TØI report 1657/2018

Fredrik Alexander Gregersen
Erik Bjørnson Lunke

Institute of Transport Economics
Norwegian Centre for Transport Research

The application of cellular network data for travel behavior research

TØI Report
1657/2018

The application of cellular network data for travel behavior research

Fredrik Alexander Gregersen
Erik Bjørnson Lunke

Front-page photography/illustration: Pixabay.com

ISSN 2535-5104 Electronic
ISBN 978-82-480-2175-9 Electronic
Oslo, October 2018

Tittel:	Anvendelsen av nettverksdata fra mobiltelefoner til reisevaneforsking	**Title:**	The application of cellular network data for travel behavior research
Forfattere:	Fredrik Alexander Gregersen Erik Bjørnson Lunke	**Authors:**	Fredrik Alexabder Gregersen Erik Bjørnson Lunke
Dato:	10.2018	**Date:**	10.2018
TØI-rapport:	1657/2018	**TØI Report:**	1657/2018
Sider:	18	**Pages:**	18
ISSN elektronisk:	2535-5104	**ISSN Electronic:**	2535-5104
ISBN elektronisk:	978-82-480-2175-9	**ISBN Electronic:**	978-82-480-2175-9
Finansieringskilde:	Norges forskningsråd	**Financed by:**	Research Council of Norway
Prosjekt:	4617 – MobilForsk	**Project:**	4617 – LITMOB
Prosjektleder:	Fredrik Alexander Gregersen	**Project Manager:**	Fredrik Alexander Gregersen
Kvalitetsansvarlig:	Tom Erik Julsrud	**Quality Manager:**	Tom Erik Julsrud
Fagfelt:	12 Reisevaner og mobilitet	**Research Area:**	12 Travel behaviour and Mobility
Emneord:	Mobiltelefoner, reisevaneforskning, nettverksdata, teleselskaper, big data	**Keyword(s):**	Mobile phones, network data, travel behavior research, literature review, call detail records

Sammendrag:

Når en mobiltelefon kobler seg til telenettet lagres informasjon om posisjonen til mobilmasten som telefonen er koblet til. Videre lagres blant annet informasjon om tekst, tale og datatrafikk. Slike data kalles nettverksdata fra mobiltelefoner og kan benyttes til reisevaneforskning. Det er viktig å merke seg at slike data vil generes ikke bare når telefonen blir brukt aktivt, men også når telefonen er slått på og koblet til mobiltelefonnettet. Denne rapporten oppsummerer litteraturen knyttet til bruk av slike data til reisevaneforskning. Litteraturen peker på at slike data har mange bruksområder innen reisevanefeltet. Flere studier har blant annet brukt slike data til trafikktellinger og tellinger av lange reiser. Det er imidlertid en del utfordringer knyttet til bruk av slike data blant annet knyttet til personvernet til respondenten, utvalgsskjevhet og utfordringer med å fastslå posisjon til mobiltelefonen.

Summary:

This report is a literature review of the application of cellular network data from mobile phones for travel behaviour research.

Language of report: Norwegian/other

Transportøkonomisk Institutt
Gaustadalleen 21, 0349 Oslo
Telefon 22 57 38 00 - www.toi.no

Institute of Transport Economics
Gaustadalleen 21, N-0349 Oslo, Norway
Telephone +47 22 57 38 00 - www.toi.no

Preface

This report is a literature review of the application of cellular network data from mobile phones for travel behaviour research. The report is written to find topics that need more investigation in order to apply this type of data for research and innovation in Norway. It is written on behalf of Delfi data, and is financed through the Norwegian Research Council and VRI Rogaland.

Oslo, October 2018
Institute of Transport Economics

Gunnar Lindberg
Managing Director

Tom Erik Julsrud
Research Director

Content

Summary

Sammendrag

1 Travel behavioural research and some challenges 1

1.1 Data from mobile phones 2

1.2 The aim of this report 3

2 Introduction to cellular network data 5

3 Positioning of the cell phone 7

4 Application of cellular network data 9

5 Conclusion and Discussion 12

6 References 14

Appendix – Important Terms 17

Summary

The application of cellular network data for travel behavior research

TØI Report 1657/2018
Authors: Fredrik A Gregersen and Erik B Lunke
Oslo 2018 18 pages English language

Travel behavior have for a long period of time primarily been studied by surveys. For the past 20 years especially abroad also data from mobile phones have been applied. In this report we give a brief overview of the application of mobile phones and in particular network data in this field of research.

Cellular network data is generated by interactions between mobile phones and the cellular network, i.e. cell towers or base stations. There are three ways in which such data is generated, the first of which relates to active use of the phone for calls, text messages or internet access. This is *actively* generated data, stored by the mobile operator for billing purposes, and is often referred to as call detail records (CDR). Data is also generated when the phone is not in active use, e.g. location updates when the phone moves between location areas. This data is collected more continuously, but is not as geographical precise as the actively generated data. A third way to obtain data is by sending a special query to a device to determine its location. The query is carried out by the telecom company and is used, for example, by emergency services to locate a phone in case of an emergency.

Telecom companies have built cellular networks with a high geographical coverage. This, combined with the fact that a high share of the population regularly uses mobile phones, opens possibilities for a range of research on mobility that was previously difficult to investigate due to scarcity of data.

Cellular network data has been applied to several research areas in the field of transportation, for instance, the counting of vehicles, monitoring of tourists movements and as a validation tool against other data sources. Several studies point in the direction that this type of data may be applicable for travel behavior research. The main challenges addressed in the literature are privacy/ethics, selection bias and the accuracy of the positioning of the phones.

