUP that is expected to make a trip over X kilometres on a specific day of the week.
USERS TRAVELING	The number of USERS PER AGE GROUP that is expected to make a trip over X kilometres on a specific day of the week.
SCALING FACTOR	The ratio between the number of INHABITANTS TRAVELING the number of USERS TRAVELING. The scaling factors applies to traveling people, because these are the people that are relevant for the OD matrix.

Appendix F: Applying the new scaling factor

The calculation of the scaling factor in the example is performed in the same sequence as the method presented in the PDD. Additionally, the colour coding can also be used to cross reference the example to the method. The scaling factor is calculated for each area, each day. Note in the example we use less age groups than specified above, but this is only for illustrative purposes. The real scaling factor will include all the age groups, i.e. starting at 0 with group sizes of 5 years and everything age 80 and over grouped together.

Table F1 a simplified and fictional example of the calculation of the OD scaling factor for a single Mezuro area for a single day. This example only includes the dimension age group (grouped by 20 years). The full method also includes the distinction between holidays, weekdays and weekends but these are omitted in this example to the improve readability.

		Age groups						
		1	**2**	**3**	**4**	**5**	**Total**	**Calculation**
	Constants							
A	Active users	4000						
B	Present inhabitants	15000						
C	Age groups	0-19	20-39	40-59	60-79	80+		
	Market share							
D	% Age group in area	15%	40%	15%	20%	10%	100%	
E	% With mobile phone	50%	90%	84%	62%	19%	-	
F	Inhabitants in area	2250	6000	2250	3000	1500	15000	B*D
G	Inhabitants with phone	1125	5400	1890	1860	285	10560	E*F
H	% Market share	38%	38%	38%	38%	38%	-	A/SUM(G)
	Users traveling							
I	Multiplier	2.8%	13.6%	4.8%	4.7%	0.7%	-	D*E*H
J	Normalized multiplier	10.7%	51.1%	17.9%	17.6%	2.7%	100%	I/SUM(I)
K	Users	426	2045	716	705	108	4000	A*J
L	Chance of trip >10 km	5%	50%	40%	30%	10%	-	
	Scaling factor							
M	Inhabitants traveling	112.5	3000	900	900	150	5062.5	F*L
N	Users traveling	21.3	1022.7	286.4	211.4	10.8	1552.6	K*L
O	Scaling factor OD	3.26						SUM(M)/ SUM(N)

We now elaborate the calculation of the scaling factor in table F1. First, the distribution of the population in the Mezuro area over the age groups is determined (row D). Second, the mobile phone penetration per age group is determined (row E). Third, the market share is calculated in three sub steps. The inhabitants in the Mezuro area are distributed according to the population distribution (row F). Then the number of inhabitants with a mobile phone is determined (row G). Subsequently, the market share of the telecom provider in the selected Mezuro area is calculated by dividing the number of active users by the sum of all present inhabitants with a mobile phone (row H).

The next step is to distribute the active users across the age groups. We obtain the multiplier by multiplying the chance per age group (row I). The multiplier represents the chance that a random person from the present population of the Mezuro area is in our sample, i.e., an active user. We normalise this multiplier in row J. The normalised multiplier represents the chance that a random person from our sample belongs to that age group. The next step is to multiply the normalised multiplier by the number of users. This results in the distribution of active users over the age groups (row K). The next is to determine the likelihood that someone of a certain age group makes a trip greater than 10 kilometres (row L), which is depicted in figure 4.10.

The last three steps are used to determine the travelling users, the travelling population and the ratio between them, i.e., the scaling factor. The travelling population is determined by multiplying the inhabitants per age group by the likelihood that someone of a certain age group makes a trip greater than 10 kilometres (row M). The users travelling per age group is calculated by multiplying the active users per age group by the likelihood that someone of a certain age group makes a trip greater than 10 kilometres (row N). Finally, the scaling factor is calculated by dividing the number of inhabitants travelling by the number of users travelling (row O). Note that this example given does not include the dimensions weekend, week and holiday in order to improve the readability of the example. However, the method as formalised in our PDD (see figure 6.5) does include these dimensions.

Due to limited availability and reliability of data from market research on the market share of the provider per age group we had to take another approach. This made us choose to calculate the market share based on the ratio between the active users and the total population with a mobile phone. The underlying assumption here is that everyone with a mobile phone has a provider. By calculating the market share this way, regional differences in the provider market share are compensated for. However, this means that from a mathematical point of view the inclusion of the market share in the calculation of the scaling factor has merely become irrelevant. That is because the resulting scaling factor only reflects differences in proportions between age groups and by including the market share as an equal value for all age groups there are no differences between age groups. Hence, fluctuations in the market share no longer affect the resulting scaling factor. Although differences in the market share per area are not reflected in a change of the scaling factor, we like to keep it in the method for the purpose of completeness. Moreover, when in the future more data becomes available on the market share per age group this can easily be added to the method.

Appendix G: Attributes in the mobile phone data

Table G1, attributes of the mobile phone dataset relevant for trip motive prediction.

ATTRIBUTE	DESCRIPTION
ORIGIN **DESTINATION**	Area of origin / destination of a trip. The area may be a municipality or subpart of a municipality. For densely populated areas, e.g. Amsterdam, there are several areas in one municipality.
TIME OF DEPARTURE	The last CDR inside the range of the cell tower at the origin.
TIME OF ARRIVAL	The first CDR inside the range of the cell tower a person at the destination.
TIME AT DESTINATION	The total time after a trip before the next trip starts.
HOME BASED	This attribute states if the person is either departing home, leaving home, or traveling between non homebased places. Home is where a person spends most evenings (between 20h and 7h) during weekdays and weekends (Geerts, 2014). This is calculated per person on a monthly basis and is specified to a level of detail of a four digit postal code, which is smaller than a municipality or an area.

Table G2, descriptions of the attributes used in modelling and evaluating trip motive prediction.

ATTRIBUTE	DESCRIPTION
TRIP MOTIVE	The motive of a trip. In this research we use “zakelijk bezoek in werksfeer” for business trips, “van en naar het werk” for home-work trips, and merge all other motives in OVIN and call it “other”.
ARRIVAL / DEPARTURE TIME	This is the start and end time of a trip. Trip start and end are both expressed in minutes from 0:00.
FIRST TRIP START / END	Start and end of the first trip for a person. This attribute is included because, for example, people starting the day early might be more likely to have business trips later in the day.
LAST TRIP START / END	Same as the previous. When a person ends a day of travelling might indicate what type of traveller a person is and so help predict trip motives of other trips during the day.
HOMEBASED	This attribute states if a person is leaving home, arriving home, or has a trip not involving his/her home location.
HOLIDAY	Whether or not it is a holiday at the persons home location.

Appendix H: Correcting for travel time

We have a total of three assumptions to check. These are that the residuals are normally distributed, that there is no correlation between consecutive residuals, that the residuals should have the same distribution over the range of predicted values, and there is no correlation between parameters (Field et al., 2009). We test this each time with a QQ-plot, a plot where the lagged residuals (t-1) are plotted against the residuals (t0), and a plot with on the x-axis the predicted values and on the y-axis the corresponding residuals. The correlation between parameters is evaluated with VIF-scores and will be presented in a table (Field et al., 2009). In the three sections below we will present the assumption checks for correcting travel time (H.1), correcting travel time std (H.2), and correcting people on the road (H.3) corresponding to sections 9.2.2 through 9.2.4, respectively.

H1. Assumptions: Correcting travel time

Normality of the residuals

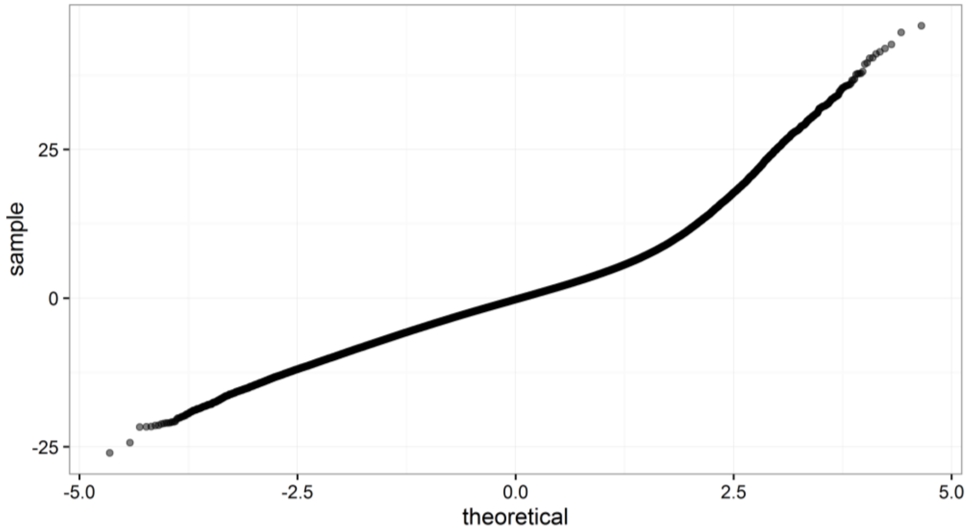


Figure H.1, QQ-plot showing the theoretical residuals plotted against the residuals in the sample.

From figure H.1 we see an approximately straight line, which would indicate the presence of a normal distribution. When we look closer, however, we see an unnatural dip around 2.5 for the theoretical residuals. This implies the distribution of residuals, though close, is no perfectly normally distributed.

Independence of the residuals

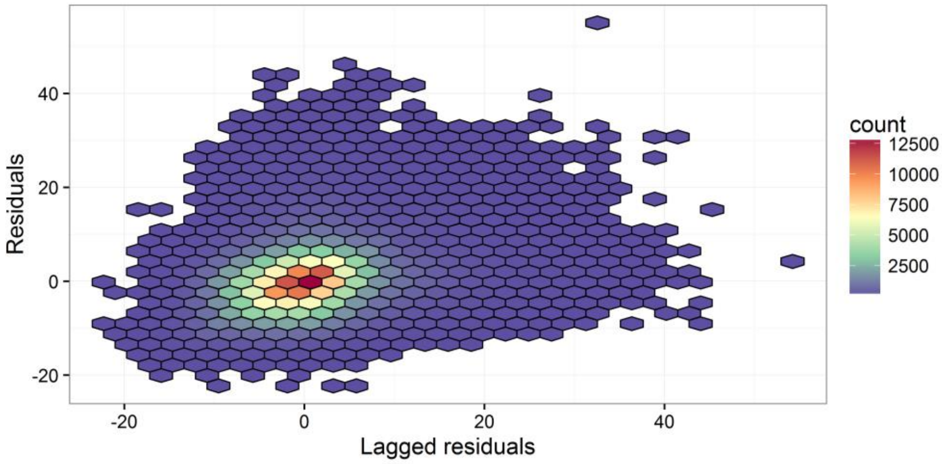


Figure H.2, residuals plotted against the lagged residuals.

