DEPARTMENT OF TRANSPORT & PLANING, FACULTY OF CIVIL ENGINEERING AND GEOSCIENCES, DELFT UNIVERSITY OF TECHNOLOGY

Use of Mobile Phone Data for Planning a Road Network: Application to the Country of Senegal

Thesis Report

Yihong Wang

June 2015

TU Delft Delft University of Technology

Thesis Author

Yihong Wang BSc

Civil Engineering MSc student

Transport & Planning track

Delft University of Technology

Graduation Committee

Prof.dr.ir. Erik de Romph (chair)

Faculty of Civil Engineering and Geosciences

Delft University of Technology

Dr.ir. Gonçalo H. A. Correia (daily supervisor)

Faculty of Civil Engineering and Geosciences

Delft University of Technology

Dr.ir. Bruno F. Santos (external supervisor)

Faculty of Aerospace Engineering

Delft University of Technology

Ir. Paul Wiggenraad (coordinator)

Faculty of Civil Engineering and Geosciences

Delft University of Technology

Table of Contents

List of Abbreviations

AATR	Autonomous Agency of Road Work of Senegal
ANSD	National Agency of Statistics and Demography of Senegal
CDR	Call Detail Records
D4D	‘Data for Development Senegal’
GDP	Gross Domestic Product
GIS	Geographic Information System
GPS	Global Positioning System
OD	Origin-Destination
OSM	Open Street Map
RMSE	Root Mean Square Error

List of Tables

List of Figures

Acknowledgements

The thesis is the final product of my graduation project. The idea for this project originated from the 'Data for Development Senegal' challenge (D4D challenge) where the anonymous data of Orange users were open for researchers to solve various problems in different categories, such as transport, agriculture and energy. Some important transport issues in Senegal are encouraged to be explored, of which one is about how to identify the mobility patterns for road infrastructure development. Focusing on this subject matter, Prof. Erik de Romph, Dr. Gonçalo Correia and I participated in the challenge and finally won the Transport Prize. The thesis is an extended version of the paper we submitted to the D4D challenge.

I would like to thank all the committee members for their academic guidance, advice and support. My special thanks go to Prof. Erik de Romph, without whom I would never have found such a great opportunity to conduct research on what I am really passionate about. I am sincerely thankful to Dr. Gonçalo Correia, who inspired me with his great ideas constantly and provided very detailed feedback on my drafts. Furthermore, I wish to thank Dr. Bruno Santos for answering all my questions about the road network design problems and Ir. Paul Wiggenraad for giving me valuable advice at the meetings.

And last, but certainly not least, I would like to thank my family, my girlfriend Huipeng and all my friends for their endless love and support. Especially, I owe my gratitude to my parents and grandmother, who support my graduate studies financially and emotionally. This work is dedicated to them.

Yihong Wang

Delft, May 10th 2015

Summary

Use of Mobile Phone Data for Planning a Road Network: Application to the Country of Senegal

The thesis is the outcome of a research on using mobile phone data to understand mobility patterns in a country and to give advice about decisions on how to design the national and regional road network. It provides not only several practical solutions to the existing road network design problems in Senegal but also a methodological approach, consisting of three main parts: filtering mobile phone traces to derive mobility patterns, estimating a new form of gravity model to understand how mobility patterns can be influenced, and optimizing road network based on the new gravity model. The formulation of the new gravity model indicates that the telecommunication patterns (i.e., the aggregate number of calls and text messages between zones) can be a better proxy than the distribution of population to predict mobility patterns and to measure the accessibility of a road network, allowing transport planners to make better decisions on road network design. The presented methodology is useful and reproducible not only to the country of Senegal but also to other countries where mobile phone data are available to transport planners.

Background

In a country like Senegal where the mobility data is scarce, it is difficult for the government to understand the mobility patterns well and to make good decisions on how to plan the road network. Even in those countries where the mobility data is available, the traditional way of collecting mobility data by surveys is criticized for being too costly, time-consuming and static. In recent years, mobile phone data is used more and more as a proxy for mobility patterns at a regional scale. It shows great potential for not only understanding mobility but also making decisions on transport planning in many existing studies for many specific purposes. However, to the author's knowledge, few studies have been made on road network planning by using mobile phone data, to which this research is expected to contribute.

The 'Data for Development Senegal' challenge (D4D challenge) organized by Orange, a mobile phone network operator, made the data of Orange users available, giving the opportunities to solve the practical road network design problem in Senegal, where the fundamental and applied knowledge can be accumulated on the use of mobile phone data for understanding mobility patterns and planning a road network, especially in a developing country where mobility information is scarce.

A filtering algorithm

A trip is defined as a one-way movement, from where a person starts (origin) to where the person is going (destination). To have a picture of mobility patterns in a target area during a

given period, the most common way is to build the origin-destination (OD) trip matrix, containing the aggregate number of trips between each OD pair made by people during the given period. In the datasets provided by the D4D challenge, the traces where and when travellers used their mobile phones were recorded. Connecting those geographical traces in a time sequence, most trajectories of each individual during one year can be easily reproduced on the map. However, the information of where each trip started and ended was missing.

To solve this problem, an algorithm was produced to find the OD information by filtering out all the unimportant traces of where users have passed by and did not stay and keeping the important traces, namely origins or destinations. The idea of this algorithm is based on a reasonable assumption that if one user, as observed in the mobile phone traces, stayed at one place for a very short period, which was assumed as less than two hours in this study, this place should not be an origin or destination for the user. After applying the filtering algorithm, the trajectories were made approximate to a series of trips, named as the filtered trajectories.

In this way, the monthly OD filtered trajectory matrices were estimated between each two departments[1] of sampled users in 2013 by using the provided dataset of mobile phone traces. Those matrices do not contain the exact number of trips made by population, but contain the aggregate number of filtered trajectories, which helped estimate the number of trips made by sampled users. They were named as the relative OD matrices, since it was assumed that they can serve as proxies for the real OD trip matrices of the population and reflect the mobility patterns in Senegal proportionally.

The estimation of a traditional gravity model

Making the decisions on road network planning should depend on mobility patterns in that country. However, given the fact that any road connection changes will in return influence the mobility patterns, having proxies for historical mobility patterns seems not sufficient. It is necessary to have a good understanding of how mobility patterns can be influenced by potential road network changes. A traditional gravity model can help us understand it in this way: the mobility between two departments is proportional to the product of the populations in these two departments and inversely proportional to the travel time between two departments[2]. By fitting the data of the estimated relative OD matrices, departmental population and the skim matrices of shortest travel time, the parameters of the traditional gravity model were estimated for the country of Senegal. As a result, the fitness was not

[1] Department is a spatial/administrative unit used in Senegal as well as the spatial unit which was focused on in this case.
[2] Here, the travel time between two departments can be simply assumed as the travel time between the two capitals of the two departments.

satisfying, and especially, the model seemed not be able to explain the distribution of long-distance travel demand very well.

The relation between telecommunication and travel

One advantage of mobile phone data over other mobility data sources is its potential for revealing social interaction reflected in people's calling and messaging behavior, which can sometimes serve as a good reference to understand human mobility. Based on a literature study on the relations between travel and telecommunication, such relations were explored empirically by comparing the estimated relative OD matrices and the mobile phone interaction matrices. The mobile phone interaction matrices contain the aggregate number of mobile phone interactions (i.e., calls and text messages) of all Orange's users from one department to another department, extracted from the provided dataset of communication between antenna towers. It was found that the aggregate number of filtered trajectories made by sampled users between each two departments is almost proportional to the aggregate number of mobile phone interactions and inversely proportional to the travel time between departments.

The formulation and the estimation of a new form of gravity model

The observed relation between telecommunication and travel gave inspiration to construct a new form of gravity model, based on the aggregate number of mobile phone interactions instead of population which the traditional gravity model is usually based on. The parameters of this new model were estimated, which gave a model to predict elastic travel demand patterns for potential road network changes in Senegal.

It was found that the aggregate number of mobile phone interactions can be a better proxy than population for predicting travel demand since the new model fits the data better. Especially, the new model based on mobile phone interaction is more capable of modeling long-distance mobility than the traditional model. In the author's opinion, this might be because that the mobile phone interactions can reflect the social ties, which influence travel demand to a larger extent than population does.

Road network planning for Senegal

In the final step, the national and regional road network was designed by using a bi-level multi-objective optimization model, where two kinds of action can be performed to make changes on road network: the construction of a new road; and the upgrading of an existing road. At the lower level of the optimization model, the new gravity model was used to predict the distribution of travel demand, responding to any road network changes, and then traffic was assigned to different routes by following the 'all-or-nothing' principle. At the higher level, the chosen objective indicators were assessed for each road network change based on the corresponding distributed travel demand and assigned traffic, and moreover, a weighting method was used to calculate an overall value based on the values of different objective indicators with different weights, which can be given to reflect the relative importance of each objective. This overall value indicates to what extent the objectives are achieved. A local search algorithm was used to generate different solutions (i.e., road network changes) in an efficient way, and to find which solution would lead to the best result, in terms of the maximization of the overall value, under the assumed budget.

Conclusions and recommendations

The created tool gained good insight into where and how to expand the Senegal network. Under the assumed budget constraints, the model suggested that the focus of road development should be on the western part of Senegal for the efficiency objective, whereas on the south-eastern part for the equity objective. In addition, the Trans-Gambia ferry service was strongly recommended to be replaced by a bridge connection for all objectives.

It is believed that the presented methodology can be applied not only to the country of Senegal but also to other countries where mobility information is scarce and mobile phone data is available, for understanding mobility and planning road networks. Firstly, the filtering algorithm can perform well to extract mobility information from mobile phone traces. Secondly, the formulated new gravity model can serve as a new tool to model mobility patterns and can be widely used for other transport planning purposes. Thirdly, the optimization model based on the new mobile-phone gravity model can help the government to make better decisions on national and regional road network planning using mobile phone data. Based on the actual planning goal, the government can determine the weights of different objectives and the actual available budget in the model by themselves, in order to obtain the best solution under a certain scenario.

Furthermore, having additional traffic information available is recommended to validate the estimated mobility patterns and the estimated new gravity model. In this way, the actual number of trips, instead of the number of filtered trajectories made by sampled users, can be approximated in the best conceivable situation. For further improving the optimization model, more detailed network data, such as road capacity, can be collected, and then, instead of the simple assumption of 'all-or-nothing', the effect of congestion can be taken into consideration when travel demand is assigned to different routes in the model.

Keywords: **mobile phone data; OD matrix estimation; new gravity model; road network design**

1. Introduction

1.1 Background and Objectives

In recent years, significant changes have been happening in many aspects of transport planning. One of the greatest changes is a growing number of potential new data sources (e.g., mobile phone data, public transport smart card data and GPS data). They are being used more and more to reflect mobility patterns (Demissie, de Almeida Correia, & Bento, 2013a; Zheng, Li, Chen, Xie, & Ma, 2008), which transport planners need to understand in order to make a better planning. Such real-time data meets the needs of transport planners quite well since they always agree on that the zonal data collected once over several years is too outdated, and moreover, the collection of traditional mobility survey data costs much time and money.

Among those new data sources, mobile phone data becomes quite popular for mobility research today mainly because (1) the mobile penetration is rather high even in a less developed country like Senegal (Eto, 2012); and (2) a mobile phone device can play the role as a wearable sensor to collect the data representing the geographic locations of individual user (Ratti, Williams, Frenchman, & Pulselli, 2006). To that extent, the fact that the majority of people bring mobile phones with them in their daily lives allows transport planners to trace almost every individual's movements in a target area over a certain period. Moreover, another advantage of mobile phone data is its potential for revealing social interaction reflected in people's calling and messaging behavior, which can sometimes serve as a good reference to understand human mobility (Cho, Myers, & Leskovec, 2011; Wang, Pedreschi, Song, Giannotti, & Barabasi, 2011).

Under the influence of the trends in mobile phone data, the 'Data for Development Senegal' challenge organized by Orange, a mobile operator sharing the largest market in Senegal, was dedicated to the use of mobile phone data to contribute to the socio-economic development and welfare of the population, for which one priority subject matter is transport planning. Anonymous data of Orange's mobile phone users in Senegal was available for researchers to help address society development questions in novel ways (de Montjoye, Smoreda, Trinquart, Ziemlicki, & Blondel, 2014). The challenge allowed researchers to explore the mobile phone data in Senegal for understanding the mobility and thus making good decisions on transport planning. This research focuses on the specific topic of national and regional road network planning in Senegal, since it is regarded by the local authorities and the international experts as one of the most important transport issues worth exploring in the country.

When most people explore a country on Google maps, one of the components they would notice at first sight is the road network, which connects different parts of a country to satisfy travel demand. For long-term development, the government makes decisions on whether they should add new roads or upgrade the existing ones to improve the level of service provided of roads. This task is especially urgent for the government of Senegal, since it has been found that population growth is outstripping road development there (World Bank, 2004). In a less developed country like Senegal, it is particularly important to consider the cost efficiency of road network planning due to the strong trade-off between increasing demand and budget limitations. This goal can be achieved by using the technique of road network optimization, which is regarded as one of the most challenging transport topics (Yang & Bell, 1998), and in this case study, there is added complexity due to the scarcity of mobility data in Senegal.

Traditionally, in the process of road network optimization, mobility data should be collected through the costly and time-consuming mobility surveys to help transport planners understand mobility in a country and then decide where and how to expand the road network to meet those transport demand. In this case, the D4D challenge gives an opportunity to understand mobility by using the mobile phone data instead of the traditional mobility survey data, which is often scarce in a less developed country like Senegal.

Mobile phone data shows great potential for exploring human mobility and making decisions on transport planning. It has been widely used for many specific purposes, including estimating origin-destination trip matrices (White & Wells, 2002), transportation mode inference (Wang, Calabrese, Di Lorenzo, & Ratti, 2010), road traffic status detection (Demissie, de Almeida Correia, & Bento, 2013b), traffic management and control (Astarita & Florian, 2001), etc. However, to the author's knowledge, few studies have been made on road network planning by using mobile phone data, to which this research is expected to contribute.

In summary, one of the aims of the thesis is to solve the practical road network design problem in Senegal by using the mobile phone data provided by the D4D challenge. More important, this case study is aimed to present a methodology of using mobile phone data for understanding mobility patterns and planning a road network in a country, especially in a less developed country where traditional mobility information is scarce.

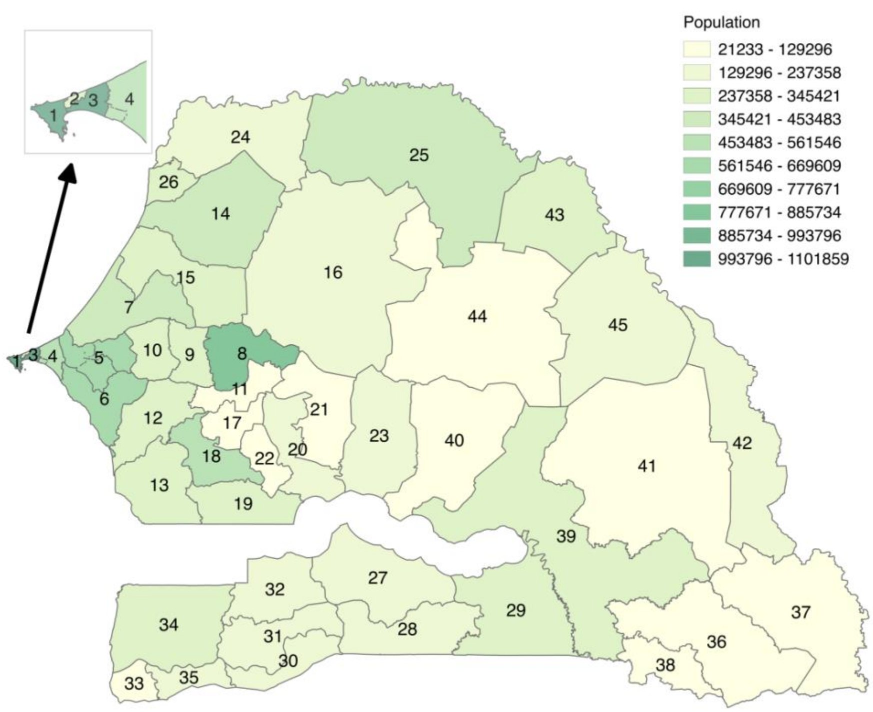
In the following sections (i.e., Section 1.2 and Section 1.3), the mobility-related information of the target country in the case study and the mobile phone datasets used for research in that country are introduced.

1.2 Introduction of the Country and Its Spatial Information

As the target country in the case study, Senegal, located in West Africa, covers a land area of almost 197,000 square kilometers, and has an estimated population of about 13 million (Wikipedia, 2015c). Senegal is a stable country and has substantially strengthened its democratic institutions since it was independent from France in 1960 (World Bank, 2015).

Senegal ranks 22th (out of 54) on the list of African countries by estimated nominal GDP in 2015 (Wikipedia, 2015a), and ranks 25th (out of 54) and 153th (out of about 190) respectively on the list of African countries and world countries by estimated GDP per capita in 2010 (Wikipedia, 2015b). It is noteworthy that the geographic disparities are very pronounced in Senegal, with almost 2 out of 3 residents poor in rural areas, especially in the south, versus one in four in Dakar (World Bank, 2015).

The D4D challenge provides the geographic information system (GIS) shapefile of Senegal, which contains the information (e.g., name, shape and location) of administrative divisions of Senegal (arrondissement, department and region, ordered by size from small to big). Among all kinds of administrative divisions, department was the basic spatial unit used in this study. To that extent, the case study focuses on the mobility between departments in Senegal, and the road network is planned to satisfy the travel demand between departments. The population and area data of Senegal, collected per department, were found on the website of the National Agency of Statistics and Demography (2013). The population distribution of Senegal in the scale of department and the names and codes of the departments are shown in Figure 1. Apart from Dakar, the major cities in Senegal include Touba, Thies, Rufisque, Kaolack, etc., all of which are close to Dakar and in the western part of the country.



ID	Name
1	DAKAR
2	GUEDIAWAYE
3	PIKINE
4	RUFISQUE
5	THIES
6	MBOUR
7	TIVAOUANE
8	MBACKE
9	DIOURBEL
10	BAMBEY
11	GOSSAS
12	FATICK
13	FOUNDIOUGNE
14	LOUGA
15	KEBEMER
16	LINGUERE
17	GUINGUINEO
18	KAOLACK
19	NIORO
20	KAFFRINE
21	MALEM HODDAR
22	BIRKELANE
23	KOUNGHEUL
24	DAGANA
25	PODOR
26	SAINT-LOUIS
27	MEDINA YORO FOULAH
28	KOLDA
29	VELINGARA
30	GOUDOMP
31	SEDHIOU
32	BOUNKILING
33	OUSSOUYE
34	BIGNONA
35	ZIGUINCHOR
36	KEDOUGOU
37	SARAYA
38	SALEMATA
39	TAMBACOUNDA
40	KOUPENTOUM
41	GOUDIRY
42	BAKEL
43	MATAM
44	RANEROU
45	KANEL

Figure 1 the population distribution of Senegal and the names and codes of the departments

The information of the road network in Senegal was found on the website of Digital Logistic Capacity Assessment (2013). The roads in Senegal can be classified into five levels: national roads, regional roads, department roads, urban way and classified tracks. National roads provide long-distance connections between several administrative regions and with neighboring states. Regional roads provide connections between different departments of the same region. The other three levels of roads mainly provide the connections within the departments. Since department is the basic spatial unit used in the case study, the focus of this study is mainly on the network of national and regional roads which connect the different departments in the country. Note that several departmental roads should be taken into account as well since they cross different departments for some reason.

A GIS layer of Senegal road network in 2002 was found on the website of ArcGIS (2013) provided by the D4D challenge, including 1139 roads of different levels, namely national, regional and departmental road levels, shown in Figure 2. The source of road network information is the Autonomous Agency of Road Work of Senegal (AATR).

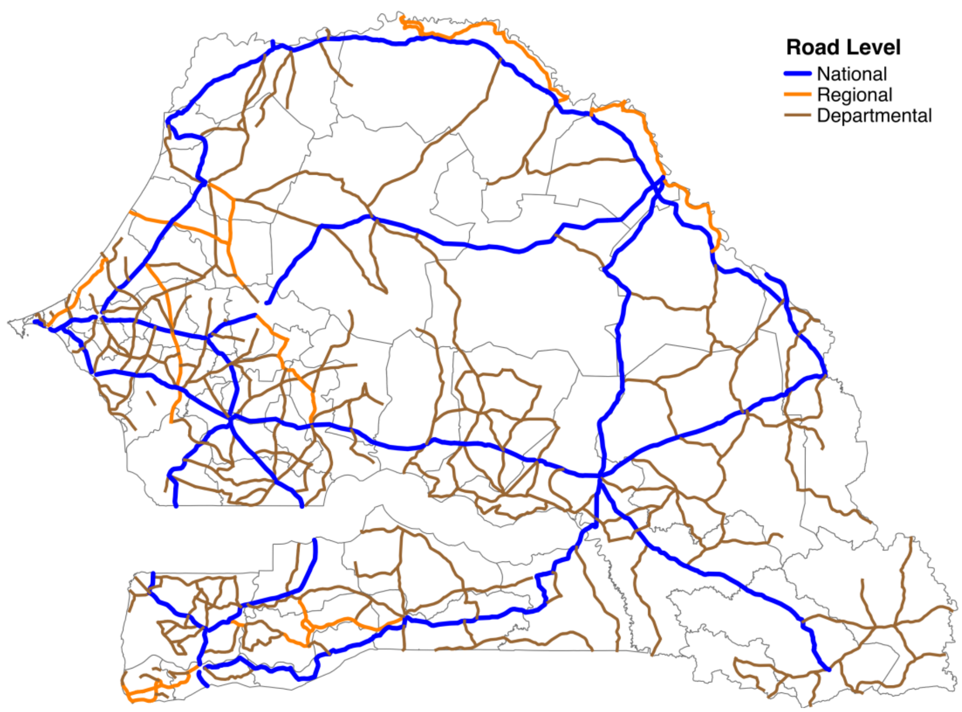


Figure 2 all roads of different levels in Senegal in 2002 according to AATR

In addition, Open Street Map (OSM) was used as a layer in GIS software to show more details in the country, and the latest road network information can be complementary to the layer of outdated Senegal road network in 2002. One of the most significant changes occurred in the meantime is the construction of a toll highway in Dakar region, which was opened to traffic in three phases: the Dakar (downtown)-to-Patte d'Oie was opened first in 2008, and then Patte d'Oie-to-Pikine section was opened in 2011, followed by the opening of the Pikine-to-Diamniadio section on August 1st, 2013 (Eiffage, 2013; Invest in Senegal, 2015). The locations of those highway sections are illustrated in Figure 3.

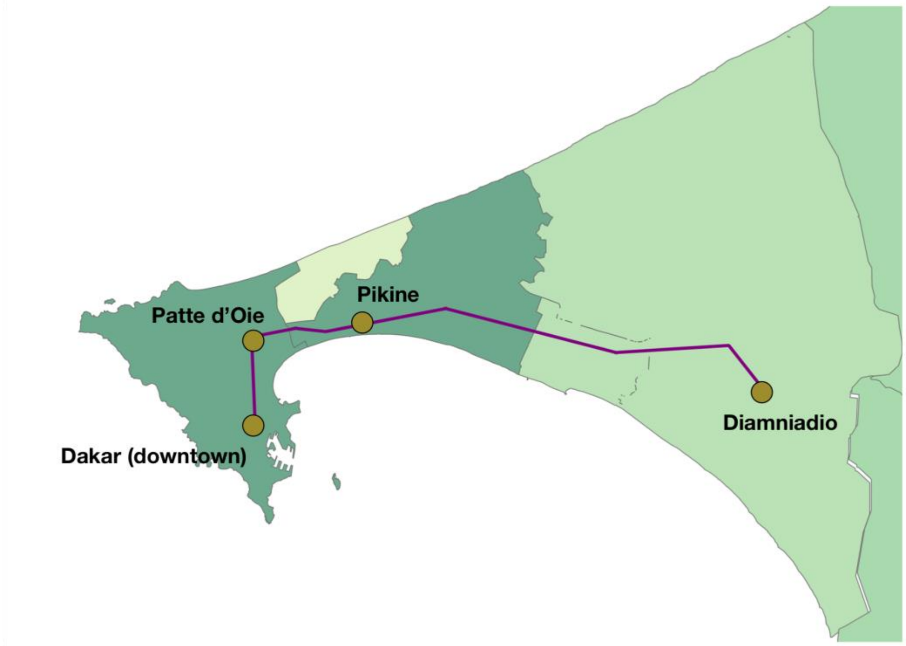


Figure 3 the newly-constructed toll highway in Dakar region

Especially, since Gambia is located as an enclave of Senegal, travellers sometimes need to cross Gambia to reach the destinations. To meet that demand, there are mainly two national roads crossing Gambia, as indicated in Figure 4. However, it can be observed that there is a river crossing those two national roads in Gambia. Currently, there are mainly two ferry services available for connecting the roads across the river.