Sammendrag

Anvendelsen av nettverksdata fra mobiltelefoner til reisevaneforsking

TØI rapport 1657/2018
Forfattere: Fredrik A Gregersen og Erik B Lunke
Oslo 2018 18 sider

Når en mobiltelefon kobler seg til telenettet lagres informasjon om posisjonen til mobilmasten som telefonen er koblet til. Videre lagres blant annet informasjon om tekst, tale og datatrafikk. Slike data kalles nettverksdata fra mobiltelefoner og kan benyttes til reisevaneforskning. Det er viktig å merke seg at slike data vil generes ikke bare når telefonen blir brukt aktivt, men også når telefonen er slått på og koblet til mobiltelefonnettet. Denne rapporten oppsummerer litteraturen knyttet til bruk av slike data til reisevaneforskning. Litteraturen peker på at slike data har mange bruksområder innen reisevanefeltet. Flere studier har blant annet brukt slike data til trafikktellinger og tellinger av lange reiser. Det er imidlertid en del utfordringer knyttet til bruk av slike data blant annet knyttet til personvernet til respondenten, utvalgsskjevhet og utfordringer med å fastslå posisjon til mobiltelefonen.

1 Travel behavioural research and some challenges

The transport system is an intricate and complex system, and to understand and assess the impact of changes therein is an important, but challenging task. While there are regularities to travel patterns at an aggregated level, e.g. rush hours, individual travel patterns are characterized by greater amounts of variability related to choice and adaptation, e.g. transport mode, departure/arrival time, route choices. Some trips are compulsory, such as work or school trips, whereas others are discretionary and can be cancelled or rescheduled. Although travel behavior is habitual, it is also irregular in that travel needs and choices differ among individuals. The nature of travel makes it difficult to accurately predict travel demand and the effects of policy intervention. The system is highly dynamic, and it is interrelated with external societal factors including the availability of transport alternatives and urban patterns, as well as factors such as season, climate and weather.

Countries and cities have traditionally assessed travel through conventional household surveys. There are, however, some problematic issues with travel surveys, such as low response rates, non-response (missing data), sampling errors, and quality control (Pearson et al., 2009). Problems with low response rates are not unique to travel surveys. Travel surveys are, however, a time demanding method in which respondents self-report highly detailed information on all trips taken on the registration day. In a time in which surveys for marketing and other purposes are abundant, peoples' willingness to participate in surveys is declining. Inevitably, researchers need to consider methods for improving the accuracy of travel surveys (Pearson et al., 2009).

Several countries have National Travel Surveys (NTSs) that have been conducted regularly, among them are Norway, Sweden, Denmark and the UK. The Norwegian NTS has been conducted in 1985, 1992, 1998, 2001, 2005, 2009, 2013/14, and the most recent survey, started in 2016, will continue until 2019. This survey covers personal travel of all types of transport modes and travel purposes. Especially due to falling response rates over time, it has been debated whether or not one should consider to support or replace parts of the survey with other types of data (Gregersen and Weber, 2016). In Norway, the response rate on NTS was 77 % in 1985, but as low as 20 % in 2013 (Hjorthol et al. 2014). For the same reason, Sweden has also started to test the correlation between cellular network data and the Swedish NTS on long distance travel (Gundlegård and Rydergren, 2018). Cellular network data is generated from mobile phones connecting to the cellular network from telecom companies. Grundlegård and Rydergrenthe use NTS as "ground truth" and find that the two data sources match quite well. We will return to this study in the application chapter (chapter 4).

Despite falling response rates, it has been difficult to prove that the quality of the surveys have declined over time. Engebretsen (2018) tested the correlation between the Norwegian NTS 2013/14 and traffic counts from loop detector around Oslo, the capital of Norway. He found a strong correlation between the two data sources. There are, however, few studies in general that test the validity of the Norwegian NTS.

While the willingness to share travel information through surveys is declining, people are sharing more data than ever. We leave electronic traces of our behavior through the use of various internet applications and mobile phones. In addition, automatic monitoring of road traffic and passengers yields detailed data on traffic flows. New data sources come with new challenges and new sources of error and bias. In addition, the actual volume of the data requires new tools, methods and models. The use of new data sources in the transport sector has received a lot of attention internationally, and the journal Transportation Research Part C: Emerging Technologies dedicated a special issue to Big Data in 2015.

For the past 20 years, data from mobile phones have increasingly been applied to research on travel behavior. Data generated from mobile phones for travel behavior research may broadly be categorized in two groups – sensor based smartphone data and cellular network data (Wang et al., 2018).

1.1 Data from mobile phones

There are two main types of mobile phone data – sensor based data and cellular network data. Both types are described briefly in this paragraph, while the latter is the main subject of this report.

1.1.1 Sensor based data

Modern smartphones contain a number of sensors that are used to record the location and motion of the device. Location based sensors include GPS receivers, Wi-Fi and Bluetooth sensors, and motion based sensors such as accelerometers and magnetometers (Wang et al., 2018). In order to utilize these sensors, numerous mobile applications have been developed.

Applications used in travel behaviour research are usually divided in two groups. The first group consists of popular applications that provide location-based services for users, while the second group are applications developed specifically to collect travel behaviour data for research purposes.

The popular applications are typically social media applications in which users share their location with friends and contacts. The geographical tracking can be done continuously during certain activities such as work out (using apps like Strava and Runkeeper), or on rarer occasions through so called "check-ins" at specific locations such as restaurants and museums (Yue et al., 2014).

The second group of applications, those that are designed specifically for research purposes, differs from the first group in that they are built to record all movements the user undertakes, not just movements during selected activities (such as working out). Research based applications run in the background on the user's phone. These applications can use the phone's sensors to imply which transportation mode is being used. For example, if a phone is moving at a high speed on long distances, the user is most likely using a car or public transportation. If the speed is lower and the phone's accelerometer records more frequent movement, the user is most likely walking or running. Nitsche et al. (2014) have used data collected by smartphones as a support for traditional travel surveys in a study in Austria. A similar study was conducted in Norway, using a smartphone application as well as a traditional web survey to study cycling behaviour in Norwegian cities (Lunke et al., 2017a, b, c).