In figure H.2 we see there is no apparent relation between the current and lagged residuals. A Pearson correlation coefficient between the two axis returned a value of 0.25 indicating there is only perhaps a minor relation between consecutive residuals.

Despite the slight correlation we do not see any effects that would strongly bias our model. We, therefore, state the assumption of independence of the residuals is met.

Homoscedasticity of the error variance

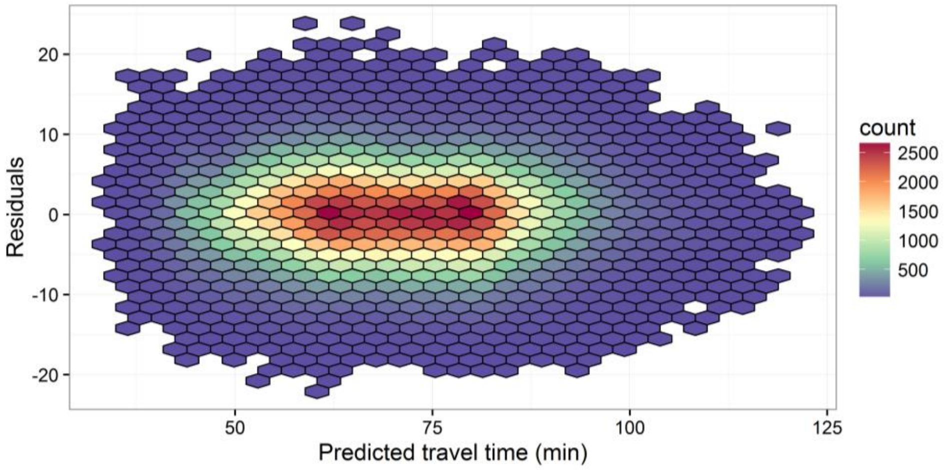


Figure H.3, residuals around the predicted travel time in minutes.

In figure H.3 we observe there is no strange behaviour regarding the residuals over the spectrum of predicted values. Within the range of 50 to 80 minutes the residuals appear to be more stretched out. However, this is also what one would expect as there are simply far more observations there. Hence, the chance of having more extreme outliers is also greater.

No multicollinearity

Table H.1, VIF scores relating to our LM that predicts travel time based on external factors.

VARIABLE	VIF-SCORE
Distance	1.54
Night	1.03
Rain * distance	1.04
Wind * distance	1.62

As shown in table H.1 all VIF scores are good, i.e. they are close to 1 and way below the threshold of 10 after which the model becomes unreliable (Field et al., 2012).

H.2 Assumptions: Correcting travel time reliability

Normality of the residuals

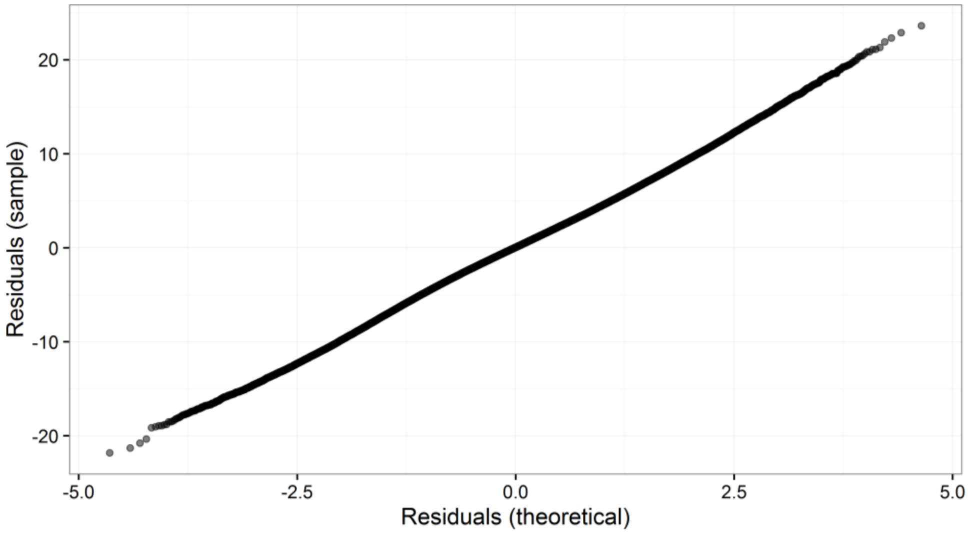


Figure H.4, QQ-plot showing the theoretical residuals plotted against the residuals in the sample.

From figure H.4 we infer the residuals are normally distributes. We thus have normality of the residuals as required by a LM (Field et al., 2012).

Independence of the residuals

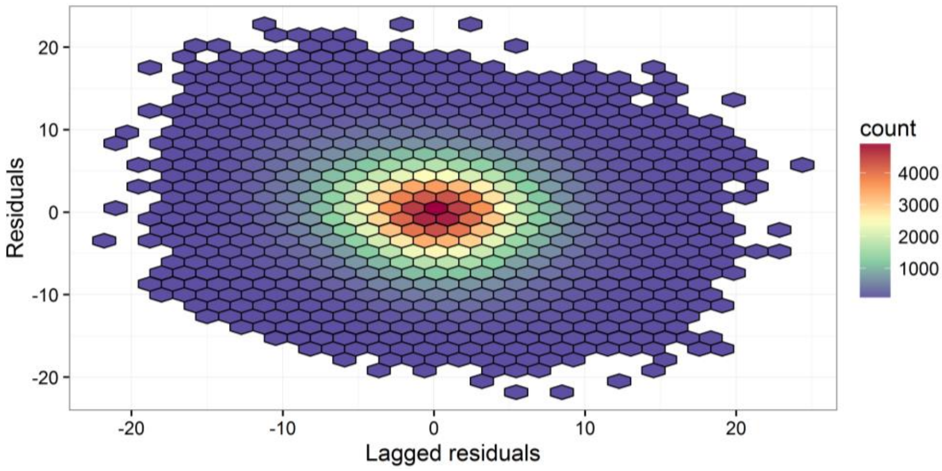


Figure H.5, residuals plotted against the lagged residuals.

In figure H.5 we see there is no relation between the current and lagged residuals. We see a nice centre of all observations that spread out evenly in all directions indicating there is no autocorrelation. Note we did apply Cochrane-Orcutt estimation prior to creating our final model, of which the residuals are shown in figure H.5. In the classic LM we did found a relation, hence the Cochrane-Orcutt estimation.

Homoscedasticity of the error variance

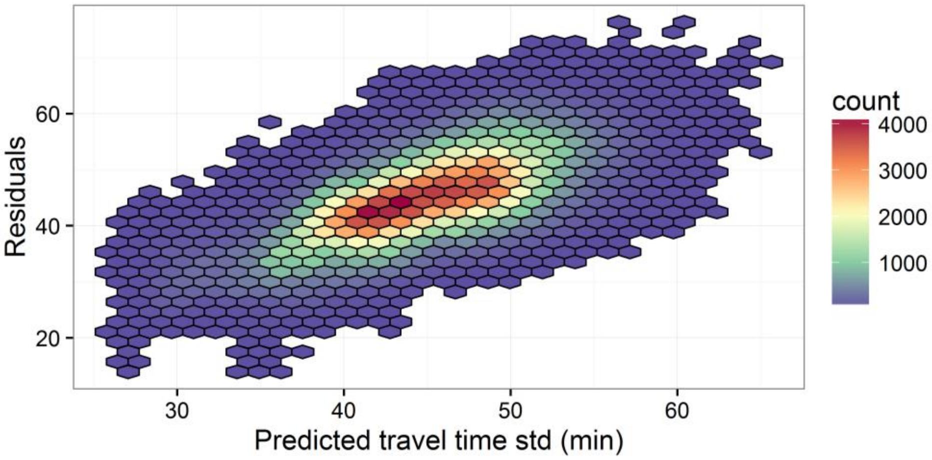


Figure H.6, residuals around the predicted travel time standard deviation in minutes.

In figure H.6 we observe there is no strange behaviour regarding the residuals over the spectrum of predicted values. Perhaps the outliers are more common towards the left and right end of the predicted values where there are less observations, but nothing too worrisome is found here.

No multicollinearity

Table H.22, VIF scores relating to our LM that predicts travel time standard deviation based on external factors.

VARIABLE	VIF-SCORE
Distance	1.54
Night	1.03
Rain * distance	1.04
Wind * distance	1.62

As shown in table H.2 all VIF scores are good, i.e. they are close to 1 and way below the threshold of 10 after which the model becomes unreliable (Field et al., 2012).

H.3 Assumptions: Correcting people on the road

Normality of the residuals

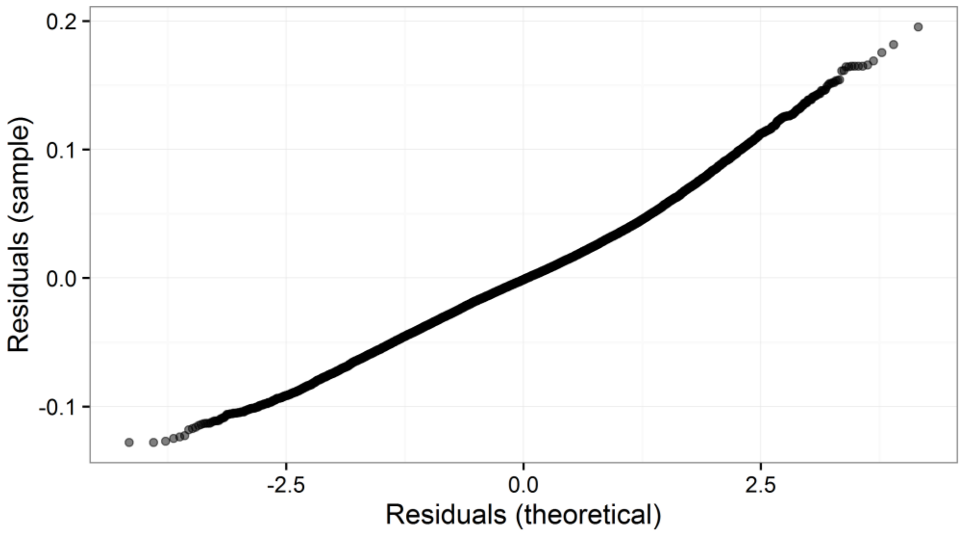


Figure H.7, QQ-plot showing the theoretical residuals plotted against the residuals in the sample.

From figure H.7 we can infer the residuals are normally distributes as all points are on a straight line. We thus have normality of the residuals as required by a LM (Field et al., 2012).