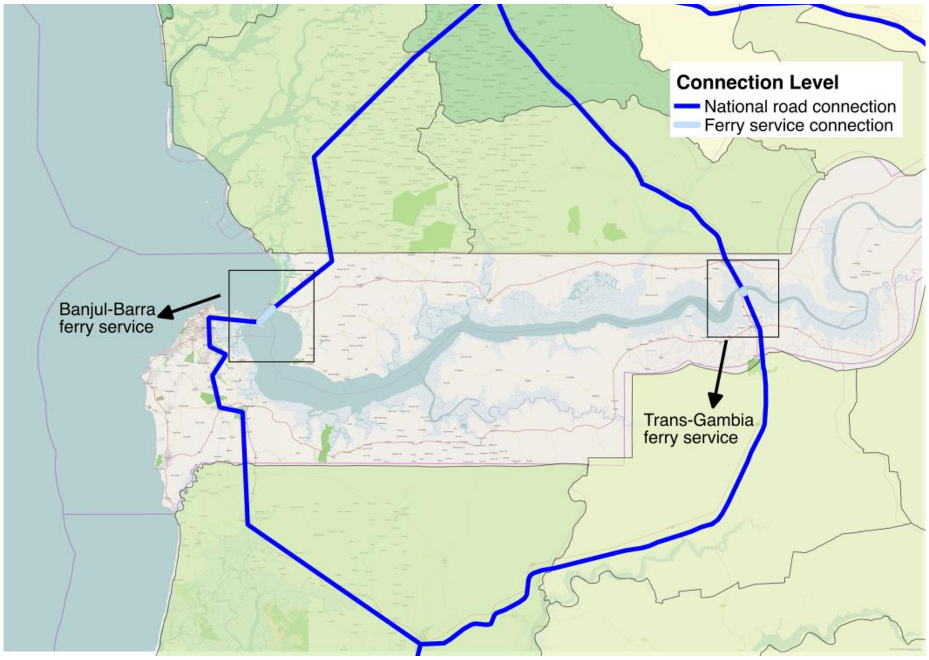


Figure 4 ferry service locations

Besides satisfying the travel demand inside the country, as a part of Trans-African highways, three national roads can serve as the connections to the other parts of Africa, as shown in Figure 5.

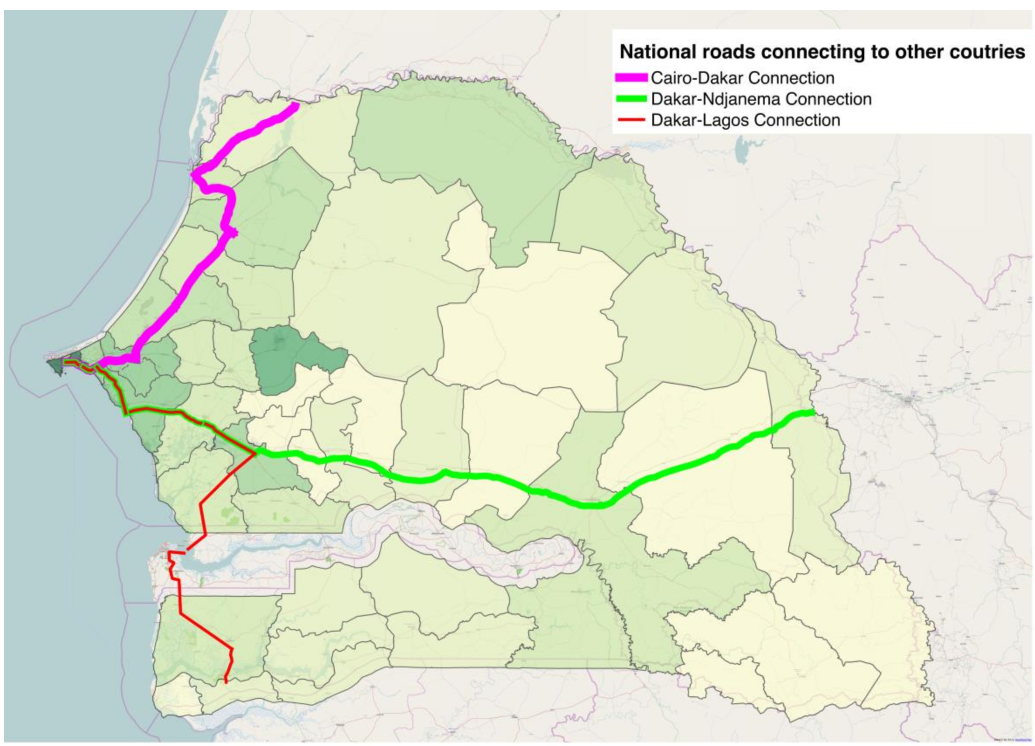


Figure 5 international road connections

More detailed information about influencing factors on mobility in Senegal can be found in Appendix A.

1.3 Introduction of the Original Mobile Phone Datasets

The datasets provided by the D4D Challenge are based on Call Detail Records (CDR) of phone calls and text exchanges between more than 9 million of Orange's customers in Senegal between January 1, 2013 and December 31, 2013. A CDR is a data record produced by a telephone exchange or other telecommunications equipment that documents the details of a telephone call or other communications transaction (e.g., text message) that passes through that facility or device, and the record contains various attributes of the call, such as time, duration, completion status, source number and destination number (Horak, 2007).

Antenna-to-antenna traffic of calls and text messages of more than 9 million users for 1666 antennas on an hourly basis is provided as Dataset 1, which was named as mobile phone interaction data since these data contain the information of interaction intensity (both number and duration of calls and number of text messages) between two zones in the country. Table 1 shows what a file in Dataset 1 looks like.

Year	Month	Day	Hour	From (antenna ID)	To (antenna ID)	Aggregate number of calls	Aggregate duration of calls (unit: minute)
2013	1	1	0	1	1	1	54
2013	1	1	0	1	2	1	39
2013	1	1	0	1	24	1	2957
2013	1	1	0	1	186	1	56
2013	1	1	0	2	2	22	418
2013	1	1	0	2	3	2	53
2013	1	1	0	2	4	4	455

Table 1 a random example of the content of Dataset 1

Dataset 2 provides fine-grained mobility data on a rolling 2-week basis for a year at individual level for about 300,000 randomly sampled users having more than 75% of the days with interactions in one year. Once a user made a phone call or had a text message with others, the location of the antenna to which this user connected at that time was recorded. Thus, his trajectories can be captured over two weeks. Since the focus of this research was on the mobility patterns in a larger scale, this dataset was not further explored in this research.

Dataset 3 provides one year of coarse-grained mobility data at individual level for about 150,000 randomly sampled users having more than 75% of the days with interactions in one year. Once a user made a phone call or had a text message with others, the arrondissement where he connected to the antenna at that time was recorded. Thus, his trajectories can be captured over one year. A random example showing what a file in Dataset 3 looks like is given in Table 2. Dataset 2 and Dataset 3 can be called as mobile phone traces. It should be noticed that the users presumed to be machines or shared phone users were excluded in these datasets.

User ID	Year	Month	Day	Hour	Minute	Second	Arrondissement ID
136084	2013	2	24	1	50	0	3
136084	2013	2	24	0	50	0	3
106356	2013	2	23	17	30	0	3
106324	2013	2	9	10	50	0	3
106324	2013	2	9	12	40	0	3
105696	2013	2	21	6	30	0	3
105696	2013	2	25	22	20	0	3

Table 2 a random example of the content of Dataset 3

As shown in Figure 6, the sample sizes of Dataset 1 and Dataset 3 were compared with the size of population in Senegal. As can be observed, Dataset 1 can reflect the telecommunication patterns of people in Senegal to a large extent, and the sample size of Dataset 3 is rather small compared to the population.

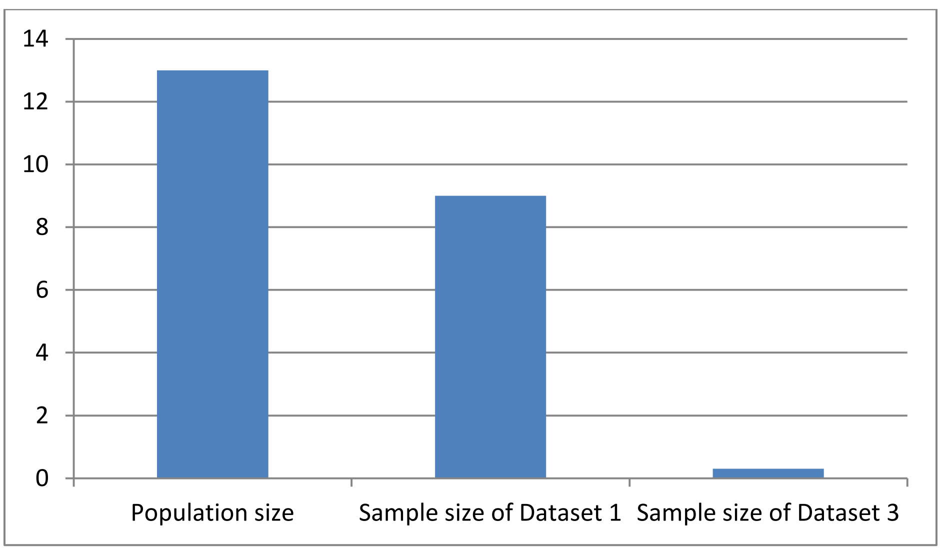


Figure 6 the comparison between population size and sample sizes of mobile phone datasets (unit: million people)

1.4 The Outline of the Remaining Chapters

Based on the formulated objectives (as marked in bold in Section 1.1), in Chapter 2, a literature review is made to define the problems of this research, and the potential solutions to those problems are found in the literature if possible. In Chapter 3, the research questions are formulated based on the defined problems and the methodology to be used for answering those questions is presented. Chapter 4, Chapter 5 and Chapter 6 focus on the case study, in which the mobile phone data, provided by the D4D challenge, is used to solve the road network design problems in Senegal. In Chapter 4, the spatial information of Senegal and the mobile phone interaction data (Dataset 1) are further explored, and mobility information is extracted from the mobile phone trace data (Dataset 3). In Chapter 5, gravity models are developed to understand the mobility in Senegal. In Chapter 6, the decisions are made on where and how to expand the existing road network in the country based on the estimated gravity model. Finally, in Chapter 7, the conclusions of the research are drawn, and the contributions of the thesis are explained. Moreover, it is discussed whether the methodology can be widely used in other situations, and several potential approaches are recommended to improve the methodology in further research.

2. Literature Overview and Problem Description

As clarified in the previous chapter, the objectives of this work are to find solutions to the practical road network design problems in Senegal and to present a methodology of using mobile phone data for understanding mobility patterns and planning a road network in a country. Therefore, in the main part of the thesis, the application of the methodological approach is illustrated in a case study involving the national and regional road network in Senegal.

To reach the objectives, the main problem addressed in this thesis should be how to use mobile phone data properly for understanding the mobility in a country and thus planning the road network better.

In order to have an insight into how the addressed problem can be solved, literature overview is made in this chapter. The conceptual framework describing the role of pervasive technology datasets in urban planning is reviewed at first in Section 2.1, followed by an adapted framework which illustrates the role of mobile phone data in road network planning in this case. Next, the literature about how to have a picture of mobility patterns by using mobile phone data is reviewed in Section 2.2, which suggests ways of exploring mobile phone traces (Dataset 3) for mining mobility information in Senegal. Some limitations of the methods and their solutions given in the existing researches are discussed. In Section 2.3, several papers are reviewed to find the importance of elastic travel demand prediction for network optimization purposes. The historical mobility patterns are not sufficient to make good decisions on road network planning because any road connection changes will in return influence the mobility patterns. To that extent, it requires understanding how mobility patterns are formed based on the network by using an unconstrained gravity model. It is discussed how to estimate such a gravity model in the current study, and the usual functional forms of the gravity model are listed. In addition, the limitation of this model is discussed. In Section 2.4, based on the literature review about the relations between telecommunication and travel, the possibilities are discussed to use the mobile phone interaction data (Dataset 1) as a proxy for current travel demand pattern or even to predict elastic travel demand pattern. In Section 2.5, a brief literature review of road network design problem is presented.

2.1 The Conceptual Framework

Calabrese, Ferrari, & Blondel (2014) established a framework to explain the role of pervasive technology datasets, such as mobile phone data, in urban planning and management. In the framework, as illustrated in Figure 7, the human behavior of people in a city reflects how citizens use the built environment and the services offered by a city. The pervasive technology datasets are able to capture human behavior and produce related datasets that contain very useful information for planning and management.

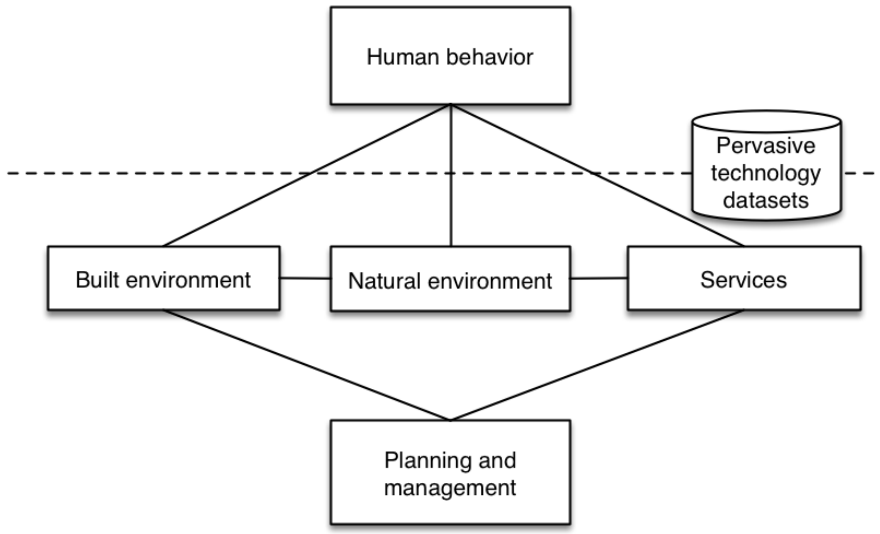


Figure 7 the role of pervasive technology in an urban scenario (Calabrese, Ferrari, & Blondel, 2014)

This conceptual framework can be adapted for the current study to describe the relations between mobility patterns, mobile phone data, road network and network planning in a country.

The available mobile phone data can help depict mobility patterns which reflect how people used road network and services (e.g., ferry services). Based on that, transport planners can have a rough idea about which road connection should be added or improved to meet the intensive demand. However, there is one complexity, which should be dealt with in this case: road network planning would always influence mobility patterns in return over the long term. Therefore, transport planners should not only know how people used the existing road network but also understand the observed mobility patterns deeply in order to further predict how people will use the upgraded road network in the future.

2.2 Origin-Destination Estimation Using Mobile Phone Traces

A trip is defined as a one-way movement, from where a person starts (origin) to where the person is going (destination). To have a picture of mobility patterns in a target area during a given period, the most common way is to build the origin-destination (OD) trip matrix, containing the aggregate number of trips between each OD pair made by people during the given period.

Due to both the fast expansion of mobile market and the availability of pervasive technology, many researchers have found the great opportunities to derive travel demand by using mobile phone data, which are regarded as a game changer to build OD matrices for various transport studies (Caceres, Wideberg, & Benitez, 2007; Nanni et al., 2014; White & Wells, 2002;

Calabrese, Di Lorenzo, Liu, & Ratti, 2011). This method could be quite efficient, compared with traditional methods like mobility surveys which are too costly, time-consuming and static.

In general, the traces where and when travelers used their mobile phones were recorded. Connecting those geographical traces in a time sequence, partial trajectories of each individual during one year can be easily reproduced on the map. The reason why the trajectories can only be partial is that it is not possible for users to use mobile phone (i.e., to be traced) constantly during the trips. In most studies, an indicator, namely the sampling rate (i.e., the time interval between consecutive location records), was used to assess how representative the partial trajectories can be to reflect the actual trajectories of individuals (Calabrese et al., 2011; Hoteit, Secci, Sobolevsky, Ratti, & Pujolle, 2014). In fact, to most users, the frequency of using mobile phone is not quite high in their daily lives, leading to the low sampling rates in most cases. This would further result in the loss of most details of people’s movement. One solution to this problem is a technique developed by Zheng, Zheng, Xie, & Zhou (2012) to reduce the uncertainty of low-sampling-rate trajectories by leveraging travel patterns inferred from the historical data. This technique is more suitable for the cases which focus on the fine-grained mobility, and is better at improving the accuracy of route depiction. In the current study, the coarse-grained (interdepartmental) mobility is focused on, and it is not necessary to depict the routes. Therefore, the required value of sampling rate can be relatively small.

Such trajectories are still not able to serve as a proxy for origin-destination travel demand of population, since the mobile phone data can only provide the trajectories without the information regarding where each trip starts and ends. To that extent, a long trip and a series of short trips might be mixed up. The most used solution is the technique of stay extraction to figure out where a traveler stays and where he or she passes by (Zheng, Zhang, Xie, & Ma, 2009). In general, the extraction of a stay point depends on two scale parameters, a time threshold and a distance threshold. A stop is regarded as a sequence of traced positions where the distance between any adjacent positions is less than a distance threshold, and time spent at these positions is greater than a time threshold (Phithakkitnukoon, Horanont, Di Lorenzo, Shibasaki, & Ratti, 2010). Usually, travelers stay at the locations of home and work in their daily lives. This is why in many studies, researchers detected the home and work locations of each individual at first and focused only on the commuting trips (Nanni et al., 2014; Csáji et al., 2013).

Another problem is that the estimated OD matrix only includes the trajectories made by the sampled mobile phone users. To check whether the sample is biased, researchers compared the density of mobile phone users’ homes and the density of population (Calabrese et al., 2011). Furthermore, researchers try to come up with an OD matrix containing the actual number of trips by using a scaling factor. Based on the OD matrix estimated by using mobile phone data, Nanni et al. (2014) suggested to weigh a factor, which depends on the marketing penetration, mobile phone ownership and mobile phone usage, to make an estimate of actual traffic flow. Iqbal, Choudhury, Wang, & González (2014) suggested to scale up the OD matrix by using the scaling factors which result best matches with the observed traffic counts.

Because of those limitations, the OD matrix derived from mobile phone traces is often questioned about its validity, especially when the quality of data (e.g., sampling rate, penetration rate, etc.) is not high enough. There are two ways to examine the validity of the estimated OD matrix. One way is to check if the estimated OD matrix can fit well to a gravity model (Csáji et al., 2013; Calabrese et al., 2011). No knowledge about the parameter values of the model is required, and a high adjusted R-squared would mean the higher validity of the estimated OD matrix. The other way is to compare the estimated OD matrix with the available Census data from existing mobility surveys (Calabrese et al., 2011). However, as mentioned before, this is not available for many countries, such as Senegal.

2.3 Traditional Ways of Modeling Origin-Destination Distribution

As explained previously, the estimation of the current mobility patterns is not adequate, and a prediction of future travel demand is needed for planning. In the paper by Santos, Antunes, and Miller (2009), it was argued that in many cases of road network optimization, travel demand is assumed to be known in advance. However, this is a poor assumption since the addition of new arcs and the improvement of existing arcs will influence travel costs and thus change the distribution of existing trips and even create the new trips. Santos, Antunes, and Miller (2009) solved this problem by applying an unconstrained gravity model to predict the elastic travel demand for all possible solutions, responding well to different possible generalized travel costs between each pair of two zones as they change with different networks. In such a case, it is often assumed that population of different regions would not be shifted to a large extent in the future.

As known, in a gravity model concerning trip distribution, the number of trips between two zones should be proportional to a trip generation indicator (e.g., population) and inversely proportional to the generalized travel cost between the zones (de Dios Ortúzar & Willumsen, 2011). A simplest version of the gravity model concerning trip distribution has the following functional form:

$$T_{ij} = K_0 \frac{P_i P_j}{{d_{ij}}^2} \tag{2-1}$$

Where, the scaling constant K_0 is the gravity constant for trip distribution; T_{ij} is the number of undirected trips between two zones; P_i and P_i are respectively the population of zone i and zone j; and d_{ij} is the Euclidean distance between zone i and zone j. The model was further generalized by assuming that the effect of distance or 'separation' could be modeled more precisely by a cost function, which can be a function of distance or travel time or generalized cost between the zones (de Dios Ortúzar & Willumsen, 2011; McNally, 2008). Also, the improvements included the use of total trip ends (O_i and D_j) instead of total population. Due to the lack of information regarding trip ends, sometimes O_i and D_j can be replaced by a power function of the population (Csáji et al., 2013). The improved model can be written as (Csáji et al., 2013; de Dios Ortúzar & Willumsen, 2011):

$$T_{ij}' = K_1 P_i^a P_j^b f(c_{ij}) \tag{2-2}$$

Where, T_{ij}' is the number of directed trips between two zones, and an easy calculation can be made to transfer the number of directed trips T_{ij}' to the number of undirected trips T_{ij}, as shown in Equation (2-3); a and b are the parameters for populations; $f(c_{ij})$ is the cost function; c_{ij} is the travel cost between i and j; The travel cost c_{ij} can be distance or travel time or generalized cost.

$$T_{ij} = T_{ij}' + T_{ji}' \tag{2-3}$$

The popular versions for cost function can be classified into exponential function, power function and combined function, which are formulated respectively as follow (de Dios Ortúzar & Willumsen, 2011):

$$f(c_{ij}) = e^{-\beta c_{ij}} \tag{2-4}$$

$$f(c_{ij}) = c_{ij}^{-n} \tag{2-5}$$

$$f(c_{ij}) = c_{ij}^{n} e^{-\beta c_{ij}} \tag{2-6}$$

Where, β and n are the exponential parameter for cost function and the power parameter for cost function respectively.

One problem is that this kind of gravity model may not perform well all the time. It was found that if the distance between two zones is larger than 150 kilometers, the number of trips no longer depended on the actual distance (Csáji et al., 2013). Also, population census is sometimes questioned regarding its accuracy, which might lead to the inaccuracy of the gravity model. Moreover, in this kind of gravity model, the social interaction between two zones is not taken into consideration. For example, imagine that two densely populated areas are close to each other, while the people in these two areas use different languages. It can be assumed that there would not be as many trips as the gravity model predicts.

2.4 The Relation between Telecommunication and Travel

As a communication tool, a mobile phone is able to record not only the trajectories of its own user, but also the interactions he or she makes, either by calls or by text messages, with other people. The revealed telecommunication patterns can serve as a good reference to understand human mobility, and some researchers tried to explore the relations between telecommunication and travel, between which it was found again and again that there exists a complementarity effect (Mokhtarian, 2002; Calabrese, Smoreda, Blondel, & Ratti, 2011; Kamargianni & Polydoropoulou, 2013), especially at an aggregate level (Plaut, 1997; Calabrese et al., 2011; Hsiao, 2007). Our question is whether this kind of relation can help understand more regarding the mobility patterns in Senegal.

It was found in a case study in Belgium that the total call duration between two zones was proportional to the product of population of two zones and that an inverse-square law decrease was found between the call duration and the distance, and a gravity model concerning the intensity of telecommunication was then estimated (Krings, Calabrese, Ratti, & Blondel, 2009):

$$I_{ij} = K_2 \frac{P_i P_j}{d_{ij}^{\ 2}} \tag{2-7}$$

Where, the scaling constant K_2 is the gravity constant for a time span of 6 months of calling activity, and I_{ij} is the undirected communication intensity (total call duration) between two zones, and P_i and P_j are respectively the population of zone i and zone j, and d_{ij} is the Euclidean distance between zone i and zone j. It should be noted that this gravity model of communication intensity was fit to the reality in Belgium, where the distance between every two zones is not large.

The equation indicates that the intensity of telecommunication would not change in the future if it is simply assumed that population would not change.

Combining the gravity model regarding intensity of telecommunication, given as Equation (2-7), and the simplest gravity model regarding trip distribution mentioned in the previous section, given as Equation (2-1), it results in a linear relationship between the intensity of telecommunication and the number of trips between two zones:

$$\frac{T_{ij}}{I_{ij}} = \frac{K_0}{K_2} \tag{2-8}$$

If this linear relationship holds true, it indicates that the intensity of telecommunication between two zones could play a role as a proxy for current travel demand between two zones. However, this is based on the simplest assumption that the number of trips is proportional to the product of populations of two regions and inversely proportional to the Euclidean distance between the two regions. It is obvious that if we combine the gravity model regarding intensity of telecommunication (Equation (2-7)) and an improved gravity model regarding trip distribution (e.g., Equation (2-2)), the ratio of number of trips to telecommunication intensity would not be a constant, and might be dependent on generalized travel cost and population as well. Since the exponential parameter of population is usually close to 1 in a gravity model regarding trip distribution, it can be assumed that the relation between travel demand and intensity of telecommunication does not depend on population to a large extent. On the other hand, the impedance of generalized travel cost is much likely to be different for intensity of telecommunication and travel demand. It can be reasonably hypothesized that the relation between travel demand and intensity of telecommunication is strongly dependent on generalized travel cost, as illustrated in Equation (2-9).

$$\frac{T_{ij}}{I_{ij}} = F(c_{ij}) \tag{2-9}$$

2.5 Road Network Design Problem

The road network design problem is usually formulated as a bi-level problem, where the network designer makes the decisions on how to plan road network based on the mobility patterns of travellers, and the travelers adapt themselves to the new network by changing their travel behavior to maximize the utilities, forming the new mobility patterns that influence the network designer's decisions in return (Snelder, Wagelmans, Schrijver, van Zuylen, & Immers, 2007). The higher-level problem addresses the question of where new arcs should be constructed or which existing arcs should be upgraded. The lower-level problem concerns the estimation of travel demand in the network (Yang & Bell, 1998). Regarding the lower-level problem, travel demand should be considered not only for trip distribution but also for traffic induction. Traffic assignment is usually made according to the user-equilibrium principle or the 'all-or-nothing' principle, which indicates that travelers would always choose to follow the shortest path.

