Big Data Fusion to Estimate Driving Adoption Behavior and Urban Fuel Consumption

by

Adham Kalila

B.Eng, McGill University (2012)

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Master of Science in Transportation

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2018

Signature redacted

Author .....
Department of Civil and Environmental Engineering
May 18, 2018

Signature redacted

Certified by .....
Marta C. González
Associate Professor of Civil and Environmental Engineering
Thesis Supervisor

Signature redacted

Accepted by .....
Jesse Kroll
Professor of Civil and Environmental Engineering
Chair, Graduate Program Committee

Big Data Fusion to Estimate Driving Adoption Behavior and Urban Fuel Consumption

by

Adham Kalila

Submitted to the Department of Civil and Environmental Engineering
on May 18, 2018, in partial fulfillment of the
requirements for the degree of
Master of Science in Transportation

Abstract

Data from mobile phones is constantly increasing in accuracy, quantity, and ubiquity. Methods that utilize such data in the field of transportation demand forecasting have been proposed and represent a welcome addition. We propose a framework that uses the resulting travel demand and computes fuel consumption. The model is calibrated for application on any range of car fuel efficiency and combined with other sources of data to produce urban fuel consumption estimates for the city of Riyadh as an application. Targeted traffic congestion reduction strategies are compared to random traffic reduction and the results indicate a factor of 2 improvement on fuel savings. Moreover, an agent-based innovation adoption model is used with a network of women from Call Detail Records to simulate the time at which women may adopt driving after the ban on females driving is lifted in Saudi Arabia. The resulting adoption rates are combined with fuel costs from simulating empty driver trips to forecast the fuel savings potential of such a historic policy change.

Thesis Supervisor: Marta C. González
Title: Associate Professor of Civil and Environmental Engineering

Acknowledgments

I could not have completed this masters without the love, encouragement, and support of my family, friends, and advisors.

To my thesis advisor and mentor, Marta C. González, who took a chance on me and patiently taught me the ins and outs of not just research but MIT and the academic world in general, thank you.

To my unofficial advisor, Sarah Williams, who continues to inspire in me a love of public transportation and mapping, thank you.

To my friends, my Cantabrigian family, your support through the ups and downs of such a unparalleled place as MIT was not only wonderful but essential.

To my friends from Cairo, my shella sha2eya, your faith in my abilities kept me going even when I had lost faith in myself.

To my family, especially my mother Mona, your life inspires me to achieve more and more. To my brother Amir, I will always be grateful for the ways you have enabled me to be here.

To the Massachusetts Institute of Technology, I am forever grateful for your lasting effect on my professional, personal, and academic life.

Contents

1 Introduction **15**

1.1 Overview and Motivation 15

1.2 Literature Review 17

1.2.1 Travel Demand and Call Detail Records 17

1.2.2 Fuel Consumption Estimation Models 18

1.2.3 Social Diffusion Adoption Models 19

1.3 Thesis Outline 21

2 Calibrating the Fuel Consumption Model **23**

2.1 Introduction 23

2.2 StreetSmart Model Sensitivity Analysis 24

2.3 Results and Energy Indices 25

3 Fuel Consumption Application for Traffic Congestion Policies **31**

3.1 Methodology 32

3.1.1 From GPS Data to Speed Profiles 32

3.1.2 Fuel Consumption Results 35

3.2 Discussion and Conclusion 40

4 Fuel Effects of Women Adopting Driving in Riyadh **43**

4.1 Introduction 43

4.2 Data Description: CDR and Gender labeled Users 44

4.2.1 Data description 44

4.2.2 Gender Labeling 45
4.2.3 Expansion Factors 45
4.3 Trip Generation and Fuel Consumption Methods 46
4.3.1 Trip Generation and flows 46
4.3.2 Driver Trip Simulation 50
4.4 Calculating Fuel Consumption 51
4.4.1 Fuel Consumption Model 51
4.4.2 Verification of Aggregate Fuel Results 52
4.4.3 Model of Adoption of Driving by Women 54
4.5 Discussion and Conclusion 56
4.5.1 Limitations and Assumptions 56
4.5.2 Policy Recomendations 58

5 Conclusion **59**

5.1 Summary 59
5.2 Limitations of the Framework 60
5.3 Future Work 61

List of Figures

1-1 Stay and pass-by identification from filtered points showing home and work location examples. Source: (1) 18

1-2 Everett Rogers' innovation adoption curve showing the difference between early, middle, and late adopters. Source: (40) 20

2-1 Values of Energy Indices k resulting from the regression of fuel consumption and speed profiles in the Illinois experiment 25

2-2 Sensitivity Analysis of k parameters. Values of k are varied individually while the other indices are set such that fuel efficiency is at 20.5 mpg and the result on fuel efficiency is graphed 26

2-3 Speed profiles showing idle and moving times. Source: StreetSmart experiment (50) 27

2-4 Department of Energy Fuel Efficiency vs Average Speed Curve (52) 28

2-5 (a) FTP-75 EPA's standard speed profile used for calculating the reported inner-city fuel economies of cars. (b) The distribution of fuel economies recreated by the StreetSmart model shows the same distribution as that of the reported fuel economies. (c) The distribution of fuel economies based on Riyadh's fleet of cars compared to those of Poland and the UK shows that the distributions are similar but shifted from one another. The car fleet of Riyadh is less fuel efficient than that of Poland which is less than that of the UK 28

3-1 Average hourly taxi trip production rates in Riyadh in Ramadan, non-Ramadan, and combined. 33

3-2 Data Verification figures using trips in the morning peak time period of weekdays from 8 - 9 AM. (a) Histogram of Reported and calculated Trip Distances. (b) Histogram of free flow travel time and Observed travel time in matched trips. (c) Histogram of Fuel economies using constant speed, speed profiles, and 1 bin and all bins. 35

3-3 Choropleth Maps of fuel consumption rates [Liter/meter.hour] by the StreetSmart model on streets matched with GPS data for typical time periods morning peak (8 - 9 AM) weekdays, midday off-peak (12 - 13 PM) weekdays, evening peak (17 - 18 PM) weekdays. 37

3-4 Fuel consumption estimates at urban scale. (a) Comparison of the travel time of the routes in the constant speed model via travel demand vs. the input used in our method using GPS data (b) Sample speed profiles in two routes used to estimate fuel consumption overlaid with the constant speed used for comparison (c) Estimates of fuel consumption and fuel economy in the morning peak via our method (speed profiles) and the base line method (constant speed) (d) Random and targeted fuel savings vs. number of reduced trips. 39

4-1 Expansion Factors of CDR users, (a) Choropleth map of expansion factors for all CDR users, (b) Choropleth map of expansion factors for gender-labeled CDR users, (c) distribution of expansion factorsfor all CDR users, (d) scatter of CDR population vs census population for every TAZ in blue and after expansion in red, (e) ditribution of expansion factors for gender-labeled CDR users, (f) scatter of CDR population vs census population for every TAZ by gender 47

4-2 Features of Mobility by Gender, (a) a Joyplot showing the trip departure time distribution for male and female users by trip purpose HBW,HBO,an NHB compared to the distribution from the NHTS, (b) A map of the home and work locations of users colored by gender, (c) outer: Average locations visited per day by gender, inner: Lth most visited locations by gender, (d) Radius of Gyration and Average Stay Duration distributions by gender, (e) Mobility Diversity distribution by gender 48

4-3 Driver Trip Simulation, (a) Departure time distribution of empty driver and essential female trips, (b) diagram of the relation between empty driver and essential female trips 51

4-4 Morining Trip Simulation Verification Diagrams, (a) Boxplots of trip simulation times [min], distances [km], and fuel consumption [liter] with median shown, (b) distribution of fuel efficiency of each trip, (c) scatter of GPS-recorded total trip time vs ITA congested trip times 53

4-5 Network and Adoption Scenario Results, (a) degree distribution of the female gender-labeled CDR users network of communications and the largest component in the network visualized, (b) Adoption Scenarios and their associated Ratio of Women driving and fuel savings potential 56

List of Tables

2.1 A Comparison of Fuel Consumption Estimates from the StreetSmart Model and the DOE Fuel Economy Fit on Data From the Experiment Conducted by (51)) 27

2.2 Results of the Calibration of the StreetSmart Model. Ranges of k_i Parameters for Each Bin of Fuel Economy 29

4.1 Percent of Trips by Purpose and Time of Day Compared to NHTS and MHTS 49

4.2 Fuel Consumption Estimation Totals for Male and Female and Empty Driver Trips 53

Chapter 1

Introduction

1.1 Overview and Motivation

This thesis is the product of several motivations on my part and on the part of my advisor, Prof. Marta Gonzalez. First, it is an attempt to build upon the travel demand computed from Call Detail Records (CDRs) to estimate fuel consumption and show that the ubiquity of data being produced today has uses far beyond its original intention. Second, it is a scientific exercise in modeling the effects of the liberalization of restrictions on women driving in Saudi Arabia. The results show how we can start from several disparate data sources, combine them with a framework and human simulation model to produce tangible results that are then verified against state of the art techniques and ground truths.

The reason the methods outlined below are pertinent and realistic is thanks to the wealth of data being collected from mobile phones today. With penetration rates of over 85% in developing countries and up to 100% in some developed countries, mobile phones are constantly and precisely recording our movement and our lives. This is a far cry from the level of data that is used in a household travel survey which depicts a typical day and cannot robustly consider the effects of weather, holidays, and other day-today factors on travel demand. In order to harness useful insights from such new data, we must develop techniques that utilize its potential. Apart from straight forward origin-destination (OD) travel demand which is used in infrastructure decisions, environmental studies, etc. such data can also be used to answer specific policy questions through simulation and modeling

human behavior.

The motivation to compute travel demand from Big Data sources such as mobile phone traces or CDRs comes as a response to the difficult alternative of building discrete choice models and conducting expensive and time-consuming stated preference surveys. Previous work has shown that these new techniques, such as TimeGeo (1), are faithful to the state of the art methods and verified against real surveys and accepted truths. We extend the current framework which identifies stays, generates trips, and estimates travel demand, to add the ability to estimate fuel consumption. This energy consideration comes at a time when oil-producing countries are suffering from the low prices of oil since 2016. Saudi Arabia's economy is undergoing immense changes to diversify its economy and decrease its reliance upon the price of one commodity. For this reason, we have developed a framework to accurately estimate urban fuel consumption.

The methods are also applied to simulate the possible fuel savings potential of women starting to drive in Saudi Arabia. Since the 1970s, women have been banned from driving for cultural and supposedly religious reasons. There were some protests by activist women in the 1990s but they were immediately put down. With the recent increase in young Saudis returning after studying abroad, the ruling family, especially the young crown prince Mohamed Bin Salman, is looking to appease the new generation and improve Saudi Arabia's international reputation. One of the new changes is the promise that the ban on women driving will be lifted in June 2018. We take advantage of the fuel consumption framework we developed as well as gender-identified subset of users to simulate the empty driver trips generated around women's mobility needs. With these we apply a diffusion model similar to the spread of disease or the adoption of a new technology in the market to relate the fuel savings with time under several scenarios.

The methods used utilize and contribute to the current knowledge in three distinct areas: Energy consumption with GPS, travel demand from CDR, and adoption modeling for women driving for the first time. The contributions of this thesis are summed up as follows: First, we calibrated a previously developed fuel consumption model (StreetSmart) and applied it to varying fuel efficiencies in car fleets; Second, the model was used along with travel demand estimated by CDR-based traffic assignment to approximate fuel consumption

rates in Riyadh, Saudi Arabia; Third, we examined the effects of the proposed method by comparing the effects on fuel consumption of different traffic relief policies; Fourth, we model the associated empty driver trips made to accommodate the ban on women driving in Saudi Arabia and model the fuel savings potential of the adoption of driving after the ban is lifted. The following section outlines the previous knowledge and where the current thesis fits into it.

1.2 Literature Review

1.2.1 Travel Demand and Call Detail Records

The traditional engineering method of estimating travel demand used travel diaries and surveys and followed a four-step model. Trip generation, trip distribution, mode choice, and trip assignment produced an origin-destination flow matrix aggregated by Traffic Analysis Zone (TAZ). Since then, improved computing has enabled trip-based modeling to capture more idiosyncrasies and individual-level variations in the data. Ever increasing storage capacity and faster processing, both local on hand-held devices as well as cloud-based, has resulted in incredibly massive amounts of data such as location, elevation, purchases, tweets, check-ins, and even heart-rate to be stored and logged for billions of people around the planet.

Data collected by telecommunications providers for billing purposes has been used to shed light on human urban mobility since it was found that humans have very predictable routines and are slow to discover new places (2). For the purposes of transportation, algorithms and methods for stay extraction (3), OD extraction and validation (4–7), travel speed estimation (8, 9) and activity modeling (10, 11) have been developed to benefit from CDR data. An example of *home* and *work* identification from stay and pass-by locations is shown in Figure 1-1.

CDR typically contains a timestamp, location coordinates, duration of the call or text, and a unique identifier for the user. The coordinates in the CDR records obtained from the city of Riyadh, Saudi Arabia for this project were of the cell towers used by the phone and not of the phones themselves. This resulted in an OD model aggregated at the TAZ

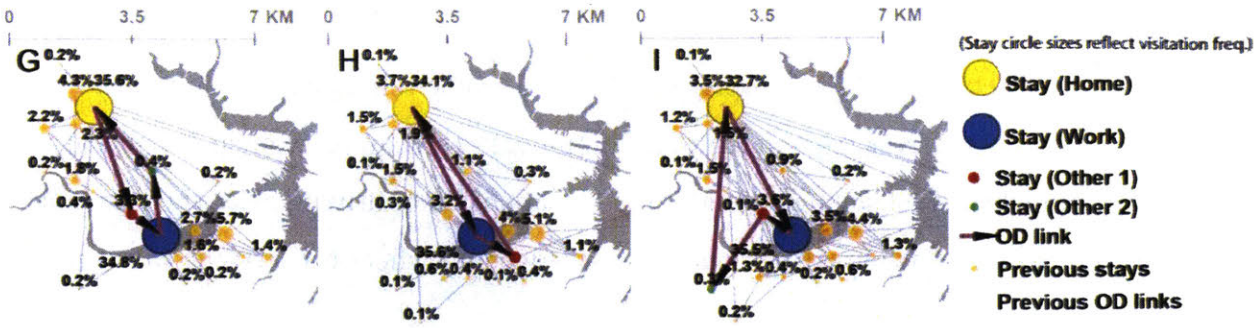

Figure 1-1: Stay and pass-by identification from filtered points showing home and work location examples. Source: (1)

level and greatly simplified the analysis. The TimeGeo framework (1), which in turn was built on previous work on CDR (12, 13), is the basis from which the OD flows used in this project are derived. A summary of how trip generation is derived from the CDR records is as follows. First, stays are identified apart from pass-by points as locations where the user spent a significant time within a threshold radius. Based on the time and frequency of the stay locations, they are labeled as one of *home*, *work*, or *other*. This label feeds into a heuristic that informs the trip generation. For example, users start and end each day at their home location and they commute to work with a departure time drawn from local and national household travel surveys. Once all trips have been generated the flows that they represent, based on how often a user is observed in the dataset, are then iteratively assigned to roads and the resulting traffic flows are calculated. This work describes using the resulting flows in conjunction with speed profiles observed from a high frequency GPS dataset to estimate fuel consumption on roads.