Independence of the residuals

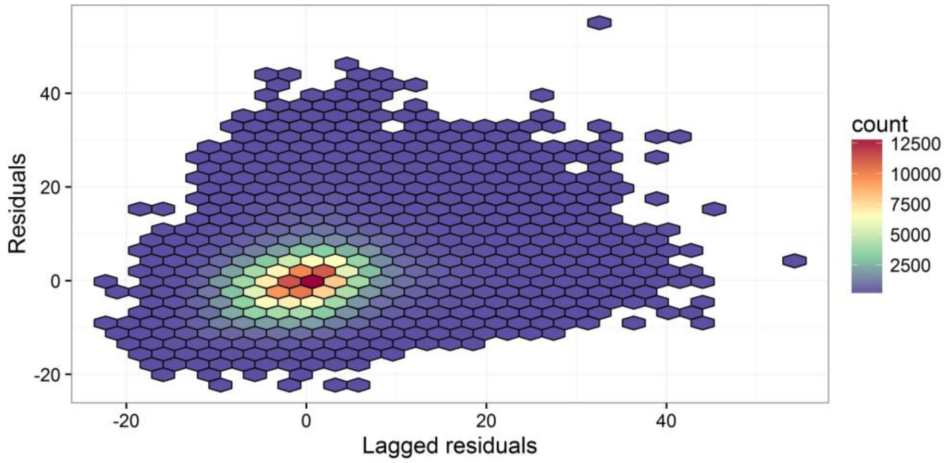


Figure H.8, residuals plotted against the lagged residuals.

In figure H.8 we see there is no apparent relation between the current and lagged residuals. A Pearson correlation coefficient between the two axis returned a value of -0.12 indicating there is only perhaps a minor relation between consecutive residuals. Despite the slight correlation we do not see any effects that would strongly bias our model. We, therefore, state the assumption of independence of the residuals is met.

Homoscedasticity of the error variance

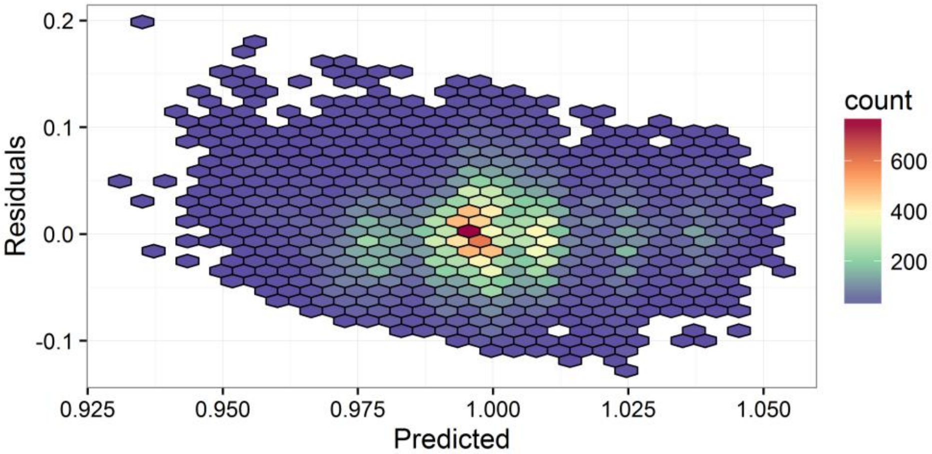


Figure H.9, residuals around the predicted ratio of people on the road during roadwork as compared to the baseline.

In figure H.9 we observe there is no strange behaviour regarding the residuals over the spectrum of predicted values. On the lower end of the predicted values we see the residuals tend to increase while they centred near zero from 0.975 onwards. This might indicate a slight relation between the residuals, though it may also be outliers. Where the majority of the observations are found, i.e. between 0.975 and 1.025, no strange behaviour is detected. We also found there is no correlation between the predicted values and the residuals (Pearson correlation coefficient of 0.0). We, therefore, do not expect the residuals to depend on the predicted values.

No multicollinearity

As shown in table H.3 all VIF scores are good, i.e. they are close to 1 and way below the threshold of 10 after which the model becomes unreliable (Field et al., 2012).

Table H.3, VIF scores of our LM to correct the number of people on the road for external factors.

ATTRIBUTE	VIF SCORE
Wind	1.12
Temperature	1.16
Rain, weekend	1.09
Rain, not weekend	1.26

Appendix I: Assumptions road closure model

Normality of the residuals

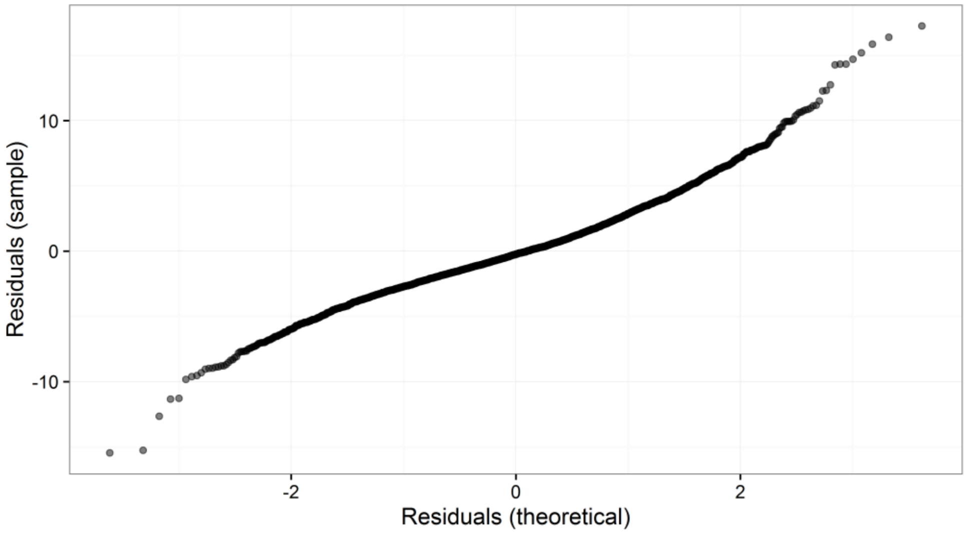


Figure I.1, QQ-plot showing the theoretical residuals plotted against the residuals in the sample.

From figure I.1 we see an approximately straight line, which would indicate the presence of a normal distribution. There are a few outliers at either end. However, even these outlier do not deviate much from the line in figure I.1. The residuals are thus normally distributed.

Independence of the residuals

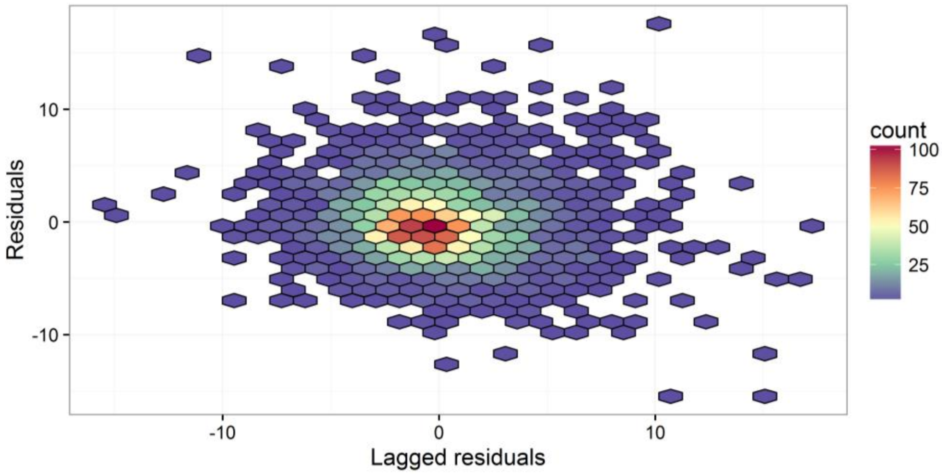


Figure I2, residuals plotted against the lagged residuals.

In figure I.2 we see there is no relation between the current and lagged residuals. We thus have independence of the residuals.

Homoscedasticity of the error variance

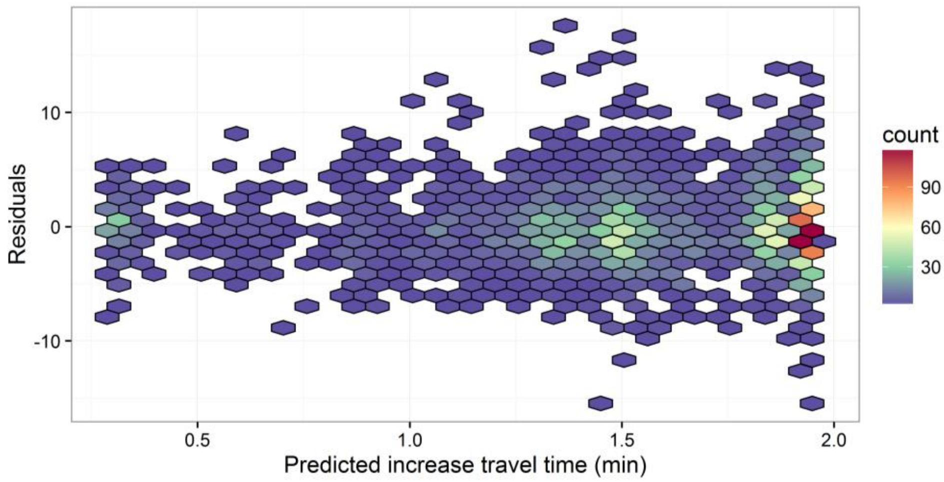


Figure I.2, residuals around the predicted increase in travel time in minutes.

In figure I.3 we observe there is no strange behaviour regarding the residuals over the spectrum of predicted values. We thus meet the assumption of homoscedasticity of the error variance.

Note we did not have to calculate VIF-scores as there is only one variable is our model.

Know Your Bias: Scaling Mobile Phone Data to Measure Traffic Intensities

Johan Meppelink, Jens van Langen, Arno Siebes, Marco Spruit

Abstract – Mobile phone data is a novel data source to generate mobility information from Call Detail Records (CDRs). Although mobile phone data provides us with valuable insights in human mobility, it is often a biased picture of the traveling population. This research, therefore, focusses on correcting for biases and suggests a new method to scale mobile phone data to the true traveling population. Moreover, the scaled mobile phone data will be compared to roadside measurements at 100 different location on Dutch highways. We infer vehicle trips from the mobile phone data and compare the scaled counts with roadside measurements. The results are evaluated for October 2015. The proposed scaling method shows very promising results with near identical vehicle counts from both data sources in terms of monthly, weekly, and hourly vehicle counts. This indicates the scaling method in combination with mobile phone data is able to correctly measure traffic intensities on highways. Nevertheless, there are still some discrepancies, for one during weekends, calling for more research. This paper serves researchers in the field of mobile phone data by providing a proven method to scale the sample to the population, a crucial step in creating unbiased mobility information.