Regarding the higher-level problem, the objective of network design problem is to optimize a given system performance measure. Some system performance measures can be: efficiency (to maximize the weighted average accessibility), robustness (to maximize the weighted reserve capacity of the network), equity (to limit the computation of accessibility to the zones with the lowest accessibilities) (Santos, Antunes, & Miller, 2009) and environmental objectives (to minimize carbon monoxide emissions) (Cantarella & Vitetta, 2006). In some studies, the minimization of the total costs of road investments can also be the objective of the network design problem (Snelder et al., 2007). However, this is most considered as a constraint of network design problem (Yang & Bell, 1998). A multi-objective evaluation of solutions can be made by using the well-known weighting method (Cohon & Rothley, 1997). The weights for the objectives can be determined by decision makers according to the relative importance.

Historically, the network design problems had two kinds of solution: a discrete form dealing with the additions of new arcs or roadway segments to an existing road network, and a continuous form dealing with the optimal service improvement of existing arcs (Yang & Bell, 1998). However, this classification has been challenged by a number of recent studies. Firstly, these two forms can be combined. To that extent, the existing arcs can be upgraded and the new arcs can be added at the same time (Santos et al., 2009). In addition, it was argued that an important issue of the real-world road network planning is the multilevel discrete nature of service improvement. A discrete form dealing with the optimal service improvement of existing arcs or potential new arcs was suggested (Santos et al., 2009). However, solving the problem of such discrete form is rather difficult, requiring heuristic methods (Yang & Bell, 1998).

3. Research Questions and Methodology

3.1 Research Questions

The main problem addressed in this research is how to use the available mobile phone data properly for understanding the mobility in a country and thus planning a road network to better satisfy the travel demand. To solve the problem, the literature review was made in the previous chapter to gain the background knowledge about each part of the research. Based on that, the research questions of the current study were formulated as follows:

The first research question focused on the use of mobile phone traces as a proxy for the mobility patterns in a country, given as follows:

1) *How can the mobility patterns in a country be derived from mobile phone traces?*

To answer this research question, we had to come up with a technique which can be used to process the data of mobile phone traces to depict the trajectories of travellers and to estimate the places of origins and destinations by detecting where travelers stayed and where they passed by. Moreover, we need to deal with the question whether the mobility patterns reflected in the mobile phone traces of sampled users can represent the mobility patterns of the whole population in the country. Those problems lead to several sub-questions which should be answered as well:

- Which technique should be used to process the data of mobile phone traces for extracting mobility information?
- How can this technique help detect where travelers stayed and passed by?
- How can the movement patterns, reflected in the mobile phone traces of sampled users, represent the mobility patterns of the population?

The second research question focused on how to understand mobility patterns by using the mathematical models, such as a gravity model, formulated as follows:

2) *How can we understand and predict mobility patterns by using gravity models, and what is the use of mobile phone interaction data for developing those models?*

To answer this research question, as many researchers did in their studies, a traditional gravity model based on population could be used to understand mobility patterns in the country. In such a model, the relation between the mobility patterns, population data and the generalized travel cost can be found. However, in this research, we also try to explore if the mobile phone interaction data can be used for understanding and predicting mobility in a model. To this end, some sub-questions should be answered:

- How does a traditional gravity model based on population fit the estimated mobility data?
- What is the statistical relation between the aggregate number of mobile phone interactions and the estimated number of trips made by sampled users empirically found in this study?
- Which is the best model to predict elastic travel demand in this study, a predictive model based on mobile phone interaction data (if any) or the traditional gravity model based on population?

The third research question focused on road network planning by using mobile phone data, given as follows:

3) How can the decisions be made to improve the road network in a country based on the gravity model and by using mobile phone data?

Road network planning is the ultimate objective of this research. An optimization model is the main tool used for road network planning. Such an optimization model should be based on travel demand of travellers, especially elastic travel demand since the prediction of it is important for design road network, as pointed out in the literature. Therefore, we should come up with an optimization model for road network design with the built-in travel demand model.

3.2 Methodology

The methodology to answer the research questions proposed above is illustrated in Figure 8. The main steps are listed as follow.

Firstly, the census data, the GIS data and the original mobile phone datasets were explored respectively. The population of departments could be simply collected from the census data. In Section 4.1, based on the current national and regional road network, the fastest path network analyses could be conducted under two scenarios, without and with the newly-opening Pikine-Diamniadio highway section. The skim matrices of shortest travel times between each two departments could be generated under these two scenarios. In Section 4.2, the mobile phone interaction matrices of all users derived from Dataset 1 were aggregated at a department scale for twelve months in 2013, and in Section 4.3, the estimated inter-departmental OD matrices of sampled users for twelve months in 2013, named as the relative OD matrices, can be derived from Dataset 3. In Section 4.3, we moreover examined the monthly fluctuations of the estimated mobility data. If there were no obvious seasonal fluctuations observed during the whole year, in order to apply the cross-validation technique for model validation afterwards, we classified the relative OD matrices into two groups as the training set and the test set, which were respectively the matrices under the first scenario (without the newly-opening Pikine-

Diamniadio highway section, before August 1st, 2013) and the matrices under the second scenario (with the newly-opening Pikine-Diamniadio highway section, after August 1st, 2013).

We estimated a traditional gravity model based on population using the training set of the relative OD matrices, the shortest travel time calculated under the first scenario and population of each department. This part of the work is presented in Section 5.1. We explored the relations between the mobile phone interaction matrices and the estimated relative OD matrices in Section 5.2. In Section 5.3, we explored if the mobile phone interaction data can help predict elastic travel demand. If it was proved that the mobile phone interaction data can be used to predict elastic travel demand, we could furthermore build a new predictive model based on mobile phone interaction by fitting the model to the data of the training set, and compare this model with the traditional gravity model based on population regarding their model performance of predicting the test set of the relative OD matrices. This comparison is made in Section 5.4.

After we determined whether to use the traditional gravity model based on population or to use the new predictive model based on mobile phone interaction to predict elastic travel demand in order to solve the road network design problem, we started the work of road network planning. A detailed flowchart for methodology regarding road network planning is shown in Chapter 6.

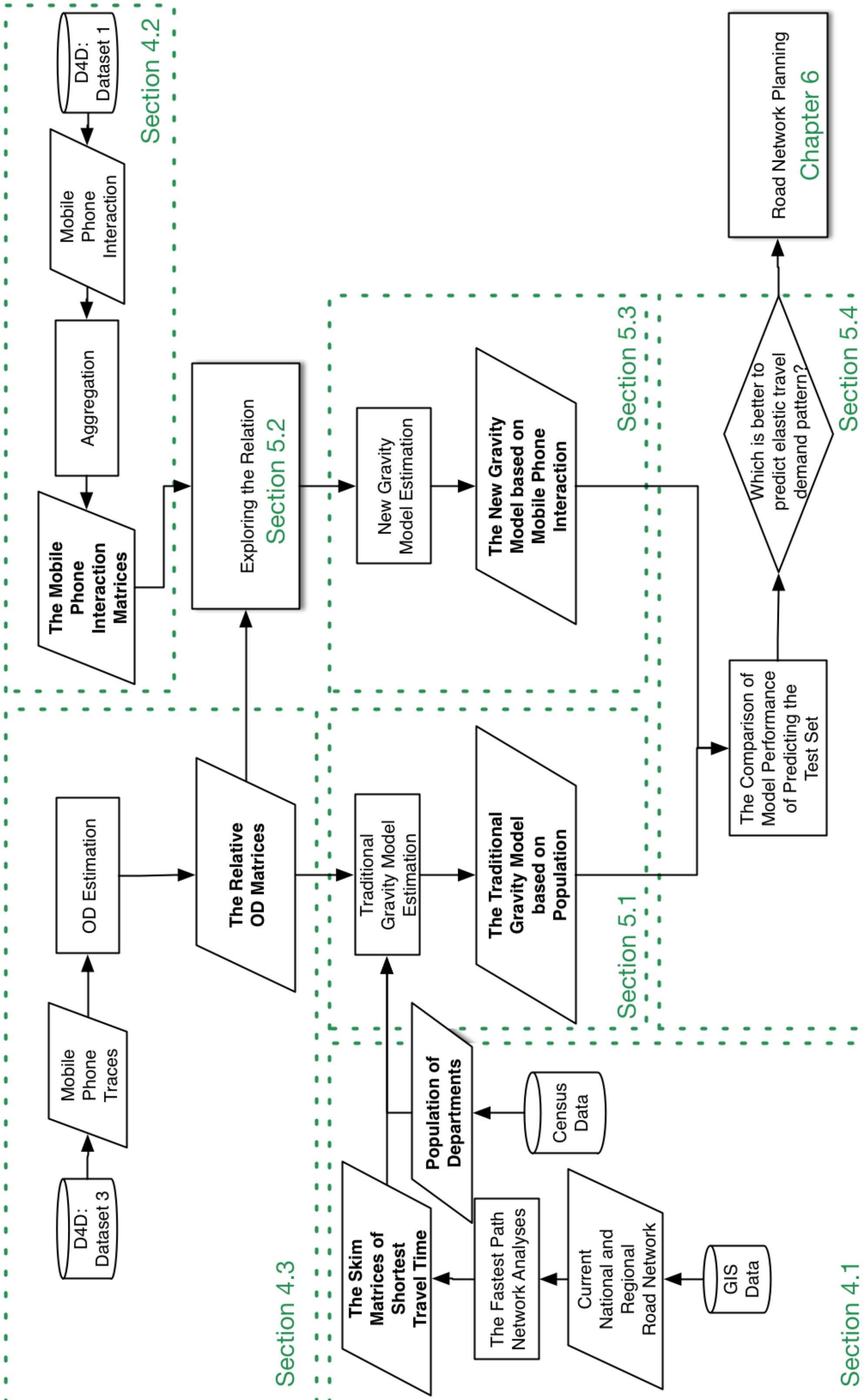


Figure 8 the flowchart for the methodology of the thesis

4. Exploring Spatial Information and Mobile Phone Datasets

4.1 Network Analysis

As a part of the flowchart for the methodology of the thesis shown in Figure 8, the flowchart showing the methodology of this section is illustrated in Figure 9.

Figure 9 the flowchart for the methodology of Section 4.1

The road network shape shown in Figure 2 is simplified in Figure 10. First of all, only the national and regional roads in the network were kept and the lower levels were removed. We defined department center as the traffic generation centroid of each department. In most of them we chose the capital of the department for generating the traffic. The nodes of the network include the centroids of departments and the intersections of roads. The separated links were merged if they were just part of the same road, and some low-level roads such as departmental roads are added only if they are necessary for inter-departmental connections. It should be noticed that roads are not extended to the foreign countries except the country of Gambia which is an enclave of Senegal. The newly-constructed Dakar-Diamniadio toll highway and the ferry service at the Banjul-Barra crossing point and at the Trans-Gambia crossing point are complementary to this network. Note that the Pikine-to-Diamniadio section was open on August 1st 2013.

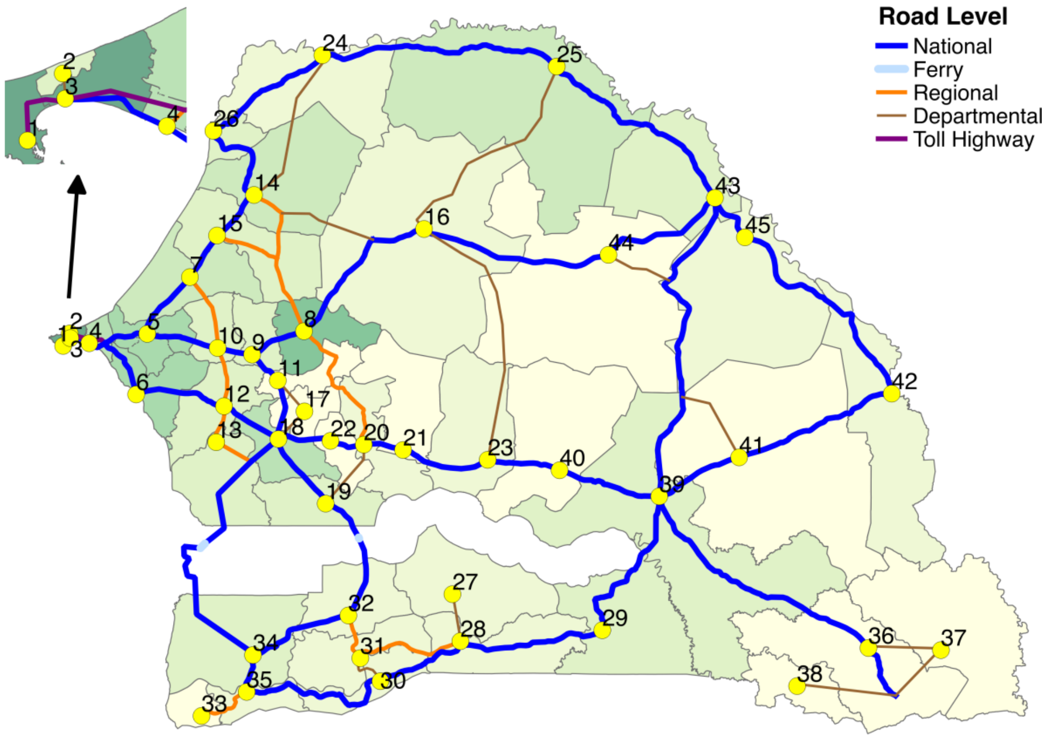


Figure 10 simplified road network and population distribution in the study area

In order to calculate the skim matrices of shortest travel time, we should know average travel speed, which is influenced by speed limits, capacity and traffic volume to a large extent. Since knowledge is limited about speed limits, capacity and daily traffic on these roads, given the available information, 60 km/h, 45 km/h and 30 km/h were simply assumed as the average service speeds on national, regional and departmental roads respectively. 80 km/h was assumed as the average service speed on Dakar-Diamniadio toll highway though in the reality this value is even higher. The reason of making this assumption was to take into consideration the effects of the road toll. In addition, it was assumed that it would take people around 4.5 hours and 3.5 hours (including travel time, waiting time and effects of ferry tariff) to take the ferry services at the Banjul-Barra crossing point and the Trans-Gambia crossing point respectively.

Based on the simplified road network in the country shown in Figure 10, the Dijkstra's Algorithm was applied to calculate the shortest travel time between each two departments. Due to the opening of the highway section in August, the calculation of the skim matrices of shortest travel time were made without and with this new section.

The skim matrices of shortest travel time calculated without the new section were used to estimate predictive models by using the training set of the relative OD matrices, and the ones calculated with the new section were used to test how accurately the estimated models can predict the test set of the relative OD matrices.

4.2 Aggregation and Analysis of the Mobile Phone Interaction Data

As a part of the flowchart for the methodology of the thesis shown in Figure 8, the flowchart showing the methodology of this section is illustrated in Figure 11.

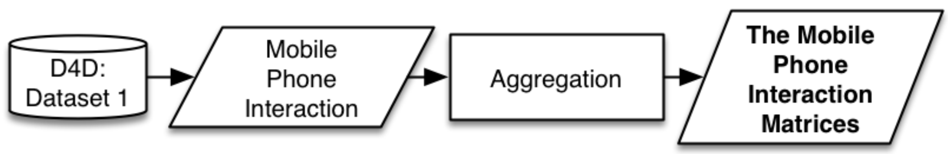


Figure 11 the flowchart for the methodology of Section 4.2

As explained previously, the number and the total duration of calls as well as the number of text messages between each two antennas of all the Orange's mobile phone users in 2013 are provided per hour.

According to the document provided by the D4D Challenge (de Montjoye et al., 2014), the number of Orange's mobile phone users has reached 9 million, and the population of Senegal is about 13 million, yielding a penetration rate of nearly 70% of the population. We have no available information regarding how these 9 million users are distributed in the different departments. However, we can somehow regard this dataset as a representative sample of the interaction pattern in Senegal because of the considerable penetration rate.

We combined both the number of calls and the number of text messages as the intensity of telecommunication used in this study. Compared with the intensity of telecommunication used in Krings et al.'s study (2009), which was defined as the total call duration between two zones, our intensity of telecommunication is more comprehensive since it includes not only the interaction by calls but also by text messages. Moreover, we can simply assume that the duration per call is approximately constant. To that extent, it is sufficient to use the number instead of the total duration to indicate intensity of interaction.

We aggregated the data and built the mobile phone interaction matrix in which every cell presents the aggregate number of one-year calls and text messages from one department to another. The probability distribution of this aggregate number between each two departments is plotted in Figure 12, where it can be observed that the value of skewness is rather high. This is because the mobile phone interaction between departments in Dakar region is rather intensive compared to the average intensity of telecommunication in the rest of the country.

This high value of skewness makes the variable not suitable for certain statistical analyses, where a normal distribution of variables is often required. Since a transformation can be used to reduce the skewness, we transformed the mobile phone interaction data in a log 10 scale. As shown in Figure 12 and Figure 13, the value of skewness becomes close to zero, and the distribution tends to be a normal one, which is preferred for statistical analysis.

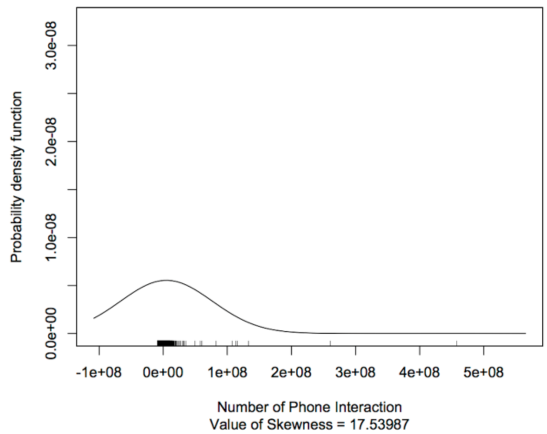


Figure 12 probability distribution of the aggregate number of one-year directed calls and text messages

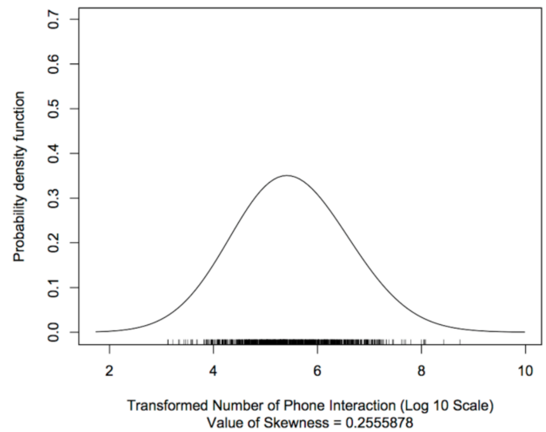


Figure 13 probability distribution of the aggregate number of directed mobile phone interactions over one year in a log 10 scale

It can be observed that the directed matrix of the aggregate number of mobile phone interactions between departments is almost symmetric. This means that the number of mobile phone interactions aggregated in one year from one department to another is almost equal to the one in return.

The mobile phone interaction matrix can be further aggregated. The numbers of outgoing and incoming mobile phone interactions of each department over one year can be calculated respectively. In Figure 14, the y-axis represents the numbers of outgoing and incoming mobile phone interactions of each department in a log 10 scale, and two different colors are used to distinguish the number of outgoing mobile phone interactions and the number of incoming mobile phone interactions. The x-axis represents the population of each department in a log 10 scale. It can be observed that with the increase of population, the aggregate numbers of outgoing and incoming mobile phone interactions have a power law increase with an exponent close to 1.4, which can be explained by two possible reasons. The first possible reason is that people in the densely populated area of the country might use mobile phone more frequently, and the second possible reason is that there are more mobile phone users in the densely populated area of the country. Both of the two reasons sound reasonable. In addition, it can be observed that the two regression lines are almost overlapped each other, which proves again the symmetry of people’s telecommunication patterns during the year.

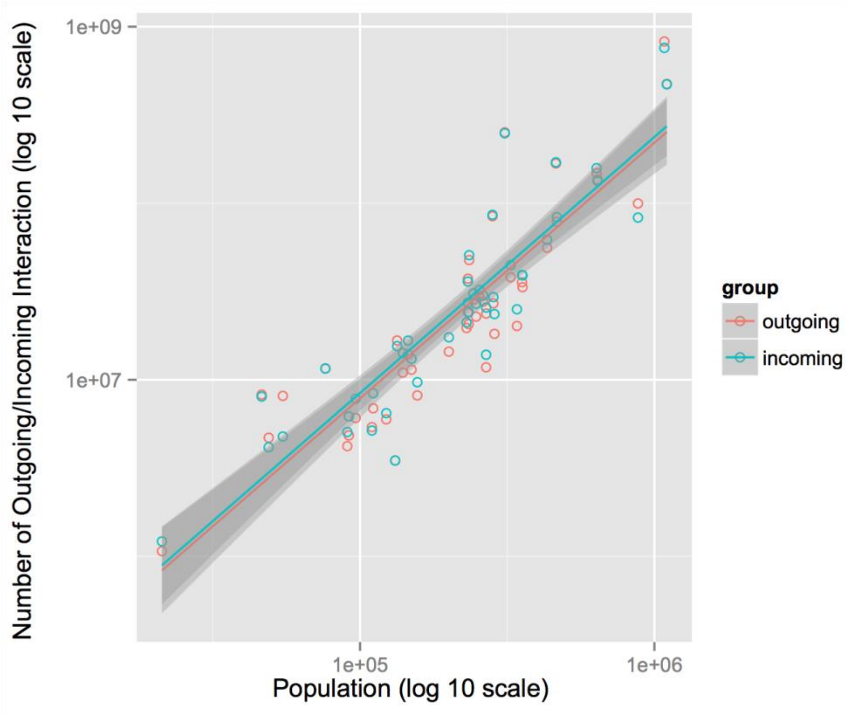


Figure 14 the relation between the aggregate number of outgoing and incoming mobile phone interactions of each department and the population of each department

Next, twelve mobile phone interaction matrices can be built respectively for each month in 2013. We firstly examined the fluctuation of the total number of monthly mobile phone interactions, shown in Figure 15. It can be observed that the number of mobile phone interactions increased in the second half of the year, and especially rocketed in August. However, the correlation coefficient between mobile phone interaction matrices of each two months is higher than 0.99, which seems to indicate that the mobile phone interaction pattern between departments in Senegal keeps almost the same.

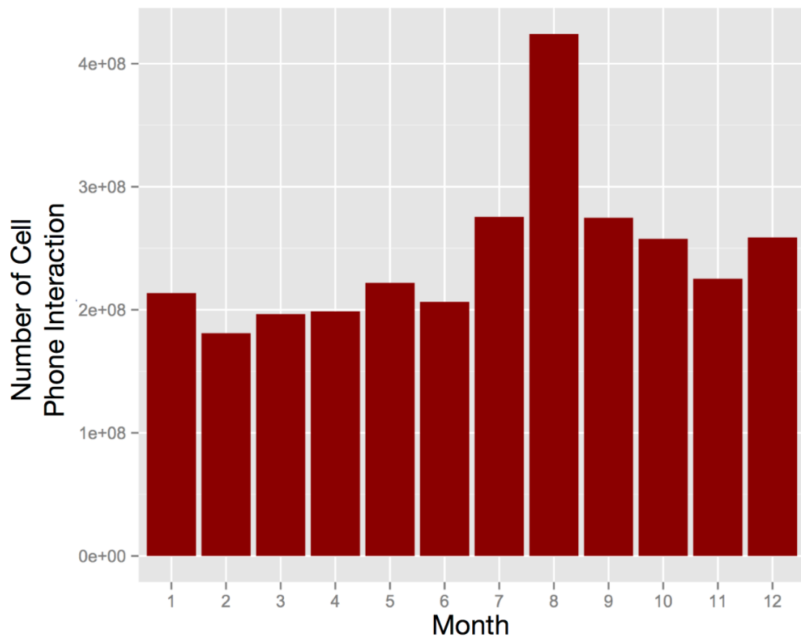


Figure 15 the aggregate number of monthly mobile phone interactions (i.e., calls and text messages) per month

4.3 The Estimation of Mobility Patterns Using Mobile Phone Traces

As a part of the flowchart for the methodology of the thesis shown in Figure 8, the flowchart showing the methodology of this section is illustrated in Figure 16.

Figure 16 the flowchart for the methodology of Section 4.3

The foundation of this research is to use the mobile phone traces for depicting mobility patterns. At this point we should note the fact that travellers may use different modes of transportation, while we would like to only focus on the trips by road transport in this study. In a report by the World Bank (2004), it was said that road passenger share in Senegal was above 99%, and road freight share was above 95%. It can be confirmed that most of the inter-departmental trips are made by road transport.

As discussed in the literature review, the trajectories of each user can be easily traced, but the OD information is missing, which is essential for understanding human mobility. In many existing studies, researchers tried to extract the information regarding where travellers stayed from the mobile phone traces. Furthermore, some of them detected the most important places where travellers stayed (e.g., home location and work location) as the origins and destinations and focused only on the commuting trips between OD pairs. As mentioned in Section 2.1, there are some other limitations of using mobile phone traces to derive mobility patterns. One limitation is that trajectories might not include the traces where travellers went without using mobile phones. Another limitation is that the trajectories made by sampled mobile phone users might not reflect the mobility patterns of the whole population in the country. In the following sections, it is explored how we can estimate mobility patterns by using the available mobile phone traces and it is discussed how the aforementioned problems can be solved in this case.