1.2.2 Fuel Consumption Estimation Models

The estimation of energy consumption from location data is prevalent in the literature with several models that utilize different factors. Since smartphone market penetration is almost complete in the Transportation Networking Companies (TNCs) industry (14), GPS tracking has been successfully used to estimate air pollution (15), instantaneous fuel consumption (16–20), and traffic conditions (21–25). Most studies that use GPS data to estimate fuel consumption have focused on individual user fuel consumption for route optimization(18, 20).

Fuel consumption and emissions models have been extensively developed in the literature (26–33). They are generally split between models that estimate the fuel consumption by balancing the engine's carbon intake and combustion and those that attempt to use mode-specific variables, such as speed and acceleration, to fit a model that estimates fuel consumption (32). Of those models that use instantaneous mode-specific variables, some estimate air pollution and emissions(15, 27, 28, 30), fuel consumption (18–20, 31) or both (16, 17, 29). Most previous attempts at estimating instantaneous fuel consumption and emissions do not incorporate GPS data but rely instead on On-Board Diagnostics devices (OBD-II) that measure fuel consumption and emissions (16–18, 29). The models that have attempted to use GPS data to estimate fuel consumption do so without consideration to the different fuel economies found in today's cars (19, 20, 31) and do not account for the total demand.

Our foray into fuel consumption aims to contribute to this evident lack in the framework. GPS data is used to compute fuel consumption rates per street for a variety of car fuel economies. This is then combined with traffic flows from an optimization algorithm to give a comprehensive picture of urban fuel consumption for an entire city, by time of day. This framework is then applied to Riyadh for verification and analysis.

1.2.3 Social Diffusion Adoption Models

The study of diffusion stems from rural sociology in the early 20th century and was solidified by a study on the spread of hybrid seed corn in rural Iowa (34). In 1969, Bass published a paper outlining the detail of what will become the most popular diffusion model. It assumes that potential adopters are influenced into adopting by innovation and imitation. Innovation can be the result of a media campaign whereas imitation comes from the interaction with other people. The resulting S-curve shows a distinction between the adoption of the phenomenon by early adopters, early and late majority adopters, and finally laggards. The Bass diffusion model takes as input aggregate numbers and its results are also aggregated and thus do not take into account the individual interactions between people or nodes. To make use of the social networks that are represented in telecommunication records (phone, text, email or social media), several studies have been conducted to simulate diffusion on an agent

based level (35–39). This new framework has been proposed in the literature which leverages network information for the imitation parameter, the threshold of probability above which a node decides to adopt. These have been more successful than the differential equation framework at describing diffusion with the Bass diffusion model.

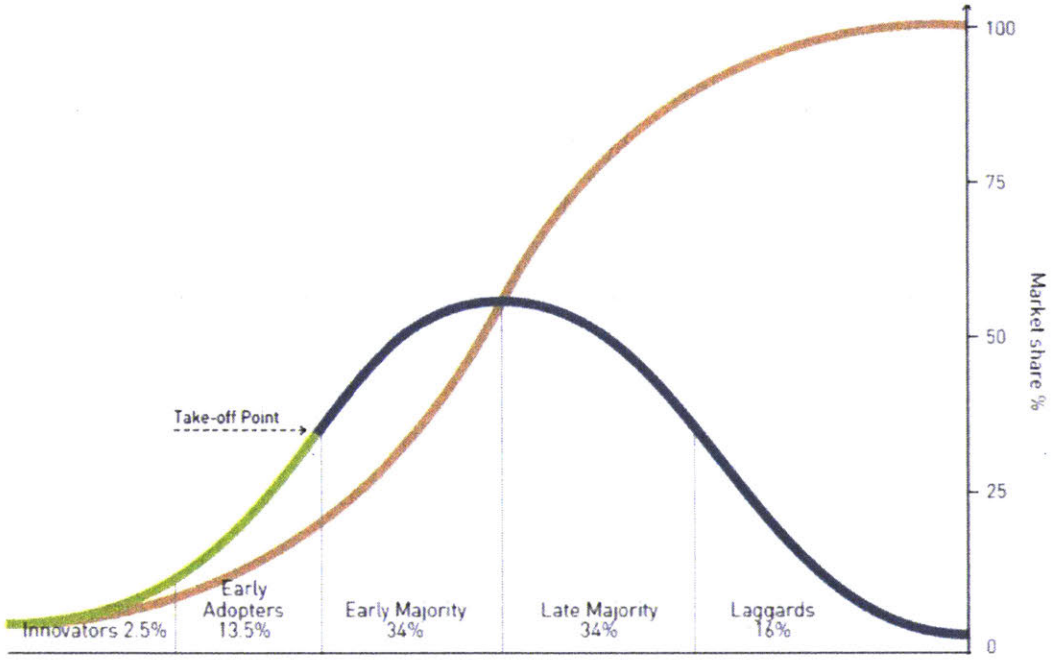

Figure 1-2: Everett Rogers' innovation adoption curve showing the difference between early, middle, and late adopters. Source: (40)

The parameters used for the Bass diffusion curve are the main area of debate when using the model for a particular product or phenomenon. Since the adoption of driving for the first time has never been recorded, electric car sales offer a reasonable proxy for its parameters. A summary of the sources of parameter values was compiled by Massiani and Gohs (41). They have found that a considerable number of papers and studies use previous parameters than are based on little more than conjecture by an expert (42–45). Other sources of parameter values can be non-peer reviewed such as Masters theses or reports from the private sector (46). Finally, when observable data is unavailable for a certain product in a specific market, similar markets or products are used as a proxy and parameters are fitted to observed sales data (47–49). Based on the results of the study by Massiani and Gohs, several scenarios of parameters were chosen to represent the adoption of Saudi women. A couple

of scenarios were added to mathematically model the results of surveys where women were asked about when they intended to drive, if at all, once the ban is lifted. This is presented in Chapter 4.4.

1.3 Thesis Outline

The remainder of this thesis presents a framework to utilize GPS and CDR to compute fuel calculation and test several policy application in Riyadh, Saudi Arabia.

Chapter 2 describes the method used to calibrate a model which uses speed profile (or driving schedules) to estimate fuel consumption given certain parameters. The process uses the results of an experiment where several cars outfitted with On Board Diagnostic Devices were driven around a track and their fuel consumption as well as their speed profiles were recorded. The model was calibrated for every type of fuel efficiency by categorizing vehicles, including motorcycles, buses and trucks, into 14 fuel efficiency ranges or bins. Parameter values for each bin were derived and later used on a real world application in Riyadh.

Chapter 3 presents the framework developed for fuel estimation from the model and applied to several traffic reduction policy scenarios. It was published as a paper in the Transportation Research Record in 2018.

Chapter 4 applies the fuel consumption framework on a subset of women and the empty driver trips that they incur. The travel demand of the city of Riyadh is used to compute the total urban fuel consumption which is verified against official reports and comparable cities. Moreover, driver trips were simulated and their fuel consumption calculated. The Adoption of driving by women is modeled in 4 different scenarios and the associated fuel savings from avoided empty driver trips are computed.

Chapter 5 presents the conclusion and summary of findings as well as areas of potential future work.

Chapter 2

Calibrating the Fuel Consumption Model

2.1 Introduction

Fuel consumption is the result of the function of the power needed by the vehicle to overcome resistance integrated over time. This relationship between fuel and the variables that describe the vehicle and its movement is fit into a simple yet comprehensive model. The StreetSmart model was developed in 2011 at MIT to utilize accurate mobile phone sensor data to predict fuel consumption. It is part of a growing body of work that leverages GPS and Wi-Fi sensors on mobile phones to measure traffic. It goes beyond other efforts by using speed profiles with acceleration information over just average speed along a trip. The model fits data on the speed profiles to approximate four energy indices that capture different aspects of fuel consumption in a vehicle trip. The variables in the power function that are dependent on a vehicle's characteristics such as area, mass, accessories, etc. are replaced by these energy indices which are estimated by regression. For this project, we used the measured fuel consumption and speed profiles reported by an experiment conducted at the University of Illinois to get initial values for these energy indices.

The Illinois experiment analyzed and smoothed the speed profile data of four cars driven around a track with stop-and-go movements. The dataset contained raw video data as well as files with measured performance properties such as speed, acceleration, fuel efficiency, air mass flow rate etc. For our purposes, the data used were speed, time, and fuel consumption. The experiment consisted of 9, and sometimes 10, different cars driven around

a 30 meter diameter track. The next section will provide details on how we used the raw data to estimate energy indices of the StreetSmart model for 14 ranges or bins of fuel efficiency in cars.

2.2 StreetSmart Model Sensitivity Analysis

The StreetSmart model was developed by measuring the energy required by vehicles for various movement conditions. It estimates the fuel consumption with data from GPS coordinates from smartphones and ground truth fuel consumption data from On Board Diagnostics II (OBD-II) devices. Using the details of a trip's speed profile, the model successfully predicts fuel consumption with over 96% accuracy [26], a substantial improvement over models that only consider constant average speeds. Average speed estimations do not account for the stop and go effect of traffic, which is a significant factor leading to an increase in fuel consumption. In other words, average speed simplification results in lower, or more optimistic, fuel consumption estimates since drag is lower at low speeds.

After testing different variables for their use in predicting fuel consumption, the model employs a combination of four variables to predict fuel consumption as shown in 2.1. The first term accounts for energy wasted while the car is idling with the engine turned on; the second accounts for energy used with time spent moving; The third accounts for energy used due to acceleration and deceleration over a distance; the fourth accounts for energy used with distance traveled. Each term is multiplied by a specific energy index, k_i, which depends on the vehicle efficiency, such that:

$$FC = k_1 T_{idle} + k_2 T_{move} + k_3 \int |a| dx + k_4 L; \tag{2.1}$$

where FC is fuel consumption in US gallons and k_1,k_2,k_3, and k_4 are the energy indices calibrated with data separately for each bin of vehicle efficiency. T_{idle} and T_{move} are time spent idling and moving respectively in seconds, a is acceleration in m/s^2, and L is the distance driven in km.

The energy indices varied between the different models of cars. A chart of their difference is shown in Figure 2.1. To understand the effect on fuel consumption of each

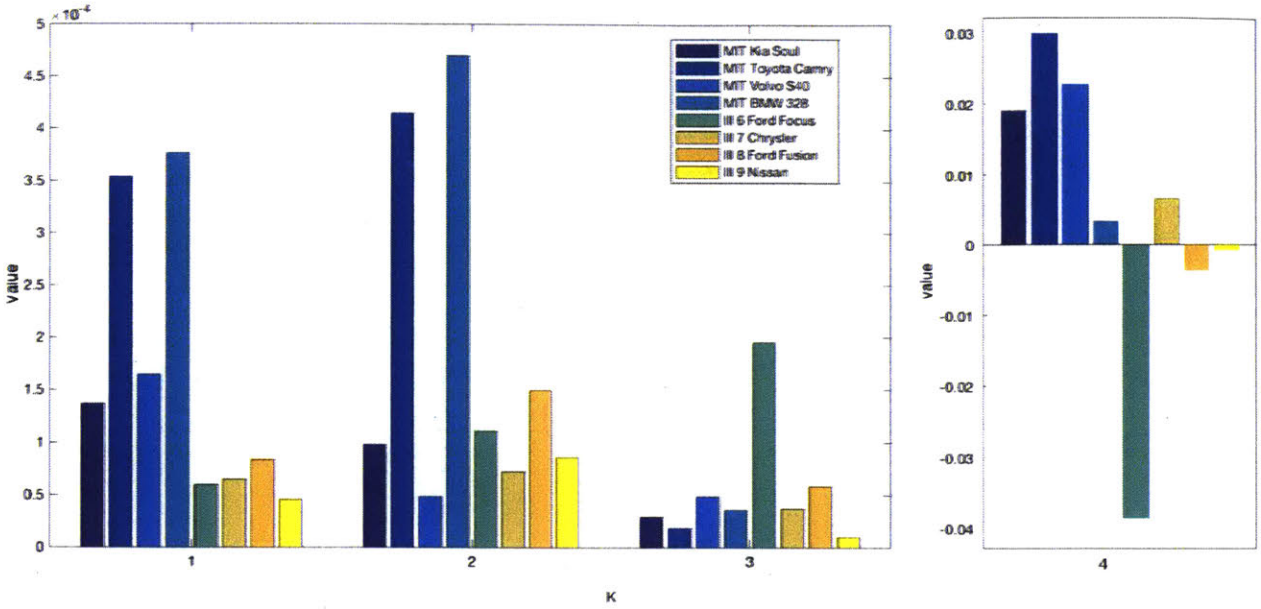

Figure 2-1: Values of Energy Indices k resulting from the regression of fuel consumption and speed profiles in the Illinois experiment

energy index alone, a sensitivity analysis was conducted where each index was varied between its observed minimum and maximum while keeping the other 3 such that fuel efficiency is at the average of 20.5 miles per gallon (mpg). The results on the fuel efficiency are shown in Figure 2.2. It was found that the change in fuel efficiency for each index was 3.5, 15.1, and 26 mpg respectively. Evidently, fuel efficiency is less sensitive to k_1 and k_3 and more sensitive k_2 and k_4.

2.3 Results and Energy Indices

Following (50), we further tested the StreetSmart model's indices by regression of the idle fuel consumption and moving fuel consumption separately using data from an experiment conducted at the University of Illinois by [27]. A graph of speed profiles showing idle and moving times is showin in Figure 2-3. To verify the benefit of using the model, its results were compared to a baseline estimation using average speed and the United States' Department of Energy's (DOE) graph of fuel economy variations by speed as shown in Figure 2-4. The StreetSmart model achieved inaccuracies of about 4% while the baseline method had inaccuracies of up to 29%. The details of the comparison are summarized in Table 2.1.