Keywords: call detail record (CDR) data, mobile phone data, human mobility networks, trajectory data mining, roadside measurements, scaling method

1. Introduction

Samples are strongly influenced by the information present in the environment (Fiedler, 2000). In practice this means that samples are practically never truly random, which leads to biases in resulting judgements (Fiedler, 2000). Furthermore, humans often reason through heuristics explaining a simplified version of the world (Tversky & Kahneman, 1974). Although these simplifications are useful they can result in severe systematic errors (Tversky & Kahneman, 1974). To create an unbiased view of the population these biases have to be addressed and corrected for (Fiedler, 2000). In this paper we, therefore, address and provide a method for correcting structural biases in mobile phone data, i.e. mobility data generated from Call Detail Records (CDRs) of mobile providers and scale it to the traveling population.

Mobile phone data is a hot topic in the field of human mobility studies. A vast amount of recent research has been performed that investigates the use of mobile phone data to gather better insights into behaviour, social networks and mobility patterns of the masses (Daas et al., 2009; Snijkers, 2009; Ahas, Aasa, Roose, Mark, & Silm, 2008; Eagle, Pentland, & Lazer, 2009; Becker et al., 2011; Palchykov, Kaski, Kert´esz, Barabási, & Dunbar, 2012). Mobile phone data is regarded a prime candidate to replace traditional mobility measurement techniques such as surveys and roadside measurements (Cáceres et al., 2007; Astarita & Florian, 2001). Mobile phone data can provide us with mobility information on unprecedented scale (Cáceres et al., 2007; Toole et al., 2015). Where surveys provide a snapshot of the lives of a few thousand inhabitants, mobile phone data can provide 24/7 information on millions of people (Cáceres et al., 2007; Astarita & Florian, 2001). This new data source is already providing us with new and unique insights into human mobility (González, Hidalgo, & Barabasi, 2008).

Provided the increasing use of this new data source, making sure the results are unbiased also becomes more important. Studies that employ mobile phone data, however, generally pay little attention to the bias in mobile phone data while scaling the sample to the population. The latter is a must when comparing values on different scales. Scaling is in essence performed for the same purpose as calibrating in any other type of measurement device. To exemplify, mercury in a thermometer will expand when temperature increases. However, unless the thermometer is calibrated the only conclusion can be that it became warmer or colder. We do not know by how much temperature increased nor are we able compare with measurements of other thermometers.

Studies such as the one by Jiang, Ferreira and González (2015) recognize this fact. Jiang et al. (2015), for one, scale the sample to the population based on the ratio of sample to population for different geographic areas. Coole et al. (2015) take a similar approach with the extension of removing users with few events, for whom they argue no trustworthy origins and destinations can be established. Iqbal et al. (2014) built Origin Destination (OD) matrices from CDRs and calibrated their OD matrices using road side measurements. The disadvantage of calibration is having to rely on a second source of information limiting its use in practice, especially for areas where less information is available. In this study we will build upon the approaches by Jing et al. (2015) and Coole et al. (2015) by also compensating for the bias in the sample, which requires only census data unlike the calibrations such as the one performed by Iqbal et al. (2014).

In the present study we address a bias identified in the mobile phone data. The bias we identified is introduced by demographic discrepancies between the traveling population and the sample from which mobile phone data is created. The bias is due to a difference in mobile phone possession and travel behaviour across age groups (Telecompaper, 2015). To illustrate, young children have both less chance to own a mobile phone as well as a lower chance to make a trip greater than 10 km, especially during schooldays, compared to 30 to 40 year olds (Telecompaper, 2015). Without compensating for this sampling bias we would get a biased view of the population, where especially age groups that travel less are underrepresented. Thus, when employing mobile phone data to answer mobility questions one needs to adjust for the bias to prevent erroneous conclusions.

In this research we will present and evaluate a method to scale the sample in mobile phone data to the population. To validate our proposed scaling method we will estimate vehicles present on a multitude of Dutch highways and compare the outcomes to roadside measurements. The advantage of comparing to roadside measurements rather than Origins and Destinations (ODs) from surveys is that the roadside measurements generally produce unbiased and highly accurate traffic counts (Technical University Delft, 2006; Nihan, Wang, & Zhang, 2002). The main goal of this study is to provide a validated scaling method for mobile phone data that could be used in future studies to get an unbiased view of the population.

In this paper we will adhere to the following structure. In section 2 we discuss the study area and provide a description of the data used in this study. Section 3 will elobarate upon the methodology used in this research. Here also our scaling method is presented. Section 4 encompasses the results and a discussion of these results. Finally, the conclusions of this research are presented in section 5.

2. Study area and data description

The data available for this study covers the Netherlands as a whole. The Netherlands currently has 17 million inhabitants and spans 41.526 square km (CBS, 2003). The country is approximately 200 km wide and 300 km tall and shares borders with Germany and Belgium. In 2015 80% of the population had a mobile phone of which the vast majority are smartphones (Statista, 2016a; Statista, 2016b). The latter is relevant as the more people connect to the network, e.g. for calling but also receiving e-mail and browsing the web, the more events are generated and thus the more data points we have per inhabitant. This trickles down to better mobile phone data.

In the following sections we will elaborate on the CDRs, census data, survey data, and roadside measurement data used in this study.

Call Detail Records & cell tower network

CDRs are the basis of mobile phone data. These are records describing when a mobile phone connected to the network by sending or receiving voice, text, or other data via a provider's network. The records consists of a time stamp, a cell code relating to a cell tower in the network and a one-way hashed id created from a mobile phone number. At the time of writing about 370 million CDRs are generated per day by 3 million subscribers.

The network of cell towers of the provider consists of approximately 50.000 unique cells covering every part of the country. Geographic characteristics of the cell towers, i.e. their location and their radius, angle, and orientation are also provided by the provider. These geographic properties in combination with the CDRs enable us to pinpoint people at a certain moment in time.

The data used in this study covers October 2015.

Census data & geographic zones

Census data forms the basis of the scaling method we present in this research. As previously explained, there is a bias in the sample mobile phone data provides due to demographic discrepancies. Census data provides us with crucial information about the population that allows us to assess our sample in comparison to the true population. In the Netherlands the Central Bureau of Statistics (CBS) reports census data. Age distribution on four digit postal code level is acquired from the CBS.

Where roadside measurement data as used by Iqbal et al. (2014) might not be available everywhere, census data nearly always is. In addition, census data is often gathered in task of the government and is hence typically freely available (CBS Isreal, 2016). As the census data is widely available and often accurate, the scaling method, which mainly relies on this source of information, could be implemented not only in the Netherlands, but also in many other countries all over the world. This greatly benefits the generalizability and enhances the overall value of the proposed method.

Mobility survey data

In essence the mobile phone data measures the movement of mobile phones and thus the people carrying the devices rather than the vehicles on the road. For a fair comparison, a translation is required to go from people to vehicles.

To get a good estimate of the number of people per vehicle, survey data from 5 years of Onderzoek Verplaatsingen in Nederland (OViN) are used, starting at 2010 and ending at 2014. OViN is chosen as it is the largest free nationwide mobility survey in the Netherlands (CBS, 2010). When combined, the surveys contain 97.432 trips that are comparable to those in the mobile phone data. To be able to compare the data, we only selected trips 10 km or longer as we are interested in trips on highways, which are often longer than 10 km. Furthermore, we only selected trips with average travel velocities below 145 km/h, deeming them unrealistic and untrustworthy. For both calculation the distance taken is the distance as the crow flies from the origin postal code to the destination postal code.

An overview of the distribution of trips greater than 10 km per mode of transportation is shown in table 2.1. The two largest groups by a margin are car (driver) and car (passenger). The third largest group bus (public transport) accounts for nearly 5% of all trips. The chance of being a driver for this class (11%) and the class motor is extracted from literature on Dutch public transport (Otten, Hoen & Den Boer., 2014). Bus (private) is assumed to be similar to bus (public transport). All other classes combined including taxi and freight truck only account for just over 1% of all trips. The assigned chances of being a driver for these classes are based on personal experience. Given their low share amongst all trips, the possible impact of mistakes due to guess work is considered to be negligible.

Table 2.1, different modes of transportation, their prevalence on the Dutch roads and the change of being a driver.

Means of transportation	% of trips	Chance of being a driver
Bus (public transport)	4.93%	11%
Bus (private)	0.57%	11%
Camper	0.05%	50%
Car (driver)	43.67%	100%
Car (passenger)	17.09%	0%
Delivery van	0.52%	95%
Motor	0.44%	87%
Taxi	0.44%	50%
Freight truck	0.06%	100%
Other / not on highway	32.23%	-

We find the motive of a trip, rather than hours of the day or day of the week, can provide very stable people per vehicle ratios. In figure 2.1 the people per vehicle ratio, i.e. the inverse of the chance of being a driver, is depicted for each of the three motives, i.e. work, business and other, and hour of the day. As can be observed from figure 2.1, the ratio people per vehicle is stable over the majority of the day. The unstable pattern in the early morning can be attributed to the small sample size in the early hours of the day. Note that we included the motive "other" twice, once for workdays and once for weekends. We did so because there is a clear difference in people per vehicle between the two day types for this motive. During the weekend there are generally more people per vehicle for non-work and business related activities than during workdays. The people per vehicle ratios applied are 2.02 for other weekend, 1.64 for other workdays, 1.08 for business, and 1.10 for work trips.

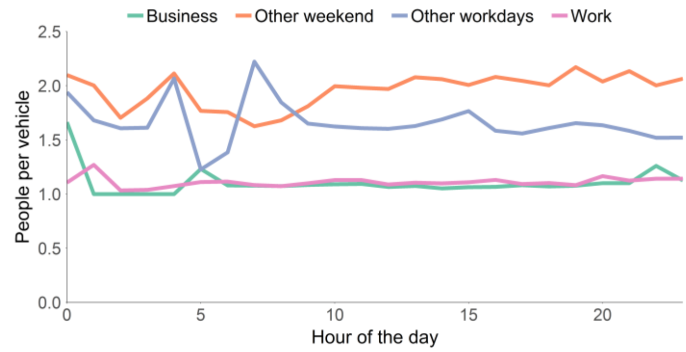


Figure 2.1, people per vehicle ratio plotted for different trip motives averaged per hour of the day.

Mobile phone penetration

Information about the mobile phone penetration by age group was acquired via two sources. The first being Telecompaper (2015), which contains the majority of the information used. Gaps, for people below age 12 and over 80, are filled by data from Offermans et al. (2013). The latter refers to a study by CBS who assessed the representativeness of the sample of the mobile phone data used in this study, though with a two year gap. They also included information about the penetration of people having a smartphone at the telecom provider by age. This information, however, is not always available and hence we chose to use data from Telecompaper (2015) who provide a more nationwide assessment of smartphone penetration by age group. Moreover, they also provide information on a large variety of countries (Telecompaper, 2015).