The main advantage of smartphone sensor-based data is the high geographical detail. Using GPS, Wi-Fi and other sensors, the applications can locate the device very accurately, with a margin of error of just a few meters. The applications can also record continuously, which gives detailed information about route choice during a trip. While running in the background at all times, they also avoid the problem of forgotten trips (or parts of trips) which is common in traditional travel surveys.

The biggest challenge with data from smartphone applications is sampling bias, as mentioned by Wang et al. (2018). When working with popular applications, which are used by certain groups (such as applications for athletes), as well as research based applications, recruiting a representative sample of the population in an area can be challenging.

Another challenge when recording movement is the battery capacity of smartphones. This is especially evident with research based apps that record movement at all times, which can drain the battery quite quickly.

1.1.2 Cellular network data

Cellular network data is generated from mobile phones connecting to the cellular network from telecom companies. When a connection occurs due to phone calls, text messages or internet access, the data is referred to as call detail records (CDRs) or actively generated data. It may also be generated by the phone simply being switched on and, therefore, searching for the nearest cell available, referred to as ”idle” or passive mode. Telecom operators also measure, for instance, number of calls, text messages and internet usage (Chen et al., 2016; Steenbruggen et al., 2013). The telecom operators may also actively send a query to the phone even when the phone is in idle mode (Ahas et al., 2007).

Telecom companies have built cellular networks with a high geographical coverage. This, combined with the fact that a high share of the population regularly uses mobile phones, opens a range of research possibility on mobility that was previously difficult to investigate due to scarcity of data.

1.2 The aim of this report

Researchers, telecom companies and start-up initiatives in Norway are all looking into how cellular network data can improve our knowledge about travel behavior. In a society where almost everyone bring their mobile phone wherever they go, the phone becomes an useful device for massive sensing of mobility patterns. The utilization of such data, however, is not without challenges.

First, the sensitive nature of data about an individual's mobile usage and travel habits, brings to light concerns about privacy and the need for aggregation techniques that ensures anonymity of the data. Second, the gap between traditional travel survey data, with well-established statistical techniques for sampling and analyses on one side, and cellular network data with more unknown sources of error and bias on the other side, raises the need for validation and comparison studies.

The Institute of Transport Economics (TØI) was invited by the company Delfi Data to carry out a literature study to find the research gaps that need to be addressed in order to utilize cellular network data in Norway. Despite the literature on cellular network data from mobile phones being large, there is a paucity of review papers in this field (Wang et al., 2018). This report summarizes the literature on cellular network data in travel behavioral research. Travel behavioral research investigates how people move in space and use different transportation modes. The literature on cellular network data is large and to include all papers is not possible within the frame of this report. We started out by looking into the pioneering work performed in Rein Ahas' mobility lab in Estonia and followed the papers that sited their work. Furthermore, we tried to include the main issues raised in the two last Mobile Tartu conferences. We also summarized the main issues addressed in (Wang et al., 2018) who summarize the literature on application of mobile phones, both sensor and network based data, for travel behavioral research. As such, this is a non-systematic review of the literature, and the aim is to provide researchers and policymakers unfamiliar with the literature an overview of the main issues addressed.

The next chapter elaborates on cellular network data. Chapter 3 describes how the positioning of the cell phone is determined with network data, while chapter 4 gives a brief description of the application of network data. Finally, we conclude and present some areas that would benefit from more research.

2 Introduction to cellular network data

Mobile operators collect and store data on the communication between mobile phones and the mobile network: so called cellular network data. The access point of a mobile device to the network is a Base Transceiver Station (BTS, also referred to as base tower or cell tower). The area covered by one BTS is called a cell. A BTS can cover a smaller or larger area depending on geographic context. In urban areas the radius of a BTS is generally smaller than in rural areas (Wang et al., 2018).

In most cases the phone is connected to the nearest BTS. The mobile operator knows the position of the BTS and in most cases, the coverage areas of each BTS as well. The geographical radio limit for the GSM (Global System for Mobile communications) network antenna (BTS antenna) is 65km, i.e. the maximum radius of the BTS under normal conditions (Ahas et al., 2007). In most circumstances the cells' actual coverage areas are much smaller. In China, Liu et al. (2009) report that the coverage areas vary with a radius of 100 meters to 500 meters in urban areas and from 400 to 10,000 meters in rural areas.

To further specify the location of the phone, two techniques may be applied: received signal strengths and/or triangulation. Signal strengths applies the fact that the closer the phone is to the cell the stronger the signal is. Triangulation is done by combining the signal strength from three or more cells to locate a the mobile device (Steenbruggen et al., 2013). Despite these techniques, the location of the phone still has some uncertainty. In most cases both triangulation and signal strength are seen as too time consuming or the data is not available to the researcher and, therefore, the location of the cells are applied as opposed to a more precise positioning of the phones (Wang et al., 2018).

The characteristics and level of detail of network data depends on whether the studied cell phones are in active or passive (idle) mode. Active mode is when a cell phone is being used, for example during a call, while sending a text message or during down- or uploading of internet content. When a phone is not in use, it is in passive mode. Even when the phone is in passive mode, but connected to the network, the location and movement of the phone is still being registered, but on a less detailed level than during active mode, see section 3.

This type of data differs from traditional datasets (such as travel survey data) in many aspects: high marked penetration rates and the ubiquitous use of mobile phones yields massive datasets. The data is collected passively without any effort from the users, and since people tend to carry the mobile phone with them wherever they go, it is an excellent device for capturing spatiotemporal movement patters from the wider population.