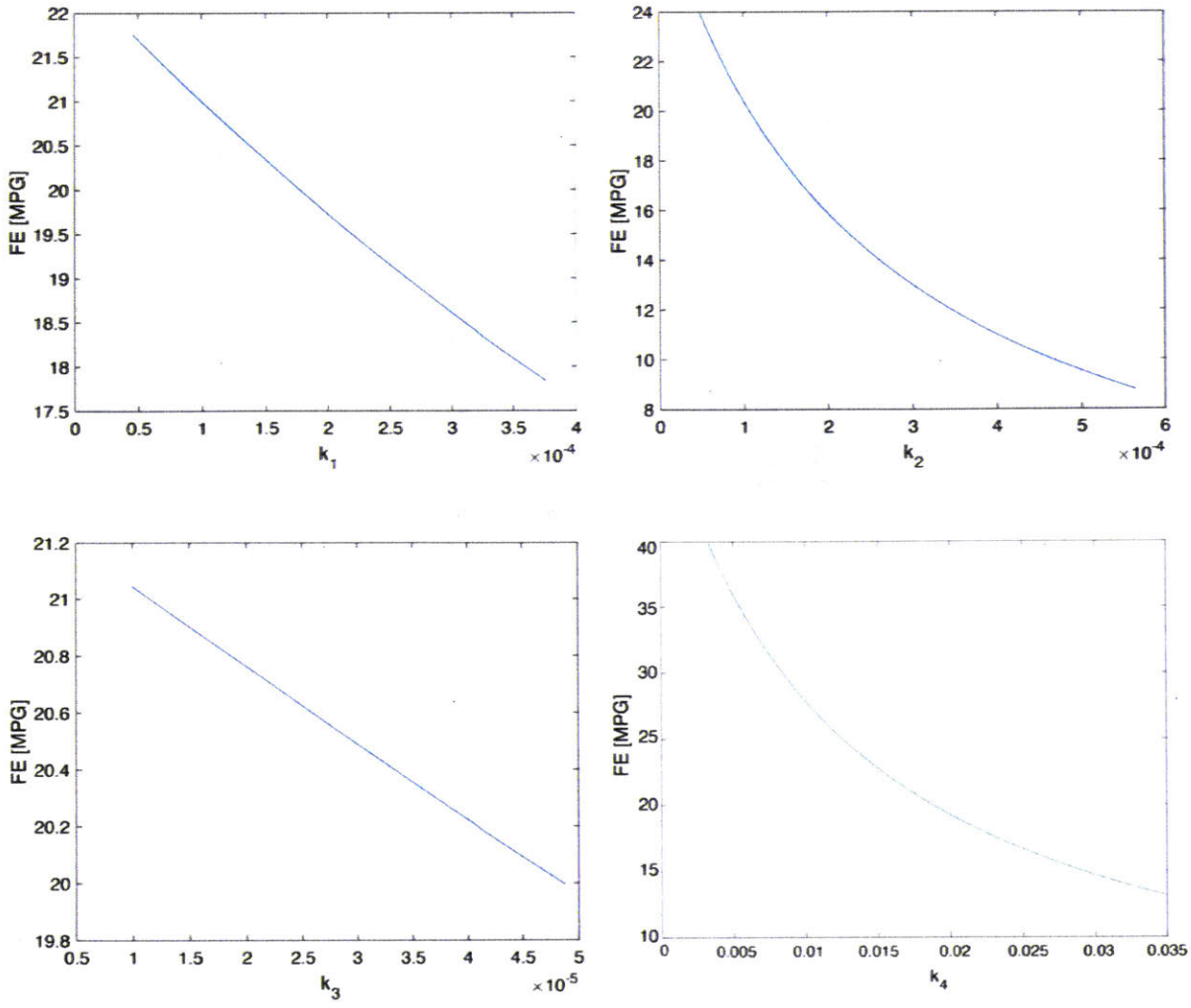

Figure 2-2: Sensitivity Analysis of k parameters. Values of k are varied individually while the other indices are set such that fuel efficiency is at 20.5 mpg and the result on fuel efficiency is graphed

To apply the StreetSmart model on a known distribution of car fuel economies, the first step is to relate the energy indices specific to each car fuel efficiency or economy. The variables with the highest influence over fuel consumption were the second and fourth indices representing time moving and the distance traveled respectively. This suggests that the first and third terms, representing the influence of speed profiles, have a lower impact on the overall estimate of fuel consumption. However, results shown in section 3.3 show that all terms are useful since speed profiles improve the accuracy of estimation.

To calibrate the model and get energy index values for use on the scale of a city, we used the fuel economies reported by the Environmental Protection Agency's (EPA) 2016

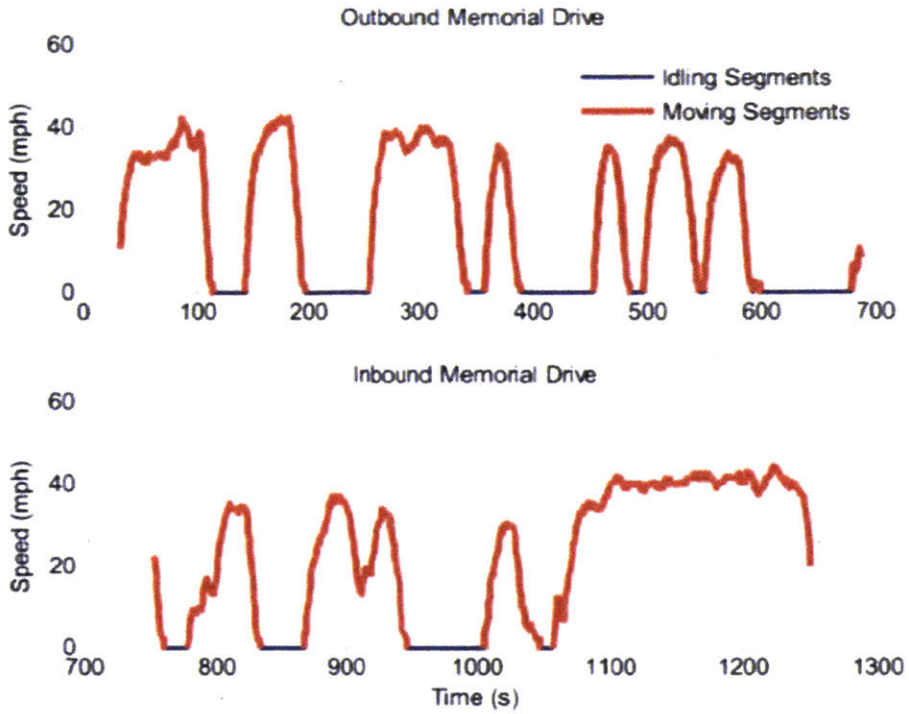

Figure 2-3: Speed profiles showing idle and moving times. Source: StreetSmart experiment (50)

report to arrive at ranges of each index for different cars, categorized by their fuel economy [FE guide]. The EPA uses a standard speed profile to test for a car's urban fuel economy, the FTP-75, shown in Figure 2a [EPA test procedure]. We used the mode specific variables from the FTP-75 speed profile with the reported fuel economy to calibrate the energy index ranges for each bin shown in Table 2.2. A standard deviation of 1 mpg was used to create Gaussian distributions of fuel economies to set the range of index values for each bin. This resulted in a smooth transition between the fuel economies when the entire city's fleet efficiency profile

Table 2.1: A Comparison of Fuel Consumption Estimates from the StreetSmart Model and the DOE Fuel Economy Fit on Data From the Experiment Conducted by (51))

Car No. (Illinois Test A) :	6	7	8	9
OBD-II FC [US Gal]	0.0246	0.0220	0.0356	0.0211
DOE Fitted Curve [US Gal]	0.0247	0.0253	0.0252	0.0245
StreetSmart FC [US Gal]	0.0243	0.0230	0.0371	0.0210
% diff. StreetSmart	-1.2%	4.2%	4.2%	-0.6%
% diff. DOE Fitted Curve	0.4%	14.7%	-29.3%	16.0%

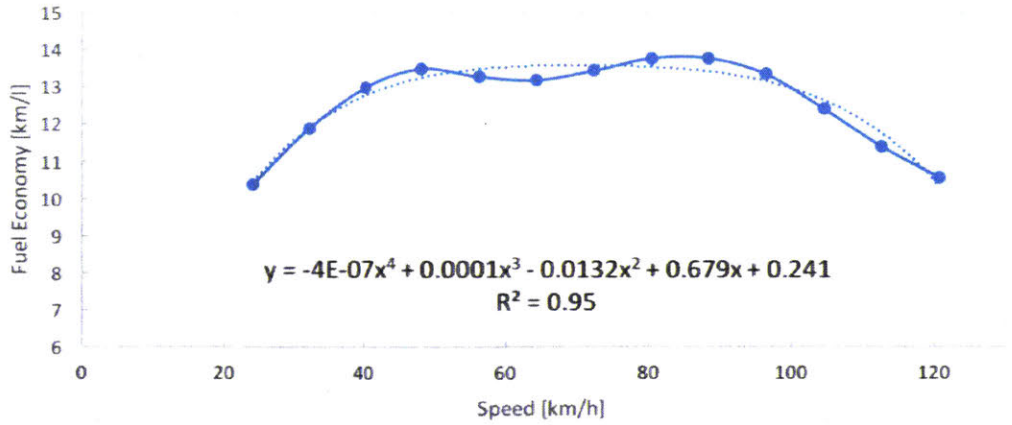

Figure 2-4: Department of Energy Fuel Efficiency vs Average Speed Curve (52)

was recreated. Using these ranges and the distribution of fuel economies found in the EPA report, we successfully recreated the distribution of fuel economies (shown in in Figure 2-5b) with the StreetSmart model

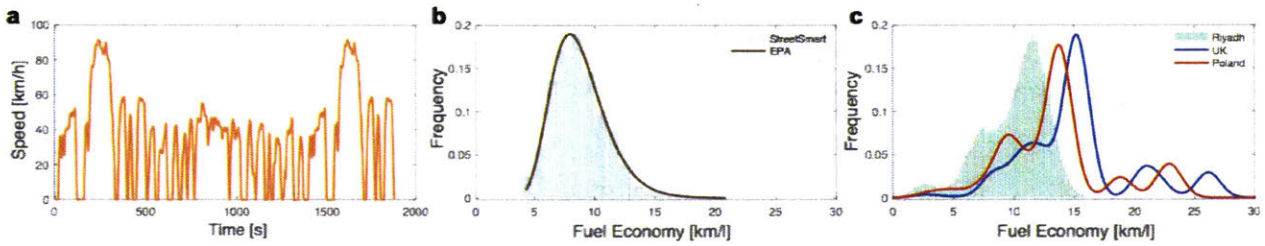

Figure 2-5: (a) FTP-75 EPA's standard speed profile used for calculating the reported inner-city fuel economies of cars. (b) The distribution of fuel economies recreated by the StreetSmart model shows the same distribution as that of the reported fuel economies. (c) The distribution of fuel economies based on Riyadh's fleet of cars compared to those of Poland and the UK shows that the distributions are similar but shifted from one another. The car fleet of Riyadh is less fuel efficient than that of Poland which is less than that of the UK

For verification, the energy index ranges for each bin are then used to recreate the fuel economy distribution of the fleet of cars in Riyadh using the car makes and models from motor vehicle crash statistics data provided by the city of Riyadh. Data on car crashes from January 2013 until October 2015 in the city of Riyadh were used as a proxy for Riyadh's fleet composition. The different energy indices, which produce different fuel consumption rates, are combined in proportion to Riyadh's fleet bin distribution. For validation, the fuel economy distribution of Riyadh's fleet was compared to two cities in Europe chosen on the basis of similar population size or Gross Domestic Product and the results show similar trends

Table 2.2: Results of the Calibration of the StreetSmart Model. Ranges of k_i Parameters for Each Bin of Fuel Economy

Bin	FE Range [MPG]	FE Range [km/l]	k_1	k_2	k_3	k_4
1	[10 - 12)	[4.25 - 5.10)	37.0	30 - 34	1 - 4.8	2000 - 2300
2	[12 - 14)	[5.10 - 5.95)	34.6	23 - 32	1 - 4.8	1300 - 2300
3	[14 - 16)	[5.95 - 6.80)	31.9	21 - 26	1 - 4.8	1100 - 1900
4	[16 - 18)	[6.80 - 7.65)	29.5	17.5 - 24	1 - 4.8	1000 - 1600
5	[18 - 20)	[7.65 - 8.50)	26.9	15 - 22	1 - 4.8	1000 - 1250
6	[20 - 22)	[8.50 - 9.35)	24.3	13 - 18	1 - 4.8	980 - 1250
7	[22 - 24)	[9.35 - 10.20)	21.7	12 - 16	1 - 4.8	850 - 1200
8	[24 - 26)	[10.20 - 11.05)	19.0	12 - 15	1 - 4.8	750 - 1050
9	[26 - 28)	[11.05 - 11.90)	16.3	11 - 14	1 - 4.8	780 - 900
10	[28 - 30)	[11.90 - 12.75)	14.0	10.5 - 12.5	1 - 4.8	710 - 900
11	>30	>12.75	5.0	5 - 14.5	1 - 4.8	500 - 1000
12 (Bus)	6.3	2.68	30.0 - 37.0	30 - 75	1 - 4.8	2000 - 8000
13 (Truck)	17.27	7.34	29.0 - 30.0	12 - 27	1 - 4.8	500 - 2200
14 (Motorcycle)	43.5	18.49	5.0	6 - 10	1 - 4.8	500 - 600

exist in the relative variety of fuel economies in all cities. The data on fleet compositions of the two European countries were acquired from an in-depth study of the fleets of all European countries [31]. As can be seen in Figure 2-5c, the comparison in the distributions of fleet fuel economies between the three areas indicates that the car fleet of Riyadh is less energy efficient those of Poland and the UK. The usage of country level fleet composition for Poland and the UK compared to city level for Riyadh represents a limitation in this comparison but the results adequately verify the credibility of the fleet composition used in this study. With the relative fuel economies of Riyadh's fleet and the calibrated StreetSmart model, we discuss next how we integrate speed profiles from GPS data to estimate fuel consumption at the urban scale.

Chapter 3

Fuel Consumption Application for Traffic Congestion Policies

In many oil producing countries with substantial fuel subsidies, a fall in oil revenue and increasing domestic consumption has put increasing strain on government budgets (53). Countries in the Gulf Cooperation Council, including Saudi Arabia and the UAE, have launched programs to reduce government expenditure on energy subsidies (54). In Saudi Arabia, energy subsidies are estimated at 9.3% of GDP, with 1.4% for petroleum subsidies alone. Decreasing energy subsides can be achieved in a number of ways but congestion relief offers a simple and direct path to lower fuel consumption. As the burden of fuel subsidies continues to grow, it has become increasingly important for these countries to find simple and accurate methods to quantify the effects of policies on congestion relief and fuel consumption in cities. Recent technological advances in collecting and analyzing big data offer a potential method to measure how policy changes impact fuel consumption. With the advent of ubiquitous sensing devices, Transport Network Companies (TNCs), and new methods of estimating flow, we propose a method to answer such questions that can be further applied anywhere in the world and extended to model emissions and air pollution.