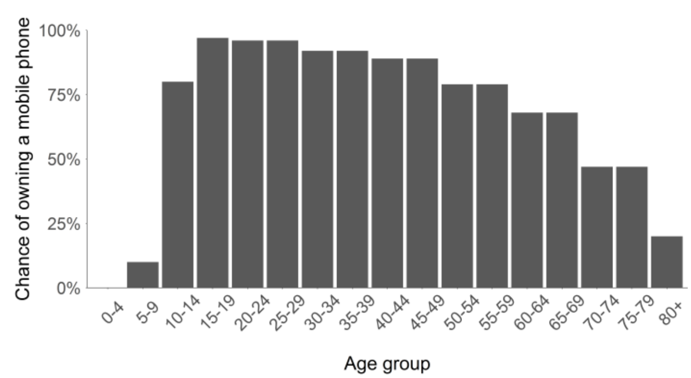


Figure 2.1, chance of owning a smartphone by age group.

Roadside measurement data

In order to validate our scaling method we compared vehicles inferred from the mobile phone data with vehicles measured on the road. The number of vehicles measured on the road is obtained via road side measurements provided as open data by Nationale Databank Wegverkeersgegevens (NDW). NDW is a governmental organization in the Netherlands that collects the measurement data from different parties such as Rijkswaterstaat. Most of the roadside measurements are collected by using an inductive-loop measurement devices placed on or in the road's surface with a self-reported accuracy upwards of 99% (NDW, 2015a; NDW, 2015b; NDW, 2015c). An overview of the types of roadside measurement devices, the number of occurrences on the Dutch road network, and the mean self-reported accuracies of each type of device is shown in Table 2.2 (NDW, 2015b).

Table 7.1, roadside measurement devices in the Dutch road network.

Measurement device	Counts	Accuracy
Inductive-loop vehicle detector	21.711	99%
Automatic Number-Plate Recognition	1.484	95%
Bluetooth	1.409	Unknown
Infrared	948	100%
Floating Car Data (from navigation systems)	24	Unknown

The information represented in table 7.1 covers all measurement sites in the Netherlands, including information on the use of parking lots and gas station et cetera. Moreover, the measurements from these measurements sites contain predominantly raw data. For major roads Rijkswaterstaat cleaned the raw data by, for one, removing outliers. The algorithm processing the raw data is called Monibas, an algorithm that is proven to be highly accurate (Technical University Delft, 2006). The data processed by Monibas are also included in the raw data as a separate measurement sites. In total there are 13.693 measurement sites to which Monibas is applied, all of which employ raw data from inductive-loop vehicle detectors.

3. Methodology

Our methodology encompasses the steps taken to evaluate our scaling method. Part of the methodology is also the scaling method we present in this research. There are two main processes leading up to the final comparison of vehicle counts. These are represented in figure 3.1. On the top are three steps displayed regarding mobile phone data. This goes from creating the mobile phone data, to scaling the data to the population and vehicle counts on the road. At the bottom the steps are shown relating to processing and filtering the roadside measurement data. The latter is done to ensure only the most trustworthy sites are used in our final comparison in which vehicle counts from both sources are compared.

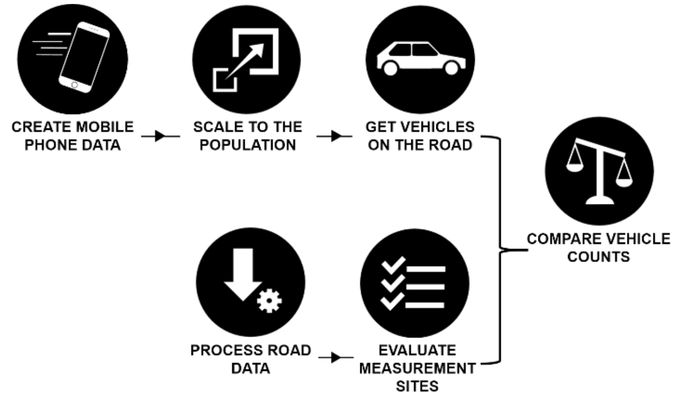


Figure 3.1, overview of the methodology applied to compare vehicle counts from scaled mobile phone data and roadside measurement data.

Creating mobility data

The basis of the mobile phone data is the definition of a destination. A destination we define as a place where a person resides for 30 minutes or more. Furthermore, we divided the Netherlands in a total of 1.259 areas each of which could be a destination of a trip. These areas are created from census data in such a way that each area is defined as a municipality. Large municipalities such as Amsterdam are divided by hand based on four digit postal codes to provide more detailed information. Whether a person is located within one area depends on whether that person had events with cell towers covering the area for at least 30 straight minutes. A more detailed description of the algorithm used to extract origins and destinations from CDR data can be found in the studies by Keij (2014) and Van Kats (2014).

There is, however, one noteworthy change between the algorithm Keij (2014) and Van Kats (2014) used versus the one used to create mobile phone data for this study. We proposed to discard the CDRs that occurred with cell towers having a radius of 12.5 km and higher. Cells with a radius larger than 12.5 km provide only very limited information about a person's location. In addition, due to the design of the algorithm, tends to increase noise in recorded movements. The events handled by cells larger than 12.5 km are the minority and as such discarding them still leaves us with 94% of the total events (table 3.1).

Table 3.1, percentage of events compared to the maximum allowed cell size

Cell radius up to (km)	Percentage of events
2.5	60%
5	77%
7.5	86%
10	91%
12.5	94%
15	95%

To determine the impact of discarding cells larger than 12.5 km we compared the destinations in the mobile phone data with a GPS trace. To make sure that the data was comparable, we created destinations from the GPS data with a destination being defined as a circle of 5 km in diameter in which a person has to reside for 30 minutes or more. We observed a total of 52 trips. The GPS trace covers the entire month of February 2015. The analysis will thus cover 28 days of measurements. The GPS trace is from one of the employees of Mezuro, i.e. the company who creates the mobile phone data used in this study. Typically, there are privacy limitations that imply we can only observe aggregated trips when there are at least 16 unique individuals involved. For the employee of Mezuro the privacy limitations are lifted upon formal request to allow us to perform these types of analyses. In total we compared three different sets of mobile phone data: with all events, only those with cells smaller than 12.5 km and cells with smaller than 10 km.

By taking only cells with radii smaller than 10 km and 12.5 km we find the algorithm can correctly determine 92% and 96% of the destinations, respectively. Using cells of all sizes we came to a mere 72% with the majority of the errors consisting of assigning a destination to a neighbouring area. Thus, setting a threshold on cell radius helps improve accuracy. Furthermore, 12.5 km performs better than 10km as a threshold. Plausibly this is because more events are removed with the 10 km threshold, also those that help improve the mobile phone data's accuracy.

When compared to OViN we see a similar distribution of the number of trips on longer distance classes, but diverging trip counts for trips below 10 km (figure 3.2). The comparison with OViN is made by setting the total trips over 26 km as a baseline (100%) and comparing this to the total number of trips including shorter distances. 26 km is chosen as the largest destination in our mobile phone data is 26 km in diameter. Hence, a person is able to travel at most 26 km without being recorded traveling.

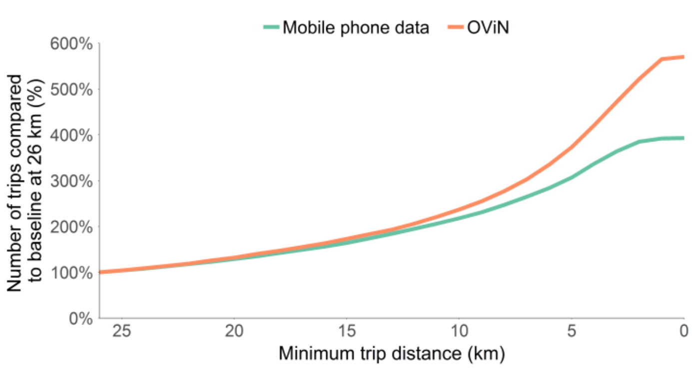


Figure 3.2, comparison between OViN and Mezuro of the relative number of trips per travel distance. Both datasets have their number of trips set to 100% at a travel distance of 26 km.

We find there are about as many trips in both datasets over 10 km, though below 10 km OViN appears to record far more trips. Hence, we choose to only take trips over 10 km into account as we have an underrepresentation below 10 km. This is also a bias that is hard to correct for provided it depends on the physical cell tower infrastructure. In areas with larger reaching cell towers one might never be able to detect short trips, whereas in cities with smaller cells one might. Hence, we only look at trips over 10 km and focus on measuring mobility on highways, where trips are typically greater than 10 km.

Scaling mobile phone data

The scaling method comprises both a bias correction as well as expanding the counts to approach those in the true population. This method corrects for demographic differences in the population that affect the probability that a person is recorded in our sample when performing a trip. The probability of being observed traveling depends on two factors: (1) the chance of being in the sample and (2) the chance of traveling. Both these factors are dependent upon inhabitant demography and type of day. Children under the age of 10, for example are both less likely to own a mobile phone and to be found traveling on highways (Telecompaper, 2015). Consequently they are both less likely to be present in the sample and less likely to be found traveling over 10 km. Moreover, the chance children are traveling longer distances differs between schooldays, holidays and weekends. The likelihood of a person taking a trip of at least 10 km or more for a range of age groups is shown in figure 3.3. Note we focus on trips over 10 km as trips below 10 km are underrepresented in the mobile phone data. Hence, we choose not to use that information. From figure 3.3 can be observed there are strong differences in travel behaviour between age groups. Our scaling method is designed to take these effects into account and adjust trip counts accordingly.

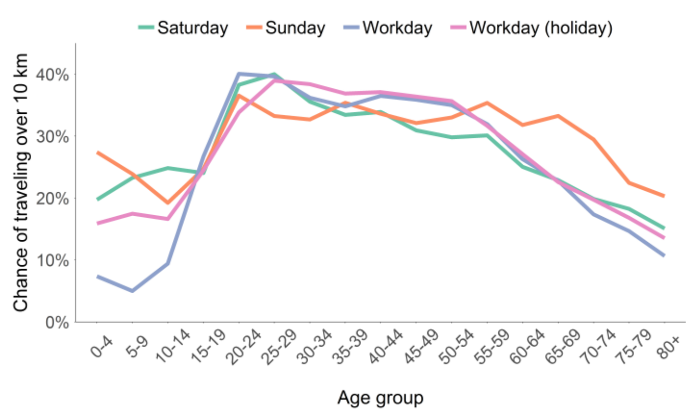


Figure 3.3, chance of observing a person making a trip over 10 km during a day depending on the age group and type of day.

We make a distinction between four types of day. These are workdays, Saturdays, Sundays, and workdays during holidays. These day types provide insights into the distinct travel behaviours, with differences mainly observed for the young and old inhabitants. Further distinctions in day types did not provide additional information and would decrease the number of observations per group to the point where the group sizes are too small to extract stable and trustworthy values.