Importantly, cellular network data follows phones, not individuals. This is contrary to data from most NTSs that typically ask for the movement of the respondents for the past 24 hours. Furthermore, most studies only include data from one telecom operator. In contrast to data collected by researchers or transport agencies for the purpose of transport monitoring and modelling, network data is not collected with this intent. Selection bias in the sample is, therefore, an issue that has been raised in several studies (Ahas et al., 2011).

Cellular network data is owned by the telecom companies. However, mobile operators are looking into how their data can supply society with the information needed for better planning and development, and there have been efforts towards making cellular network data available to the general population. The telecom company Orange has launched two research challenges, Data for Development (D4D), with anonymous network datasets from Côte d'Ivoire (2012/2013) and Senegal (2014/2015). Transport and planning was one of the prioritized topics for which there have been publications in the academic literature based on these datasets: Gundlegård et al. (2016) uses network data for travel demand estimation in Côte d'Ivoire and Senegal, and Kujala et al. (2016) estimates the travel times between cities in Senegal. Examples of others that have used network data (from different operators) are Järv et al. (2014) who investigate variations in monthly travel behaviour, Toole et al. (2015) who develop software system for travel demand estimation based on network data, and Secchi et al. (2015) who investigates urban mobility patterns in the city of Milan.

In travel surveys, data is provided by informed content. In most cases this is not the case for network data from mobile phones. This raises questions regarding the privacy of the users included in the sample (Ahas et al., 2011; Steenbruggen et al., 2013). Julsrud and Krogstad (2018) find that more than half of the population of Oslo is concerned that positioning data from mobile phones, i.e. network data, represents a threat to individual privacy.

Cellular network data does not contain any specific information about the purpose of the trips or transport mode performed by the phone holder. Despite this, some researchers have tried to apply algorithms to estimate the trip's purpose and mode of transport. For example, the location of the phone during night is assumed to be the user's home location, while daytime location is their work place (Wang et al., 2018). To count traffic, so called hand-over data may be applied. This refers to a situation where the phone is switching an on-going call to a different network cell. By following such switching along roads, several researchers have been able to estimate the number of vehicles passing (Demissie et al., 2013).

3 Positioning of the cell phone

In general, the data obtained from the cellular network normally contains the position of the BTS where the phone is connected and a time stamp of the connection. Furthermore, the data may contain information on the signal strength to the tower at the time of connection and the type of information exchanged between the tower and the phone. The strength of the signal may indicate the distance from the tower to the device. Based on this information, one may get an indication of the location of the phone at the time of connection (Ahas et al., 2010). The information on a device's location is stored in a system of databases (Steenbruggen et al., 2013). However, compared to GPS data, the exact location is normally very uncertain. This is so partially because that the phone may "jump" between cells even though the user is not moving. In addition, a phone may be connected to a cell tower other than the nearest one. This typically happens when many people congregate, e.g. for a festival or other events, leading to an overload of the nearest cell tower's capacity (Traag et al., 2011; Wang et al., 2018).

The most accurate network data is obtained when the device is in use, e.g. during a phone call or while sending a text message. At those times, when the device is in "active" mode, it is connected to only one cell tower. The margin of error when implying the device's location depends on the cell's radius, which vary depending on the context (e.g. urban or rural areas). However, most of the time the phone is turned on, it is in "idle" mode and not being actively used. While in idle mode, there is no need for the phone to be directly connected to one cell. Instead, the phone is linked to multiple cells within its range (its Location Area), so that it can be connected if for example a call is to be made. In idle mode, the location of the device is, therefore, less certain, and the margin of error is larger than while in active mode (Caceres et al., 2008; Steenbruggen et al., 2013).

When studying travel behavior using network data, the researcher must therefore decide between using data from active and/or passive phones. The use of active phone data increases the precision, but at the same time the sample size is greatly reduced compared to the passive data.

A third type of network data has been used by Rein Ahas et al. (see Ahas et al., 2008; Ahas et al., 2010), who define "active" data differently than Caceres et al. (2008) and Steenbruggen et al. (2013). In Ahas' definition, active mobile positioning data are data "in which the location of the mobile phone is determined (asked) with a special query using a radio wave" (Ahas et al., 2008, 470). By applying this approach, mobile phone companies can locate the device more precisely, even though it is in "idle" mode and not in active use. This type of positioning is typically used by authorities during emergency calls. However, it is normally not allowed to collect this data for research purposes without obtaining informed consent from the users.

In addition to locating the device, network data may be used to record movement. This is typically done using *handover* and *cell dwell time (CDT)* data (Steenbruggen et al., 2013). Handover data is recorded when a device switches from one cell to another during a call in order to retain network coverage when moving away from a cell's range. CDT, on the other hand, is a measure of how long a phone is connected to one cell, i.e. the time between two handovers. Information about the average CDT of a cell tower at different times of the day, can for example be used to measure traffic congestion (Steenbruggen et al., 2013).

Another source of data is the telephone traffic of a cell, which is measured in Erlang. One Erlang equals one person-hour of phone use. The information on one cell's traffic can be combined with the approximate average traffic per subscriber to assume the number of subscribers in the cell's area (Caceres et al., 2008).

The network data (including handover and CDT data) may be applied to measure and study traffic along a specific road. However, to be certain that the data collected is from users travelling along a specific road, some criteria must be in place. First, there needs to be a sufficient number of cells along the road. In that way, the handover and CDT data can indicate the travel speed along the road. Second, in order to study that specific road's traffic, there should not be other roads within the cells' range. For example, a bicycle road running along a highway may give inaccurate data when studying car traffic. As shown in Figure 1, in such a scenario, one cannot be certain which of the roads a registered device has been travelling on (Caceres et al., 2008).