4.3.1 Primitive Estimation to Trace the Trajectories of Each User

In Dataset 3 provided by the D4D challenge, the traces are recorded at an arrondissement scale. First of all, we can sort the traces of each user in a time sequence and aggregate the traces at a department scale. Then the consecutive traces at the same department of each user can be fused together. To that extent, every inter-departmental move of an individual user can be observed if it was detected that he used his mobile phone in one department, and later he used his mobile phone in another department. The trajectories of each individual can be traced on the map in a department scale. It can be roughly known how long the user stayed at one department since the time of the first and last trace of staying at one department can be observed. An example of how the primitive estimation works is shown in Figure 17.

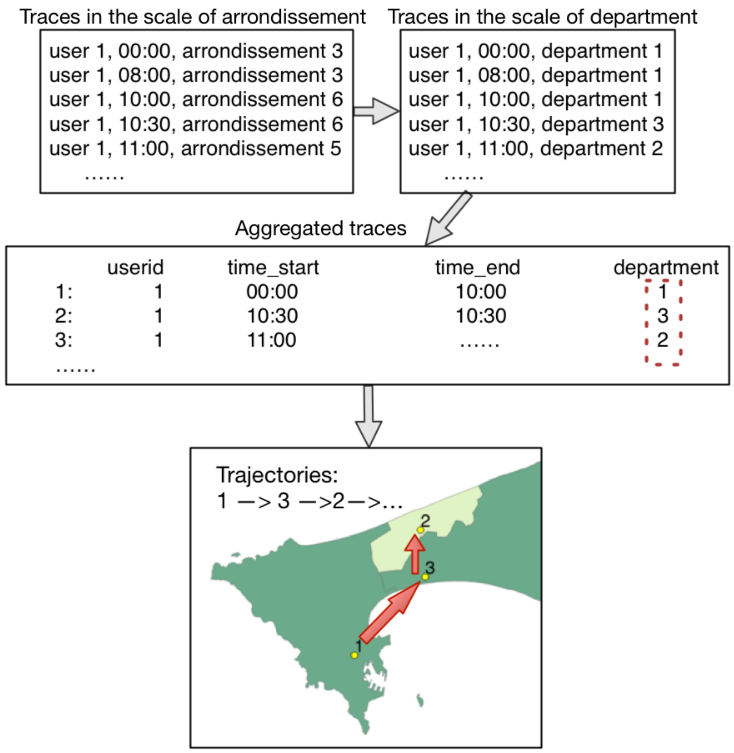


Figure 17 the steps of the primitive estimation to trace the trajectories of each user

However, the trajectories traced in this way cannot reflect the mobility patterns of the whole population in Senegal, and it cannot even reflect the real mobility pattern of sampled users because of some limitations of this method we have mentioned in Section 2.1. The solutions to those limitations in this case are given in the following sections.

4.3.2 A Filtering Algorithm to Find the OD Information

The main limitation of the traced trajectories is that the trajectories lack of the OD information, or in other words, the trajectories are not the real trips. Based on the trajectories, we might mix up a long trip with many partial trips. If a user passed by a department and used his or her mobile phone there, this department should not be a real origin or a real destination. A possible solution which can be found in many studies is to identify the places where travellers stayed for the longest time (i.e., home location and work location) and then only focus on the commuting trips (Nanni et al., 2014; Csáji et al., 2013). It was assumed in these studies that the place where a user was most frequently traced to stay and the place where he or she was second most frequently traced to stay should respectively be the location of his or her home and the location of his or her work. However, this method cannot capture non-work trips and the patterns of weekday and weekend as well as seasonal variations (Csáji et al., 2013). There is another problem in our study if we only focus on the commuting trips. Except the departments in the Dakar region of which the area is relatively small, a department in Senegal has an area of larger than 1000 square kilometers. Since the typical size of a department is very large, it is quite possible that most people live and work in the same department. To examine if this problem exists in this case, we explored the data to find the locations of each user's home and work. We found that the probability of being traced (using mobile phones) in the 'home departments' over one year is higher than 80% for about 75% sampled users. The result seems to indicate that most of inter-departmental trips in Senegal belong to irregular trips instead of commuting trips.

In this study, we made an attempt to filter the traces by using a threshold of least duration at one department plus travel time passing the department. This method is based on a reasonable assumption that if a user only stayed at one department for a very short time, it can be derived that this department should not be an origin or a destination and should be where he or she passed. We can simply assume that the least duration of one user at one department as a reasonable value, which is two hours we used in this case as the threshold. Based on these ideas, we applied an algorithm, of which the approach can be understood as follows:

We started from the trajectories, which are the outcomes of the primitive estimation. For every sampled user u, the r th trace that he had is at department D_{ur}, and the time he made the first interaction at D_{ur} is at FT_{ur}, and the time he made the last interaction at D_{ur} is at LT_{ur}. As assumed, the least duration of u at D_{ur} should be 2 hours. The shortest travel time between $D_{u(r-1)}$ and D_{ur} is $t(D_{u(r-1)}, D_{ur})$, and the shortest travel time between D_{ur} and $D_{u(r+1)}$ is $t(D_{ur}, D_{u(r+1)})$. It is simply assumed that user u always made phone calls and had text messages at department centers. Then, the interval $FT_{u(r+1)}$ - $LT_{u(r-1)}$ should be larger than the sum of least duration at department D_{ur}, $t(D_{u(r-1)}, D_{ur})$ and $t(D_{ur}, D_{u(r+1)})$.

Therefore, if $FT_{u(r+1)} - LT_{u(r-1)} < t(D_{u(r-1)}, D_{ur}) + t(D_{ur}, D_{u(r+1)}) + 2$ (unit: hour), the r th trace of the user u should be removed from the traces since department D_{ur} is proved as where u passed in this way. Table 3 shows the components of the filtering algorithm.

Trace ID for each user	user ID	Time_start (the first time of using mobile phone)	Time_end (the last time of using mobile phone)	Department ID
……				
$r-1$	u	$FT_{u(r-1)}$	$LT_{u(r-1)}$	$D_{u(r-1)}$
r	u	FT_{ur}	LT_{ur}	D_{ur}
$r+1$	u	$FT_{u(r+1)}$	$LT_{u(r+1)}$	$D_{u(r+1)}$
……				

Table 3 the components of the filtering algorithm

In this algorithm, $u \in \{1,2,...,160000\}$; $r \in \{2,3,...\}$; $D_{ur} \in \{1,2,...,45\}$; FT_{ur} and LT_{ur} are in the form of yyyy-mm-dd hh:mm:ss; and the value of $t(D_{u(r-1)}, D_{ur})$ or $t(D_{ur}, D_{u(r+1)})$ can be derived from the shortest travel time matrices calculated in Section 4.1.

We applied the filtering algorithm to the whole dataset. A great number of traces are eliminated. Then we fused the consecutive traces at the same department again. Afterwards, we applied the filtering algorithm, and then fused the traces and then applied the filtering algorithm again, and so on and so forth. As a result, about 30% traces were eliminated. Connecting the filtered traces, we could depict the filtered trajectories including the OD information, which can help us estimate the number of trips made by the sampled users from one department to another. We came up with the monthly OD trip (filtered trajectory) matrices between each two departments of sampled users in 2013. An example of the conversion of the filtered trajectories into matrices is shown in Figure 18. Those matrices do not contain the exact number of trips made by the population, but contain the aggregate number of filtered trajectories, which helped us estimate the number of trips made by sampled users. We named them the relative OD matrices, since we assumed that they can serve as proxies for the real OD trip matrices of the population and reflect the mobility patterns in Senegal proportionally.

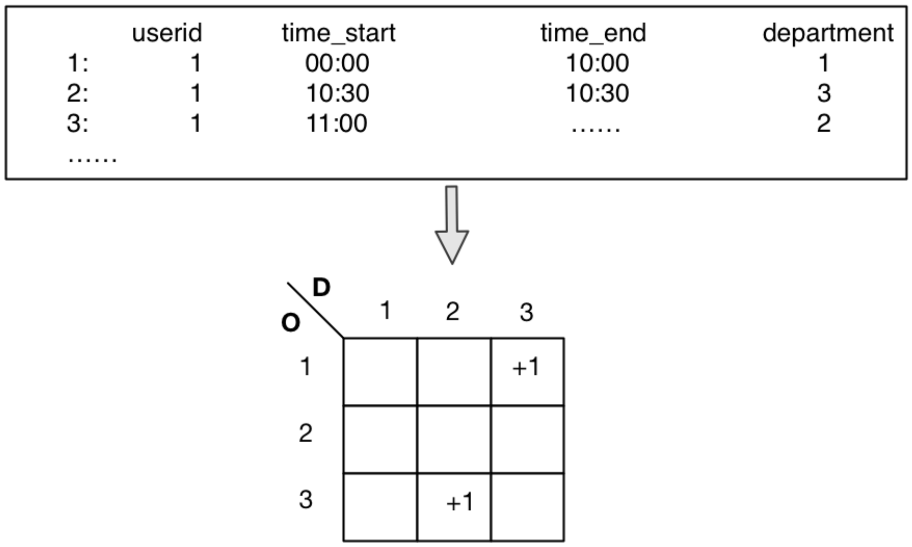


Figure 18 an example of the conversion of filtered trajectories into a relative OD matrix

The filtering algorithm can solve the aforementioned problem efficiently. A good example showing the effect of the algorithm is shown in Figure 19. In this example, it can be observed that the user 3 experienced a trip from east to west. The trajectories as the outcome of the primitive estimation indicate that travellers moved from department 39 to 23, to 22, to 18, to 10 and to 5. It is obvious that he did not make a series of trips from department 39 to 23, from 23 to 22, from 22 to 18, from 18 to 10 and from 10 to 5. In fact, as shown in the map, department 23, 22, 18 and 10 are the departments where the user 3 must pass by if he made a trip from department 39 to 5. The filtering algorithm helped us eliminate the traces at department 23, 22, 18 and 10 and keep the traces at department 39 and 5, which we derived as the origin and the destination.

Moreover, we found that the algorithm is able to solve two more problems which are specific in this case to some extent. One problem is regarding sensing errors (there are some impossible traces in the original dataset, e.g., moving too fast). The other one is that some very short trips around the boundary between departments might have been sensitively recorded as inter-departmental trips.

The trajectories as the outcome of the primitive estimation

Trace ID	User ID	The first time of using mobile phone	The last time of using mobile phone	Department ID
"122"	3	"2013-04-15 18:30:00"	"2013-04-15 18:30:00"	21
"123"	3	"2013-04-16 21:40:00"	"2013-04-18 11:50:00"	28
"124"	3	"2013-04-19 17:40:00"	"2013-04-28 05:40:00"	39
"125"	3	"2013-04-28 07:40:00"	"2013-04-28 08:00:00"	23
"126"	3	"2013-04-28 09:40:00"	"2013-04-28 09:40:00"	22
"127"	3	"2013-04-28 11:40:00"	"2013-04-28 12:00:00"	18
"128"	3	"2013-04-28 13:20:00"	"2013-04-28 13:20:00"	10
"129"	3	"2013-04-28 15:00:00"	"2013-05-14 14:50:00"	5
"130"	3	"2013-05-14 20:40:00"	"2013-05-21 10:20:00"	26
"131"	3	"2013-05-21 14:00:00"	"2013-06-04 21:00:00"	5

The filtered trajectories as the outcome of applying the filtering algorithm

Trace ID	User ID	The first time of using mobile phone	The last time of using mobile phone	Department ID
"107"	3	"2013-02-03 05:50:00"	"2013-04-15 13:20:00"	5
"108"	3	"2013-04-15 18:30:00"	"2013-04-15 18:30:00"	21
"109"	3	"2013-04-16 21:40:00"	"2013-04-18 11:50:00"	28
"110"	3	"2013-04-19 17:40:00"	"2013-04-28 05:40:00"	39
"111"	3	"2013-04-28 15:00:00"	"2013-05-14 14:50:00"	5
"112"	3	"2013-05-14 20:40:00"	"2013-05-21 10:20:00"	26
"113"	3	"2013-05-21 14:00:00"	"2013-06-04 21:00:00"	5
"114"	3	"2013-06-05 00:10:00"	"2013-06-09 19:50:00"	8
"115"	3	"2013-06-09 21:10:00"	"2013-06-10 09:50:00"	5
"116"	3	"2013-06-10 16:10:00"	"2013-06-11 16:40:00"	19

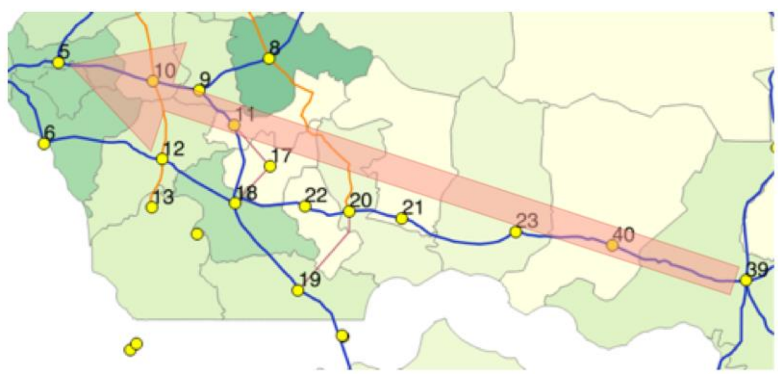


Figure 19 an example showing the effect of the filtering algorithm

4.3.3 Sampling Rate

The basic idea of deriving mobility patterns from mobile phone data is that once a user made a call or had a text message in a different department, he had a trajectory. We used a filtering algorithm to extract the OD information from the trajectories previously. However, it could be argued that some trajectories would be missing since the user might not use mobile phone during those trips, and then the extracted OD information is incomplete. In most of studies, sampling rate was examined at first (Calabrese et al., 2011; Hoteit et al., 2014). If sampling rate is high enough, we can say that there is enough evidence to only focus on the recorded trajectories. In addition, the acceptable value of sampling rate should be related with the area of the spatial unit. For example, the frequency of inter-departmental trips made by one person should be much lower than the frequency of his intra-urban trips. It can be concluded that the sampling rate in this study is high enough since we only focus on the inter-departmental trips and the sampled users are active enough (having more that 75% days with interactions in one year).

4.3.4 Sample and Population

In Dataset 3, to guarantee the sampling rate of every user, only 0.16 million Orange's users, who are the most active users, are sampled. The question is whether the movements of them can reflect the travel demand pattern of the whole population in Senegal. To answer this question, we followed the idea of Calabrese et al. (2011) to compare the population distribution and the home location distribution of sampled users. It can be assumed that people often stay at home from 18 in the evening until 7 in the morning. The department where one user is traced most frequently during that night interval in the whole year is detected as the location of that user's home. Therefore, the number of the sampled users' homes of each department is known. A linear regression in a log 10 scale is made to find the relationship between population and the number of sampled users' homes of each department. In Figure 20, a power law increase with an exponent close to 0.5 can be observed. It is indicated that in the densely populated departments, there are relatively more active Orange's users who are sampled, which prove that this sample is biased to some degree.

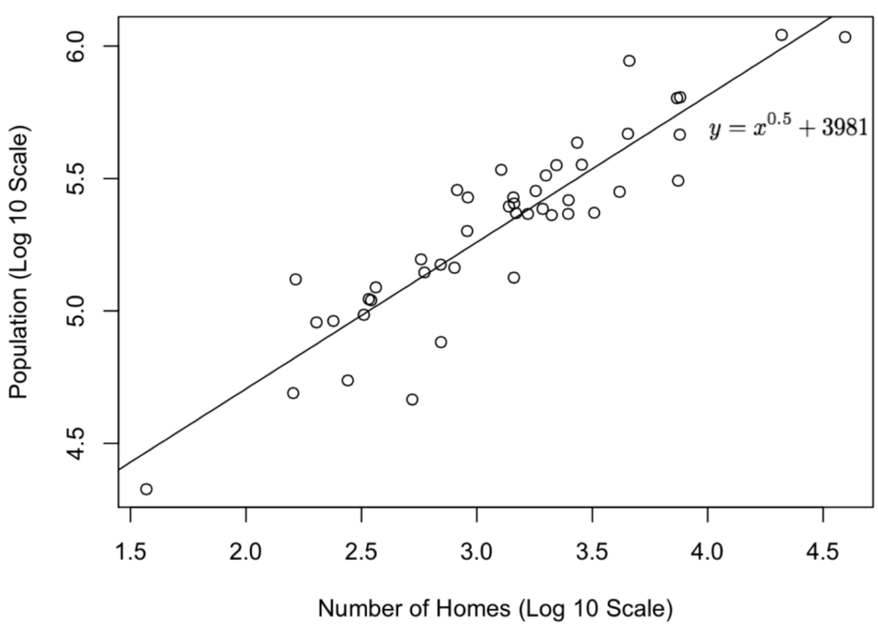


Figure 20 the regression between population and number of detect homes in a log 10 scale

It was recommended by Nanni et al. (2014) that a factor which depends on the marketing penetration, mobile phone ownership and mobile phone usage can be weighed to make an estimate of actual traffic flow. However, it has been found that people who have more mobile phone interaction with others will generate more trips (Kamargianni & Polydoropoulou, 2013; Nobis & Lenz, 2009). This conclusion made us believe that simply multiplying by a factor would even make the calibrated traffic flow more biased. Moreover, based on this conclusion, we can somehow assume that the sampled users are the most active travellers in each department since they are the most active mobile phone users there. To that extent, the movements of sampled users are representative to some degree.

Due to the development of tourism in Senegal, the travel demand of external visitors should be considered as well, and their travel demand might be much different to the travel demand of residents in Senegal. However, it would be rather difficult to capture their traces since they might not use the mobile service provided by the local mobile companies. In this case, we simply assumed that the number of those external visitors is relatively small, and the available mobility data sample can somehow reflect their mobility patterns as well.

Since this sample is the best mobility data of Senegal which can be obtained, we can estimate the mobility in Senegal based on nothing but this sample. However, it should be kept in mind that the estimated relative OD matrix records the filtered trajectories of sampled users, which can somehow reflect the mobility pattern of the population in a relative way, or in other words, this estimated relative OD matrix cannot provide the actual traffic flow.

4.3.5 Results and Analysis of the Estimated Mobility Patterns

As a result, a one-year relative OD matrix containing the aggregate number of filtered trajectories can be estimated, and the symmetry of the matrix is observed. The probability distribution of the number of filtered trajectories from one department to another is plotted in Figure 21. It can be observed that the value of skewness is quite high, which can be explained by the fact that there are quite many filtered trajectories in the Dakar region than in the other regions. We transformed the data in a log 10 scale for statistical analysis, shown in Figure 22, where an approximately normal distribution can be observed.

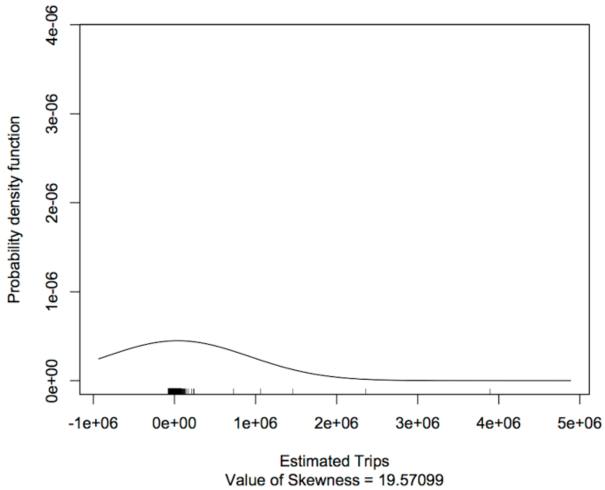


Figure 21 probability distribution of the aggregate number of one-year directed filtered trajectories

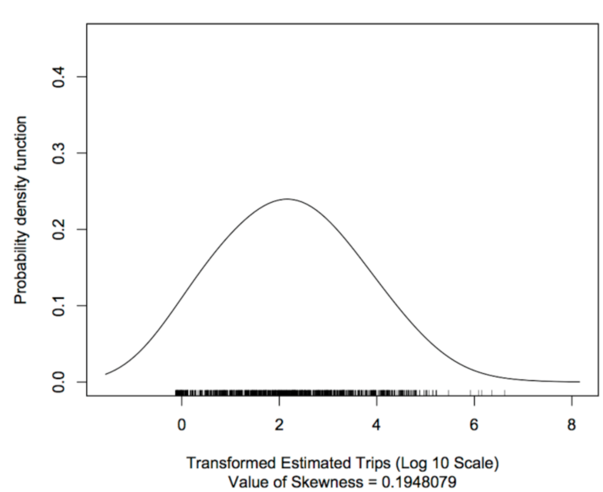


Figure 22 probability distribution of the aggregate number of one-year directed filtered trajectories in a log 10 scale

We also made a bar plot to indicate the distribution of estimated trip (filtered trajectory) length, in Figure 23. We classified the length of trips into extra-short trip (of which travel time is between zero and 3.3190 hours), short trip (between 3.3190 and 5.6793 hours), long trip (between 5.6793 and 8.1858 hours) and extra-long trip (over 8.1858 hours) by using a quantile classification method. As a result, in 2013, there are 38300 extra-long trips, 124846 long-trips, 270420 short-trips and 13983798 extra short trips.

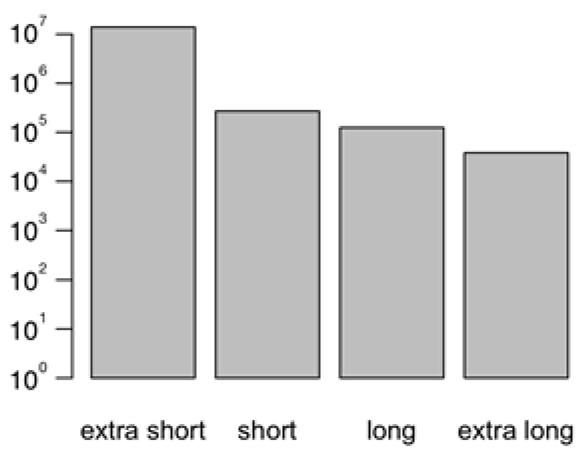


Figure 23 the distribution of trip length

In the same way, twelve relative OD matrices can be estimated for each month as well. The total estimated number of trips (filtered trajectories) made by sampled users per month was calculated, and it can be observed in Figure 24 that the fluctuation is smooth and reasonable. For example, the number in February is the smallest. The correlation coefficient between relative OD matrices of each two months in 2013 is higher than 0.99, which indicates that the relative mobility pattern keeps almost the same during the whole year.

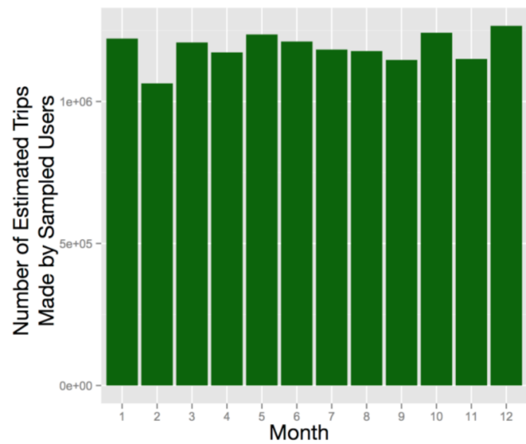


Figure 24 the total number of filtered trajectories

The relative OD matrix can be further aggregated. The numbers of production and attraction of trips (filtered trajectories) of each department over one year can be calculated respectively. In Figure 25, the y-axis represents the numbers of trip production and trip attraction from and to each department in a log-10 scale, and two different colors are used to indicate the number of production and attraction respectively. Note that most of red points cannot be observed in the figure since they are overlapped by the green ones. This observation indicates that the filtered trajectories over one year are quite symmetric. The x-axis represents the population of each department in a log-10 scale. It can be observed that with the increase of population, the aggregate numbers of trip production and attraction have a power law increase with an exponent about 1.5, which can be explained by two possible reasons. One of the reasons is that the people in the densely populated area tend to travel more, which makes sense since people in the densely populated area have more activities, and those activities would lead to the trips to a large extent. The other reason is that more mobile phone users are sampled in the densely populated area, as explained in Section 4.3.4.

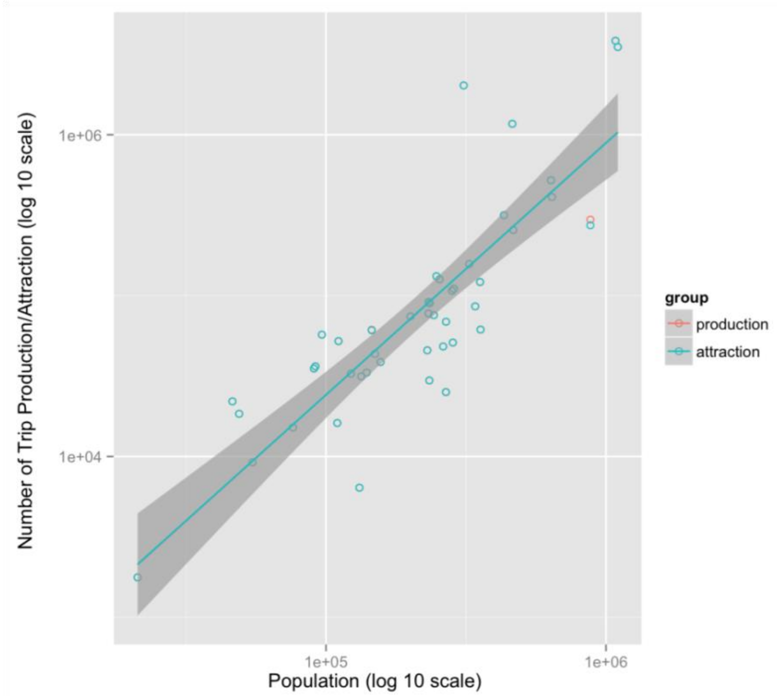


Figure 25 the relation between production and attraction of the filtered trajectories and population

As mentioned in Section 1.2, the Pikine-to-Diamniadio highway section was open on August 1st, 2013. It did not lead to any significant changes of overall mobility pattern in Senegal because this section is short compared to the total distance of road network, and it is parallel to an existing national road. However, the travel demand induced by the newly-opened highway can be observed if we only focus on the sub-region around this new section.