The proposed scaling method is formally represented in a Process Deliverable Diagram (PDD) (figure 3.4). In a PDD the processes are shown on the left and the product of the action on the right (Van de Weerd & Brinkkemper, 2008). The PDD is accompanied by two tables, one describing the processes (Appendix A, table A1) and one describing the products (Appendix A, table A2). The next paragraphs will provide a brief overview of the steps and products in the scaling method.

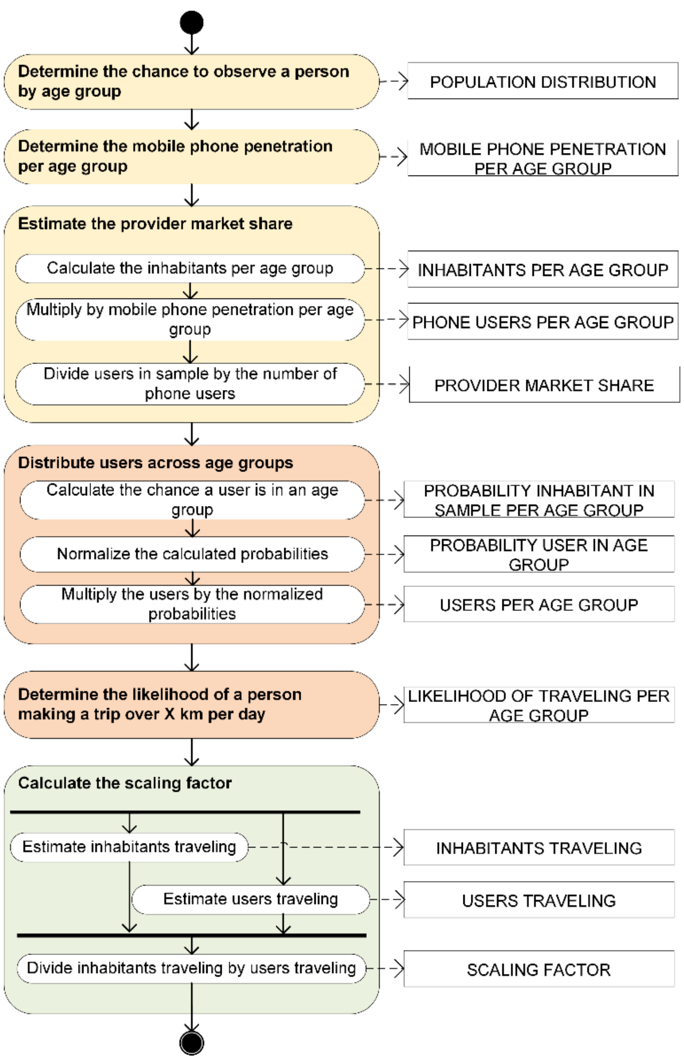


Figure 3.4, a PDD showing the scaling method that will provide a scaling factor used to expand the sample to the population..

As shown in the PDD with the different colour codes there are three types of steps in the scaling method. The first is represented with yellow concerns acquiring information about the population for all relevant areas. In addition we also estimated the number of mobile phone users living in each area by assigning a home location to each user. This is done using a slight variation of the home assignment algorithm described by Toole et al. (2015). Whereas Toole et al. (2015) assigned home as the most frequent stay location between 8 pm and 7 am on workdays we also include data about weekends, but the differences are minor. A user is assigned one home location for the month October 2015. Home locations are key to determine the distribution of users across the country and hence the ratio mobile phones in the sample over the inhabitants for all areas.

The second step in our method shown in red is the difference between traveling and general population. Here we first estimated the age distribution within our sample by looking at the a-priori information about the population distribution in the area's population and the chance of owning a smartphone per age group. This helped us to calculate the number of users per age group. For each of these age groups we can predict how likely they are to be found traveling for a specific day type.

The final step is what differentiates our scaling method from those in previous studies. Here we calculate how many inhabitants and mobile phone users in our sample are expected to travel over 10 km. In the end we divide the inhabitants we expect to be traveling by those we expect to be traveling within our sample. The calculated ratio will be the scaling factor for a specific area on a specific day.

Validating the scaling method using roadside measurements

There are four essential steps required to measure vehicles on highways using mobile phone data: (1) scale the mobile phone data to the population, (2) focus on the vehicle trips, (3) select the vehicle trips going over a specific road, and (4) convert trips into vehicles. The first step, i.e. scaling, was covered in the previous section and will not be discussed here. The latter three will be discussed in the following paragraphs.

The algorithm by Keij (2014) is implemented to distinguish train trips from vehicle trips. Note that there might also be other categories, e.g. cycling, that are not included. This is a slight shortcoming, though given our focus on longer trips in combination with highways we expect the majority not to be cycling. The basis of the algorithm is detecting whether events occur with cells covering only highways or only rail tracks (Keij, 2014). Especially on longer distances the algorithm is shown to provide good results (Keij, 2014).

We are interested solely in a subset of all car trips. These are the ones passing over specific road section where also roadside measurement data is available. Hence, we need to make an educated guess on the route people travel from their origin to destination. Once we got a good idea of the routes travelled we can select all origin destination combinations that cross a specific road section. This provides an estimate of all people who travel over the road network and plausibly cross a specific road. Next, we will assign a time stamp to these trips defined as the middle of their trip, e.g. trips leaving at 8:00 and arriving at 9:00 will cross the road section at 8:30. This is not very precise, but as some trips will in reality cross the road section earlier and other later we are confident the majority of the uncertainty will level out.

A major step in the above process is acquiring routes for origin destination pairs. There is a large volume of published studies describing how to assigning vehicle trips from Origin Destination pairs (OD-pairs) (Prato, 2009; Ortúzar & Willumsen, 2011). A key assumption that is often made is that people are rational and take the route that minimizes their travel cost (Ortúzar & Willumsen, 2011). Travel cost can be seen as a combination of multiple factors such as travel time, distance, cost of fuel, congestion charges, et cetera (Ortúzar & Willumsen, 2011). The most important factors, explaining 60% to 80% of all route choices in practice, are travel time and distance (Ortúzar & Willumsen, 2011). Methods taking the shortest path or k shortest paths as the possible route choices between OD-pairs account for the largest group of path generation methods (Prato, 2009). At the moment the dataset available for this study uses Dijkstra's shortest time path algorithm to link OD-pairs to road sections (Dijkstra, 1959). The shortest time path is chosen by time rather than distance. This is done by taking into account the maximum speed allowed on a road section based on the information from Open Street Maps (OpenStreetMap contributors, 2014). We provided a 20% discount for travel times on highways provided people prefer to stay on highways when the time benefits are minimal (Ziebart, Maas, Dey & Bagnell, 2008). The 20% will account for this.

The conversion factor is applied as calculated in section 2.3. Note here we said the people per vehicle ratio depends on trip motives. Trip motives are inferred from the mobile phone data based on trip characteristics (Van Langen, 2016). A detailed description of the algorithm used can be found in Van Langen (2016). After applying the algorithm to our dataset we could simply use our conversion factors to go from trips to vehicles.

Prepare roadside measurement data

Raw data is provided by NDW (NDW, 2015b) in xml format and converted to a manageable csv using software written in Python (Van Rossum & Drake, 1995). The csv contains information about the measurement site as a whole for 15 minute periods, e.g. the average vehicle count or average velocity. In addition, it contains information about the minimum and maximum vehicle counts and velocities as well as information on the number of trucks passing by. To keep data manageable, information about the independent lanes are aggregated to a single vehicle count.

Filtering untrustworthy roadside measurement data

To ensure data quality we constructed two criteria the roadside measurement sites have to meet before taking them in contention for the comparison with mobile phone data.

The first criteria is that a 15 minute interval may at most lack 6 minutes of erroneous data within the roadside measurement data. Standard deviations of on average 4.5% between differences in vehicle counts are found by comparing values from consecutive measurement sites, i.e. where no off-ramp or on-ramp separates the measurement points. The latter is done such that the same amount of vehicles are destined to travel by both measurement sites. As a result the difference in reported vehicle counts can only be explained by errors in measurement. The 4.5% is found by comparing consecutive measurement sites at 30 locations. When the error in measurement of each site is normally distributed and equal amongst sites, the standard deviation for each measurement site becomes approximately 3.2% following the variance sum law. This is very acceptable and provides us with a good indication of how accurate the roadside measurements are.

The second criteria is slightly more complicated and relates to our suspicion that the location of measurement sites are not always accurately represented in the database. To test whether sites are on the reported location we performed a test based on the following premise: measurement sites that are located on the same road near each other will produce more similar vehicle counts than measurement sites further apart. The idea is to (1) cluster measurement sites based on their vehicle counts, here we opted for k-means clustering, and (2) check if sites within each cluster are located on the same road, e.g. the A4, as reported by the NDW (Hartigan & Wong, 1979).

For clustering we used vehicle counts of 145 hours during October 2015 where all 4.775 sites are without missing data. The vehicle counts are then normalized per hour. For clustering we set the number of clusters equal to 240, leaving us with approximately 20 measurement sites per cluster. The results are shown in figure 3.5. We observe the majority of the clusters have over 70% of the measurement sites on the same road and also headed in the same direction, e.g. North. The 70% we use as a threshold, when less than 70% of the sites on one cluster are located on different roads we are unsure about their locations. Only the sites belonging to clusters with the majority (70%) of the measurement sites on the same stretch of road meet our second criteria.

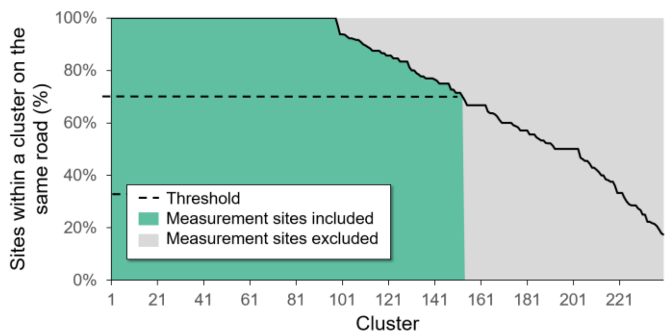


Figure 3.5, percentage of sites at the same road and on in the same direction for all 240 clusters.

In the end both criteria leave us with 1.761 measurement sites. The majority are discarded for having more than 6 minutes of erroneous data within 15 minute intervals for 95% of the recorded hours. These measurement sites are less accurate in general and hence are not the gold standard we are looking for. The final 140 measurement sites are discarded due to the second criteria as we are not sufficiently confident that their reported location is correct.

Comparison of vehicle counts

The final comparison will compare vehicles measured from the scaled mobile phone data in contrast to the vehicles recorded by the NDW, i.e. the roadside measurement data.

In total 100 comparison sites are nearly randomly selected. We put one constraint on the measurement sites to ensure we get a better sample of the Netherlands in general. As the government is most interested in traffic intensities on busier roads, there are many more data points near and between the larger cities in the Netherlands. If we would randomly select measurement sites, the majority would be near and in between these cities. Hence, we ensured there are no measurement sites on the same road, headed in the same direction, within 10 km from other measurement sites in the sample. Sites are randomly drawn from the 1.761 measurement and added to our sample if there is no overlap within 10 km with other sites in the sample. The resulting measurement sites are depicted in figure 4.1.