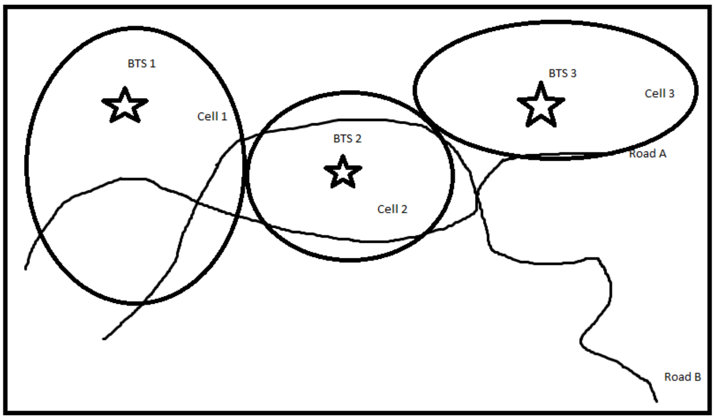

Figure 1 Multiple roadway links within a cell

4 Application of cellular network data

Cellular network data has been applied on a number of different areas in behavioral research, including the counting of vehicles, disaster management, and study of long distance travel, and commuting. This chapter will present some of the studies.

Traffic parameters:

By applying hand-over data from cellular networks from Lisbon, the capital of Portugal, Demissie et al. (2013) try to estimate the number of vehicles and compare the numbers to traffic counters (loop detectors). They find a high correlation, but the correlation varies substantially for the different counters included in the study. They, therefore, apply multinomial logit models to improve the models.

Steenbruggen et al. (2013) summarizes the literature on estimating traffic parameters by mobile phone data. They find that several studies are better able to estimate travel speed for vehicles over longer distances as compared with shorter ones (10 km), and that the measurement error is large on shorter distances. Furthermore, BTSs are not always optimally located to monitor traffic. Most studies investigate a small part of the road network, typically a stretch of road and not the entire network.

Long distance travel:

Gundlegård and Rydergren (2018) are comparing data from the Swedish NTS on long distance travel (above 100 km) and data from the mobile operator Telia. Due to confidentiality concerns they do not describe the data they apply fully. The results show that the NTS and mobile data do match quite well. There are, however, some problems regarding the definitions of the start point and end point of the journey; it may be difficult to match the data sources according to the same definition.

Tourist statistics:

Raun et al. (2016) applies roaming data from Estonia. Roaming occurs when a phone is registered in one country and is applied in another, in this case Estonia. They assume that phones that applies roaming must be applied by visitors. The methods applied in the paper were later adopted by the Estonian tourist board.

Ahas et al. (2011) summarizes the experiences with applying network data to provide official tourism statistics in Estonia. They conclude that the wide spread use of mobile phones do enables cost effect collection of mobility data for tourists. On the other hand they do raise some concerns regarding the privacy of the respondents and difficulties related to access to the data processing procedures by the telecom companies.

Commuting and daily mobility:

Ahas et al. (2010) studies the daily rhythms of suburban commuters' movements in the Tallinn metropolitan area by network data. They find that the respondents have a similar temporal rhythm related to work and leisure. As the different activities are primarily carried out at different points in time, the researchers claim to be able to map "functional differences" in the city.

Csáji et al. (2013) studied mobile phone data from 100 000 individuals in Portugal. Using communication data, they were able to identify the individuals' home and work locations with a high accuracy. This information could in turn be used to measure commuting distances. The results show a concentration of locations in and around Porto and Lisbon, the two largest cities in Portugal. In addition, commuting distances are shorter in the urban areas.

Alexander et al. (2015) have done a similar study, applying CDR data from the Boston metropolitan area. They developed methods to identify daily trips, as well as locations (such as home and work) and trip purposes. To validate the applicability of the methods, the results were compared with locations and trips reported in local and national travel surveys.

Events:

In an early study, Candia et al. (2008), applied mobile phone records to study anomalous events. They concluded that the rich information provided by mobile phone data was well suited to detect concentrations of mobile phone users, which served as a proxy for non-regular events.

In a later study, Traag et al. (2011) applied call data from a European country to develop a methodology to detect social events. They study the activity on base stations located near a football stadium, music festival, rural area and tourist destination. The researchers find clear peaks in network activity when events such as football matches and the music festival takes place. They also see clear seasonal changes in the activity around the tourist destination. Finally, they identify a peak in activity on Christmas eve. The study suggest that network data can be used to detect events, and to determine which users participated in the event.

Other applications:

To optimize the location of docking stations for electric vehicles in Milan in Italy, Vantini et al. (2012) applies CDR data. They exploit the Telecom Italia database and conclude that the data may give valuable insight into location of the population. This may again be applied to find optimal location of docking stations.

Mooses et al. (2016) applies CDRs from Estonia to measure ethnic segregation in the country by comparing traveling during public holidays. They compare the Russian and Estonian speaking part of the population. They assume that the language the mobile user prefers to communicate with the telecom provider represents his or her ethnic identity. They find considerable differences in travel behavior between the two ethnic groups. The study found that citizens of Tallinn that are Russian speaking stay in the city more during holidays compared to the Estonian speaking population.

By applying anonymized CDRs from Grameenphone, the largest mobile phone operator in Bangladesh, Lu et al. (2016) investigated how citizens evacuated during a cyclone. They conclude that "…mobile network data provides a novel tool to quantify directionality and seasonality of migration patterns on both local and national scales".

Urban land-use types are usually differentiated either by their physical characteristics or social functions. Pei et al. (2014) investigate how mobile phone data can be used to map residents' activities, which in turn can be used to indicate the social function of land use. The study, although preliminary, indicate some promising results, which have been most successful in areas with high heterogeneity in land use, indicating a need for more research in this field.

To summarize, the application of network data may be applied for a range of different research areas. Most papers apply data from one telecom operator, and this may lead to selection bias. Furthermore, it may be difficult to compare network data with traditional surveys as the start and end points of a journey can have different definitions in the two approaches. The accuracy of the positioning of the phones are also addressed in all the above mentioned papers. Furthermore, privacy of the mobile phone users is debated in several of the above mentioned papers.