Call Detail Records (CDRs) produced passively by mobile phones represent a cutting edge method to estimate travel demand. Most traffic studies currently use local and national household travel surveys to estimate the rate of trip production between different zones of the city but such surveys are expensive to conduct and only cover a small sample of

the population. Leveraging on previous work (12, 55, 56), CDRs can provide simple and effective methods for estimating Origin-Destination (OD) flows using location data collected from millions of individual mobile phone users.

This chapter is structured as follows. In section 2.1, we describe the cleaning of the GPS data and the extraction of speed and acceleration profiles on each street in specific time windows in a typical week to represent different snapshots of traffic throughout the road network of Riyadh. In section 2.2 we describe the application of the model combined with flow data to visualize the fuel consumption across different time periods. Moreover, we present the results of the targeted and random flow reductions and their effects on fuel consumption via the presented model vs. baseline estimates. Finally, in Section 3 we discuss the results and the conclusions derived from the study.

3.1 Methodology

3.1.1 From GPS Data to Speed Profiles

We extracted speed profiles from a large dataset of GPS tracking points of taxi trips from a local Saudi Arabian TNC company over the period of May 2015 until December 2016. Speed profiles are the result of both driving styles and traffic conditions. Driving styles of taxis may be different from those of personal vehicles, affecting the results in a small extent, but we assume speed profiles are mainly a reflection of traffic conditions which would affect taxis and personal vehicle trips similarly. The dataset included trip duration and length, pick up and drop off times, and a chronologically ordered list of GPS coordinates. To ensure that the traffic is representative of year-round conditions we compared the rate of trip production during Ramadan of 2015 and 2016 with non-Ramadan trip production rates. We found that Ramadan trip production rates are much fewer so their impact on the average traffic speed profile for a specific street is negligible. For this reason, we kept the Ramadan trips in the analysis to benefit from the higher amount of data on the street level. A graph of the average number of trips per hour during Ramadan, non-Ramadan, and combined can be seen in Figure 3-1.

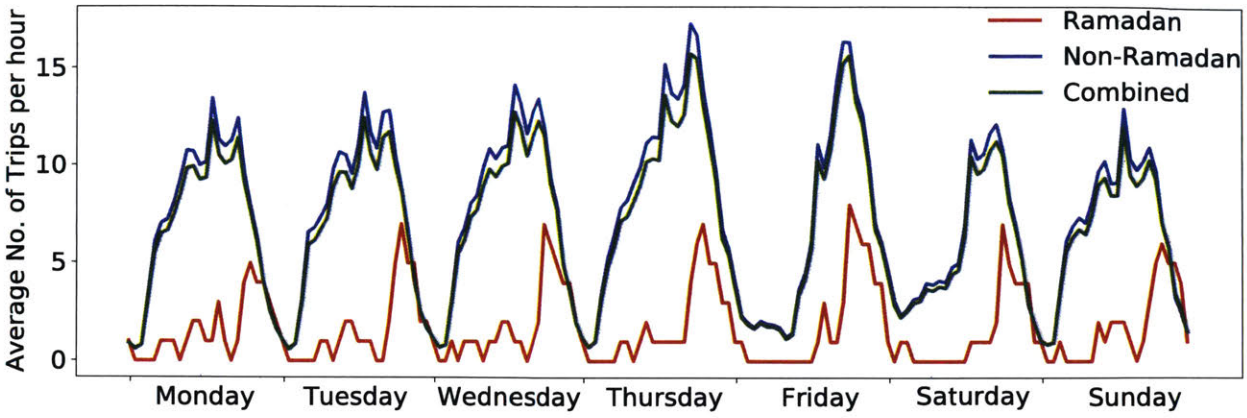

Figure 3-1: Average hourly taxi trip production rates in Riyadh in Ramadan, non-Ramadan, and combined.

Before use, the data was filtered to remove trips that were outside Riyadh and the GPS routes were cleaned and modified to correct for measurement errors. We detected errors in the GPS points and fixed them using the following algorithm. Since the GPS coordinates were given as an ordered list without a timestamp, and are known to be collected at regular intervals, the interval or frequency of recording was calculated as the total duration of the trip divided by the number of points recorded. Errors were detected as spikes in speed that are greater than 160 km/h since the taxi's fleet and the road conditions typically preclude speeds above that. Two different causes of error were detected and repaired using different methods. In the first case, errors were caused by missing points due to a lack of network signal. This would result in a spike in speed for one segment but not the following segment, which would revert to realistic speeds and GPS points. The number of skipped points was estimated from the average of the speed before and after the single speed spike. Second, for errors caused by a GPS point that is in an obviously anomalous location, the speed spike occurs in two simultaneous segments, one to jump to the wrong location and another to return to the realistic location. This error was fixed by removing the erroneous point. This simple method was able to adequately correct the GPS coordinates.

We also cleaned the taxi data using an algorithm developed by Jiang et al. (57) to detect long periods of immobility, or stays, that can be interpreted as parking in our dataset. This may occur if a client asks to keep the meter running while they finish an errand. The reason for the splitting is not to affect the recorded speeds on the streets where

the taxi is effectively parked. We chose the minimum thresholds for idling time and distance by inspecting the distribution of stay durations throughout the week. A minimum value of $2,200s$ (around 37 min) was chosen to allow for the majority of traffic stays during peak and off-peak hours. A maximum distance for a trip to be split around a stay was 10 m. This number reflects the relatively high precision of the GPS points. In other words, a trip that remained within a 10 m radius for longer than 37 min was split into two trips, removing the 37 min or longer section of immobility.

After filtering and cleaning GPS routes for the city of Riyadh, we obtained nearly $43,000$ trips that were analyzed by a custom mapping algorithm to assign full GPS routes to edges in the road network. For more accurate mapping, longer edges in the road network were split into shorter segments to ensure that nodes are no more than 10 m apart. The mapping algorithm was implemented using the following procedure:

1. For each point i in the GPS trajectory, we identify the set of nodes (N_i) in the road network that fall within a 25 m radius.

2. We constructed a path network G consisting of the nodes N_i.

3. For each point i in the GPS trajectory, we used the Dijkstra algorithm to find the fastest route from every node in N_i to every node in N_{i+1}. For each route, we added a representative edge to G with the route's total travel time as the weight.

4. For each edge, we added a time penalty based on the distance of the target node to the original GPS coordinate at a rate of 1 second per additional meter past the closest node.

5. Any gaps in G were identified to determine contiguous sequences of paths that represent segments of potential routes.

6. For each contiguous sequence, we identified the fastest path in G as the most likely route taken by the vehicle.

For verification purposes, trip distances, free flow and observed travel times as well as fuel economy estimates are plotted in Figure 3-2. Fuel economy is defined as the distance

traveled per liter of fuel consumption. In Figure 3-2a, we verified that the reported distances were generally consistent with the sum of the distances calculated between every two consecutive GPS points in each trip using the Haversine formula. In Figure 3-2b the free flow times, computed as the sum of the free flow times of every matched segment in each trip is compared to the observed flow times as reported in the taxi data. The comparison only considered the observed total times from trips that were successfully matched with streets. The figure shows consistent results with observed travel times during the morning peak hour being slower than the free flow times of each trip. As a baseline comparison to using speed profiles from taxi data, we calculate fuel economy with the StreetSmart model assuming a constant speed and one fuel economy bin based on the Hyundai Elantra, the most common car according to the accident data (Figure 3-2c). We plotted the results if speed profiles are used with only one fuel economy bin and if all bins are used in proportion to the fleet of Riyadh from the crash statistics data. It shows a very close distribution to the simpler assumption that all cars are the most common car fuel economy, and in high contrast to the result given by not using the speed profiles. Incorporating speed profiles in the model results in higher fuel consumption, or much lower fuel economy, which is more accurate and important for policy projections. This constitutes the core benefits of integrating the more accurate fuel consumption model.

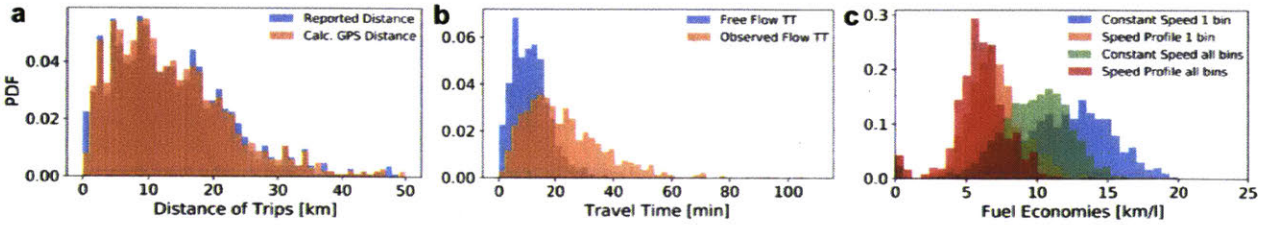

Figure 3-2: Data Verification figures using trips in the morning peak time period of weekdays from 8 - 9 AM. (a) Histogram of Reported and calculated Trip Distances. (b) Histogram of free flow travel time and Observed travel time in matched trips. (c) Histogram of Fuel economies using constant speed, speed profiles, and 1 bin and all bins.

3.1.2 Fuel Consumption Results

With the StreetSmart model calibrated and speed profiles extracted from GPS data, we can now estimate fuel consumption for three typical time periods representing distinct traffic

conditions. The time periods are morning peak (8 - 9 AM), midday off-peak (12 - 13 PM), and evening peak (17 - 18 PM) during weekdays. For every edge in the road network, we calculated fuel consumption per car based on each speed profile matched to that edge. The estimates were computed for each bin of fuel economy in the StreetSmart model. For each edge, the average fuel consumption rate per car was multiplied by the flow of cars per hour as computed by a version of the Iterative Traffic Assignment (ITA) algorithm to get a fuel consumption rate per hour, with car volumes included.

The flow was calculated from OD matrices previously derived from CDR data in Riyadh by (12), in cars/hour over morning, midday, and evening time periods. Following (12), a factor of 1.5 was applied to the average morning flow of the morning and evening time periods to determine peak hour demand. The congested time estimates correlated successfully with the ones acquired from Google Maps (see comparisons in (56)).

Fuel Consumption Rate

As presented in (56), the ITA algorithm assigns trips in a series of four increments. It incrementally assigns 40%, 30%, 20% and 10% of the OD flows to the fastest routes in the road network. After each iteration, the congested time of each edge is updated so that the effects of congestion can be factored into the assignment. The resulting flow volume is multiplied by the fuel consumption rate per car to arrive at hourly fuel consumption as shown in Equation 3.1 which is normalized by edge length.

$$FC = \frac{flow_e[\frac{car}{hr}] \times fcr_e[\frac{liter}{car}]}{L_e[m]}, \tag{3.1}$$

where FC is the rate of fuel consumption in $[\frac{liter}{m.hr}]$, $flow_e$ is the flow on edge e as estimated by the assignment algorithm, fcr_e is the rate of fuel consumption per car on edge e as estimated by our application of the StreetSmart model, and L_e is the length of the edge in meters.

The results of the model are shown in the choropleth maps in Figure 3-3 below. For the fuel consumption rate per street, we used a weighted average of fuel consumption by bin proportional to Riyadh's fleet. The maps show that the most fuel consuming streets are the grid highways of the city. As expected, the flow of cars and total fuel consumption per

meter of road is found to be higher in the peak periods of the morning and evening than the midday off-peak. The evening peak shows slightly higher fuel consumption than the morning peak, indicated by more red streets. We used quantile breaks on the fuel consumption rate values of all time periods combined to display the streets that are the most fuel consuming. A high fuel consumption rate per meter of road is due to high fuel consumption rate per car and high car flow values. We observe that King Fahd Road is by far the most fuel consuming road in the city, especially in the area bounded by the old city center from the south and the Northern Ring road from the north.

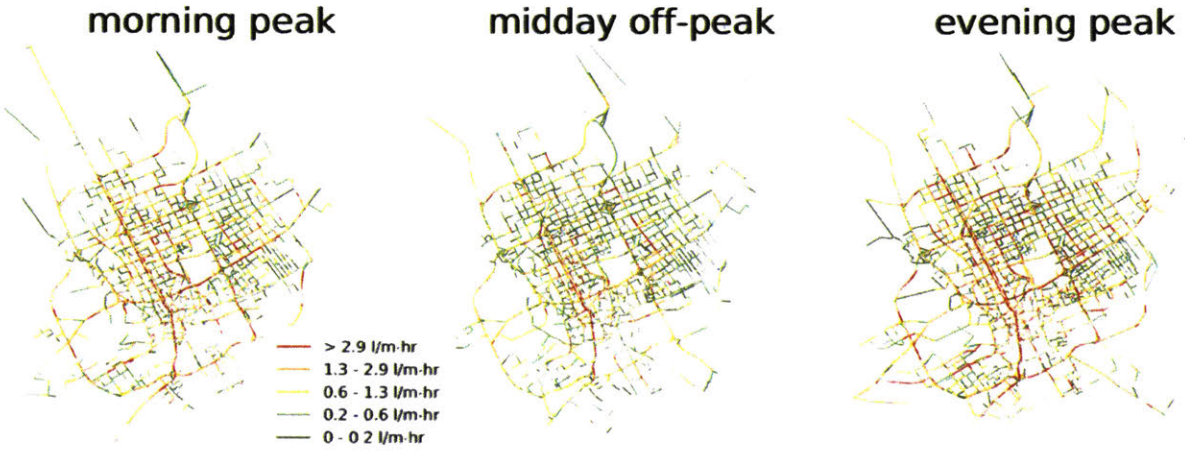

Figure 3-3: Choropleth Maps of fuel consumption rates [Liter/meter.hour] by the StreetSmart model on streets matched with GPS data for typical time periods morning peak (8 - 9 AM) weekdays, midday off-peak (12 - 13 PM) weekdays, evening peak (17 - 18 PM) weekdays.

The taxi GPS data necessary to calculate a fuel consumption on each street did not cover the entire network of streets. Similarly, the traffic assignment of ODs did not use all of the city's streets and the majority of roads used were also matched with at least one speed profile from the taxi GPS data. Specifically, 56%, 57%, and 63% of streets were covered for the morning,midday, and evening time periods respectively.

Relevance of integrating GPS data

For the morning time period, a simulation of all OD trips in the city is made and the path, defined as the sequence of edges taken from origin node to destination node, is defined using the shortest path algorithm as in the ITA algorithm. In the simulation, OD pairs with flow values that are less than one car/hour are omitted to ensure each flow represents a discrete

trip. We assign a car fuel economy bin at random to each trip in proportion to the probability of that bin in Riyadh's fleet. Using the StreetSmart model, we estimate fuel consumption and trip time as the sum of the their values on each edge in the trip's path. For verification, total trip times derived via the demand model which were previously verified against Google Maps estimates (56) are plotted against the trip times derived via the GPS data. As shown in Figure 3-4a, travel times via GPS data are generally higher than their counterparts from the simulation but their correlation is satisfactory.