Vehicle counts will be compared on measurement site, weekday, and hour level. The main metric using in the analysis is the 'Extra vehicles on the road by NDW (%)'. This is a value is calculated as described in equation 1, where E stands for 'Extra vehicles on the road by NDW (%)', NDW is the number of vehicles measured via roadside measurement devices, and M stands for vehicles inferred from the scaled mobile phone data.

$$E = \frac{NDW - M}{NDW}$$

4. Results

The first analysis we performed is on measurement site level. In figure 5.1 the extra vehicles on the road by NDW (%) is represented for all 100 measurement sites. The majority of locations show the vehicle counts are nearly equal between the two data sources.

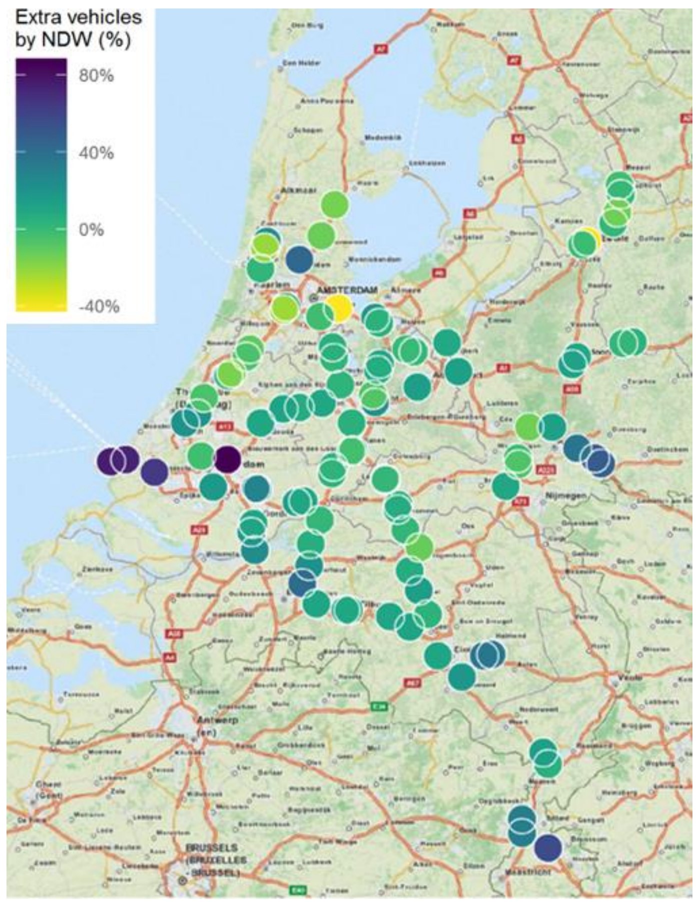


Figure 4.1, extra vehicles measured by NDW (%) for all 100 selected comparison locations.

Unfortunately, there are differences in vehicle counts on several locations on the Dutch highways. The mobile phone data underestimates the number of vehicles on the road near the country's borders. Where the country borders Belgium and Germany we see darker bluish colours indicating there are less vehicles measured via mobile phone data. This may be due to people switching off their phones while being or going abroad, resulting in an underrepresentation of the sample.

Furthermore, on the left there are a few dark spots where land meets ocean. Here the routing algorithm in combination with the relatively low spatial resolution is the culprit. There are few routes crossing roads near the country borders and hence few trips will be assigned to these roads. With the current dataset the mobile phone data will not provide good estimates of traffic on roads near the border, which is a limitation we have to take into account. Nevertheless, it does work very well in the majority of the country.

Finally, we see two outliers in the form of one dark dot near Rotterdam and two yellow dots, one near Amsterdam and one near Zwolle. Provided they are surrounded by measurement sites that do show good results we can only imagine some bad roadside measurement sites slipped through our filters or the routing algorithm made some very local mistakes. In the further analyses the measurement sites where the the extra vehicles on the road by NDW (%) is over 20% or below -20% are removed either because they are outliers or because they show deviations not related to faults in the scaling method under investigation. This reduces our sample by 26 locations to a total of 74.

The second comparison made relates to differences between different weekdays. Each datapoint in figure 5.2 represents the extra vehicles measured by NDW (%) per weekday per comparison location. The information is visualized in box whisker plots with outliers shown as dots outside the box whisker plots. Outliers here are defined as values deviating more than 1.5 times the Inter Quartile Range (IQR). For weekdays the scaled mobile phone data shows solid results. During Mondays and Tuesdays especially the number of vehicles inferred from the mobile phone data near perfectly matches the number of vehicles measured by roadside measurement devices. There is some variation, although this could also be the result of deviations from the shortest time path in the applied routing algorithm.

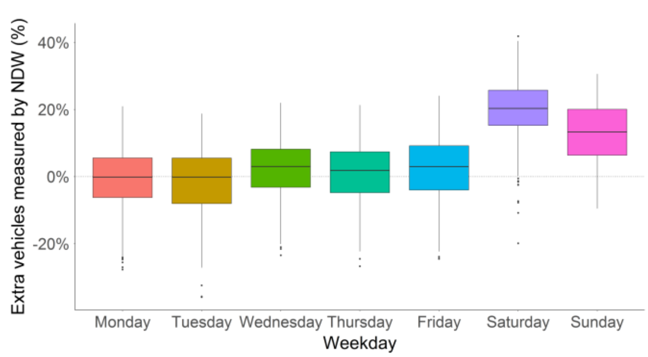


Figure 4.2, differences in extra vehicles measured by NDW (%) per weekday.

What is striking though, is the large deviations during weekend days. On Saturday especially the extra vehicles measured by NDW (%) is more often than not over 20%. For this we do not yet have a solid explanation, merely a view hypotheses. The most logical hypothesis, also in the trend of this paper, would be a bias in our sample. We expect there are people who use their mobile phone only for work purposes. These people might leave there phone at work or at home during the weekend. As we would expect some of these mobile phone would be traveling we overestimate the expected number of trips. As a result the scaling factor, i.e. the ratio expected traveling inhabitants over mobile phones traveling, is lower than it should be during the weekend. Alternatively, this may have to do with a mistake in our people per vehicle ratio, which is much higher during the weekend. There are less working people, and much more recreational traffic where OViN tells us there are more people per vehicle on average. The exact reason for the lower counts during weekends will have to be investigated in future research.

The final comparison we made is between hourly vehicle counts. This is done to get a grasp on how vehicle counts compare by hour. For this we selected a week in October 2015 and aggregated the counts per data source on hourly level. The results are shown in figure 4.3.

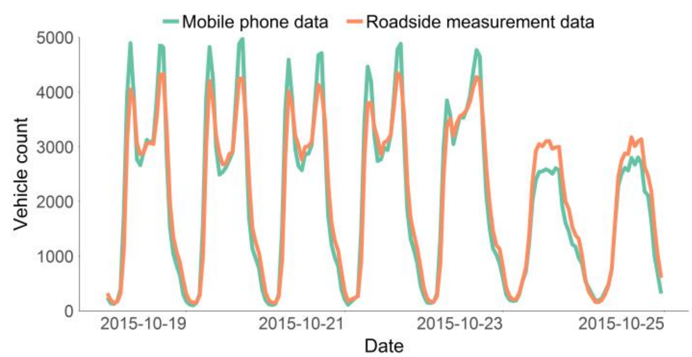


Figure 4.3, aggregated vehicle counts for one week in October 2015 for both data sources.

On hourly level the vehicle counts between both data sources follow the same general pattern indicating that not only on weekday level, but also hour level, the scaling method correctly scales sample to the population. In fact on hour level the Pearson's correlation coefficients are consistently above .87 and are .95 on average. Nevertheless, there are some discrepancies. The vehicle count during weekend days, as we know from the previous analysis, is slightly too low in the mobile phone data. This, however, appears to primarily affect the day time. Night time values for both work and weekend days are very evenly matched. Moreover, there appears to be a structural overestimation of vehicle counts during workday rush hours. Rush hour peaks during workdays are higher in the mobile phone movement data on every workday with the differences being approximately 20%. Note, however, during hours surrounding the peaks there are more vehicles measured via roadside measurements. Our best guess is that our time stamping technique, i.e. middle of the trip assignment, is to blame. When people get stuck in traffic and travel times increase our time stamping technique becomes less accurate. When people leave work at 5:00 and get home by 7:30 we guess they passed a road section by 6:15 while in fact this may be much earlier or later. For each specific point the delay may be skewed one way or the other. In our aggregation this would lead to more spread and lower peaks in traffic intensities as observed in figure 4.3.

5. Conclusion

The results of this research are very promising. With the exception of road segments near the country border, the vehicles inferred from trips in the mobile phone data matches those from roadside measurements. This proves the scaling method we present and implement in this research allows us to measure mobility patterns and traffic intensities on highways using mobile phone data. Nevertheless, there is still some work to do with a slight underestimation of traffic intensities during the weekend and overestimation during workday peak rush hour traffic.

In this research we aimed to complement the vast amount of studies investigating and reporting the potential and possible applications for mobile phone data (Cáceres et al., 2015; Iqbal et al., 2014; Jiang et al., 2015; Ahas et al., 2008). We did so by focussing and refining a crucial step in translating the sample to the traveling population. We identified a bias in that people more likely to travel are also more likely to own a phone, leading to potential overestimations of the traveling population. To correct for the bias we devised and presented a new scaling method that can be implemented in future studies. Using this scaling method showed mobile phone data can accurately measure traffic intensities without post calibration with, for example, roadside measurements. The latter implies the created scaling method is able to explain and account for the majority of biases in the data.

The results of this study provided answers, but also introduced new questions to investigate. During the daytime on weekend days, for one, the scaled mobile phone data appeared to underestimate the roadside measurements by 5% to 25%. Our best guess is that business users will have a separate phone for private use and will leave their work phone at home or at work during the weekend. This might explain why we underestimate trips during the weekend. Alternatively, the scaling method might be correct and the people per vehicle ratio extracted from OViN, which is much higher during the weekend, might cause this. How to correct for the bias in the weekend is to be answered by future research. Furthermore, mobile phone data has an underrepresentation of trips on shorter distances. Hence, we focused on highways and only included trips over 10 km. Although some research has been performed on this topic, we think it is worthwhile to further investigate this and add the required steps to correct for the bias in the scaling method proposed in this research.

The proposed and evaluated scaling method presented in this research can help us to get a better grasp of the traveling population. Demographic biases are accounted for leaving future researchers and policy makers with a less biased view of the population allowing them to make better judgements. When traffic models are created or calibrated with mobile phone data the scaling method could help to add provide more accurate information. Furthermore, the scaling method relies solely on information that is often widely available. When there is no other data source for calibration, e.g. in underdeveloped parts of the world, mobile phone data can still provide proper insights in the population's mobility patterns.