5 Conclusion and Discussion

In theory, the sky is the limit in terms of what types of data may be applied for travel behavioural research. This report has focused on cellular network data. Other data may also be applied for mobility research, such as travel cards for public transport, traffic loop detectors and data from bike or car sharing services.

Travel cards have been developed in order to automatically collect fares for public transportation, and their records can give information about the passengers travel behaviour. The information recorded in travel cards from various providers differ, although they often contain details on time, mode, station and route of trips, as well as personal data about the users (Yue et al., 2014). Transport systems using a flat fare – e.g. a monthly fare – does not track every single boarding, giving less detailed information about each user's trips. In addition, the linking of trip and personal information poses challenges regarding the privacy of the respondents. Several studies show that this data can be used to study the dynamics and flows of transport in an urban public transportation system (Gong et al., 2012; Roth et al., 2011)(Utsunomiya et al., 2006).

Another possible data source is loop detectors that are located in several places in larger cities. These are built to count vehicles, mainly cars, buses and bicycles. Although the counters don't follow individual vehicles over longer distances – they only record the number of vehicles passing at specific points – they can be used to measure traffic flows. Loop detector data has been used in several countries, including Norway (Tennøy et al., 2016) and New Zealand (Tin et al., 2012), to study traffic volume and traffic flows.

All data sources have their strengths and limitations – the most suitable one depends on the research questions being investigated. Comparing cellular network data with data from traditional travel surveys provides different information on the population investigated. While travel surveys generally follow small, and in many cases representative, samples over short periods of time, network data usually follows a large part of the population, typically all costumers from a telecom company, over a longer period of time. Furthermore, travel surveys may ask the respondents about the purpose of the trips. Although network data, in most cases, cannot detect trip purpose, the use of algorithms can help researchers arrive at approximate determinations of trip purpose (Wang et al., 2018).

It can be difficult to completely replace traditional travel surveys with cellular network data. However, network data can be more suitable for other, more defined studies. As mentioned before, network data has been successfully used to identify certain types of activity, such as social events (Traag et al., 2011) and commuting patterns (Ahas et al., 2010).

Perhaps the most positive aspect of network data from a research perspective is the wide spread use of mobile phones in society, which generates an abundant and continuous stream of mobility data (Ahas et al., 2011). This enables researchers to test complex transportation models with real-time data. In this respect the network data can inform increasingly complex mobility research (Steenbruggen et al., 2013). Furthermore, data collection is already conducted by the telecom operators. However, such data is owned by the telecom provider and gaining access may, in many cases, be difficult for researchers (Steenbruggen et al., 2015).

Despite the large amount of data generated from network data, there is still some uncertainty regarding the validity of the data generated. More research on the noise and bias generated in such types of data should be discussed in greater detail. "Unlike travel surveys during which subjects self-report locations visited and modes of transportation used, passively generated datasets lack ground truth to be validated against. Probably due to this reason, only a few studies using passive data have addressed this to some extent" (Chen et al., 2016). Therefore, more research should be conducted in order to validate the network data to further understand the strength and limitation of the data source.

Furthermore, privacy issues for the respondents are still not fully solved (Ahas et al., 2011). There is an inherent conflict between society's need for knowledge and empirical evidence of citizens' travel habits in order to make accurate predictions and good decisions, while at the same time protecting the individuals' right to privacy (Boyd and Crawford, 2012; Kitchin, 2013). There remain a significant amount of unresolved issues and nuances when it comes to data storage, proprietary rights and user consent.

To summarize, the main challenges and advantages of using network data are as follows:

Main challenges:

- Selection bias
- Uncertainty regarding the location of the phone
- Ethical issues regarding privacy and anonymity to the phone holder
- Changes in phone use and the cellular network over time

Main advantages:

- Covers a large part of the population
- Long time series of data
- Cost effective
- May give information about respondents that do not want to participate in surveys

6 References

Ahas, R., Aasa, A., Mark, Ü., Pae, T., Kull, A., 2007. Seasonal tourism spaces in Estonia: Case study with mobile positioning data. Tourism management 28, 898-910.

Ahas, R., Aasa, A., Roose, A., Mark, Ü., Silm, S., 2008. Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. Tourism Management 29, 469-486.

Ahas, R., Aasa, A., Silm, S., Tiru, M., 2010. Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: Case study with mobile positioning data. Transportation Research Part C: Emerging Technologies 18, 45-54.

Ahas, R., Tiru, M., Saluveer, E., Demunter, C., 2011. Mobile telephones and mobile positioning data as source for statistics: Estonian experiences. presentation for NTTS.

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, Part B, 240-250.

Caceres, N., Wideberg, J., Benitez, F.G., 2008. Review of traffic data estimations extracted from cellular networks. IET Intelligent Transport Systems 2, 179-192.

Candia, J., González, M.C., Wang, P., Schoenharl, T., Madey, G., Barabási, A.-L., 2008. Uncovering individual and collective human dynamics from mobile phone records. Journal of physics A: mathematical and theoretical 41, 224015.

Chen, C., Ma, J., Susilo, Y., Liu, Y., Wang, M., 2016. The promises of big data and small data for travel behavior (aka human mobility) analysis. Transportation research part C: emerging technologies 68, 285-299.

Csáji, B.C., Browet, A., Traag, V.A., Delvenne, J.-C., Huens, E., Van Dooren, P., Smoreda, Z., Blondel, V.D., 2013. Exploring the mobility of mobile phone users. Physica A: statistical mechanics and its applications 392, 1459-1473.

Demissie, M.G., de Almeida Correia, G.H., Bento, C., 2013. Intelligent road traffic status detection system through cellular networks handover information: An exploratory study. Transportation research part C: emerging technologies 32, 76-88.