Two sample speed profiles used by the StreetSmart model are plotted with the constant speed assumed by the baseline comparison. As expected, not all trip times are equivalent to their speed profile counterparts and the constant speed is generally lower than the peaks of the speed profile. Extracted speed profiles do not always end at 0 km/h speed since the originally matched trips on these edges may not be stopping at that edge. This overestimation of speed at the end of a trip represents a negligible increase in fuel consumption estimation. The lower constant speed assumption used in the baseline comparison results in fuel consumption estimates that are consistently lower than those from using speed profiles from GPS data. Thus, the effect of the accelerations and deceleration on the StreetSmart model are observed to be significant and not negligible which shows the benefit of using GPS data in the fuel consumption model. The distributions of fuel consumption per trip using speed profiles and constant speed are shown in Figure 3-4c along with the fuel economies. It is clear that the acceleration and detailed speed profiles are not negligible in the estimates. These results bring to the urban scale the results of Figure 3-4c.

The effect of reducing flows on overall fuel consumption is shown in figure 3-4d. Trips were removed from the simulation described above via three rankings and the resulting fuel saving potential is shown, normalized by the total fuel consumption of each method. Random trip reduction results in a perfectly linear fuel saving effect. In contrast, we see the best case scenario, where trip reduction of the worst fuel consuming trips per meter are ranked. As expected, targeted reduction results in higher fuel savings with the same number of reduced trips. More importantly, when comparing the speed profile estimates (method proposed here) vs. constant speed fuel estimates (baseline) differences emerge. Constant speed targeting shows a higher return on fuel savings because the variance of the

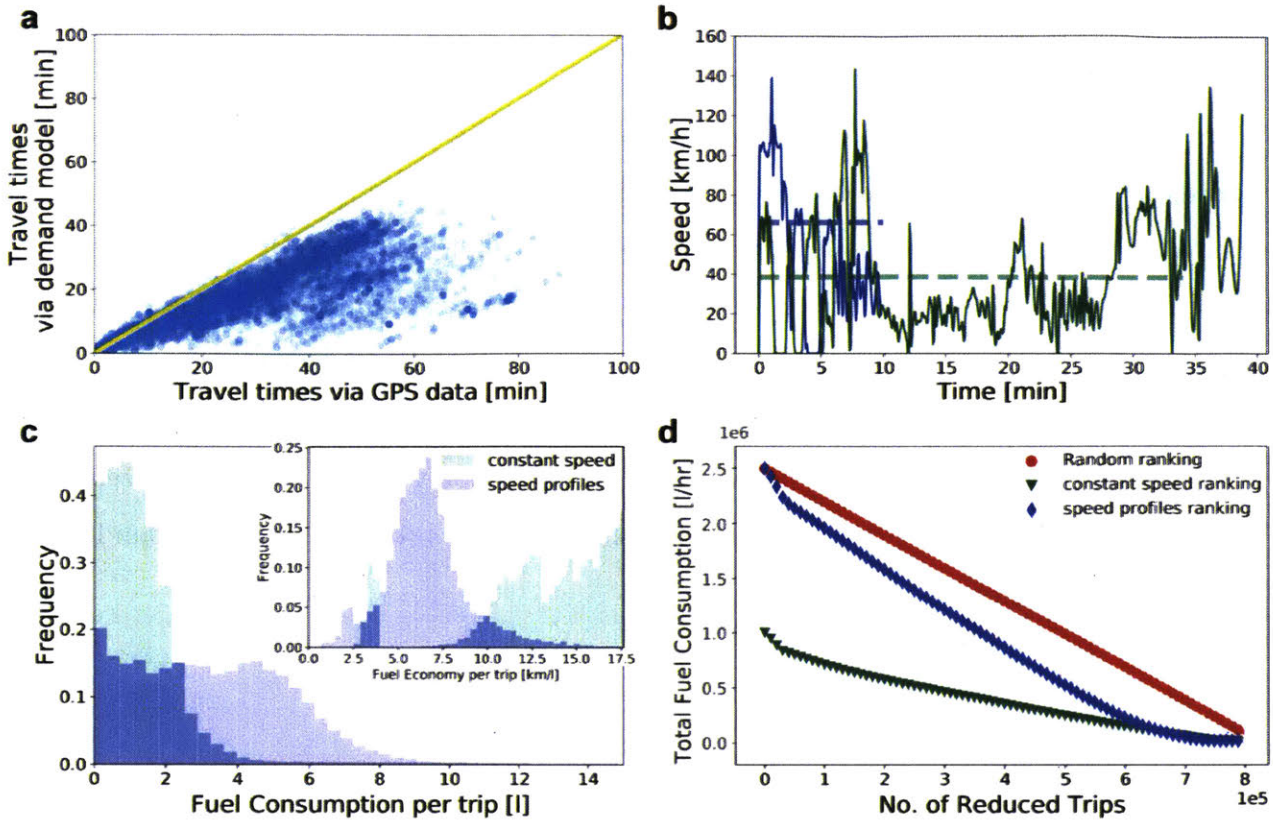

Figure 3-4: Fuel consumption estimates at urban scale. (a) Comparison of the travel time of the routes in the constant speed model via travel demand vs. the input used in our method using GPS data (b) Sample speed profiles in two routes used to estimate fuel consumption overlaid with the constant speed used for comparison (c) Estimates of fuel consumption and fuel economy in the morning peak via our method (speed profiles) and the base line method (constant speed) (d) Random and targeted fuel savings vs. number of reduced trips.

distribution of fuel consumption estimates using constant speed is less than that of the speed profiles. In other words, the fuel economy distribution of the constant speed estimates shows higher proportions at the high end of the distribution than the high end of the speed profile distribution shown in Figure 3-4c. This would explain the higher fuel savings observed when the trips are ranked by constant speed fuel consumption per meter. However, since the constant speed estimates are not as accurate as those derived using speed profiles, the fuel savings are spurious.

The real gain in using speed profiles to estimate fuel consumption over the baseline constant speed assumption is in the city-wide total fuel consumption per hour estimate. The constant speed assumption underestimates the city-wide fuel consumption for the morning

hour by 60% compared to using the speed profiles.

The relative gain in fuel saving potential of targeting the highest fuel consuming trips over random trip reduction is approximately 10% if 14.5% of trips are reduced. In other words, if 14.5% of trips are reduced, the policy targeting the highest fuel consuming trips per m would save 10% more fuel for every morning peak hour. When 14.5% of targeted trips are reduced, 25% of fuel consumption for the morning peak hour is saved. These results are encouraging but the ratio is not as high as the effect of similar targeting policies on household energy consumption (58), where the increase over random is 51%. These differences can be explained by the more normal distribution of fuel economy of the trips which renders the effects of targeting to be much less dramatic than the more broad distribution found in house energy consumption.

3.2 Discussion and Conclusion

We present a data fusion method to estimate fuel consumption at the urban scale. We leverage a travel demand model that uses mobile phone data and integrate speed profiles from taxi GPS data that covered most of the street network of Riyadh. To identify fleet distribution, we used car crash statistics data as a proxy with the assumption that the distribution of car makes and models is representative of Riyadh's car fleet distribution. The method developed here and the calibration of the StreetSmart model for fuel consumption can easily be extended to any other region. It is significantly faster than if speed profiles were simulated by software which would require a large amount of time and computing resources (59).

The fuel consumption model was verified to produce better results than than the DOE fuel economy by speed graph and tested on real OBD-II fuel consumption measurements before being calibrated to be used on any car fuel economy. The resulting calibration results are presented in Table 2.2. We used the calibrated StreetSmart model with speed profiles from GPS data to produce estimates which were compared with the baseline without GPS data and assuming constant speed. The results show that the differences of using speed profiles is significant, justifying the introduction of the more elaborate model in policy

estimates.

As a proof of concept, we applied the calibrated StreetSmart model to test a policy of flow reduction, both random and targeted to the least fuel efficient trips and simulated the effect on the rate of fuel consumption in the city. We showed that the difference in fuel consumption reduction between targeted and random schemes was around 10% more fuel savings for 14.5% trip reduction. Interestingly, 25% of city-wide fuel savings potential can be achieved by removing only 14.5% of trips ranked by the worst fuel consuming trips per meter .

While this project has demonstrated the potential of data-driven models to estimate the effects of policies on fuel consumption, it can benefit from further study to understand the impacts of several simplifications and assumptions. There are three main areas which require further research:

1. To correctly assess the benefit of our straightforward approximation over the computationally costly alternative, speed profiles of each car across every origin-destination trip should be simulated and the results compared.

2. The GPS data used in our model covered most but not all of the street network. A more accurate fuel consumption calculation can be achieved with more data, covering a higher proportion of flow over a longer time period.

3. The trip assignment method used is an efficient and reasonably effective method to estimate overall congestion based on a simplified method of route choice. However, the baseline estimates would benefit from trip assignment derived from a dynamic traffic assignment model.

We hope the method used to obtain the results of the fuel reduction strategies can be implemented for other fuel or emissions reduction strategies, assuming the conditions remain relatively unchanged. For example, the composition of the Riyadh fleet remains similar and that streets remain unchanged. For future implementations, the composition of the fleet can easily be adjusted since the method uses bins to account for all car types. However, a significant change in the road network layout or the traffic conditions would require new

datasets (Road network and GPS routes for speed profile extraction) to recompute modular road-by-road fuel consumption estimates for each bin of car type.

Current fuel consumption models are adapting to the wealth of data available from ubiquitous sensing devices such as smart-phones. Our project aims to show the relative gain in accuracy of incorporating both fuel economy distribution considerations and speed profiles derived, in our case, from GPS devices used by a local taxi company. The method developed in this paper can be adapted to also measure emissions. It can be further augmented to analyze the economic and environmental impacts of policies targeting specific trips by conducting a network analysis to identify the affected trips. The tools developed in this work have the potential to assess the consequences of a variety of policies under different circumstances and in any region of the world.

Chapter 4

Fuel Effects of Women Adopting Driving in Riyadh

4.1 Introduction

In a recent policy change, it was announced that women will be allowed to drive in Saudi Arabia in June 2018 (60). This change will certainly have significant effects on traffic, fuel, and emissions in the middle eastern kingdom. In 2017, women made up just over 16% of the labor force (61). The trend since the early 1990s shows a 3-fold increase from just over 5.4% in 1992 to current levels. This trend is expected to continue going forward with the government's commitment to easing social constraints and increasing the participation of women in the labor force to 30% by 2030. With 70% of the Saudi Population under the age of 30 and many having studied abroad, pressure is mounting for them to enjoy some of the social freedoms taken for granted outside the kingdom.

A major strategy outlined in Vision 2030, a general reform plan set forth by Mohamed Bin Salman (MBS), the country's young crown prince, is to decrease Saudi Arabia's heavy financial reliance on oil. In 2018, subsidized retail gas prices in Saudi Arabia were almost doubled from 0.75 to 1.37 riyals per liter (from 0.76 USD to 1.38 USD per gallon) as part of the ongoing effort to promote more efficient use of the government's resources. Using a tool previously developed at MIT for estimating fuel consumption, we attempt to gauge the effect on fuel consumption of allowing women to drive in Riyadh, Saudi Arabia's capital city.

To our knowledge no other projects have attempted to quantify the adoption rates of women driving and its consequences on fuel consumption at the urban scale.

Until June 2018, male drivers, whether family or hired, must accompany all commute trips made by women to work or university. We differentiate between essential trips made by women, with both the women and the driver in the car, from empty return trips made by the drivers alone. We combine Call Detail Records (CDR)-based travel demand modeling with a model of how drivers arrive in time for a woman's essential trips to produce trip flows. This enables us to simulate empty driver as well as essential trips for a subset of the users whose gender has been identified and compute their fuel consumption. No other work has recorded the adoption of driving by a population with previously no access to driving. We model the adoption of driving by women on an agent-based Bass diffusion curve (35) and assume that once driving is adopted by a woman, all non-essential empty driver trips can be avoided and thus estimate the possible savings of fuel over time. Our method does not take into account the latent increased potential for trips that may be made once the women is free to make trips without a male driver. This would need surveys and a stated preference travel demand model which is a possible next step.

The structure of the chapter is as follows. Section 4.2 describes the datasets used in our model. Section 4.3 describes our methods for trip generation and for simulating empty driver trips which compliment essential trips made by women as well as verification of the resulting trip flows. Section 4.44 contains a description of the fuel consumption calculation and adoption model for driving and its repercussions. Finally, section 4.5 will discuss policy conclusions and next steps.

4.2 Data Description: CDR and Gender labeled Users

4.2.1 Data description

CDRs are extremely valuable in producing travel demand estimates. Following the tools developed in (2, 56, 57), we used the extracted travel demand from CDRs from a mobile telephone company in Riyadh over a period of one month in December 2012. Each CDR

record contains an anonymous user ID, coordinates, and timestamp for every instance of a phone call or SMS. The coordinates recorded are those of the nearest cell tower used. Fuel consumption was computed following the method described in Chapter 3, using GPS speed profiles from a local Transportation Networking Company (TNC) and a model that was calibrated for use on cars of any fuel efficiency. We also derived the composition of fuel efficiencies of Riyadh's car fleet using as a proxy the motor vehicle crash statistics from January 2013 to October 2015 provided by the city of Riyadh. These data were combined to predict the fuel consumption in Riyadh during a typical weekday. This project was aided by the identification of a subset of women in the CDR population thus enabling a simulation of the additional empty driver trips needed to accommodate the mobility needs of women hitherto banned from driving in the kingdom.

4.2.2 Gender Labeling

The gender of 20,000 CDR users (10,000 women and 10,000 men) were labeled by manually identifying the points of interest (POIs) that are restricted to only women or men as described in detail in (62). Since the work done in (62), an expanded set of POIs that are restricted to women only, like Princess Noura University (PNU) during class hours, have yielded more users identified as females. Other locations include female/male only universities, a football stadium, and female-only shopping malls in Riyadh city. Additional measures were taken to ensure that the users identified were of women and not male residents in the PNU dormitories. Only cell towers near the lecture buildings were used and users who stayed overnight or during non-lecture times were filtered out since they could be custodial staff or faculty. Finally, we identify the gender of the users based on the number of distinct gender-specific locations visited. The resulting CDR users were confidently identified to be female and were used in the rest of the project to simulate empty driver trips for our fuel consumption estimations.