References

Ahas, R., Aasa, A., Roose, A., Mark, U., & Silm, S. (2008). Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. Tourism Management, 29 (3), 469–486.

Astarita, V. & Florian, M. (2001). The use of mobile phones in traffic management and control. In Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE (pp. 10-15). IEEE.

Becker, R. A., Cáceres, R., Hanson, K., Loh, J. M., Urbanek, S., Varshavsky, A., & Volinsky, C. (2011). A tale of one city: Using cellular network data for urban planning. IEEE Pervasive Computing, 10 (4), 18–26.

Cáceres, N., Wideberg, J. P., & Benitez, F. G. (2007). Deriving origin destination data from a mobile phone network. Intelligent Transport Systems, IET, 1(1), 15-26.

CBS (2003, December 22). Ontwikkeling bodemgebruik in Nederland 1996-2000. Retrieved May 14, 2016, from https://www.cbs.nl/nl-nl/achtergrond/2003/52/ontwikkeling-bodemgebruik-in-nederland-1996-2000

CBS Isreal. (2016). Censuses Around the World. Retrieved May 11, 2016, from http://www.cbs.gov.il/census/census/pnimi_page_e.html?id_topic=5

Centraal Bureau Statistiek (CBS). (2011). Onderzoek verplaatsingen in Nederland 2010. Retrieved from http://www.cbs.nl/NR/rdonlyres/D1D6A7E6-A1C94329-9908-70EA47E1D8F6/0/2010ovinonderzoeksbeschrijvingart.pdf

Daas, P., Roos, M., de Blois, C., Hoekstra, R., ten Bosch, O., & Ma, Y. (2011). New data sources for statistics: experiences at statistics Netherlands.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. Numerische mathematik, 1(1), 269-271.

Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. Psychological review, 107(4), 659.

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. L. (2008). Understanding individual human mobility patterns. Nature, 453(7196), 779-782.

Hartigan, J. A. & Wong, M. A. (1979). A K-means clustering algorithm. Applied Statistics 28, 100–108.

Iqbal, M. S., Choudhury, C. F., Wang, P., & González, M. C. (2014). Development of origin–destination matrices using mobile phone call data. Transportation Research Part C: Emerging Technologies, 40, 63-74.

Jiang, S., Ferreira Jr, J., & González, M. C (2015). Activity-Based Human Mobility Patterns Inferred from Mobile Phone Data: A Case Study of Singapore. In Int. Workshop on Urban Computing.

Keij, J. (2014). Smart phone counting: Location-Based Applications using mobile phone location data (Unpublished master's thesis). Delft University of Technology, Delft, The Netherlands.

Nationaal Databank Wegverkeersgegevens (NDW). (2015a). MST - Dit bestand geeft aan waar de NDW meetlocaties zich bevinden. Nationaal Databank Wegverkeersgegevens. Retrieved from http://www.ndw.nu/documenten/

Nationaal Databank Wegverkeersgegevens (NDW). (2015b). 201411281300_MST_NDW01_MT_662.xml. Retrieved from http://www.ndw.nu/documenten/nl

Nationaal Databank Wegverkeersgegevens (NDW). (2015c). Datalevering. Retrieved from http://www.ndw.nu/pagina/nl/103/datalevering/

Nihan, N. L., Wang, Y., & Zhang, X. (2002). Evaluation of dual-loop data accuracy using video ground truth data (No. WA-RD 535.1,). Transportation Northwest, Department of Civil Engineering, University of Washington.

Offermans, M., Priem, A., & Tennekes, M. (2013). Rapportage project impact ict; mobiele telefonie. Programma Impact ICT, 1 (9).

Ortúzar, J. D. D., & Willumsen, L. G. (2011). Modelling Transport. Modelling Transport. http://doi.org/10.1002/9781119993308

Otten, M.B.J., 't Hoen, M.J.J., Den Boer, L.C., (2014). STREAM personenvervoer 2014 VERSIE 1.1. Retreived from: http://www.ce.nl/?go=home.downloadPub&id=1478&file=CE_Delft_4787 _STREAM_personenvervoer_2014_1.1_DEF.pdf

Prato, C. G. (2009). Route choice modeling: past, present and future research directions. Journal of Choice Modelling, 2(1), 65-100.

Snijkers, G. (2009). Getting data for (business) statistics: What's new? What's next. In European conference for new techniques and technologies for statistics (ntts) (pp. 18–20).

Statista (2016a). Mobile phone users in the Netherlands 2011-2019 | Forecast. Retrieved May 10, 2016, from http://www.statista.com/statistics/274751/forecast-of-mobile-phone-users-in-the-netherlands/

Statista (2016b). Smartphone penetration 2012-2015 | Netherlands. Retrieved May 10, 2016, from http://www.statista.com/statistics/488353/smartphone-penetration-netherlands/

Technische Universiteit Delft, Faculteit Civiele Techniek en Geowetenschappen, Afdeling Transport & Planning. (2006). Evaluatie en Analyse van Reisinformatie. Retrieved from http://rijkswaterstaat.nl/rws/corporate/Wegen%20naar%20de%20toekomst /floating%20car%20data%20%20evaluatie%20en%20analyse%20van%20reisinformatie.pdf

Telecompaper. (2015, Q3). Smartphone penetration Netherlands 2015 q3 (Tech. Rep.). Southwark, London.

Toole, J. L., Colak, S., Sturt, B., Alexander, L. P., Evsukoff, A., & González, M. C. (2015). The path most traveled: Travel demand estimation using big data resources. Transportation Research Part C: Emerging Technologies, 58, 162-177.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. science, 185(4157), 1124-1131.

Van de Weerd, I., & Brinkkemper, S. (2008). Meta-modeling for situational analysis and design methods. Handbook of research on modern systems analysis and design technologies and applications, 35.

Van Kats, J. (2014). Supporting Incident Management with Population Information Derived from Telecom Data (Unpublished master's thesis). Utrecht University, Utrecht, The Netherlands.

Van Langen, J.J. (2016). The attractiveness of cities (Unpublished master's thesis). Utrecht University, the Netherlands.

Van Rossum, G., & Drake Jr, F. L. (1995). Python reference manual. Amsterdam: Centrum voor Wiskunde en Informatica.

Ziebart, B. D., Maas, A. L., Dey, A. K., & Bagnell, J. A. (2008, September). Navigate like a cabbie: Probabilistic reasoning from observed context-aware behavior. In Proceedings of the 10th international conference on Ubiquitous computing (pp. 322-331). ACM.

Appendix A: PDD tables for the new scaling method

The activity table and concept table describe the activities and deliverables, respectively, of the new scaling method presented in the PDD in figure 3.4.

Table A1, activity table belonging to the PDD shown in figure 3.4.

ACTIVITY	SUB-ACTIVITY	DESCRIPTION
DETERMINE THE CHANCE TO OBSERVE A PERSON BY AGE GROUP		Divide the number of people per age group by the total number of inhabitants in that area.
DETERMINE THE MOBILE PHONE PENETRATION PER AGE GROUP		Retrieve data concerning the mobile phone penetration per age group from (a) trusted source(s).
ESTIMATE THE PROVIDER MARKET SHARE	Calculate the inhabitants per age group	Multiply the total number of inhabitants by the POPULATION DISTRIBUTION. Note, the total number of inhabitants used here is adjusted for people being abroad, on a

		holiday or on a business trip (Geerts, 2014).
	Multiply by mobile phone penetration per age group	Multiply the INHABITANTS PER AGE GROUP by the MOBILE PHONE PENETRATION PER AGE GROUP.
	Divide users in sample by the number of phone users	Divide the number of users in the area by the number of phone users in the area, i.e. the sum of the PHONE USERS PER AGE GROUP.
DISTRIBUTE USERS ACROSS AGE GROUPS	Calculate the chance a user is in an age group	Multiply the POPULATION DISTRIBUTION by the MOBILE PHONE PENETRATION PER AGE GROUP and by the PROVIDER MARKET SHARE.
	Normalize the calculated probabilities	Divide the PROBABILITY INHABITANT IN SAMPLE PER AGE GROUP by the sum of the PROBABILITY INHABITANT IN SAMPLE PER AGE GROUP.
	Multiply the users by the normalized probabilities	Multiply the users in the area by the chance of observing a user in a certain age group, i.e. the PROBABILITY USER IN AGE GROUP.
DETERMINE THE LIKELIHOOD OF A PERSON MAKING A TRIP OVER X KM DURING WORKDAYS, WORKDAYS DURING THE HOLIDAY, SATURDAYS, AND SUNDAYS PER AGE GROUP		Gather information about the chance that a person of a certain age groups makes a trip longer than X kilometres on a day. We use OViN to determine this and take the differences in weekday, weekend and holiday separately into account.
CALCULATE THE SCALING METHOD	Estimate inhabitants traveling	Multiply the INHABITANTS PER AGE GROUP by the LIKELIHOOD OF TRAVELING PER AGE GROUP.
	Estimate the users traveling	Multiply the USERS PER AGE GROUP by the LIKELIHOOD OF TRAVELING PER AGE GROUP.
	Divide inhabitants traveling by users traveling	Divide the sum of the INHABITANTS TRAVELING by the sum of the USERS TRAVELING.

Table A2, concept table belonging to the PDD shown in figure 3.4.

CONCEPT	DESCRIPTION
POPULATION DISTRIBUTION	The probability that an inhabitant belongs to a certain age group.
MOBILE PHONE PENETRATION PER AGE GROUP	The probability that a Dutch citizen of a certain age group possesses a mobile phone.
INHABITANTS PER AGE GROUP	The absolute number of inhabitants per age group.
PHONE USERS PER AGE GROUP	The absolute number of inhabitants that possess a mobile phone per age group.
PROVIDER MARKET SHARE	The market share of the network provider. Hence, in this case the market share is equal for all age groups.
PROBABILITY INHABITANT IN SAMPLE PER AGE GROUP	The probability that an inhabitant is in our sample per age group.
PROBABILITY USER IN AGE GROUP	The probability that a user in our sample is in a certain age group.
USERS PER AGE GROUP	The absolute number of users per age group.
LIKELIHOOD OF TRAVELING PER AGE GROUP	The probabilities that a person of a certain age group makes a trip that is longer than X kilometre on a specific day of the week.
INHABITANTS TRAVELING	The number of INHABITANTS PER AGE GROUP that is expected to make a trip over X kilometres on a specific day of the week.
USERS TRAVELING	The number of USERS PER AGE GROUP that is expected to make a trip over X kilometres on a specific day of the week.
SCALING METHOD	The ratio between the number of INHABITANTS TRAVELING the number of USERS TRAVELING. The scaling methods applies to traveling people, because these are the people that are relevant for the OD matrix.