Engebretsen, Ø., 2018. Måling av trafikkarbeid med reisevaneundersøkelser. TØI 1647/2018

Gong, Y., Liu, Y., Lin, Y., Yang, J., Duan, Z., Li, G., 2012. Exploring spatiotemporal characteristics of intra-urban trips using metro smartcard records, Geoinformatics (GEOINFORMATICS), 2012 20th International Conference on. IEEE, pp. 1-7.

Gregersen, F.A., Weber, C., 2016. Big Data eller spørreskjemaer? Ja takk! Samferdsel.

Gundlegård, D., Rydergren, C., 2018. Resestatistik för långväga resor baserat på aggregerade och anonymiserade mobilnätsdata. Linköpings Universitet.

Gundlegård, D., Rydergren, C., Breyer, N., Rajna, B., 2016. Travel demand estimation and network assignment based on cellular network data. Computer Communications.

Hjorthol, R., Engebretsen, Ø., Uteng, T.P., 2014. Den nasjonale reisevaneundersøkelsen 2013/14 - nøkkelrapport. TØI 1383/2014

Julsrud, T.E., Krogstad, J.R., 2018. Sporing av reiser i smarte byer. Hva mener innbyggerne? TØI 2018-1658.

Järv, O., Ahas, R., Witlox, F., 2014. Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records. Transportation Research Part C: Emerging Technologies 38, 122-135.

Kujala, R., Aledavood, T., Saramäki, J., 2016. Estimation and monitoring of city-to-city travel times using call detail records. EPJ Data Science 5, 1.

Liu, L., Biderman, A., Ratti, C., 2009. Urban mobility landscape: Real time monitoring of urban mobility patterns, Proceedings of the 11th international conference on computers in urban planning and urban management. Citeseer, pp. 1-16.

Lu, X., Wrathall, D.J., Sundsøy, P.R., Nadiruzzaman, M., Wetter, E., Iqbal, A., Qureshi, T., Tatem, A., Canright, G., Engø-Monsen, K., 2016. Unveiling hidden migration and mobility patterns in climate stressed regions: A longitudinal study of six million anonymous mobile phone users in Bangladesh. Global Environmental Change 38, 1-7.

Lunke, E.B., Aarhaug, J., Fyhri, A., Hulleberg, N., Ingebrigtsen, R., Sundfør, H.B., Weber, C., Ævarsson, G., 2017a. Tellesykkel - Tromsø. Transportøkonomisk institutt, Oslo.

Lunke, E.B., Aarhaug, J., Fyhri, A., Hulleberg, N., Ingebrigtsen, R., Sundfør, H.B., Weber, C., Ævarsson, G., 2017b. Tellesykkel - Trondheim. Transportøkonomisk institutt, Oslo.

Lunke, E.B., Aarhaug, J., Fyhri, A., Hulleberg, N., Ingebrigtsen, R., Sundfør, H.B., Weber, C., Ævarsson, G., 2017c. Tellesykkel Buskerudbyen. Transportøkonomisk institutt, Oslo.

Mooses, V., Silm, S., Ahas, R., 2016. Ethnic Segregation During Public and National Holidays: A Study Using Mobile Phone Data. Geografiska Annaler: Series B, Human Geography 98, 205-219.

Nitsche, P., Widhalm, P., Breuss, S., Brändle, N., Maurer, P., 2014. Supporting large-scale travel surveys with smartphones–a practical approach. Transportation Research Part C: Emerging Technologies 43, 212-221.

Pearson, D., Hard, E., Farnsworth, S., Forrest, T., Spillane, D., Ojah, M., Womack, K., Boxill, S., Lewis, C., 2009. Improving Accuracy in Household and External Travel Surveys. Aug.

Pei, T., Sobolevsky, S., Ratti, C., Shaw, S.-L., Li, T., Zhou, C., 2014. A new insight into land use classification based on aggregated mobile phone data. International Journal of Geographical Information Science 28, 1988-2007.

Raun, J., Ahas, R., Tiru, M., 2016. Measuring tourism destinations using mobile tracking data. Tourism Management 57, 202-212.

Roth, C., Kang, S.M., Batty, M., Barthélemy, M., 2011. Structure of urban movements: polycentric activity and entangled hierarchical flows. PloS one 6, e15923.

Secchi, P., Vantini, S., Vitelli, V., 2015. Analysis of spatio-temporal mobile phone data: a case study in the metropolitan area of Milan. Statistical Methods & Applications 24, 279-300.

Steenbruggen, J., Borzacchiello, M.T., Nijkamp, P., Scholten, H., 2013. Mobile phone data from GSM networks for traffic parameter and urban spatial pattern assessment: a review of applications and opportunities. GeoJournal 78, 223-243.

Steenbruggen, J., Tranos, E., Nijkamp, P., 2015. Data from mobile phone operators: A tool for smarter cities? Telecommunications Policy 39, 335-346.

Tennøy, A., Caspersen, E., Hagen, O.H., Langeland, P.A., Landa-Mata, I., Nordbakke, S., Skollerud, K.H., Tønnesen, A., Weber, C., Ørving, T., Aarhaug, J., 2017. BYTRANS: Effekter og konsekvenser av kapasitetsreduksjon i Brynstunnelen – per 2016. Transportøkonomisk institutt, Oslo.

Tennøy, A., Wangsness, P.B., Aarhaug, J., Gregersen, F.A., 2016. Experiences with capacity reductions on urban main roads–rethinking allocation of urban road capacity? Transportation research procedia 19, 4-17.

Tin, S.T., Woodward, A., Robinson, E., Ameratunga, S., 2012. Temporal, seasonal and weather effects on cycle volume: an ecological study. Environmental Health 11, 12.

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.