4.2.3 Expansion Factors

Figure 4-1 shows the distribution of the home locations of the CDR users in the whole CDR dataset and the gender labelled subset. Following (13), *home* and *work* locations were

identified for each user as the stays with the most visits during 7 pm - 8 am and 8 am - 7 pm respectively on weekdays. The weekday in 2012 in Saudi Arabia started on Saturday and ended on Wednesday. Since our dataset does not include every resident of the city, the residents that were observed in each zone were used to represent the whole residents of that zone by expanding them by a factor equal to the zone's total population to the observed subset. Expansion factors for each Traffic Analysis Zone (TAZ) for the CDR data set are shown in Figure 4-1a, c, and d. in Figure 4-1a, a map shows the distribution of the population across the city. A low expansion factor means the ratio of users in our dataset to the residents of the TAZ, as indicated in the census, is low. A high expansion factor means there are few users in our dataset compared to the residents of the TAZ. In comparison with the expansion factors of the gender-labeled subset of users whose expansion factors have a median of 200 to 300, the whole dataset is satisfactorily representative of the population in Riyadh with a median expansion factor of 17. There is no statistical significance to expanding the small gender labeled subset since it would not be representative of the entire population.

4.3 Trip Generation and Fuel Consumption Methods

4.3.1 Trip Generation and flows

Travel demand between the TAZs was computed by following the methods outlined in (13). Users with fewer than one per week home stays were filtered out of the trip generation to ensure the CDR data adequately represents a user's travel patterns. Similarly, users who visited their potential work locations less than once a week were not assigned a work and those locations were labelled as other. In effect, 36% and 53% of male and female users respectively were found to have a work location which reflects the reality that not all citizens commute to work or school on a daily basis.

With the user's home and work locations identified, a trip was simulated between every two consecutive stay locations within an effective day (from 3 am to the following 3 am). Trip departure time was randomly sampled from the distribution found in the US National Household Travel Survey (NHTS 2009 (63)) corresponding to the day and purpose

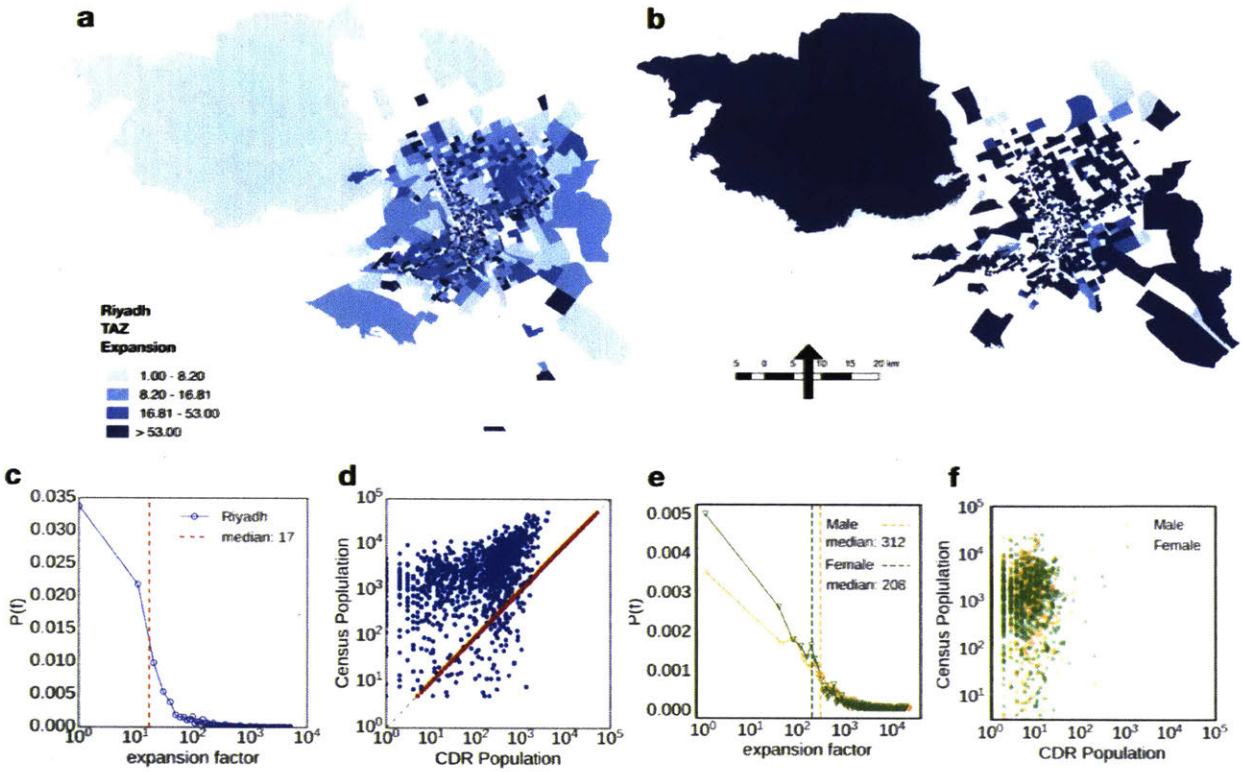

Figure 4-1: Expansion Factors of CDR users, (a) Choropleth map of expansion factors for all CDR users, (b) Choropleth map of expansion factors for gender-labeled CDR users, (c) distribution of expansion factorsfor all CDR users, (d) scatter of CDR population vs census population for every TAZ in blue and after expansion in red, (e) ditribution of expansion factors for gender-labeled CDR users, (f) scatter of CDR population vs census population for every TAZ by gender

of the trip. NHTS of the USA was used since no such data exists for Saudi Arabia. Finally, if a user did not record a stay at her/his home at the beginning (or end) of an effective day, her/his first (or last) trip was started (or ended) at the home location.

The distribution of trips by purpose and gender is shown in a Joyplot in 4-2a. For Home Based Work (HBW) trips, females have more pronounced extremes than do males whose midday is a gradual increase towards the evening peak. Both female and male HBW trips in Riyadh have a slightly later start time and earlier end time than the US NHTS which suggests a shorter work period than the US. Home based Other (HBO) trips are almost identical for females and males and generally have a morning peak only, with almost

no HBO trips made in the evening. Non-Home Based (NHB) trips represent the compliment of the HBO trips with a gradual distribution in the evening and night and none in the morning.

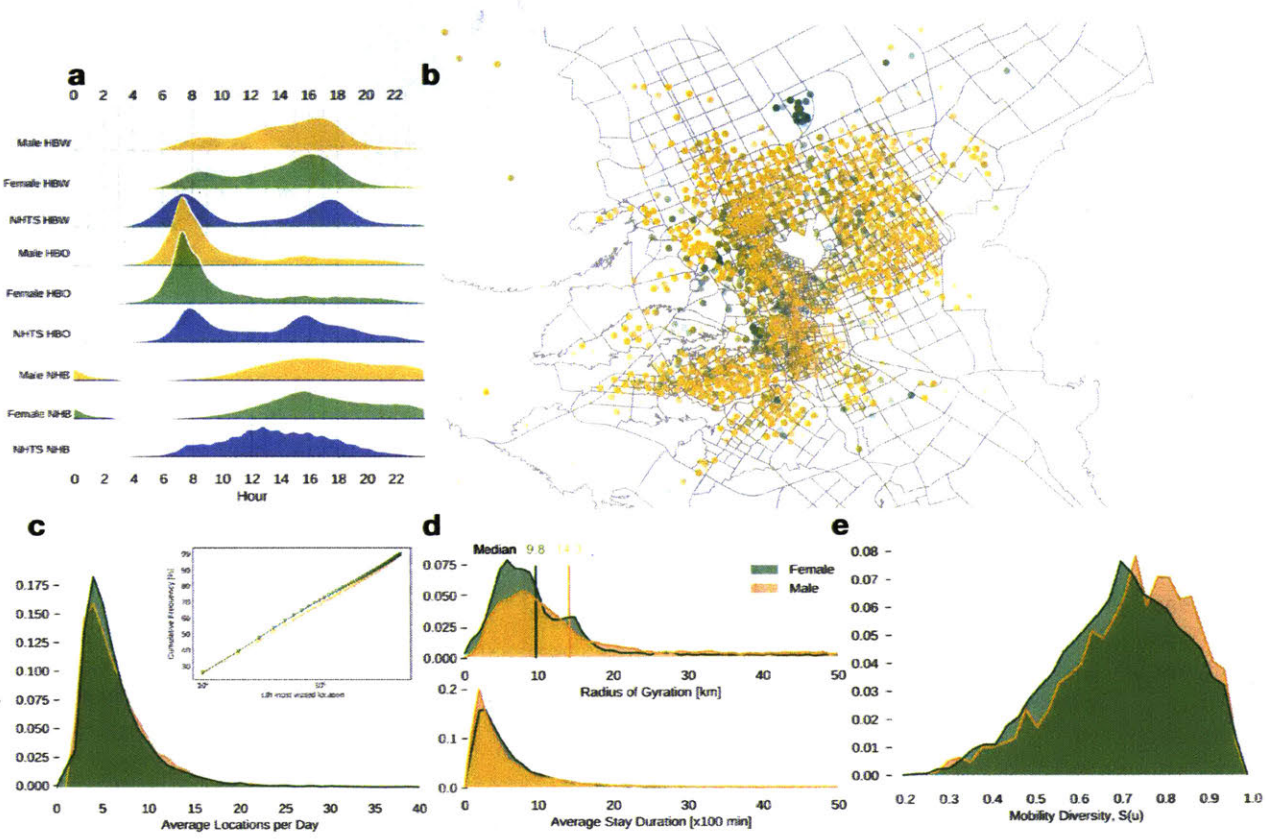

Figure 4-2: Features of Mobility by Gender, (a) a Joyplot showing the trip departure time distribution for male and female users by trip purpose HBW,HBO,an NHB compared to the distribution from the NHTS, (b) A map of the home and work locations of users colored by gender, (c) outer: Average locations visited per day by gender, inner: Lth most visited locations by gender, (d) Radius of Gyration and Average Stay Duration distributions by gender, (e) Mobility Diversity distribution by gender

Human mobility is characterized by frequent return to previously visited locations (2). This is confirmed in our dataset as shown in 4-2c inner plot, about 60% of stays occurring in the 5 most visited locations and 70% in the top 10 locations. Several features of the mobility patterns of males and females were also computed and their distributions shown in Figure 4-2. The features that showed very similar results between the genders were the Lth most visited location, average locations per day, and average stay duration. Other features showed a noticeable difference between the genders, like Mobility Diversity and more significantly

Table 4.1: Percent of Trips by Purpose and Time of Day Compared to NHTS and MHTS

	HBW [%]	HBO [%]	NHB [%]	Morning [%]	Midday [%]	Evening [%]	Rest-of-Day [%]
Male CDR	1	42	57	24	16	26	34
Female CDR	2	41	57	24	16	28	33
MHTS	12	49	39	21	34	33	13
NHTS	14	55	30	19	37	31	13

the radius of gyration (RoG). The median RoG of females is less than that of males, at 9.8 and 14.3 km respectively. This implies that female users do no venture as far from their set of visited locations as do male users. The Shannon entropy mobility diversity, defined by 4.1, was also computed for male and female users following Pappalardo et al. (64).

$$S(u) = -\frac{\sum_{e \in E} p(e) \log p(e)}{\log N} \tag{4.1}$$

where $e = (o, d)$ represents a trip between an origin phone tower and a destination phone tower, E is the set of all the possible OD pairs, $p(e)$ is the probability of observing a movement between o and d, and N is the total number of trajectories of individual u.

For verification, we compare the percentage of trips by purpose and time of day to the Massachusetts Household Travel Survey (MHTS) conducted in 2010/2011 and the National Household Travel Survey conducted in 2009. We find that the Home Based Other (HBO) of the MHTS and NHTS compare well with the observed HBO in our gender-labeled subset in Riyadh. The main departure is in the Home Based Work (HBW) trips in our gender-labeled population. This may be a result of the lack of identified work places in the subset used. The identification process used an all women university and a male-only stadium during the month of December. This bias in the identified population may lead to a smaller proportion of daily work/study commute trips, especially in December when universities usually have a semester break. As for the time of day of the trips, more trips were made in the night time in Riyadh and in the early morning. This may be explained by the difference in weather between Riyadh and the US. Riyadh's middle eastern climate means the most direct sunlight and hottest time of day is the midday and early evening hours.

The flows resulting from the trip generation described above cannot be expanded to represent the entire population of Riyadh for several reasons. First the numbers of identified users are too small to be a representative sample of the almost 6 million inhabitants of

Riyadh. The expansion factors, as shown in Figure 1e are on average 200 to 300. Such an expansion risks overestimating the trip-making tendencies based on a few select users. Another reason is that the gender-labeled subset produces about 3 times more flows in total compared to 10 random samples of the same number of phone users. This is because the method used to identify gender ensured that the users had a high usage frequency. When compared to randomly drawn users, the difference in trip generation emphasized how the gender-labeled sub-sample is not representative of the entire population. The gender specific POIs chosen may tend to be biased towards a young adult population with less children and older citizens. This bias should be considered when analyzing the results which may be skewed in favor of young adults who are highly educated, in the female case, and interested in sports in the male case.

4.3.2 Driver Trip Simulation

According to a survey conducted in 2017 by IPSOS, 47% of women are driven by a family member and another 21% have a household driver in their employ. The rest may take taxis, TNCs, or commute by riding their institution's private buses. So a fact of life for around 68% of Saudi Arabian women is that a male driver must drive them for every car trip they take. For our simulation of this reality we have made several basic assumptions based on knowledge of the local cultural norms.

As depicted in Figure 4-3a, a driver must make empty trips around the essential trips performed by a woman. The driver in these scenarios can be any male whether employed or related to the woman whose trips are being made. Before the woman's first trip in an effective day, the driver commutes to her home location from a randomly assigned driver home location across the city. Since we are only interested in the effects on fuel consumption of the empty trips made by the driver without the women, we consider the women's actual trips to be essential and thus not considered as a potential source of fuel savings. The driver then waits at the women's location unless her stay at that location is longer than a cut off time limit set to 2 hours. If the woman's stay is indeed longer than 2 hours, he returns to the women's home location immediately after dropping her off. In reality, this may be the necessary for the driver to make other household trips like run errands or drive another

female member of the same household. The same routine happens for trips of any purpose and any time of day. Finally, after the last trip of the effective day, the driver returns to the same randomly assigned driver home location and returns for the next essential trip by the women, not necessarily on the following day.