As shown in Table 1, we focused on Dakar (1), Guediawaye (2), Pikine (3), Rufisque (4), Thies (5) and Mbour (6) (note that the names of departments and their corresponding codes can be found in Figure 1), and examined the "before-after" impact of the opening of new highway section on the average estimated number of trips made by sampled users per month between these departments. It can be observed that most of these numbers are increased, and especially, a sharp increase can be observed between Guediawaye and Mbour.

OD Pair	Average Aggregate number of Undirected Trips (Filtered Trajectories) between OD Pairs Made by Sampled Users per Month	
	Before the opening of new highway section (From January to July)	After the opening of new highway section (From August to December)
1-4	54563	57932
1-5	15023	16447
1-6	14287	13980
2-4	6219	6587
2-5	1937	2186
2-6	1444	9209
3-4	87416	92972
3-5	10806	11006
3-6	8981	8452

Table 4 the 'before-after' comparison of average travel demand

For the purpose of cross-validation, two relative OD matrices were estimated respectively for the period before the opening of the Pikine-to-Diamniadio highway section and the period after the opening of the highway section. In Chapter 5, the first matrix is used as the training set to fit models that can be used to predict travel demand pattern, while the second one is used as the test set to assess the predictive power of models.

4.4 Conclusions

In this chapter, network analysis was first made to find the shortest travel time between every two departments in Senegal. Then the number of mobile phone interactions (i.e., the number of calls and text messages) was aggregated between every two departments during 2013, generating the mobile phone interaction matrices. Next, a filtering algorithm was produced to find the OD information by filtering out all the unimportant traces where users passed by and keeping only the important traces, namely origins or destinations. After applying the filtering algorithm, the trajectories were made approximate to a series of trips, named as the filtered trajectories.

As a result, the monthly OD trip (filtered trajectory) matrices between every two departments of sampled users in 2013 were estimated. Those matrices did not contain the exact number of trips made by population, but contained the aggregate number of filtered trajectories made by the sampled mobile phone users. Those matrices were named as the relative OD matrices, since it was assumed that they can serve as proxies for the real OD trip matrices of the population and reflect the mobility patterns in Senegal proportionally.

5. Formulating and Estimating Gravity Models

In the previous chapter, the relative OD matrices reflecting mobility patterns in Senegal were estimated.

In the literature, it was suggested that one of the best ways to validate them is to fit them to a gravity model and then to examine the fitness. In this gravity model, mobility is proportional to the product of population and inversely proportional to the generalized travel cost.

Also, such a gravity model allows us to have a better understanding of mobility patterns. For example, it can provide insights into the effect of the travel costs in the impedance to travel in the study area. To that extent, the gravity model can be used to predict the changes of future mobility pattern with the potential changes of generalized travel cost. From the literature, we summarized that besides generalized travel cost and population, telecommunication patterns may be strongly related to mobility patterns as well. In this chapter, we use the available mobile phone data to find which variables influence mobility most significantly, and based on the result, we propose a model which is able to help us predict travel demand in the future.

5.1 The Traditional Gravity Model Based on Population

As a part of the flowchart for the methodology of the thesis shown in Figure 8, the flowchart showing the methodology of this section is illustrated in Figure 26.

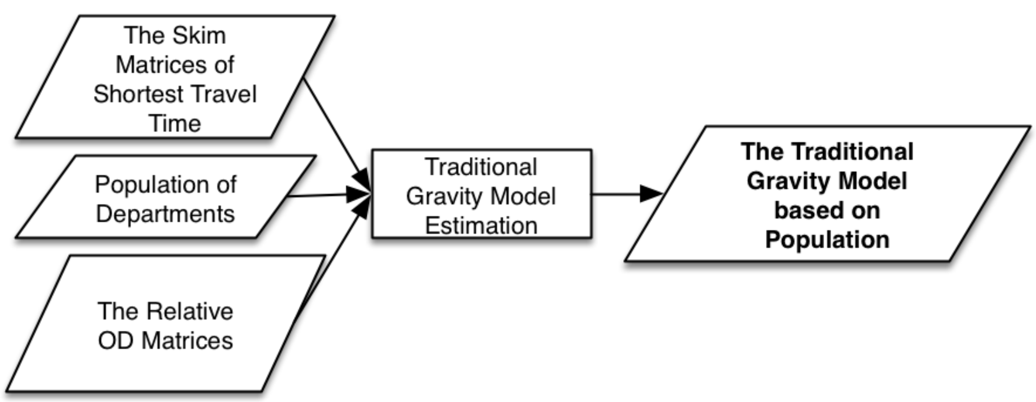


Figure 26 the flowchart for the methodology of Section 5.1

A traditional gravity model based on population, which indicates that the mobility between two zones is almost proportional to the product of population of two zones and inversely proportional to the travel cost between two zones, was used to fit the training set of the relative OD matrices (for the period before the opening of new highway section), and the parameters of this model were estimated.

The Equation (2-2) described in Section 2.3 is used as the functional form for traditional gravity model. Regarding the cost function, we follow the idea of (Csáji et al., 2013): fitting both the power law decay and the exponential decay, to find the one that provides a better fit. The cost functions with power law decay and with exponential decay are shown as Equation (2-5) and Equation (2-6) given in Section 2.3. To that extent, the functional form for traditional gravity model can be either Equation (5-1) or Equation (5-2).

$$T_{ij}' = K_1 P_i^a P_j^b t_{ij}^{-n} \tag{5-1}$$

$$T_{ij}' = K_1 P_i^a P_j^b e^{-\beta t_{ij}} \tag{5-2}$$

Where, i and j represent the department of origin and the department of destination; $i \in \{1, 2, ..., 45\}$; $j \in \{1, 2, ..., 45\}$; $i \neq j$; P_i and P_j are population of origin and destination; a and b are the parameters for population; the shortest travel time t_{ij} between i and j, as calculated without new highway section, is used as the component of cost function; β and n are the exponential parameter and the power parameter for cost function respectively; K_1 is a scaling constant for a timespan of one month. It should be noticed that we fit the gravity model by using our training set, which is the relative OD matrix for the period before the opening of new highway section; T_{ij}' is not an exact number of trips between two departments in this case, and by contrast, technically speaking, it is the average estimated number of directed trips (filtered trajectories) made by the sampled users per month before August 1st, 2013, and it can reflect in a relative way the mobility between zones during that period.

After fitting the data to models, it can be observed that the gravity model with the power parameter for cost function fits better to the training set than the one with the exponential parameter. The fitness of the better one is shown in Figure 27, and the estimated values of parameters and adjusted R-squared are listed in Table 5. The similarity between the values of a and b indicates that the trips made by the sampled users are symmetric over the year. The value of n, 2.53015, is a reasonable one which can reflect the impedance of travel costs. The value of adjusted R-squared indicates that our training set somehow fits well to the gravity model. Moreover, it can be observed in Figure 27 that when the travel time between two departments is shorter, the model fits better. The observation in Csáji et al.'s research (2013) is reproduced. These results validate our estimation of relative OD matrix to a certain degree. The estimated gravity model is formulated as follows:

$$T_{ij}' = (1.27e-09) \times P_i^{1.07067} \times P_j^{1.08714} \times t_{ij}^{-2.53015} \tag{5-3}$$

where i and j represent the department of origin and the department of destination; $i \in \{1, 2, ..., 45\}$; $j \in \{1, 2, ..., 45\}$; $i \neq j$; P_i and P_j are population of origin and destination; t_{ij} is calculated as the shortest travel time between i and j; T_{ij}' is the aggregate number of filtered directed trajectories made by users from i to j per month between January 1st, 2013 and July 31st, 2013.

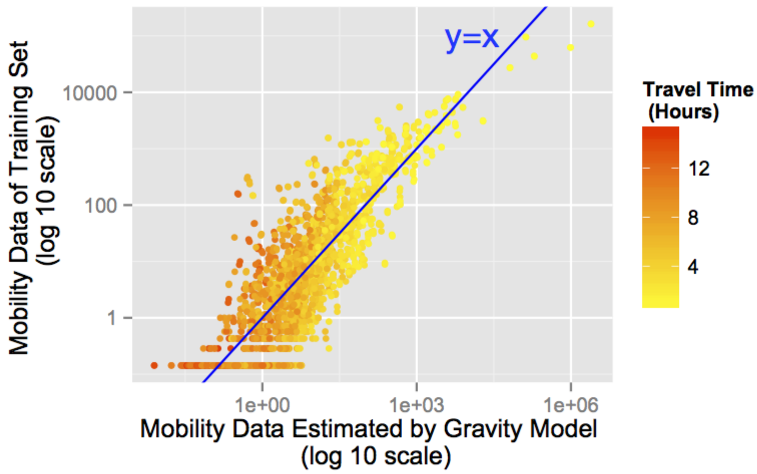


Figure 27 the fitness of the traditional gravity model

Parameter	Estimate	Std. Error	t value
$\log_{10} K_1$	-8.89705	0.34927	-25.47
a	1.07067	0.04305	24.87
b	1.08714	0.04305	25.25
n	2.53015	0.05034	50.26
Adjusted R-Squared	0.7299		

Table 5 the estimated values of parameters and adjusted R-squared of the traditional gravity model

Since it was observed that the fitness of the traditional gravity model based on population is different for shorter trips and longer trips, we fit the model for modeling shorter and longer trips respectively. We used the mean value of travel time between departments, 5.9 hours, to distinguish if a trip is shorter or longer. For shorter trips, the fitness of the gravity model is shown in Figure 28, and the estimated values of parameters and the adjusted R-squared are shown in Table 6. For longer trips, the fitness of the gravity model is shown in Figure 29, and the estimated values of parameters and the adjusted R-squared are shown in Table 7. It can be observed that the fitness of the gravity model for shorter trips is not bad, but the fitness of the gravity model for longer trips is bad. The lower value of n estimated in the gravity model for longer trips indicates that the longer the trip is, the more insignificant travel demand is related to travel time between departments.

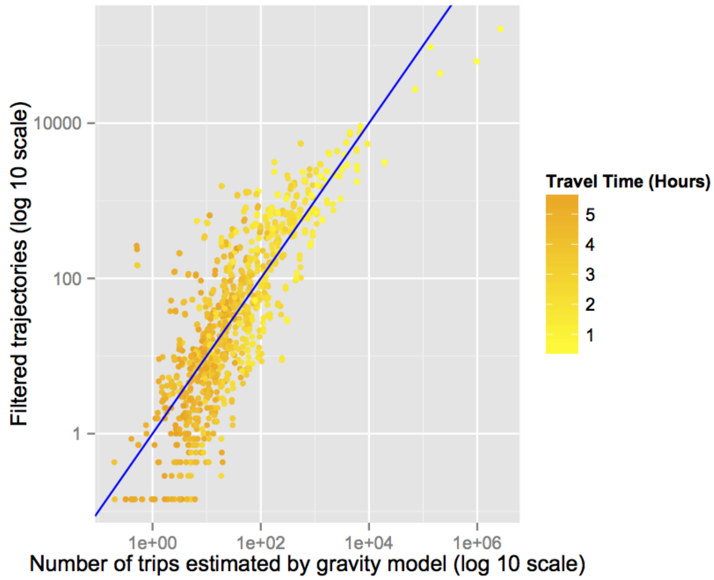


Figure 28 the fitness of the traditional gravity model for shorter trips

Parameter	Estimate	Std. Error	t value
$\log_{10} K_1$	-9.701789	0.47274	-20.52
a	1.14291	0.06316	18.10
b	1.16300	0.06316	18.14
n	-2.47574	0.07463	-33.17
Adjusted R-Squared	0.6902		

Table 6 the estimated values of parameters and adjusted R-squared of the traditional gravity model for shorter trips

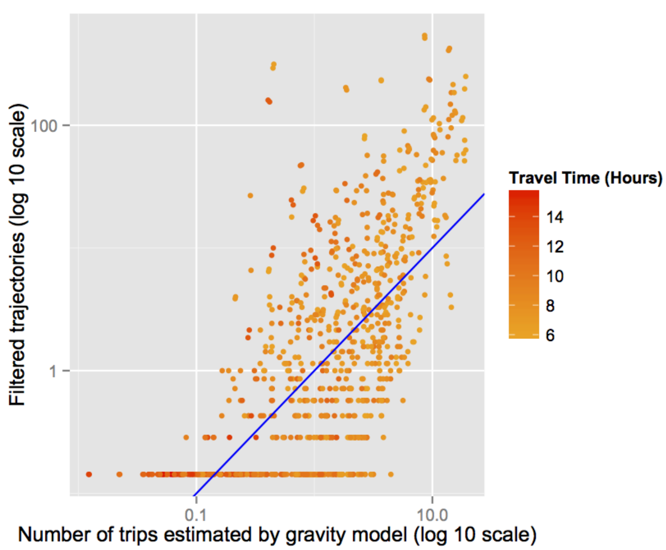


Figure 29 the fitness of the traditional gravity model for longer trips

Parameter	Estimate	Std. Error	t value
$\log_{10} K_1$	-8.94128	0.59590	-15.005
a	1.02129	0.06064	16.841
b	1.03563	0.06064	17.078
n	-1.93114	0.23258	-8.303
Adjusted R-Squared	0.4278		

Table 7 the estimated values of parameters and adjusted R-squared of the traditional gravity model for longer trips

5.2 Exploring the Relation between Telecommunication and Travel Empirically

As a part of the flowchart for the methodology of the thesis shown in Figure 8, the flowchart showing the methodology of this section is illustrated in Figure 30.

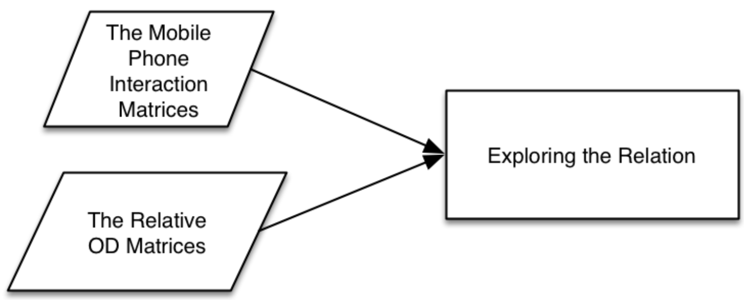


Figure 30 the flowchart for the methodology of Section 5.2

As observed in Figure 14 and Figure 25, with the increase of population, the aggregate numbers of trip production and attraction as well as the aggregate numbers of outgoing and incoming calls and text messages have a power law increase with an exponent about 1.5. It can be derived that there exists a linear relation between the aggregate numbers of trip production and attraction, and the aggregate numbers of outgoing and incoming calls and text messages, from and to one department. We proved this hypothesis in Figure 31.

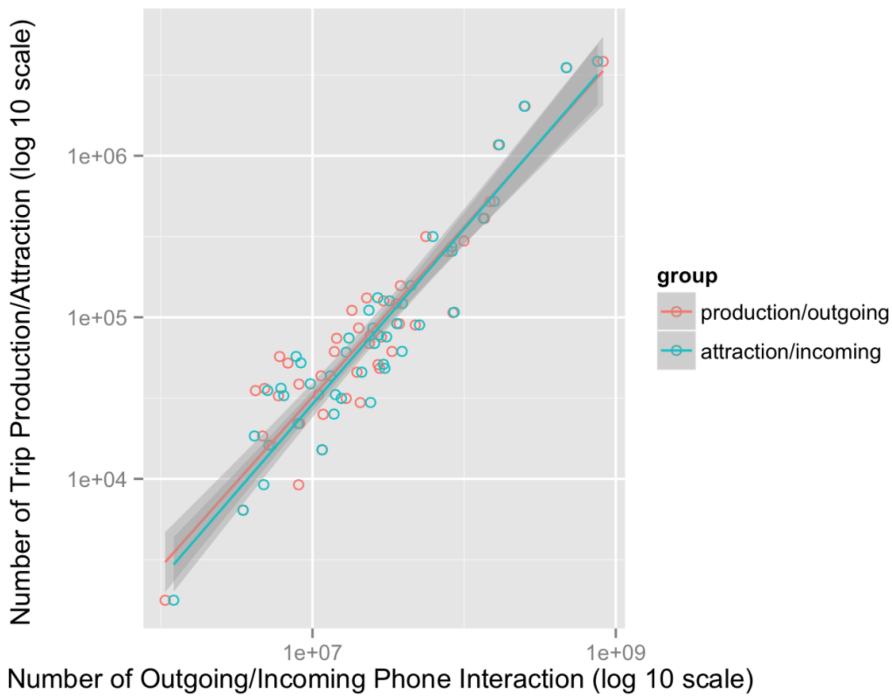


Figure 31 the relation between the aggregate number of trip production/attraction and the aggregate number of outgoing/incoming calls and text messages from and to one department

To further explore the relation between the mobile phone interaction data and the estimated mobility data, we plotted them between each two departments in Figure 32, where y-axis represents the average estimated number of undirected trips between each two departments made by sampled users per month before August 1st, 2013, and x-axis represents the average number of undirected mobile phone interaction between each two departments made by all Orange's users before August 1st, 2013, and after August 1st, 2013 in Figure 33. It should be noticed that the undirected estimated number of trips and the undirected mobile phone interaction are used here instead of the directed ones, because we assumed that the direction of mobile phone interaction would not indicate anything regarding the direction of trips. However, in this case, using directed and undirected mobile phone interaction or travel does not make any difference since both the mobile phone interaction matrix and the relative OD matrix are nearly symmetric.

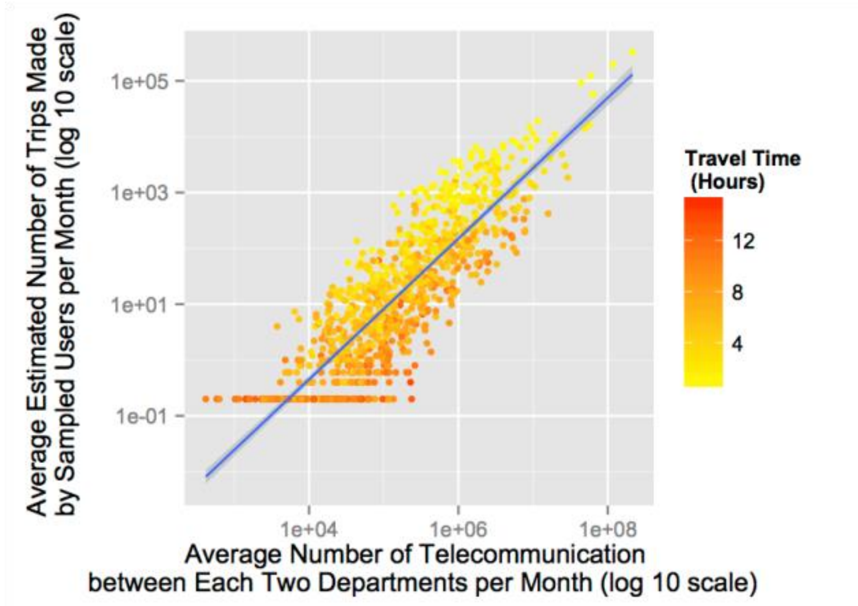


Figure 32 the relation between mobile phone interaction data and the estimated mobility data before August 1st, 2013

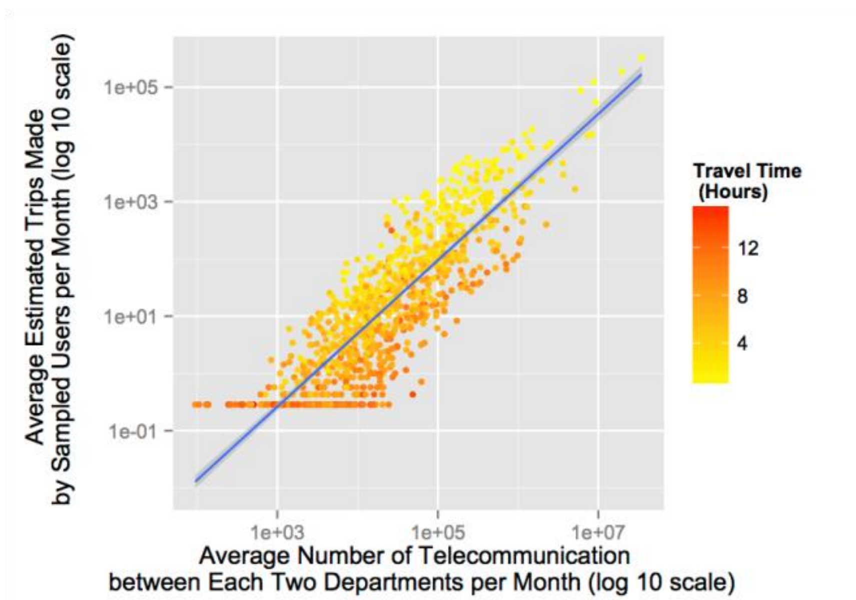


Figure 33 the relation between mobile phone interaction data and the estimated mobility data after August 1st, 2013

It can be observed again in these figures that both mobility patterns and telecommunication patterns stay unaltered after the opening of the new highway section. In addition, a power law increase with an exponent close to 1 can be observed, which indicates that there exists a close-to-linear relationship between mobility data and interaction data.

To test the hypothesis whether the relation between the mobile phone interaction data and the estimated mobility data is dependent on travel costs, we indicated the travel time by color in Figure 32 and Figure 33. It can be observed that when the travel time between departments is shorter, the ratio between mobility data and interaction data is mostly higher. **This observation proves that the hypothesis is true, and the mobility between two departments is almost proportional to the number of mobile phone interactions and inversely proportional to the travel time. To that extent, we could build a new form of gravity model by replacing the product of population of two zones with the number of mobile phone interactions between these two zones.**

5.3 The New Gravity Model Based on Mobile Phone Interaction

As a part of the flowchart for the methodology of the thesis shown in Figure 8, the flowchart showing the methodology of this section is illustrated in Figure 34.

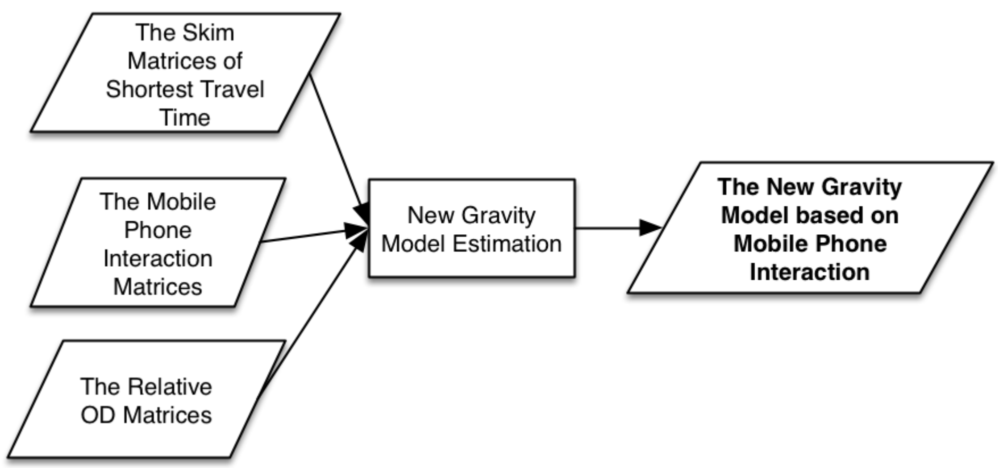


Figure 34 the flowchart for the methodology of Section 5.3

The functional forms of gravity model based on the number of mobile phone interactions (with the power parameter and with the exponential parameter for cost function) are given as follow:

$$T_{ij} = K_3 I_{ij}{}^{\alpha} t_{ij}{}^{-m} \tag{5-4}$$

$$T_{ij} = K_3 I_{ij}{}^{\alpha} e^{-\theta t_{ij}} \tag{5-5}$$

Where, i and j represent the department of origin and the department of destination; $i \in \{1,2,...,45\}$; $j \in \{1,2,...,45\}$; $i < j$; I_{ij} is the average number of undirected mobile phone interaction per month; and α is the parameter for I_{ij}. Since we have observed a close-to-linear relationship between mobility data and interaction data, we assume α would be estimated as about 1. t_{ij} between i and j, as calculated without the new highway section, is used as the component of cost function. θ and m are the exponential parameter and the power parameter for cost function respectively. K_3 is a scaling constant for a timespan of one month. T_{ij} is not an exact number of trips between two departments in this case. Technically speaking, it is the average estimated number of undirected trips made by sampled users per month before August 1st, 2013, and it can reflect in a relative way the directed mobility between zones during that period.

In the same way as we did in Section 5.1, we fit the training set to the new forms of gravity model based on the number of mobile phone interactions. The fitness of two models is shown in Figure 35 and Figure 36, and the estimated values of parameters and the adjusted R-squared are illustrated in Table 8 and Table 9.