Traag, V., Browet, A., Calabrese, F., Morlot, F., 2011. Social event detection in massive mobile phone data using probabilistic location inference.

Utsunomiya, M., Attanucci, J., Wilson, N., 2006. Potential uses of transit smart card registration and transaction data to improve transit planning. Transportation Research Record: Journal of the Transportation Research Board, 119-126.

Vantini, S., Vitelli, V., de France, É., Zanini, P., 2012. Treelet analysis and independent component analysis of milan mobile-network data: Investigating population mobility and behavior. Analysis and Modeling of Complex Data in Behavioural and Social Sciences, 87.

Wang, Z., He, S.Y., Leung, Y., 2018. Applying mobile phone data to travel behaviour research: A literature review. Travel Behaviour and Society 11, 141-155.

Yue, Y., Lan, T., Yeh, A.G., Li, Q.-Q., 2014. Zooming into individuals to understand the collective: A review of trajectory-based travel behaviour studies. Travel Behaviour and Society 1, 69-78.

Appendix – Important Terms

Term	Definition
Active / idle mode	The characteristics and level of detail of network data depends on whether the studied cell phones are in actively used or only switched on and connected to the telecom network (see Caceres et al., 2008; Steenbruggen et al., 2013). Active mode is when a cell phone is being used, for example during a call, while sending a text message or during down- or uploading of internet content. When a phone is not actively used, it is in idle/passive mode. Even when the phone is in idle mode, but connected to the network, the location and movement of the phone is still being recorded, but on a less detailed level than during active mode. It is also possible for the telecom company to actively connect to the phone even though the phone being in idle mode. This is typically applied by the emergency services when they would like to find the location of the phone. (See section 3)
Applications	Smartphone applications are used to record and collect sensor-based data from smartphones, for example GPS sensors. The phones' built-in sensors can record both location and motion, and can either record continuously or at certain moments (controlled by the user) (Wang et al., 2018).
Base Transceiver Station (BTS)	A Base Transceiver Station (BTS, also referred to as base tower or cell tower) is the access point of a mobile device to the network. BTS can cover a smaller or larger area, depending on the context. In urban areas the radius of a BTS is smaller than in rural areas (Wang et al., 2018). The area covered by one BTS is called a cell.
Call detail records	The event-driven type of network data (see Network data) is also referred to as Call Detail Records (CDR). According to Lu et al. (2016) "CDR data comes in an industry standard format, which contains for each of the mobile network operator's subscribers, the location of the closest mobile phone tower at the time of each call, text message or data download. The data is routinely collected and stored by mobile network operators".
Cell	One BTSs area is referred to as a cell (see also Base Transceiver Station and Location Area) (Caceres et al., 2008).
Cell Dwell Time (CDT)	The time a cellular phone is connected to one base station, before switching to a different station, is called the Cell Dwell Time (CDT). This information can be used to measure traffic congestion, by investigating the CDT at adjacent cells along a road (Steenbruggen et al., 2013). When a device in active mode switches to a different cell, a handover takes place.

Handover data	When a smartphone is in use (in active mode) while on the move, the phone may need to switch from one cell to another without losing the connection. This switching mechanism is called a handover, providing the user with a stable connection while moving through the network. The information about handovers (handover data) is stored in the network databases, and this can give information about the mobility of devices in the network (Steenbruggen et al., 2013).
Location area (LA)	The cells, defined by the coverage area of BTSs, are grouped in Location Areas (LA). A LA is thus a larger service area covered by a group of cells (Caceres et al., 2008).
Network data	Network data are data collected by telecom companies about cell phones connected to the Base Transceiver Stations in a network. Network data can be divided into event-driven (see CDR) and network-driven data. Event-driven data comes from cell phones in use (through communication or internet usage), while network-driven data are collected periodically unrelated to the actual use of the cell phones (Wang et al., 2018). See also Active / idle mode.
Received Signal Strength (RSS)	One way to locate a cell phone within a cell or LA is by measuring the strength of the signal from the BTS. Received Signal Strength (RSS) can be used to measure the distance from the BTS to the cell phone (Steenbruggen et al., 2013).
Routing area (RA)	With the introduction of new generations of mobile systems (3G and 4G), more information is sent from the device than in previous systems, and there is therefore a need for more accurate information about the location of mobile devices. These systems use a routing area (RA) which is defined as a subset of cells of an LA. The size of an RA is always less or equal to the LA it belongs to (Caceres et al., 2008).
Triangulation	Triangulation can be used to get more detailed information about the location of a cell phone. When a cell phone is connected to three or more BTSs, the signal strength (RSS) from each BTS can be combined to estimate the position of the cell phone (Steenbruggen et al., 2013).

Institute of Transport Economics (TØI)
Norwegian Centre for Transport Research

Established in 1964, the Institute of Transport Economics is an interdisciplinary, applied research centre with approximately 70 professionals. Its mission is to develop and disseminate transportation knowledge that has scientific quality and practical application.

A private, non-profit foundation, TØI receives basic funding from the Research Council of Norway. However, the greater part of its revenue is generated through contract research. An important part of its activity is international research cooperation, mostly in the form of projects under the Framework Programmes of the European Commission.

TØI participates in the Oslo Centre for Interdisciplinary Environmental and Social Research (CIENS) located near the University of Oslo. See www.ciens.no

TØI covers all modes of transport and virtually all topics in transportation, including road safety, public transport, climate change and the environment, travel behaviour, tourism, land use and urban planning, decision-making processes, freight and travel demand, as well as general transport economics.

Claiming copyright to its products, TØI acts independently of its clients in matters of scientific approach, professional judgment and evaluation. TØI reports are generally downloadable for free at www.toi.no.

Visiting and postal address:
Institute of Transport Economics
Gaustadalléen 21
NO-0349 Oslo

+ 47 22 57 38 00
toi@toi.no
www.toi.no