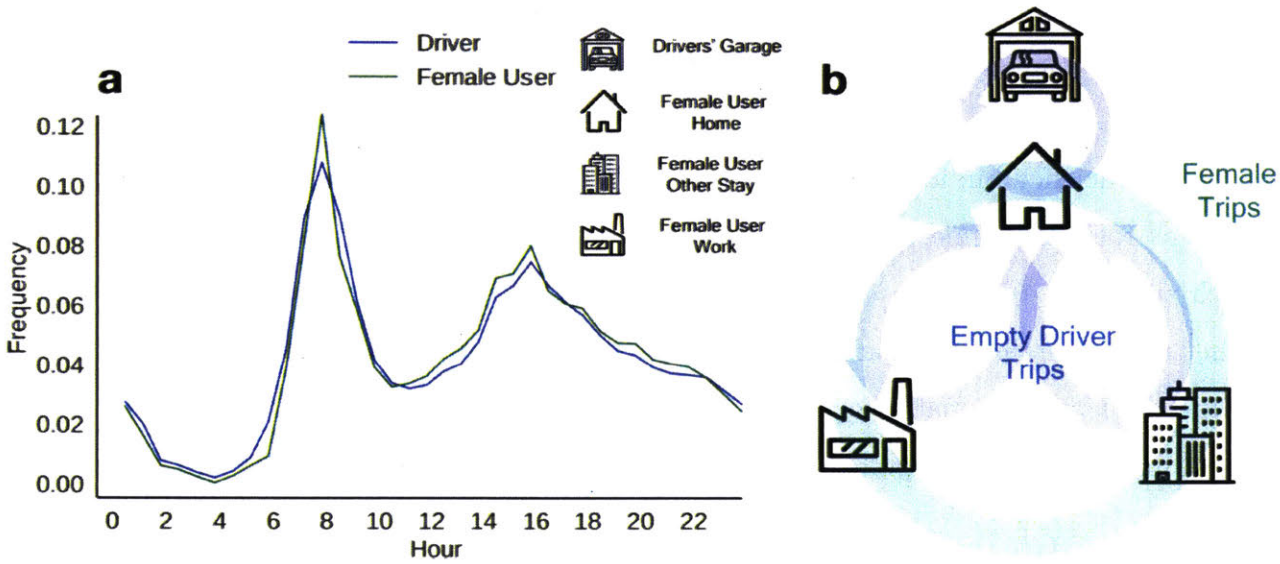

Figure 4-3: Driver Trip Simulation, (a) Departure time distribution of empty driver and essential female trips, (b) diagram of the relation between empty driver and essential female trips

4.4 Calculating Fuel Consumption

4.4.1 Fuel Consumption Model

The Fuel estimation framework can be summed up as follows. Travel demand for the entire city was calculated by expanding the flow observed by the CDR user population. The routes taken for each trip were computed using an Iterative Traffic Assignment (ITA) algorithm that staggered the assignment of cars iteratively by 40%, 30%, 20% and finally 10% and updated the congested time on each street between the iterations (56). Speed profiles for each street detected in the TNC dataset was recorded and fuel consumption was calculated for each street using the calibrated StreetSmart model. Every street matched with a speed profile received an attribute of fuel consumption per meter for every bin of fuel efficiency representing the Riyadh fleet of cars as observed in the crash statistics provided by the city.

An improvement on the previous framework developed in Chapter 3 is the now comprehensive fuel consumption rate per meter on every street in the network of Riyadh even on streets not previously matched with a speed profile. For those streets without a speed profile, the congested time of the street was used to calculate a constant speed which in turn was used in the StreetSmart model to compute fuel consumption for every bin of fuel efficiency (65).

The results of the ITA were used to simulate a trip for every flow value greater than one. Each simulated trip had a path attribute as a list of streets followed. It was assigned a car fuel efficiency bin following the distribution of Riyadh's car fleet. Finally trip properties such as total trip time, distance, fuel consumption, and fuel efficiency were computed for each simulated trip. The distributions of each property are shown in Figure 4-4a and b. The trip times, distances, and fuel consumption follow an expected distribution with mean values that are reasonable for a city of Riyadh's size. The distribution of fuel efficiency of the trips is a similar shape to the fuel efficiency of the city's fleet as shown in 3-2.

4.4.2 Verification of Aggregate Fuel Results

The fuel consumption was calculated using travel demand that was expanded to represent the entire population of Riyadh based on census population values from 2012. Streets that were not matched with GPS data in the dataset were filled with fuel rates assuming a constant speed with time equal to the congested time output by the ITA algorithm. The resulting totals of fuel, reported in 4.2, allegedly represent the urban fuel consumption of Riyadh over the course of one typical weekday. To verify that these totals are in agreement with reality, we compared them with several different sources of transportation fuel consumption in Riyadh. The computed total is 12.72 million liters (80,036 US Barrels) per day. According to the Energy Information Administration and the Saudi Arabian Monetary Agency, the transport sector in 2013 consumed 1.3 million barrels per day (66). That means our estimate for Riyadh represents 6.2% of the reported consumption for the country. Given that our method only models personal vehicle trips and does not account for either other modes like freight and airplanes or other cities in the country, the estimate of 6.2% is well within reason. The ITA results are verified against the recorded TNC travel times on the streets that were matched as shown in 4-4c.

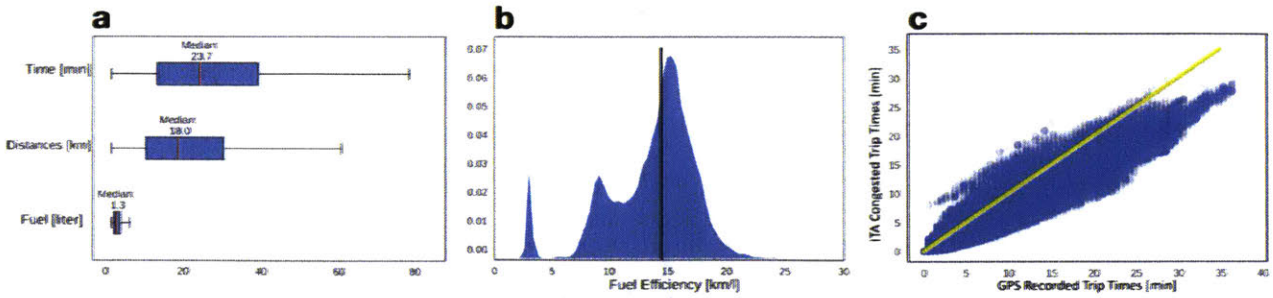

Figure 4-4: Morining Trip Simulation Verification Diagrams, (a) Boxplots of trip simulation times [min], distances [km], and fuel consumption [liter] with median shown, (b) distribution of fuel efficiency of each trip, (c) scatter of GPS-recorded total trip time vs ITA congested trip times

Table 4.2: Fuel Consumption Estimation Totals for Male and Female and Empty Driver Trips

Fuel [Liters]	Morning	Midday	Evening	Rest of Day
Female Essential	94,879	52,372	29,748	59,432
Empty Driver	166,414	69,511	43,711	88,733
Male	68,095	41,600	21,655	47,144
Total Gender labeled	329,389	163,484	95,116	195,309
Total Expanded	3,750,546	3,528,630	2,955,463	2,490,059

If we compare the calculated fuel to Motor Gasoline consumption from the US Energy Information Administration, we find that our estimate for Riyadh alone is very close to the measured estimates for counties in the region with comparable population sizes. The 12.72 million liters gives 1.86 liter per day per capita which is very close to the reported consumption of Kuwait, Bahrain, Qatar, and UAE (2.31, 1.62, 1.83, 1.86 respectively) and reasonably close to those of Singapore and Luxembourg (0.55 and 2.18). This assures us that our estimate for fuel consumption for Riyadh for the entire population is reasonably accurate and comparable to measured and reported truths. Since the same technique was used for the subset of gender-labeled users and the simulated Empty driver trips associated with the women, the fuel savings potential can be computed with confidence. The results of fuel consumption are broken down by time of day and trip occupant in 4.2.

4.4.3 Model of Adoption of Driving by Women

In order to predict the time in which these fuel savings may occur, we must first model the time it will take for women to accept driving and integrate this new phenomenon into their routines. Adoption models have been used to measure and predict markets and the spread of disease for the greater part of the 20th century. For our adoption model, we chose the Bass diffusion curve, developed in 1969 by Frank Bass, which models the adoption or sale of a new technology stochastically on the basis of two simultaneous processes. The first is innovation whereby a portion of the susceptible population spontaneously decides to adopt due to advertising or originality, and the second is imitation whereby some users decide to adopt after a certain proportion of the population adopts the technology first. This curve is used widely in the industry and the scientific community to model social adoption of phenomena as well as market penetration of new products (67). The network of female users, as represented in our dataset of CDRs, gives us an added component that can enhance the modelling capabilities of the simple Bass model.

From the phone call records of the female identified subset of users we constructed an undirected, unweighted graph of the women. Since the women were identified as such by their frequent and consistent presence at an all-women's university, it is no surprise that we found that 90.2% of the women were part of one connected component as shown in 4-5a. The colors chosen for the visualization are the result of a community detection modularity algorithm and the node sizes represent the betweenness centrality of each node. The knowledge of their relationships as observed by their phone calls enables us to use an infection algorithm, such as Susceptible-Infected-Recovered (SIR) model (68), to model the spread of the adoption of driving as the spread of a disease or a social phenomenon.

The difference between the Bass model, an aggregate system-level model, and the SIR model, an agent-based model (ABM), is the level at which the stochastic process of conversion is executed. Whereas in the Bass model, aggregate ratios of infected to susceptible nodes are used to calculate the number of newly adopted people at every time step, in the agent based model where a node's neighbors are known, the ratio used is that of infected neighbors to all neighbors. For our use, we extended the traditional SIR model, called the SIRa model, to

include a spontaneous infection parameter as described in (68). The SIRa model requires 3 parameters, α is the innovation parameter, β is the imitation parameter, and gamma is the recovery parameter (which is set to 0 for our implementation). The parameters p and q of the Bass model are not directly transferable to the parameters of the SIRa model. For this reason, in order to simulate the Bass curve with SIRa model on our network, a conversion was necessary following (69).

The choice of p and q, and subsequently α and β, follows several scenarios to represent both the values found in the literature as well as recent surveys conducted by several media firms in Saudi Arabia. The Bass curve was fitted to electric car sales in a recent study that extensively reviewed parameter choices in the literature pertaining to car sales (41). Car sales are used as a proxy to licensing and adoption of driving by women since they were the closest proxy found in the literature and carried similar weight with respect to lifestyle decisions albeit they are not as culturally sensitive as the issue of women driving in Saudi Arabia. The original estimates of the study by Massiani and Gohs found that yearly calculated adoption p values ranged from 0.000013 to 0.00322 and q values from 0.2319 to 1.2513. The surveys conducted by IPSOS and PwC, shown in 4-5b as points, also suggest the rate of women who intend to start driving in the short-term future. Using these two sources, four scenarios were used to represent the varying adoption speeds by the subset of women in our dataset. These range from ambitiously fast in Scenario 1 to modestly fast in Scenario 4.

Results of Adoption and Fuel Consumption

Under the most ambitious adoption scenario, the rate of innovation exceeds imitation, and the diffusion does not follow the traditional S-curve. Moreover, the IPSOS survey claimed that 24% of respondents intended to drive immediately, perhaps indicating they already drove outside the country previously. In 4-5b, the proportion of women in our subset who have adopted driving is related on the right-side y-axis to the associated potential of fuel savings, taking into account that only 68% of women incur empty driver trips. The maximum potential of fuel consumption is liters and is achieved at different times in each scenario. It ranges from 2 years to 14 years. The surveys we used as a starting point for the adop-

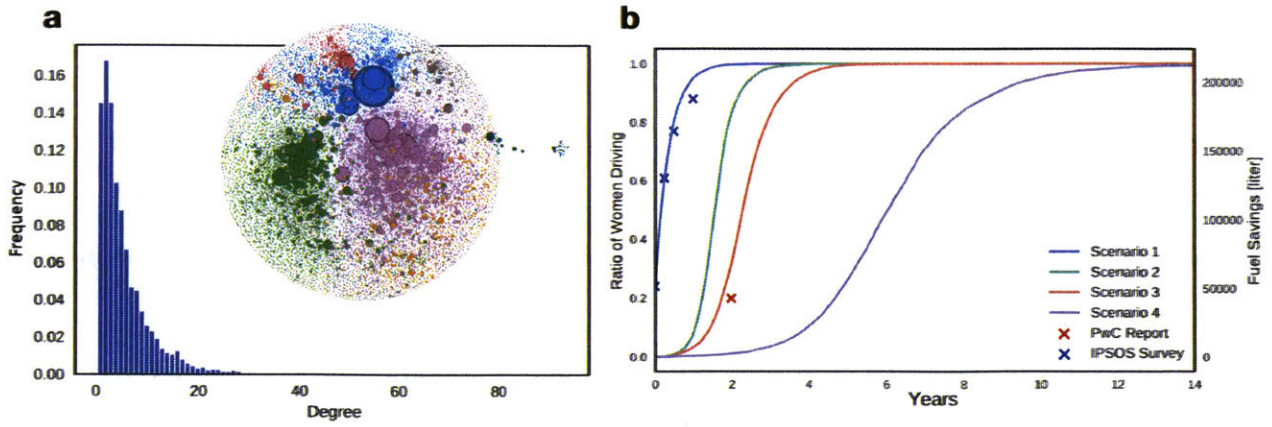

Figure 4-5: Network and Adoption Scenario Results, (a) degree distribution of the female gender-labeled CDR users network of communications and the largest component in the network visualized, (b) Adoption Scenarios and their associated Ratio of Women driving and fuel savings potential

tion scenarios were not rigorously administered and thus their representation of the average Riyadh woman is questionable. The more realistic scenario would lie somewhere in between the options modeled here.

4.5 Discussion and Conclusion

4.5.1 Limitations and Assumptions

These empty driver trips represent an unnecessary waste of fuel which can be a source of economic benefit both for the employers and the country's subsidized fuel program. The fuel consumption associated with each trip is computed following the method developed and verified in Chapter 3. The main assumption in our current framework is that the empty driver trips simulated in the section above will be cut once a woman has the right and ability to drive herself and her associated driver trip costs can be saved. There are several limitations in our method. First, the framework assumes that women will not make more trips than those currently made with a male driver and assumes no latent travel demand exists for women in Saudi Arabia. Second, we assume that once a woman is able and willing to drive, the driver's Empty Trips are removed once and for all with no reverting to pre-driving habits. Third, the effect on traffic flow and consequently driving speed profiles are not taken into account on an urban scale. These simplifications enable us to calculate the

fuel savings potential of women driving in Saudi Arabia but the nuances of the actual fuel savings are dependent on more factors than we can model at this time with the information at hand.