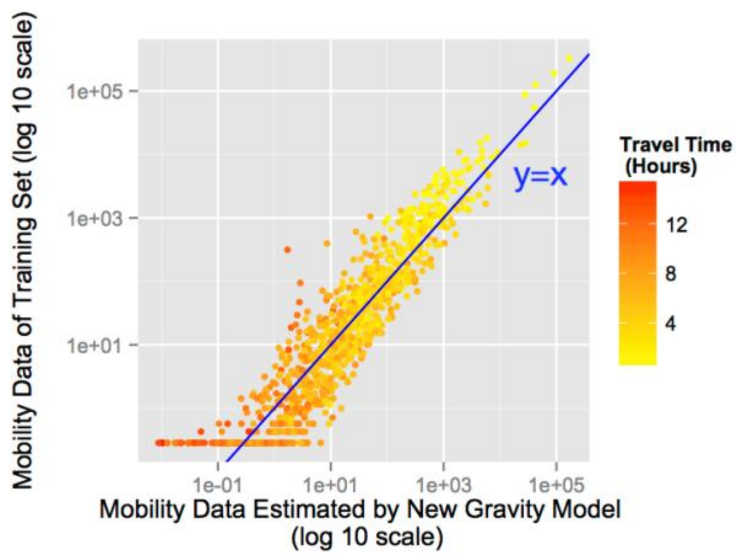


Figure 35 the fitness of the new gravity model with the exponential parameter

Parameter	Estimate	Std. Error	t value
$\ln K_3$	-5.31188	0.25115	-21.15
a	1.00108	0.02041	49.05
θ	0.34970	0.01236	28.3
Adjusted R-Squared	0.8566		

Table 8 the estimated values of parameters and adjusted R-squared of the new gravity model with the exponential parameter

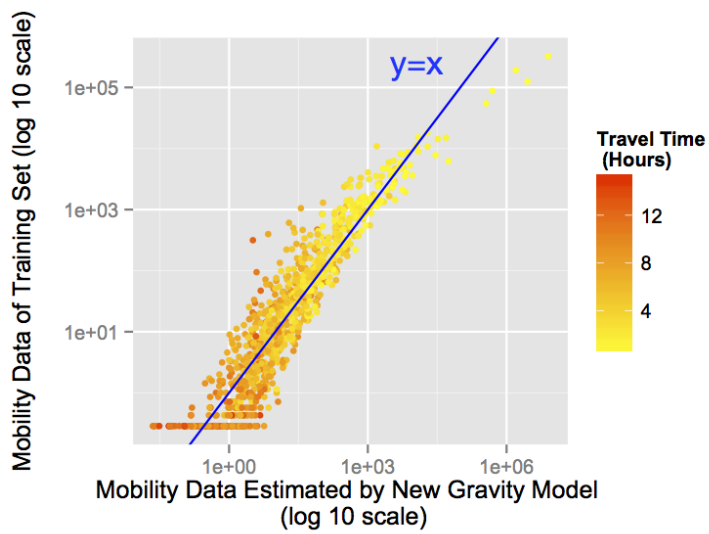


Figure 36 the fitness of the new gravity model with the power parameter

Parameter	Estimate	Std. Error	t value
$\log_{10} K_3$	-1.8021	-0.11102	-16.23
α	0.94989	0.01976	48.16
m	1.72147	0.05384	31.98
Adjusted R-Squared	0.8724		

Table 9 the estimated values of parameters and adjusted R-squared of the new gravity model with the power parameter

The fitness of both the new gravity models seems better than the fitness of the traditional gravity models, and the values of the adjusted R-squared are higher. Especially, as observed in Figure 27 and Figure 29, the traditional gravity model is not fit well when the travel time between two departments is long. By contrast, it can be observed that this problem does not exist when fitting the new gravity models.

Despite the higher adjusted R-squared value of fitting the new gravity model with power parameter, it can be observed in Figure 36 that the model overestimates the mobility between departments with the highest mobility. Therefore, we chose the new gravity model with the exponential parameter as the one to be compared with the traditional gravity model regarding their model performance. The functional form of this estimated new gravity model based on the number of mobile phone interactions is illustrated as follows:

$$T_{ij} = 0.00493 \times I_{ij}^{\;1.00108} \times e^{-0.3497 \times t_{ij}} \qquad \textbf{(5-6)}$$

The value of α , 1.00108, indicates that when travel time is the same, mobility between departments is proportional to the number of mobile phone interactions. It should be noticed that since the new gravity model based on the number of mobile phone interactions is trained using estimated relative OD matrix, T_{ij} , what this model can predict is a relative value, and actual traffic flow cannot be predicted. Since T_{ij} represents a relative value, the constant, 0.00493, is not important in this functional form.

Since the mobility patterns are in terms of the relative OD matrices, which contain the aggregate number of the filtered trajectories, reflecting the mobility patterns proportionally, and the gravity model was estimated by using the relative OD matrices, the gravity model can only predict the proportional travel demand. This fact implies that the aggregate number of mobile phone interaction, as an important input of the model, is only required to be proportionally accurate.

We also examined the fitness of the new gravity model with the exponential parameter for modeling shorter trips and longer trips. For shorter trips, the fitness of the gravity model is shown in Figure 37, and the estimated values of parameters and the adjusted R-squared are shown in Table 10. For longer trips, the fitness of the gravity model is shown in Figure 38, and the estimated values of parameters and the adjusted R-squared are shown in Table 11. It can be observed that the fitness of the new gravity model for shorter trips is excellent, and the fitness of the new gravity model for longer trips is much better than the fitness of the traditional gravity model.

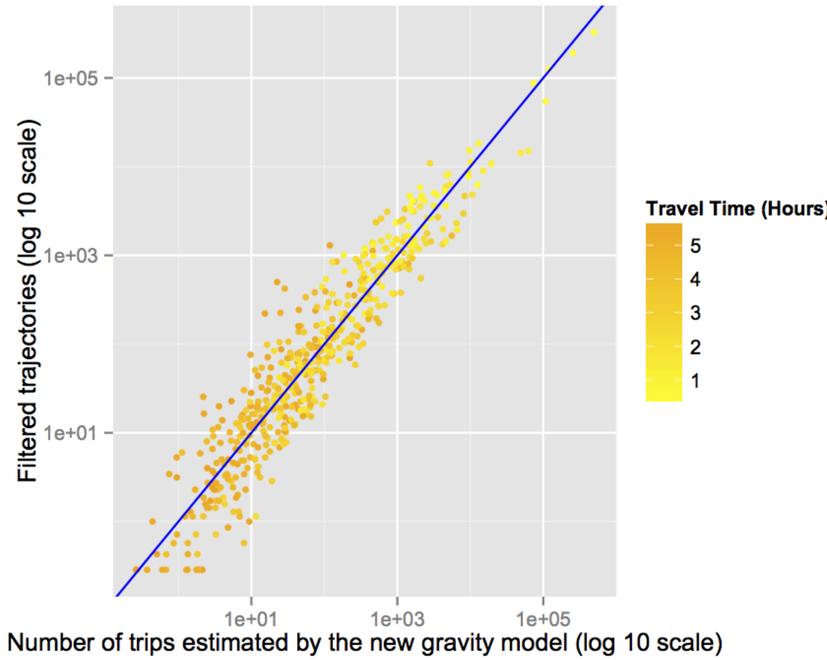


Figure 37 the fitness of the new gravity model for shorter trips

Parameter	Estimate	Std. Error	t value
$\ln K_3$	-4.98627	0.29904	-16.67
a	1.04662	0.02258	46.36
θ	0.58108	0.02787	20.85
Adjusted R-Squared	0.8937		

Table 10 the estimated values of parameters and adjusted R-squared of the new gravity model for shorter trips

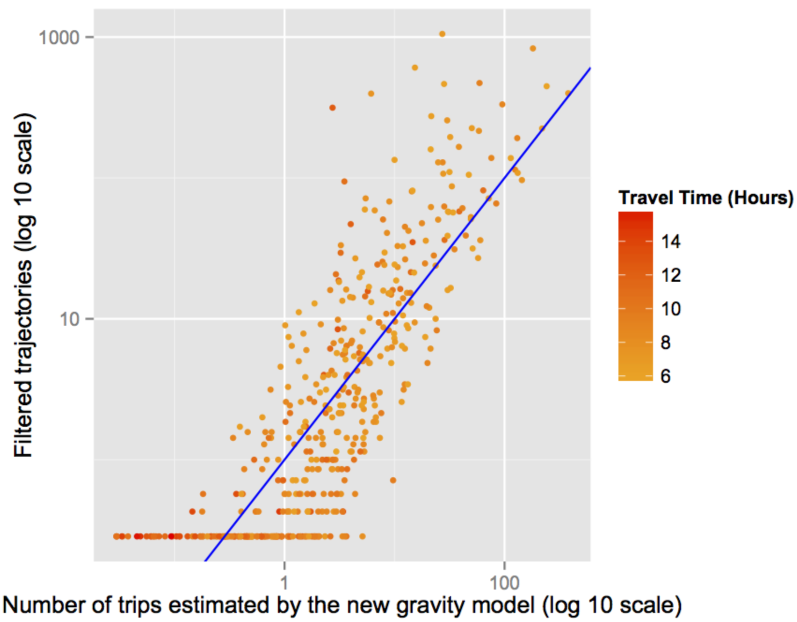


Figure 38 the fitness of the new gravity model for longer trips

Parameter	Estimate	Std. Error	t value
$\ln K_3$	-5.63971	0.41877	-13.467
a	0.87758	0.03136	27.984
θ	0.18451	0.02647	6.971
Adjusted R-Squared	0.6778		

Table 11 the estimated values of parameters and adjusted R-squared of the new gravity model for longer trips

5.4 The Comparison between the Two Gravity Models

As a part of the flowchart for the methodology of the thesis shown in Figure 8, the flowchart showing the methodology of this section is illustrated in Figure 39.

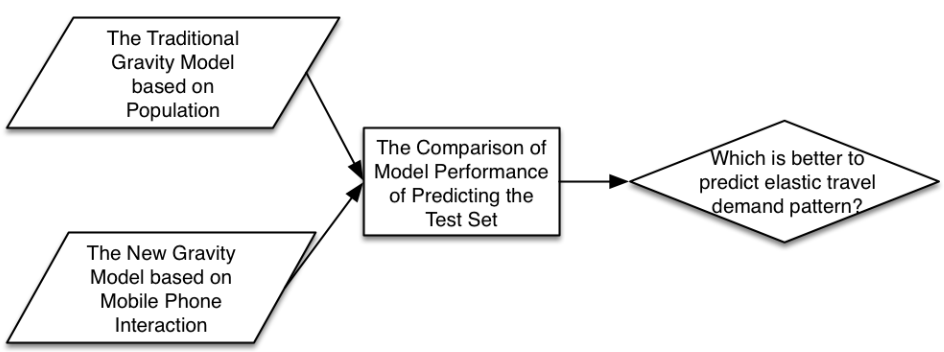


Figure 39 the flowchart for the methodology of Section 5.4

Two different estimated gravity models were used to predict the mobility patterns after August 1st, 2013. This test set was used to assess which model has a greater predictive power. The comparison of their model performance of predicting the test set is shown in Figure 40 and Figure 41. It can be observed that the traditional gravity model based on population does not perform well especially when the travel time is higher. Root Mean Square Error (RMSE) is used as the indicator to test model performance by comparing observed values and predicted values. The undirected mobility data in the test set were used as observed values, and we converted the directed travel demand predicted by the traditional gravity model into the undirected one in order to be compared with the undirected travel demand predicted by the new gravity model. As calculated, RMSE of using the traditional gravity model based on population is 157229.3, while RMSE of using the new gravity model based on the number of mobile phone interactions is only 5590.4. As a result, we chose the new gravity model based on the number of mobile phone interactions, which performs much better, in order to support the decisions on road network design.

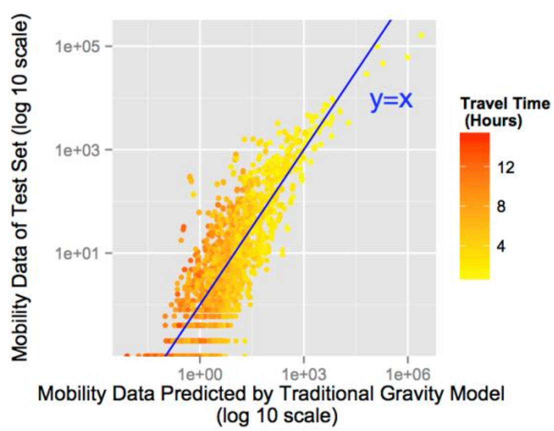


Figure 40 model performance of the estimated traditional gravity model based on population (RMSE: 157229.3)

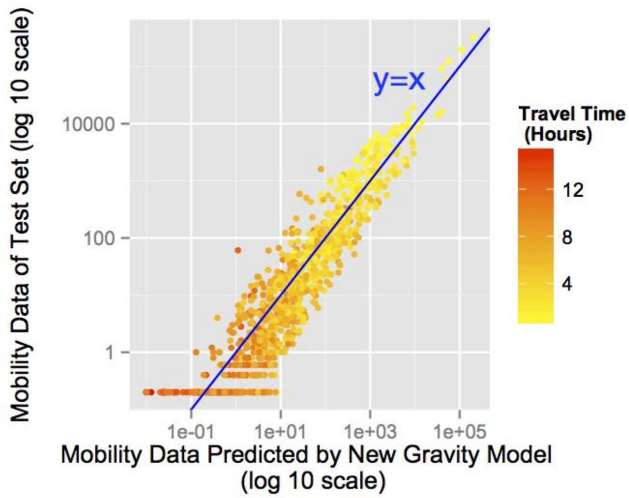


Figure 41 model performance of the estimated new gravity model based on mobile phone interaction (RMSE: 5590.407)

From our point of view, there are some possible reasons why the new gravity model based on the number of mobile phone interactions performs better:

- The mobile phone interaction data are more reliable, more precise and more updatable than the population census data.
- The mobile phone interaction data can reflect the social interaction between zones, which population cannot reflect.
- The estimated mobility information and the mobile phone interaction data are correlated inherently since they are both collected from mobile phone devices.

The first two possible reasons can positively prove that the mobile phone interaction data can be a better proxy than population to estimate travel demand.

However, the third possible reason might indicate that the significant relationship found between telecommunication and travel was a result of the fact that the mobility information was estimated using mobile phone data. In other words, the estimated mobility information and the mobile phone interaction data were correlated inherently since they were both collected from mobile phone devices. If the mobile phone interaction data and the mobile phone trace data had some sampling biases in common, then the built model, despite its good fitness, would predict the biased results. It has been found in Chapter 4 that the sample of both datasets, especially the sample of mobile phone trace data, might not be evenly distributed in the country.

It can be explained that in spite of the existing sampling biases of mobile phone trace data, it can reflect mobility patterns to a large extent since (1) the most active mobile phone users are also likely to be the most active travellers; and (2) we applied a promising filtering algorithm to improve the OD estimation. Therefore, we can still regard the mobility patterns of those travellers as the most representative ones in the country. Moreover, mobile phone traces are the best available mobility information which can be obtained in Senegal. We can derive the mobility patterns from nothing but this dataset.

Nevertheless, the doubt cannot be answered to a full extent without additional traffic information. Therefore, it is recommended that the government can use additional traffic information, such as available road counts and available mobility survey data, to validate the estimated relative OD matrices and the estimated gravity models, and to further test if mobile phone interaction is still a better proxy for predicting travel demand than population. In addition, if possible, the data of other mobile provider's users can be combined with the data of Orange's users, including mobile phone interaction data and mobile phone trace data, to improve the representativeness of the sample.

In addition, international mobility was not taken into account in both the traditional gravity model and the new gravity model. Firstly, as mentioned in Section 4.3.4, the external visitors from other countries who do not use the mobile service provided by local mobile companies are not traced in the dataset. It was simply assumed that the number of those external visitors is relatively small compared to the residents of Senegal, who dominate the mobility patterns in Senegal. Therefore, the mobility of external visitors is not modeled in this case. Secondly, the real destinations of some Orange's users might be outside the border of Senegal, but the destinations revealed in the mobile phone traces could only be the departments which are next to the neighboring countries since the traces in other countries cannot be recorded. To that extent, travel time for the international trips might have been underestimated, leading to the inaccuracy of the model. However, the number of people who travelled outside the country is also relatively small, and the estimation of the model would be influenced to a very small extent.

5.5 Conclusions

By fitting the data of the estimated relative OD matrices, departmental population and the skim matrices of shortest travel time, the parameters of the traditional gravity model regarding departmental mobility were estimated for the country of Senegal. However, the fitness was not satisfying, and especially, the model seemed not be able to explain the distribution of long-distance travel demand very well.

The relations between telecommunication and travel were further explored empirically by comparing the estimated relative OD matrices and the mobile phone interaction matrices. It was found that the aggregate number of filtered trajectories made by sampled users between every two departments is almost proportional to the aggregate number of mobile phone interactions and inversely proportional to the travel time between departments.

The observed relation between telecommunication and travel gave inspiration to construct a new form of gravity model, based on the aggregate number of mobile phone interactions instead of population which the traditional gravity model is usually based on. The parameters of this new model were estimated, which gave a model fitting better than the traditional gravity model to predict elastic travel demand patterns for potential road network changes in Senegal. Especially, the new gravity model based on mobile phone interaction is more capable of modeling long-distance mobility than the traditional gravity model.

6. Road Network Planning

In the previous chapter, we estimated a new gravity model based on the number of mobile phone interactions, which can model the travel demand better than the traditional gravity model based on population. In this chapter, we use the new model and the available mobile phone data to design the road network for the country.

6.1 Planning Approach

The approach to road network planning in this study followed the main principles listed below:

- Planning decisions include adding new links of given levels or upgrading existing links to higher levels.
- Both efficiency and equity are the objectives.
- Construction costs of adding and upgrading links should not exceed the budget.
- Travel demand is elastic with road network design.

The flowchart of the planning approach is illustrated in Figure 42.

Firstly, we started from the top of the figure. We used the road network with newly-opened Pikine-Diamniadio highway section as the current road network, as shown in Figure 10, including five types of road which are: toll highway, national road, regional road, departmental road and ferry connection. We assumed different levels of average service speed on these different types of road, and shortest travel time can be calculated based on these assumed speeds. To upgrade existing links, we improved road levels by improving the corresponding speed levels since speed is the only design characteristic of road levels to be considered in this case. In fact, besides speed, the capacity of roads should also be considered as an important design characteristic. However, in this case, actual traffic flow cannot be estimated, and we can only estimate a relative mobility pattern in Senegal, not to mention the scarcity of the information regarding capacity of roads in Senegal. Therefore, we had no knowledge about the ratio between traffic flow and capacity on each road, and thus we only used assumed average speeds, instead of capacity, as the design characteristic to indicate different service levels of different road types.

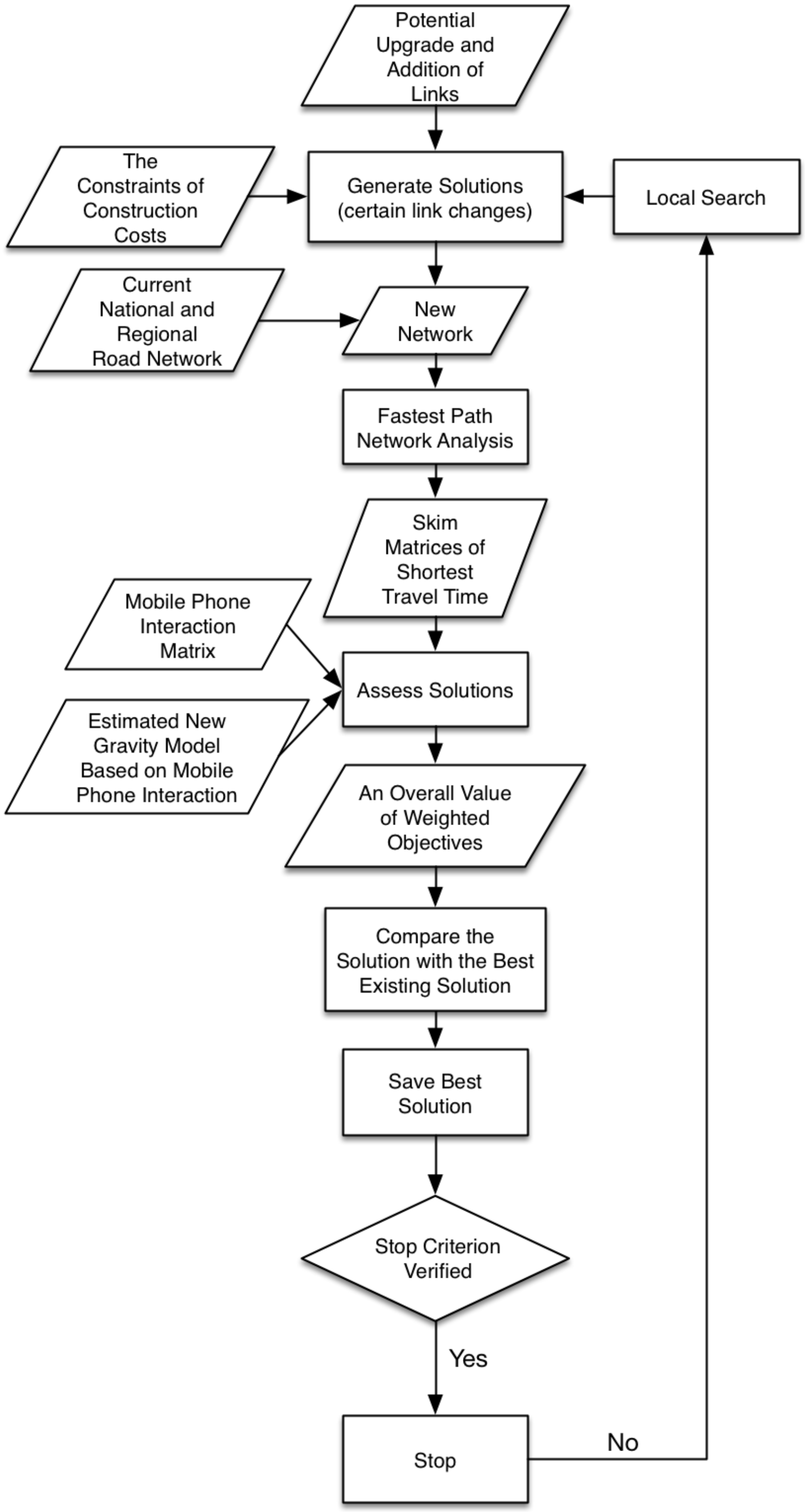


Figure 42 the flowchart for the methodology regarding road network planning

Apart from upgrading existing links, adding new links was also considered. The potential links should be determined. In this case if neighboring departments are not well connected, a potential straight link was made between them unless there are physical barriers (e.g., mountains). All potential links are shown in Figure 43. For solving road network design problems, the average service speeds of these potential links are assumed as zero. In this case, highway is considered as the supreme level of all road types since in recent years there are more projects regarding construction of new highway in Senegal. We assumed that the average speed in a highway is 80 km/h, same as the assumed average speed on toll highway in Dakar. All road types can be upgraded to highway level. In addition, regional roads can be upgraded to national roads, and departmental roads can be upgraded to regional or national roads. Potential links can be added as regional or national road or highway. It is noteworthy that we considered in this planning whether the two ferry services should be replaced by bridges. Because of long waiting time and limited capacity of ferries, it could be supposed that those ferry services are mobility bottlenecks. Thus we assumed that ferry connections can be upgraded to bridges, and moreover we assumed that the average service speed on bridges is the same as the one on national road, 60 km/h. The average service speeds of each road level and the relative unit costs for road construction and upgrading are shown in Table 12. We take the relative unit costs used in the study by Santos et al. (2009) as a reference for determining the ones used in our study.

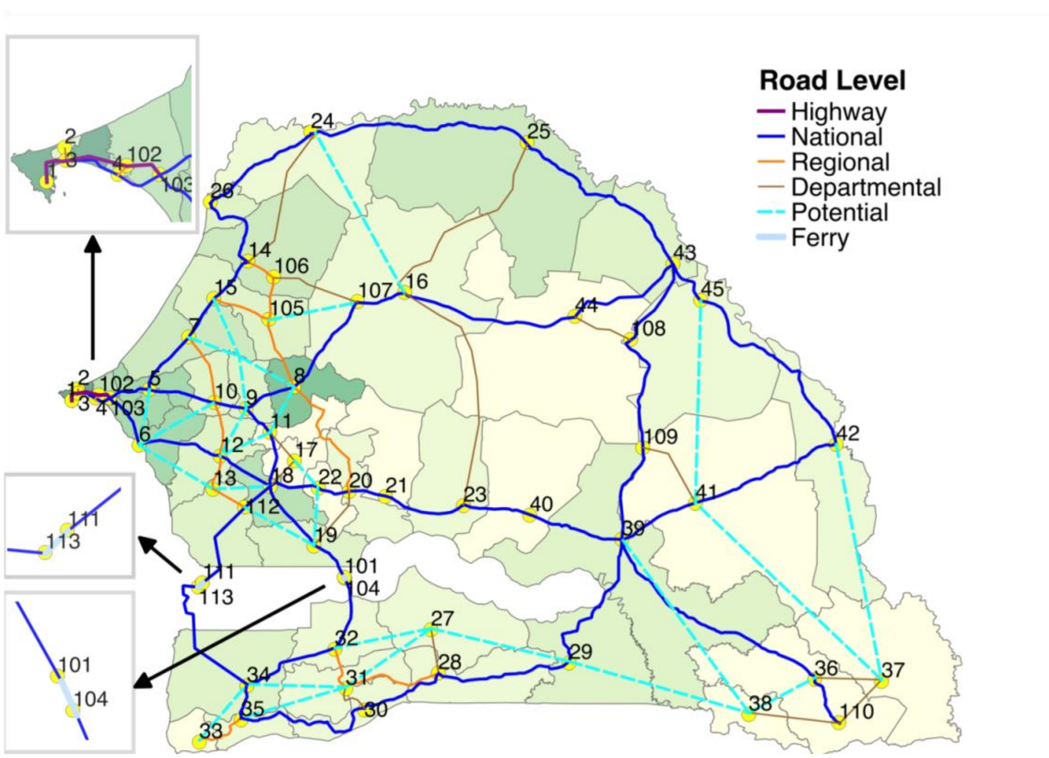


Figure 43 potential links to be added and existing links to be upgraded

	Upgraded level	Potential	Departmental	Regional	National	High way	Bridge
Existing level	Average speed	0 km/h	30 km/h	45 km/h	60 km/h	80 km/h	60 km/h
Potential	0 km/h	-	-	1.2	1.6	4	-
Departmental	30 km/h	-	-	0.2	0.6	3	-
Regional	45 km/h	-	-	-	0.4	2.8	-
National	60 km/h	-	-	-	-	2.4	-
Ferry	0-2 km/h	-	-	-	-	-	8

Table 12 design characteristics of different road levels and relative unit costs for road construction and upgrading

The best assignment of 2171 monetary units (which represents 10% of the total budget required to construct all potential links as highway and to upgrade all existing links to the highest level, assumed as the available budget in this case) was determined to improve the existing road network. Under the budget constraint, there are still millions of solutions regarding how to add and upgrade links. Different solutions would lead to different new networks, which result in new shortest travel times between departments. With these potential network changes, we applied our estimated new gravity model based on the number of mobile phone interactions to predict elastic travel demand pattern, in terms of a relative OD matrix which is meant to reflect as best as possible predicted mobility pattern. Travellers were assumed to follow the fastest paths, travelling at the average service speeds consistent with the road levels of the links included in their routes.

When the elastic travel demand was predicted for a solution of a new network, it was simply assumed that the telecommunication patterns would keep constant, and focused only on the impact of new travel time on travel demand. Some people might argue that the number of mobile phone interactions would change as well in the future. In fact, the one-year data cannot give us plenty of information to judge if the telecommunication patterns would change for a long time span. In Santos et al.'s study (2009), it was assumed that population would be constant for predicting elastic travel demand. If that is the case, given Equation (2-7), telecommunication patterns can be assumed to be constant as well. A more elaborate but complicate way to deal with this problem is to take human migration into consideration, which requires a human migration model to predict the shift of population as well as the shift of telecommunication patterns, but this method was not used in this case.

We assessed the solutions with regard to efficiency and equity objectives. Regarding efficiency, we used the maximization of the accessibility of centers in the country as the measure. According to Santos et al.'s study (2009), accessibility was defined as (proportional to) the spatial interaction between the center and all other centers, and the typical expression used to calculate weighted average accessibility based on traditional gravity model is given as follows:

$$Z = \sum_{i \in N} A_i \times \frac{P_i}{P} \text{ and } A_i = \sum_{j \in N \setminus i} P_j \times f(c_{ij}) \tag{6-1}$$

Where, Z is the measure of weighted average accessibility; N is the set of traffic generation centers; P, P_i and P_j are the total population of country, the population of center i and the population of center j; A_i is the accessibility of center i; c_{ij} is the travel cost between center i and center j, such as distance or travel time; $f(c_{ij})$ is the cost function which reflects the impedance of travel cost. The cost function estimated in the gravity model can be used here directly.

Since the new gravity model based on the number of mobile phone interactions is applied in this case, and in the estimated model, mobility between departments is proportional to the number of mobile phone interactions and inversely proportional to travel cost, a new expression used to calculate weighted average accessibility based on the new gravity model was adapted as follows:

$$Z = \sum_{i \in N} A_i \times \frac{1}{I} \text{ and } A_i = \sum_{j \in N \setminus i} I_{ij} \times f(c_{ij}) \tag{6-2}$$

Where, I is the total number of undirected mobile phone interaction between all pairs of departments per month; I_{ij} is the undirected number of mobile phone interactions between department i and department j per month; $f(c_{ij})$ is the cost function estimated as a component of the new gravity model based on the number of mobile phone interactions. In this case, the expression of this cost function can be derived from Equation (5-6):

$$f(c_{ij}) = e^{-0.3497 c_{ij}} \text{ and } c_{ij} = t_{ij} \tag{6-3}$$

Where, t_{ij} is travel time between department i and department j.

In this case, the accessibility measure of the efficiency objective is rather compatible with the new gravity model which can only predict relative mobility pattern, since the functional form of the accessibility measure based on the new gravity model does not necessarily incorporate a scaling factor, or in other words, the actual travel flow between each OD pairs is not required for calculating this measure.

Regarding equity, we used the maximization of accessibility for the centers with the lowest accessibility in the country as the measure. The expression is given as follows (Santos et al., 2009):

$$E = \sum_{i \in N_{low}} P_i \times A_i \text{ and } A_i = \sum_{j \in N \setminus i} P_j \times f(c_{ij}) \tag{6-4}$$

Where, E is the measure of equity; N_{low} is the set of centers with lowest accessibility. In this case, we focus on the 20% of department centers with the lowest accessibility (i.e., 9 department centers with the lowest accessibility).

Also, we adapted this equation based on our new gravity model (mobile phone interactions) in this case:

$$E = \sum_{i \in N_{low}} A_i \text{ and } A_i = \sum_{j \in N \setminus i} I_{ij} \times f(c_{ij}) = \sum_{j \in N \setminus i} I_{ij} \times e^{-0.3497 t_{ij}} \tag{6-5}$$

We chose efficiency as the unique objective at first, and a best solution to achieve this objective could be found. Then we took equity objective into consideration, giving different weights to accessibility and equity, leading to different solutions. Afterwards, all the solutions to achieve different objectives could be compared.

Since this road network design problem is non-linear, the optimal solutions are difficult to be found without using heuristic methods. In this study, a local search algorithm was applied to help find the best solutions efficiently. In every iteration, a new solution was generated through the local search algorithm. We compared the new solution assessed in each iteration with the best existing solution obtained in previous iterations. Once the new solution was better than the existing best solution, it would become the existing best solution, and if it was found that the existing best solution cannot be improved any more, the stop criterion indicated in Figure 42 would be verified, and the iteration would stop.

6.2 Optimization Model

To accomplish the approach explained previously, an optimization model should be solved in each iteration.

This model is illustrated as below:

$$\max V = w_z \times \frac{Z(y_{lm}) - Z_0}{Z_B - Z_0} + w_E \times \frac{E(y_{lm}) - E_0}{E_B - E_0} \tag{6-6}$$

Subject to:

$$Z(y_{lm}) = \sum_{i \in N} \sum_{j \in N \setminus i} I_{ij} \times e^{-0.3497 \times t_{ij}(y_{lm})} \times \frac{1}{2I}, \forall i, j \in N (i \neq j) \tag{6-7}$$

$$E(y_{lm}) = \sum_{i \in N_{low}} \sum_{j \in N \setminus i} I_{ij} \times e^{-0.3497 \times t_{ij}(y_{lm})}, \forall i, j \in N (i \neq j) \tag{6-8}$$

$$\sum_{m \in M_l} y_{lm} = 1, \forall l \in L \tag{6-9}$$

$$\sum_{m \in M_l} e_{lm} \times y_{lm} \leq b \tag{6-10}$$

$$T_{ij} \geq 0, \forall i, j \in N, l \in L, y_{lm} \in \{0,1\}, l \in L, m \in M \tag{6-11}$$

Where V = normalized value of a solution; w_z and w_E = weights attached to efficiency and equity objectives; Z and E = values of a solution in terms of each objective (which are not scalable); Z_B and E_B = best values obtained for each objective in previous iterations; Z_0 and E_0 = worst values obtained for each objective in previous iterations; I_{ij} = the number of mobile phone interactions between department i and department j; t_{ij} = the shortest travel time between department i and department j, which is dependent on $y\{y_{lm}\}$; $y\{y_{lm}\}$ = matrix of binary variables equal to one if link l is set at road level m and equal to zero otherwise; N = set of departments; L = set of links; M_l = set of possible road levels for link l; e_{lm} = cost of setting link l at road level m; and b = budget.

The objective function (Equation (6-6)) of this optimization model is set to maximize the normalized value of the road network planning solution. The weights w_Z and w_E, which can reflect the relative importance of accessibility and equity objectives, are given to the normalized values of the solutions. The values of the solutions are normalized using the range of variation of solutions. The values of the solutions Z and E are essentially dependent on the decisions made regarding road levels which are expressed as equation. The constraints Equation (6-7) and Equation (6-8) are the expressions of accessibility and equity based on new gravity model, which have been explained in the previous subsection. The constraint Equation (6-9) is used to guarantee that each link should be set at only one level. The constraint Equation (6-10) is used to guarantee that the cost should not exceed the available budget. Expressions Equation (6-11) gives the domain for each decision variable.

6.3 Solution Algorithm

In this study, a local search algorithm is used to find the best solution. For solving non-linear problems, a local search algorithm generates a new solution based on the current solution by applying a transformation to the current solution in every iteration. This method can prevent from exhaustively searching the entire space of possible solutions (Michalewicz & Fogel, 2004). The key to apply a local search algorithm is to find how a transformation can be applied to the current solution in a specific case. Santos (2009) introduced a specific local search algorithm for road network design problem. This algorithm includes three procedures: add, interchange and drop, which are three ways to transform the solutions. The add procedure starts with the initial network and selects the one-level upgrade link change that improves the objective measure most in successive iterations. The interchange procedure starts with the add solution and selects the combination of one-level upgrade and downgrade link changes that improves the objective most. The drop procedure starts when no further accessibility increase is possible, and it downgrades the links which are previously upgraded by one level. The flowchart for the local search algorithm is illustrated in Figure 44. First, the add search and the interchange search are made iteratively until the budget is reached, and then the drop search is used, followed by the iteration of the add search and the interchange search again, and so on and so forth. When the drop search does not influence the resulted solution much, the algorithm will stop, and the final solution is obtained.

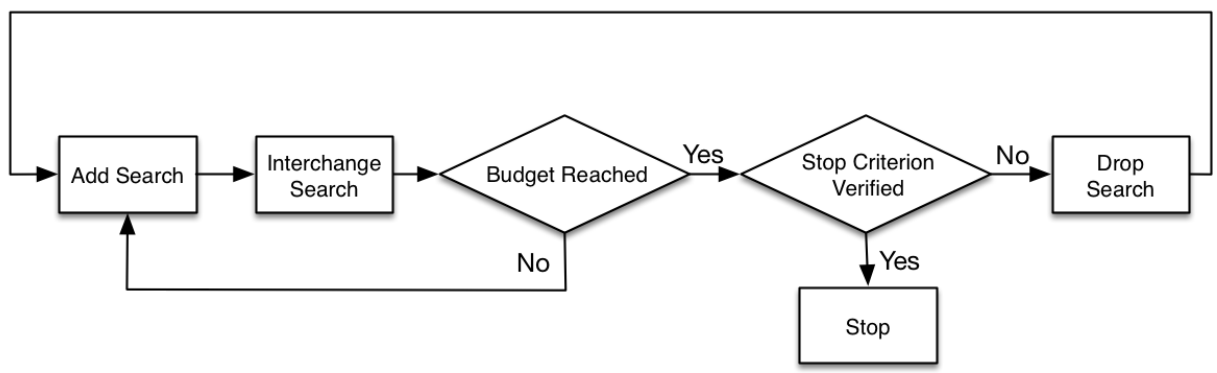


Figure 44 the flowchart for the local search algorithm

According to Santos's evaluation (2009), this local search algorithm performs decently, and the computation time is rather short compared with other algorithms such as an enhanced generic algorithm. Especially when the number of links in the network is more than 100, the solution quality becomes better, and the computation time is much shorter than the computation time of other algorithms. In this case, the number of links in the network is 107, which is one of the important reasons to choose this local search algorithm to solve the problem.

In this case, the optimization program was implemented in R, which is a programming language and software environment for computing. The computation time to generate a solution is about ten minutes on average on an Intel Core i5 processor running at 2.4 GHz.

6.4 Results

6.4.1 Results for a Single Efficiency Objective

Firstly, we only considered efficiency objective by setting w_Z as 1 and setting w_E as 0. The new network of the best solution is shown in Figure 45, and all the links which are upgraded in this solution are highlighted in Figure 46. It can be observed that three main lines of national roads originated from Dakar are suggested to be upgraded to highway in this planning solution. Along these three lines, the existing Dakar-Diamniadio highway could be extended to Dagana (24), Mbacke (8) and Bignona (34) respectively. Especially, the line extended to Bignona passes through Gambia, where Trans-Gambia ferry service is on the way. In this planning solution, a bridge is suggested to be built to replace the ferry service. The connection between Tivaouane (7) and Bambey (10) and the connection between Fatick (12) and Foundiougne (13) are found as the most important regional roads for the accessibility in the country. Thus, they are suggested to be upgraded to national roads. Moreover, a link is suggested to be added between Thies (5) and Mbour (6), and the national road between Pikine (3) and Rufisque (4), which is parallel to the newly-opened Pikind-Diamniadio highway section, is suggested to be upgraded to highway. The departmental connection between Guediawaye (2) and Pikine (3) is suggested to be upgraded to highway as well. All the links suggested to be upgraded are in the western part of Senegal, where the departments are more densely populated.

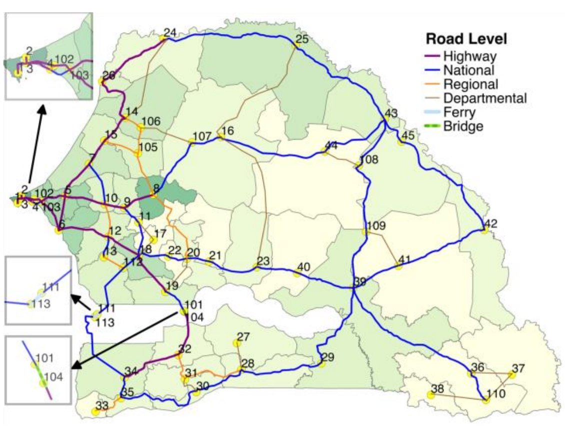


Figure 45 new network of the optimal solution to achieve the single efficiency objective

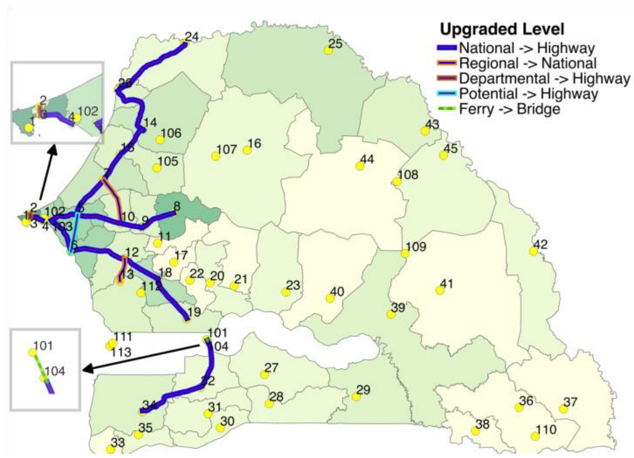


Figure 46 all upgraded links of the optimal solution to achieve the single efficiency objective

In this planning solution, the value of efficiency measure Z increases by 6.548% from the value of the current network.

It can be observed that the upgraded links of the optimal solution mainly radiate from the Dakar region, to reach Dagana (24), Mbacke (8) and Bignona (34), which are all in the western part of the country, which means that to improve the accessibility of the road network in the country, more links should be upgraded to satisfy the travel demand between Dakar and other regions in the western part of the country, since those travel demand is much more than the travel demand in the eastern part of Senegal.

As introduced in 1.2, the geographical disparities are pronounced in Senegal. The Dakar region developed much better than the southeastern regions of the country. If efficiency is the only objective considered for road network planning, this would lead to the improvement of roads next to the centers where travel demand is higher. To that extent, the dissimilarities between large and small centers' welfare will be potentially further increased. For sustainable development, equity issue is taken into account in road network planning.

6.4.2 Results for a Single Equity Objective

As mentioned in planning approach, we chose the accessibility to low-accessibility centers as our equity measure. Firstly, the full weight was given to equity objective by setting w_Z as 0 and setting w_E as 1. The best solution is depicted in Figure 47 and Figure 48. It can be observed that three main lines upgraded to highway radiate from Tambacounda (39). Two potential links are suggested to be added as national roads between Medina Yoro Foulah (27) and Bounkiling (32) and between Kedougou (36) and Salemata (38). The existing departmental link between Kedougou (36) and Saraya (37) is suggested to be upgraded to national roads. A bridge is suggested again to be developed to replace the Trans-Gambia ferry service. This is the only same link change in the two different planning solutions for different objectives. Most links suggested to be upgraded to achieve the equity objective are in the southeastern part of Senegal, where the departments are less populated.

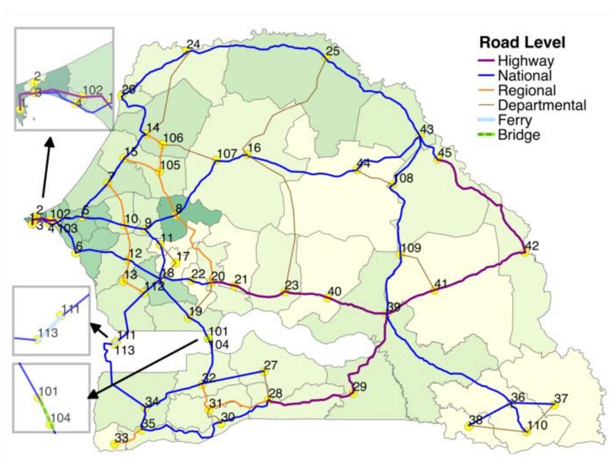


Figure 47 new network of the optimal solution to achieve the single equity objective

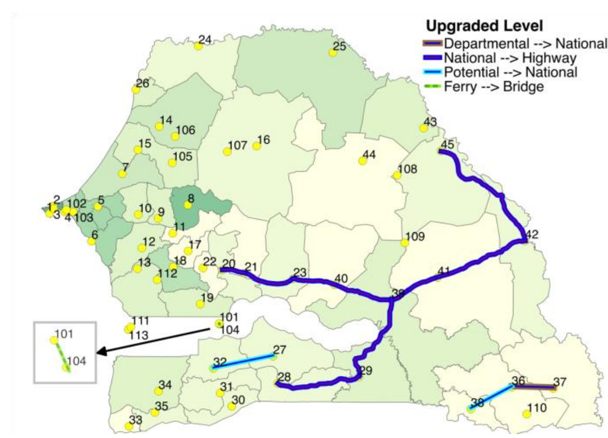


Figure 48 all upgraded links of the optimal solution to achieve the single equity objective

Different from the solution for the single efficiency objective, this solution suggests developing more links in the area where travel demand is not so much.

6.4.3 Results for Efficiency and Equity Objectives

However, it is not possible for government to only consider the equity objective since the solution to achieve the equity objective is not a good one for the efficiency objective. Therefore, to make a trade-off between the different objectives, the different weights are usually given to them. In this case, we include the efficiency objective and the equity objective, assigning equal weights (0.5) to them. The best solution obtained is depicted in Figure 49 and Figure 50.

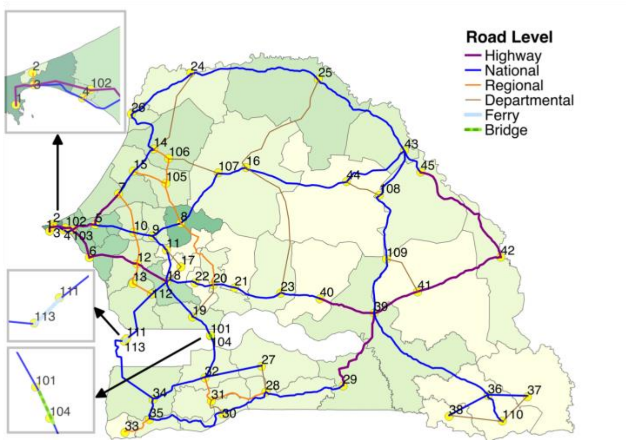


Figure 49 new network of the optimal solution to achieve both efficiency and equity objective

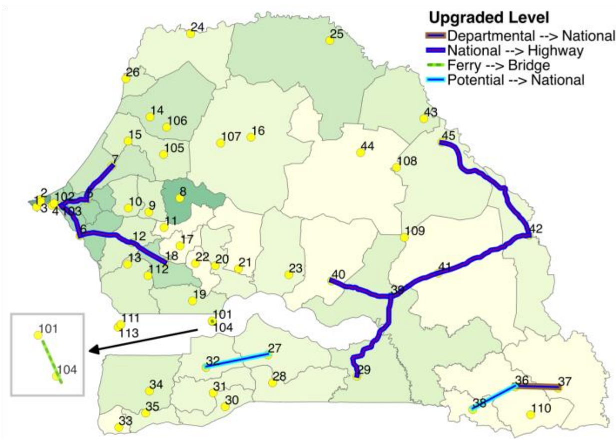


Figure 50 all upgraded links of the optimal solution to achieve both efficiency and equity objective

It can be observed that this planning solution includes the improvement of roads both in the eastern part of Senegal, where the departments are not populated, and in the western part, where the departments are populated. It is noteworthy that the Trans-Gambia ferry service is suggested again to be replaced by a bridge.

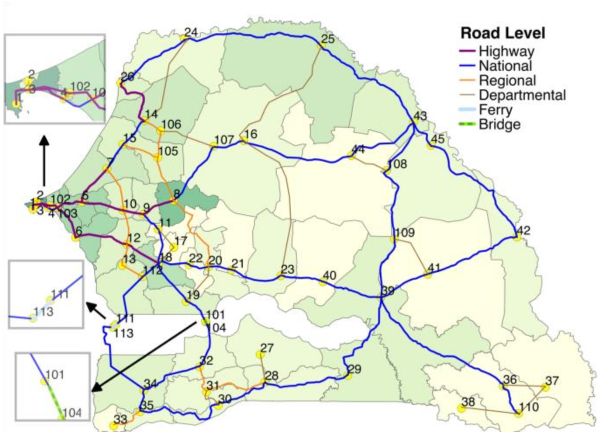
From the values of assessment measures of the current network, the value of equity measure E increases by 18.341%, and the value of efficiency measure Z increases by 3.537%.

6.5 Sensitivity Analysis

6.5.1 Sensitivity to a Budget Reduction

To test the sensitivity of the solutions to a budget reduction, the budget level is considered as 50% of the initial budget for the single efficiency objective and for the objective of 50% efficiency and 50% equity.

Under budget constraint of 1086 monetary units, the best solution for the single efficiency objective is depicted in Figure 51 and Figure 52. From the values of assessment measures of the current network, the value of efficiency measure Z increases by 4.644%.



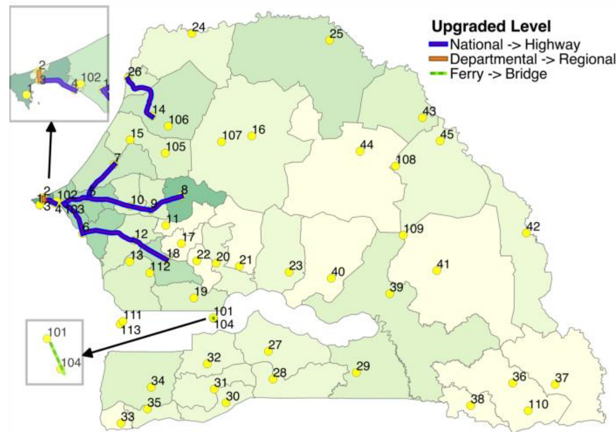
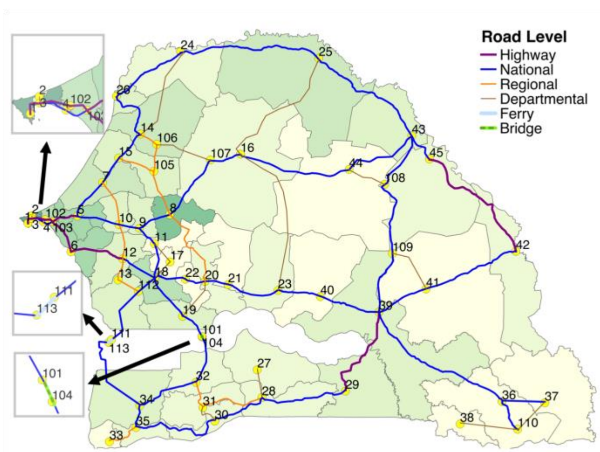


Figure 51 new network of the optimal solution to achieve efficiency objective given ½ budget

Figure 52 all upgraded links of the optimal solution to achieve efficiency objective given ½ budget

Under the budget constraint of 1086 monetary units, the best solution for both efficiency and equity objectives is depicted in Figure 53 and Figure 54. From the values of assessment measures of the current network, the value of equity measure E increases by 8.988%, and the value of efficiency measure Z increases by 2.158%.



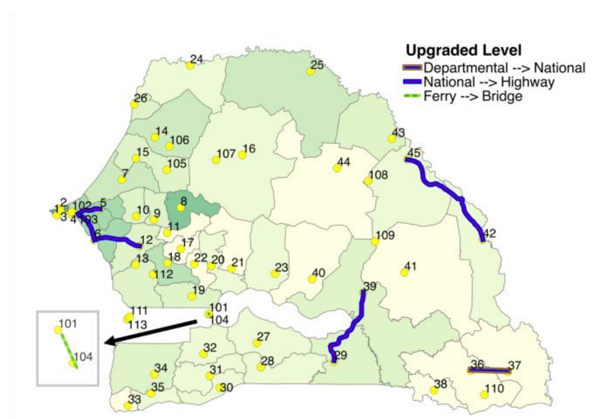


Figure 53 new network of the optimal solution to achieve both efficiency and equity objective given ½ budget

Figure 54 all upgraded links of the optimal solution to achieve both efficiency and equity objective given ½ budget

In Table 13, the increase of the assessment measure values from the current measure values are presented under different scenarios (for different objectives and under different budget constraints). It can be observed that the reduction of budget has less impact on the efficiency measure than on the equity measure. In other words, the increase of efficiency measure slows down with the increase of budget, and on the other hand, there is still much room for improvement of the equity of road network in Senegal, which explains why the use of budget is sensitive to the increase of equity measure.

Solution	Measure	Budget	
		50%	Full
For the single efficiency objective	Z (efficiency)	4.644%	6.548%
For both efficiency and equity objectives	Z (efficiency)	2.158%	3.537%
	E (equity)	8.988%	18.341%

Table 13 the increase of assessment measure values from the current measure values under different scenarios

The Trans-Gambia ferry service is suggested to be replaced by a bridge in all the planning solutions not only for the efficiency objective but also for the equity objective under different budget constraints. Thus, it was not surprising to find that the construction of a bridge has been planned for a long time, though the plan has not come to fruition (Wikipedia, 2013). In addition, the Dakar-Diamniadio highway is suggested to be extended to Thies (5) and Mbour (6) in most of the planning solutions, which is exactly similar with what the government of Senegal is planning as the phase 2 of the Dakar Toll Road Project (ADBG, 2014). The consistency between the model results and the reality validates the model to a certain degree.

6.5.2 Sensitivity to the Selection of the Gravity Model

In Section 5.4, the new gravity model based on mobile phone interaction and the traditional gravity model based on population were compared regarding their ability to predict travel demand, and the new gravity model was chosen as the model used for optimize the road network because of its better performance. In this section, the sensitivity of the selection of the gravity model is examined. Both the gravity models were used to optimize the road network for achieving the single efficiency objective. The respective solutions are shown in Figure 55 and Figure 56.

It can be observed that the upgraded links of both solutions mainly radiate from the Dakar region. However, in the solution found by using the new gravity model, the radiation of the upgraded links are suggested to reach the further regions, and on the other hand, in the solution found by using the traditional gravity model, the upgraded links are more concentrated near the Dakar region, and except the link between Ziguinchor (35) and Bignona (45), there is no road development suggested in the southwestern part of the country. Especially, the bridge is not suggested to replace the ferry service. This observation can be explained by the fact that the traditional gravity model always underestimates the number of longer trips. Therefore, the use of the new gravity model for road network planning can be justified.

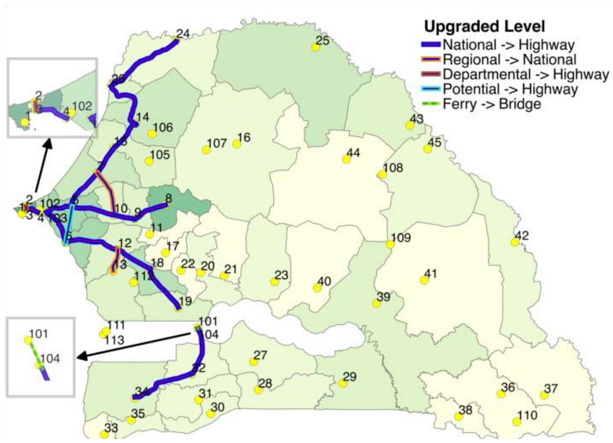


Figure 55 all upgraded links of the optimal solution to achieve the single efficiency objective by using the optimization model based on the new mobile-phone gravity model

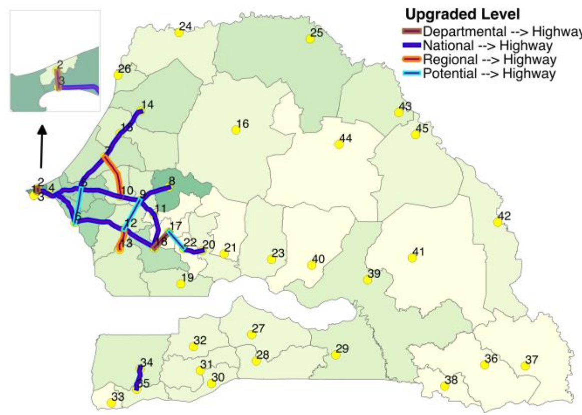


Figure 56 all upgraded links of the optimal solution to achieve the single efficiency objective by using the optimization model based on the traditional gravity model

6.6 Conclusions

The created tool allowed gaining good insight into where and how to expand the Senegal road network. Under the assumed budget constraints, the model suggested that the focus of road development should be radiated from the Dakar region to other regions on the western part of Senegal for the efficiency objective, whereas more roads should be developed on the southeastern part for the equity objective. In addition, the Trans-Gambia ferry service was strongly recommended to be replaced by a bridge connection for all objectives. This suggestion was made for improving the mobility between departments in Senegal, but the bridge cannot be built without the approval from the Gambia government.

The sensitivity of the solutions to a budget reduction was examined, and it was found that there is still much room for improvement of the equity of road network in Senegal whereas the increase of efficiency measure will slow down with the increase of budget.

The sensitivity of the solutions to the selection of the gravity model was examined as well. As a conclusion, it is better for the decision-makers to use the new gravity model based on mobile phone interaction since the traditional gravity model based on population would underestimate the number of longer trips, leading to the bias of the optimal road network planning.

7. Conclusions and Recommendations

7.1 Conclusions

In this study, by using the mobile phone interaction data and the mobile phone traces that the D4D Challenge provided us, we first applied a filtering algorithm to estimate the mobility patterns, and found that the mobility between departments is proportional to the aggregate number of mobile phone interactions between departments and inversely proportional to the travel costs between departments in Senegal. To that extent, we estimated a new gravity model based on the aggregate number of mobile phone interactions between departments, and compared it with the traditional gravity model based on population regarding the model fitness and the predictive accuracy. Because of the better model fitness and the stronger predictive power, the estimated new gravity model based on the number of mobile phone interactions was used to solve the lower-level problem of the national and regional road network planning (i.e., the problem of how travel demand would be distributed and assigned) in Senegal. Under the assumed budget constraints, we selected the efficiency and the equity as the main objectives for solving this network design problem by giving them different weights, and we adapted the functional forms of the efficiency measure and the equity measure, which were originally based on traditional gravity model, to the version based on the new mobile-phone gravity model. The model results show a consistency with some potential plans for the roads and the bridges in the near future which have been announced by the government, and further give the government a good insight into how the other improvements of national and regional road network should be made.

7.1.1 The State-of-the-Practice Contributions

It is believed that the methodology presented in this study is useful in the following ways to the government of Senegal:

- The filtering algorithm introduced in this project can be used to filter the mobile phone traces, to find the OD information for trajectories, and thus to improve the estimation of mobility patterns.
- The empirically found relation between telecommunication and travel, and the new gravity model based on mobile phone interactions, allow the government to better understand and predict mobility patters in Senegal.
- The optimization model based on the new gravity model can help the government to make better decisions on national and regional road network planning using mobile phone data. Based on the actual planning goal, the government can determine the weights of different objectives and the actual available budget in the model by themselves, in order to obtain the best solution under a certain scenario.

7.1.2 The State-of-the-Art Contributions

Moreover, we believe that the presented methodology can be applied not only to the country of Senegal but also to other countries where mobility information is scarce and mobile phone data is available, for understanding mobility patterns and planning road networks. In summary, we can give our answers to the proposed research questions as follow:

1) *How can the mobility patterns in a country be derived from mobile phone traces?*

The trajectories of each user can be directly obtained in the mobile phone trace data by connecting the geographical traces in a time sequence. Next, the developed filtering algorithm can be used to filter the traces and to find the OD information. Finally, by aggregating the number of filtered trajectories, the relative OD matrices can be created as a proxy for real travel demand proportionally.

2) *How can we understand and predict mobility patterns by using gravity models, and what is the use of mobile phone interaction data for developing those models?*

We came up with a new form of gravity model which is based on the aggregate number of mobile phone interactions (i.e., number of calls and text messages) for understanding and predicting mobility patterns proportionally, and we found that the aggregate number of mobile phone interactions is a better proxy used to model the distribution of travel demand than population.

3) *How can the decisions be made to improve the road network in a country based on the gravity model and by using mobile phone data?*

The road network optimization model is based on the prediction of elastic travel demand influenced by any changes made on the road network. Such prediction can be accomplished by the new form of gravity model we estimated. Then mobile phone interaction data can be used as a proxy for predicting elastic travel demand with the impedance of travel time. In summary, mobile phone trace data and mobile phone interaction data were used to estimate the new gravity model, which was used to make decisions on road network design.

Besides the case study in Senegal presented in this thesis, the cases in other countries can be compatible with the methodology as long as the data of mobile phone interactions and mobile phone traces can be provided in those countries. To depict the mobility patterns, this methodology is especially suitable for the countries where mobility information is scarce or outdated. Moreover, the methodology provides not only the less developed countries but also the developed countries with a way to develop a good model to predict future travel demand and to plan a road network.

7.2 Recommendations

In this study, the mobile phone traces were used to depict the mobility patterns in Senegal as best as possible, and furthermore regarded them as the ground truth to find the relationship between telecommunication and travel and to estimate the gravity models. Even though a filtering algorithm was applied to improve the OD estimation, it might still be questioned whether the filtered traces could represent the real mobility of population, and some people might furthermore argue that the strong relationship found between telecommunication and travel is a result of the fact that the mobility information was estimated using mobile phone data. Nevertheless, these questions cannot be answered without additional mobility information. Therefore, it is recommended that the government can use additional traffic information, such as road counts and mobility survey data, to validate the estimated relative OD matrices and the estimated gravity models. In this way, the actual number of trips, instead of the number of filtered trajectories made by sampled users, can be approximated in the best conceivable situation. For further improving the optimization model, more detailed network data, such as road capacity, can be collected, and then, instead of the simple assumption of 'all-or-nothing', the effect of congestion can be taken into consideration when travel demand is assigned to different routes in the model. This is especially relevant in high-demand areas like big cities.

To further elaborate the methodology and improve the model, more detailed factors are recommended to be considered in the future research. Firstly, modal split of the traced travellers can be inferred. Since we focused on the road network planning, we should have identified which individuals used road transport, even though road transport clearly dominates among other transport modes in Senegal. Secondly, the travel demand of external visitors in the country can be complimentary to the estimated mobility patterns since those visitors do not use the local mobile service and they are not traced in the available datasets. Thirdly, the international mobility can be modeled as well. Our mobile phone trace datasets cannot capture the travel across the border of the country, leading to the inaccurate estimation of the model. Fourthly, we could take human migration into account to predict the shift of telecommunication patterns and thus to predict the mobility patterns in a more accurate way. However, we did not take those factors into consideration in this research since they would make our methodology much more complicated, and we think that considering them would not improve our results to a large extent.

In further research, it is recommended that the government can combine more pervasive data sources (e.g., the data of other mobile providers' users and the mobile internet access data) to increase the sample size and thus to improve the representativeness of the data samples. With the development of the internet technology, there is a new trend that more and more people use mobile applications such as Facetime, Skype, Wechat, imessage, etc. instead of traditional mobile phone calls or text messages to contact with each other. To that extent, the GPS data recorded in those mobile applications and the mobile internet access data might have better qualities to help transport planners trace the trajectories of people more accurately, and to reflect the real telecommunication patterns of people, the use of the interaction data recorded in those mobile applications might become more important in further research.

In this study, the interdepartmental mobility in a country was focused on. In further research, the scale can be changed, and either inter-zone mobility in a city or international mobility in the world can be studied. To that extent, the outcome of the research can be the optimization of urban road network planning or the optimization of the international airline network planning. To extract the OD information in a more elaborate scale, the filtering algorithm should be improved, and a higher sampling rate of mobile phone data would be required. It can be hypothesized that the new gravity model can still be useful to predict the trip distribution based on mobile phone interactions in a different scale, but that should be proved empirically in future case studies.

Reference List

ADBG (2014). Executive Summary of the Environmental and Social Impact Assessment, Dakar Toll Road

Project - Phase 2 Diamniadio-AIBD Section. http://www.afdb.org/fileadmin/uploads/afdb/Documents/Environmental-and-Social-Assessments/Senegal_-_Dakar_Toll_Road_-_Phase_2_-_Diamniadio-Aibd_Section_-_ESIA_Executive_Summary.pdf

ANSD (2013). http://www.ansd.nl/

ArcGIS (2013). Retrieved 12-11-2014, from http://www.arcgis.com/home/item.html?id=7a4486a4903b42c3ba3ce3a750c3bcca

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

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

Calabrese, F., Dahlem, D., Gerber, A., Paul, D., Chen, X., Rowland, J. et al. (2011). The connected states of america: Quantifying social radii of influence. In (pp. 223-230). IEEE.

Calabrese, F., Di Lorenzo, G., Liu, L., & Ratti, C. (2011). Estimating origin-destination flows using mobile phone location data. *IEEE Pervasive Computing,* 36-44.

Calabrese, F., Ferrari, L., & Blondel, V. D. (2014). Urban Sensing Using Mobile Phone Network Data: A Survey of Research. *ACM Computing Surveys (CSUR), 47,* 25.

Calabrese, F., Smoreda, Z., Blondel, V. D., & Ratti, C. (2011). Interplay between telecommunications and face-to-face interactions: A study using mobile phone data. *PloS one, 6,* e20814.

Cantarella, G. E. & Vitetta, A. (2006). The multi-criteria road network design problem in an urban area. *Transportation, 33,* 567-588.

Cho, E., Myers, S. A., & Leskovec, J. (2011). Friendship and mobility: user movement in location-based social networks. In (pp. 1082-1090). ACM.

Cohon, J. L. & Rothley, K. (1997). Multiobjective methods. *Design and operation of civil and environmental engineering systems,* 513-566.

Csáji, B. C., Browet, A., Traag, V. A., Delvenne, J. C., Huens, E., Van Dooren, P. et al. (2013). Exploring the mobility of mobile phone users. *Physica A: Statistical Mechanics and its Applications, 392,* 1459-1473.

de Dios Ortúzar, J. & Willumsen, L. G. (2011). *Modelling transport*. John Wiley & Sons.

de Montjoye, Y. A., Smoreda, Z., Trinquart, R., Ziemlicki, C., & Blondel, V. D. (2014). D4D-Senegal: The Second Mobile Phone Data for Development Challenge. *arXiv preprint arXiv:1407.4885*.

Demissie, M. G., de Almeida Correia, G. H., & Bento, C. (2013a). Exploring cellular network handover information for urban mobility analysis. *Journal of Transport Geography, 31,* 164-170.

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

Digital Logistic Capacity Assessment (2013). Senegal road network information. Retrieved 10-12-2014, from http://dlca.logcluster.org/display/public/DLCA/2.3+Senegal+Road+Assessment/

Eiffage (2013). Dakr-Diamniadio Motorway. http://www.eiffage.com/en/home/le-groupe-1/our-projects/projets-acheves/autoroute-de-lavenir-au-senegal.html

Eto, David (2012). Senegal's mobile penetration passes 88 percent. http://www.itnewsafrica.com/2012/10/senegals-mobile-penetration-passes-88-percent/

Horak, R. (2007). *Telecommunications and data communications handbook*. John Wiley & Sons.

Hoteit, S., Secci, S., Sobolevsky, S., Ratti, C., & Pujolle, G. (2014). Estimating human trajectories and hotspots through mobile phone data. *Computer Networks, 64,* 296-307.

Hsiao, M. H. (2007). Predicting Aggregate Travel Demands of Japan under the Impact of Telecommunications. In (pp. 90). Eastern Asia Society for Transportation Studies.

Invest in Senegal (2015). Highway to the future Dakar. http://investinsenegal.com/Highway-to-the-future-Dakar.html

Iqbal, M. S., Choudhury, C. F., Wang, P., & González, M. C. (2014). Development of originΓÇôdestination matrices using mobile phone call data. *Transportation research part C: emerging technologies, 40,* 63-74.

Kamargianni, M. & Polydoropoulou, A. (2013). Does social networking substitute for or stimulate teenagersΓÇÖ travel? Findings from a latent class model. In.

Krings, G., Calabrese, F., Ratti, C., & Blondel, V. D. (2009). Urban gravity: a model for inter-city telecommunication flows. *Journal of Statistical Mechanics: Theory and Experiment, 2009,* L07003.

McNally, M. G. (2008). The four step model. *Center for Activity Systems Analysis*.

Michalewicz, Z. & Fogel, D. B. (2004). *How to solve it: modern heuristics*. Springer Science & Business Media.

Mokhtarian, P. L. (2002). Telecommunications and travel: The case for complementarity. *Journal of Industrial Ecology, 6,* 43-57.

Nanni, M., Trasarti, R., Furletti, B., Gabrielli, L., Van Der Mede, P., De Bruijn, J. et al. (2014). Transportation Planning Based on GSM Traces: A Case Study on Ivory Coast. In *Citizen in Sensor Networks* (pp. 15-25). Springer.

Nobis, C. & Lenz, B. (2009). Communication and mobility behaviourΓÇôa trend and panel analysis of the correlation between mobile phone use and mobility. *Journal of Transport Geography, 17,* 93-103.

Phithakkitnukoon, S., Horanont, T., Di Lorenzo, G., Shibasaki, R., & Ratti, C. (2010). Activity-aware map: Identifying human daily activity pattern using mobile phone data. In *Human Behavior Understanding* (pp. 14-25). Springer.

Plaut, P. O. (1997). Transportation-communications relationships in industry. *Transportation Research Part A: Policy and Practice, 31,* 419-429.

Ratti, C., Williams, S., Frenchman, D., & Pulselli, R. M. (2006). Mobile landscapes: using location data from cell phones for urban analysis. *Environment and Planning B Planning and Design, 33,* 727.

Santos, B., Antunes, A. n., & Miller, E. (2009). Multiobjective approach to long-term interurban multilevel road network planning. *Journal of transportation engineering, 135,* 640-649.

Santos, B. F. L. (2009). Road Network Planning With Efficiency, Equity, and Robustness Objectives.

Snelder, M., Wagelmans, A. P., Schrijver, J. M., van Zuylen, H. J., & Immers, L. H. (2007). Optimal redesign of Dutch road network. *Transportation Research Record: Journal of the Transportation Research Board, 2029,* 72-79.

Wang, D., Pedreschi, D., Song, C., Giannotti, F., & Barabasi, A. L. (2011). Human mobility, social ties, and link prediction. In (pp. 1100-1108). ACM.

Wang, H., Calabrese, F., Di Lorenzo, G., & Ratti, C. (2010). Transportation mode inference from anonymized and aggregated mobile phone call detail records. In (pp. 318-323). IEEE.

White, J. & Wells, I. (2002). Extracting origin destination information from mobile phone data.

Wikipedia (2013). Trans-Gambia Highway — Wikipedia. http://en.wikipedia.org/w/index.php?title=Trans-Gambia_Highway&oldid=562547018

Wikipedia (2015a). List of African countries by GDP (nominal). Retrieved 5-6-2015a, from http://en.wikipedia.org/wiki/List_of_African_countries_by_GDP_%28nominal%29

Wikipedia (2015b). List of African countries by GDP (PPP) per capita. Retrieved 5-6-2015b, from http://en.wikipedia.org/wiki/List_of_African_countries_by_GDP_%28PPP%29_per_capita

Wikipedia (2015c). Senegal. http://en.wikipedia.org/wiki/Senegal

World Bank (2004). Performance and impact indicators for transport in Senegal. http://www.worldbank.org/transport/transportresults/regions/africa/senegal-road-redi_final.pdf

World Bank (2015). Senegal Overview. Retrieved 5-6-2015, from http://www.worldbank.org/en/country/senegal/overview

Yang, H. & Bell, H. (1998). Models and algorithms for road network design: a review and some new developments. *Transport Reviews, 18,* 257-278.

Zheng, K., Zheng, Y., Xie, X., & Zhou, X. (2012). Reducing uncertainty of low-sampling-rate trajectories. In (pp. 1144-1155). IEEE.

Zheng, Y., Li, Q., Chen, Y., Xie, X., & Ma, W. Y. (2008). Understanding mobility based on GPS data. In (pp. 312-321). ACM.

Zheng, Y., Zhang, L., Xie, X., & Ma, W. Y. (2009). Mining interesting locations and travel sequences from GPS trajectories. In (pp. 791-800). ACM.

Appendix A. More influencing factors on mobility in Senegal

In Senegal, there are more influencing factors on mobility, which are not the most important ones in this research, but are worth mentioning in the appendix.

Climate[1]

The climate in Senegal is tropical, hot and humid. Rainy season starts from May and ends in November), with strong southeast winds, and dry season, dominated by hot, dry, harmattan wind, lasts from December to April. Northeast winter winds and southwest summer winds lead to well-defined dry and humid seasons.

Interior temperatures are higher than along the coast (for example, average daily temperatures in Kaolack and Tambacounda for May are 30 °C and 32.7 °C respectively, compared to Dakar's 23.2 °C, and rainfall increases substantially farther south, exceeding 1,500 mm annually in some areas. Extremes in annual precipitation range from 250 mm in the extreme north, to 1800mm (71 inches) in extreme southern coastal areas. In the far interior of the country, in the region of Tambacounda, particularly on the border of Mali, temperatures can reach as high as 54 °C (129.2 °F).

The climate is regarded as one of the influencing factors on mobility. Especially, the seasonal rainfall would influence mobility to a large extent mainly because (1) people tend to prevent having activities in a bad weather; and (2) the rainfall would have an impact on the capacity of roads. However, in Figure 24, no strong relation can be observed between the seasonal fluctuation of the aggregate number of users' movements and the seasonal fluctuation of rainfall in Senegal. It seems that the effect of the rainfall on mobility could not be observed from such a macroscopic perspective. In addition, in this research, road capacity is not taken into consideration.

More characteristics of national and regional roads[2]

In Section 1.2, all the national and regional roads have been shown on the map. However, more detailed information about the roads is missing. Actually, the related information is quite difficult to be found, and the only useful information which can be found on the internet is regarding the characteristics of several national road sections, described as below. Those descriptions include the length of each road section and the estimated travel time, based on which a rough estimate of average travel speed on national roads can be made.

[1] Reference: Wikipedia (URL: http://en.wikipedia.org/wiki/Geography_of_Senegal)
[2] Reference: Digital Logistic Capacity Assessment (URL:http://dlca.logcluster.org/display/public/DLCA/2.3+Senegal+Road+Assessment/)

The national road section between Dakar and Kidira, named as N1, has a length of 648 km, and the travel time of passing this section is 11 hours by car. Therefore, it can be estimated that the average speed on this national road is about 60 km/h. Heavy trucks (15 -23 tonnes) (6*6 or 6*4) with trailer can pass this road section.

The national road section between Tambacounda and Kedougou, named as N7, has a length of 233 km, and the travel time of passing this section is 3.25 hours by car. Therefore, it can be estimated that the average speed on this national road is about 70 km/h. Heavy trucks (15 -23 tonnes) (6*6 or 6*4) without trailer can pass this road section.

The national road section between Tambacounda and Ziguinchor, named as N6, has a length of 414 km, and the travel time of passing this section is 9.25 hours by car. Therefore, it can be estimated that the average speed on this national road is about 45 km/h. In dry season, heavy trucks (15 -23 tonnes) (6*6 or 6*4) without trailer can pass this road section, and in rain season, only light trucks (8 - 12 tonnes) (6*6) without trailer can pass this road section.

More characteristics of the new toll highway in Dakar region

The locations of the new toll highway sections have been shown in Figure 3. However, it is difficult to find more detailed characteristics of the new highway, among which the average travel speed on the highway is most related to this research.

It is easier to find the speed limit on the highway than to find the average travel speed. In a Youtube video where the passenger recorded a clip of the view along the new toll highway, a speed limit sign showing ‘130’ next to the highway can be captured as shown in the figure below. It can be derived that the average speed on this highway section is not low, but under 130 km/h.

Figure: the screenshot of the video ‘autoroute Dakar Diamniadio high way Dakar’ (credit: https://www.youtube.com/watch?v=-Azzrqdg8AU)

More characteristics of the ferry services

In fact, there are about ten ferry connections in total for crossing the Gambia River[1]. The most popular ferry services for vehicles are Banjul-Barra ferry service and Trans-Gambia ferry service, of which the locations are shown in Figure 4. Therefore, only these two ferry connections are included in the model.

According to the document found on the website of Digital Logistic Capacity Assessment[2], the Trans-Gambia ferry service is very poor. The ferries have a very low capacity, and the waiting time is at least 2 hours. When the ferries broke or when they had any kind of technical problems, the waiting time can even reach 20 hours.

According to the news reported on the website of Gainako[3], the frequency of the Banjul-Barra ferry service is half an hour to two hours at a time.

[1] Reference: Gambia Information Site (URL: http://www.accessgambia.com/information/barra-banjul-ferry.html)
[2] Reference: Digital Logistic Capacity Assessment (URL:http://dlca.logcluster.org/display/public/DLCA/2.3+Senegal+Road+Assessment/)
[3] Reference: Gainako (URL: http://gainako.com/?p=2221)