The first assumption of no latent travel demand is uncertain since opinion surveys have shown that women intend to work more once this policy change liberates them from their current dependence on a male driver. 55% of the respondent of the IPSOS survey think that traffic will get worse once the ban is lifted. In such a conservative society where women rarely have access to public space, a car's mobility and privacy may offer an attractive alternative to the limited and highly controlled environments found both at home and at commercial spaces. This private space may lead to unnecessarily long trips and an increase in travel demand that is not captured by our model. To incorporate such latent travel demand, a discrete choice model can be made with the results of stated preference surveys.

The second assumption is an optimistic result of a change in attitudes and culture. As fuel prices continue to increase and women start to make their own trips, hired driver's will in the long run be a less economical and attractive option. Once hired drivers are let go, the burden of rehiring them would inhibit the old empty driver habits from returning.

Finally, the third simplification is a consequence of our dataset and the framework developed to estimate traffic in Chapter 3 which uses a static dataset of GPS points to calculate acceleration and idle time on each street. To account for the differences in traffic after the lifting of the ban would require new GPS datasets not available at the time of this project.

Moreover, the flow estimated for the entire city of Riyadh as an output of the expansion of the observed users to represent Riyadh's population's flow was not changed since our gender-labeled subset is not representative of the entire female population. This is because it is effectively too small - 10k women and 10k men - to be representative of the whole population of Riyadh. Despite these simplifications our results are built on sound assumption, stated preferences, and verified against reported urban fuel estimations and travel demand values.

4.5.2 Policy Recomendations

After considering the limitations of the current framework, it is still useful to glean insight on the behavior of women and the associated economic effects of lifting the driving ban. According to our results, empty drivers incur from 132 to 176% of additional fuel consumption over essential female trips. The potential of savings here is significant and the speed with which it can be realized depends mainly on the adoption scenario of the women.

Our estimates show that the maximum fuel savings is likely to be achieved before 14 years if the conservative adoption scenario is followed. This ultimately depends on the innovation and imitation of other women in cutting empty driver trips. For imitation, we have seen that the our subset of women were very highly connected with only less than 10% not in communication with the rest of the female population. This indicates that imitation will occur effectively. What policy makers can focus on is increasing the innovation parameter. This can be achieved by simplifying the process and encouraging the population to decrease empty driver trips through advertising campaigns and positive messages in the media showing the drawbacks of empty driver trips. On the other hand, it is important to use this opportunity to encourage car pooling and mass transit among women to avoid a rise of extra trips that can cause high traffic and undo the benefits we foresee.

Chapter 5

Conclusion

Big Data is being collected from a constantly growing market of smart phone users, both passively and actively and in more creative ways every year. This treasure of data does not have to mean an end to privacy but a boon for data-driven policies and solutions. We have shown how different sources of data can be used to produce meaningful results in transportation planning; From a faster, more representative and more robust travel demand model, to the simulation of policies for traffic congestion relief and women driving in Saudi Arabia. The applications are as varied as the sources of data and the questions that are relevant at the moment. One thing is certain, however, the importance of incorporating more data into our models will only increase as the potential improvement in prediction increases as well.

5.1 Summary

The project showed the calibration of a robust fuel consumption model that uses speed profiles for use on a range of fuel economies. The resulting parameters are shown in Chapter 2 and were applied on the fleet composition of Riyadh to compute total urban fuel consumption in Chapters 3 and 4. This was aided by several sources of data such as Call Detail Records and GPS traces of a local taxi company. The CDRs were used first to estimate travel demand and to ultimately get car flow along each street of the city. They were also used after gender identification to build a network of women by their communications. The GPS points were

cleaned and matched to streets along the city and the resulting speed profiles per street were used to get fuel consumption per street.

The framework was applied successfully to test several traffic relief policy scenarios. The results showed that the efficacy of targeting the most fuel consuming trips, shown to be on the high speed peripheral highways as well as stop-and-go inner roads, yielded double the fuel savings compared to random reduction. Another application of the fuel framework was the modeling of empty driver trips around the identified subset of female CDRs. The drivers generated trips that would be rendered unnecessary when women begin to adopt driving in Saudi Arabia in the summer of 2018. The speed with which these fuel savings can be achieved was found by modeling the adoption of driving by women on an agent-based Bass diffusion curve. The findings suggest that, with the right encouragement and policies, the maximum savings can be achieved within 14 years. Our methods made several assumptions and thus have limitations that will be summarized in the next section.

5.2 Limitations of the Framework

First, the GPS data used to acquire speed profiles on the streets did not cover all of the streets of Riyadh. Therefore, our solution was to use the same fuel consumption model on those streets but revert to the average speed on those streets as computed by an iterative assignment algorithm (ITA). Furthermore, the ITA itself was not a dynamic equilibrium algorithm, but rather a static assignment with 3 updates between staggered assignment intervals. Our method can benefit from an improvement on this traffic assignment algorithm.

Second, the method used to calculate the fuel savings potential of women driving in Saudi Arabia only calculates the possible reduction of trips from removed empty driver trips. It does not model the potential increased demand in trip making on the part of women. It must be noted that the question was never *if* fuel will be saved when women drive, rather *how much* will be saved and *when*. The specific assumptions of this application are summed up in Chapter 4.5.1.

Third, our method is simple and re-usable as long as the conditions of the city remain relatively unchanged. For example, the composition of the Riyadh fleet remains similar and

that streets remain unchanged. For future implementations, the composition of the fleet can easily be adjusted since the method uses bins to account for all car types. However, a significant change in the road network or the traffic conditions would require new datasets (Road network and GPS routes for speed profile extraction) to re-compute modular road-by-road fuel consumption estimates for each bin of car type.

5.3 Future Work

We hope the method used to obtain the results of the fuel reduction strategies can be implemented for other pollutants or emissions reduction strategies. It can be further augmented to analyze the economic and environmental impacts of policies targeting specific trips by conducting a network analysis to identify the affected trips. The simulated trips give a granular picture of distances, travel times, and fuel consumption. This granularity can be used in discrete choice models as well as individual user analysis. For example, as in our case of identifying women and men and modeling scenarios of their choices, another subset of users can be identified and their individual trip properties modeled to test relevant policies. The tools developed in this work have the potential to assess the consequences of a variety of policies under different circumstances and in any region of the world.

Finally, as we encourage a healthy scientific replication of our results, we have made our code available on Github at `www.github.com/adhamkalila/riyadhfuelconsump`. The datasets used are protected against distribution so the code published was cleaned from any raw data and only results necessary for recreating figures were pickled and shared.

Bibliography

[1] Shan Jiang, Yingxiang Yang, Siddharth Gupta, Daniele Veneziano, Shounak Athavale, and Marta C González. The timegeo modeling framework for urban mobility without travel surveys. *Proceedings of the National Academy of Sciences*, 113(37):E5370–E5378, 2016.

[2] Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. Understanding individual human mobility patterns. *nature*, 453(7196):779, 2008.

[3] Yasuo Asakura and Eiji Hato. Tracking survey for individual travel behaviour using mobile communication instruments. *Transportation Research Part C: Emerging Technologies*, 12(3-4):273–291, 2004.

[4] N Caceres, JP Wideberg, and FG Benitez. Deriving origin–destination data from a mobile phone network. *IET Intelligent Transport Systems*, 1(1):15–26, 2007.

[5] Yu Nie, HM Zhang, and WW Recker. Inferring origin–destination trip matrices with a decoupled gls path flow estimator. *Transportation Research Part B: Methodological*, 39 (6):497–518, 2005.

[6] Md Shahadat Iqbal, Charisma F Choudhury, Pu Wang, and Marta C González. Development of origin–destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies*, 40:63–74, 2014.

[7] Pu Wang, Timothy Hunter, Alexandre M Bayen, Katja Schechtner, and Marta C González. Understanding road usage patterns in urban areas. *Scientific reports*, 2: 1001, 2012.

[8] Hillel Bar-Gera. Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: A case study from israel. *Transportation Research Part C: Emerging Technologies*, 15(6):380–391, 2007.

[9] Xianyuan Zhan, Samiul Hasan, Satish V Ukkusuri, and Camille Kamga. Urban link travel time estimation using large-scale taxi data with partial information. *Transportation Research Part C: Emerging Technologies*, 33:37–49, 2013.

[10] Jonathan Reades, Francesco Calabrese, and Carlo Ratti. Eigenplaces: analysing cities using the space–time structure of the mobile phone network. *Environment and Planning B: Planning and Design*, 36(5):824–836, 2009.

[11] Santi Phithakkitnukoon, Teerayut Horanont, Giusy Di Lorenzo, Ryosuke Shibasaki, and Carlo Ratti. Activity-aware map: Identifying human daily activity pattern using mobile phone data. In *International Workshop on Human Behavior Understanding*, pages 14–25. Springer, 2010.

[12] Jameson L Toole, Serdar Colak, Bradley Sturt, Lauren P Alexander, Alexandre Evsukoff, and Marta C González. The path most traveled: Travel demand estimation using big data resources. *Transportation Research Part C: Emerging Technologies*, 58: 162–177, 2015.

[13] Lauren Alexander, Shan Jiang, Mikel Murga, and Marta C González. Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation research part c: emerging technologies*, 58:240–250, 2015.

[14] Trinh Minh Tri Do, Jan Blom, and Daniel Gatica-Perez. Smartphone usage in the wild: a large-scale analysis of applications and context. In *Proceedings of the 13th international conference on multimodal interfaces*, pages 353–360. ACM, 2011.

[15] Audrey De Nazelle, Edmund Seto, David Donaire-Gonzalez, Michelle Mendez, Jaume Matamala, Mark J Nieuwenhuijsen, and Michael Jerrett. Improving estimates of air pollution exposure through ubiquitous sensing technologies. *Environmental Pollution*, 176:92–99, 2013.

[16] Wei Lei, Hui Chen, and Lin Lu. Microscopic emission and fuel consumption modeling for light-duty vehicles using portable emission measurement system data. *World Academy of Science, Engineering and Technology*, 66:918–925, 2010.

[17] Luc Pelkmans, Patrick Debal, Tom Hood, Günther Hauser, and Maria-Rosa Delgado. Development of a simulation tool to calculate fuel consumption and emissions of vehicles operating in dynamic conditions. Technical report, SAE Technical Paper, 2004.

[18] Guohua Song, Lei Yu, and Ziqianli Wang. Aggregate fuel consumption model of light-duty vehicles for evaluating effectiveness of traffic management strategies on fuels. *Journal of Transportation Engineering*, 135(9):611–618, 2009.

[19] Afonso Vilaça, Ana Aguiar, and Carlos Soares. *Estimating Fuel Consumption from GPS Data*, pages 672–682. Springer International Publishing, Cham, 2015. ISBN 978-3-319-19390-8. doi: 10.1007/978-3-319-19390-8_75. URL `http://dx.doi.org/10.1007/978-3-319-19390-8_75`.

[20] Eva Ericsson, Hanna Larsson, and Karin Brundell-Freij. Optimizing route choice for lowest fuel consumption–potential effects of a new driver support tool. *Transportation Research Part C: Emerging Technologies*, 14(6):369–383, 2006.

[21] Daniel B Work and Alexandre M Bayen. Impacts of the mobile internet on transportation cyberphysical systems: traffic monitoring using smartphones.

[22] Daniel B Work, Olli-Pekka Tossavainen, Quinn Jacobson, and Alexandre M Bayen. Lagrangian sensing: traffic estimation with mobile devices. In *American Control Conference, 2009. ACC'09.*, pages 1536–1543. IEEE, 2009.

[23] Brian Donovan and Daniel B Work. Using coarse gps data to quantify city-scale transportation system resilience to extreme events. *arXiv preprint arXiv:1507.06011*, 2015.

[24] Juan C Herrera, Daniel B Work, Ryan Herring, Xuegang Jeff Ban, Quinn Jacobson, and Alexandre M Bayen. Evaluation of traffic data obtained via gps-enabled mobile phones: The mobile century field experiment. *Transportation Research Part C: Emerging Technologies*, 18(4):568–583, 2010.

[25] Pierre-Emmanuel Mazaré, Olli-Pekka Tossavainen, Alexandre Bayen, and D Work. Trade-offs between inductive loops and gps probe vehicles for travel time estimation: A mobile century case study. 2012.

[26] Darrell P Bowyer, Rahmi Akçelik, and DC Biggs. *Guide to fuel consumption analyses for urban traffic management.* Number 32. 1985.

[27] Robert Joumard, Peter Jost, John Hickman, and Dieter Hassel. Hot passenger car emissions modelling as a function of instantaneous speed and acceleration. *Science of the Total Environment*, 169(1-3):167–174, 1995.

[28] Jose Luis Jimenez-Palacios. Understanding and quantifying motor vehicle emissions with vehicle specific power and tildas remote sensing. *Massachusetts Institute of Technology, Cambridge*, 1998.

[29] Alessandra Cappiello, Ismail Chabini, Edward K Nam, Alessandro Lue, and M Abou Zeid. A statistical model of vehicle emissions and fuel consumption. In *Intelligent Transportation Systems, 2002. Proceedings. The IEEE 5th International Conference on*, pages 801–809. IEEE, 2002.

[30] Hesham Rakha, Kyoungho Ahn, and Antonio Trani. Development of vt-micro model for estimating hot stabilized light duty vehicle and truck emissions. *Transportation Research Part D: Transport and Environment*, 9(1):49–74, 2004.

[31] V. Ribeiro, J. Rodrigues, and A. Aguiar. Mining geographic data for fuel consumption estimation. In *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, pages 124–129, Oct 2013. doi: 10.1109/ITSC.2013.6728221.

[32] Min Zhou, Hui Jin, and Wenshuo Wang. A review of vehicle fuel consumption models to evaluate eco-driving and eco-routing. *Transportation Research Part D: Transport and Environment*, 49:203–218, 2016.

[33] Arghavan Louhghalam, Mehdi Akbarian, and Franz-Josef Ulm. Carbon management of infrastructure performance: Integrated big data analytics and pavement-vehicle-interactions. *Journal of Cleaner Production*, 142:956–964, 2017.

[34] Bryce Ryan and Neal C Gross. The diffusion of hybrid seed corn in two iowa communities. *Rural sociology*, 8(1):15, 1943.

[35] Balázs Lengyel, Riccardo Di Clemente, János Kertész, and Marta C González. Spatial diffusion and churn of social media. *arXiv preprint arXiv:1804.01349*, 2018.

[36] Mark Granovetter. Threshold models of collective behavior. *American journal of sociology*, 83(6):1420–1443, 1978.

[37] TC Schelling. Hockey helmets, daylight sav ing, and other binary choices. s. 213-224 in: ders., micromotives and macrobehavior, 1978.

[38] Thomas W Valente. Social network thresholds in the diffusion of innovations. *Social networks*, 18(1):69–89, 1996.

[39] Duncan J Watts. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766–5771, 2002.

[40] Ian Dyason. Of early adopters, s-curves and the business life cycle, Feb 2015. URL `http://www.strategyinnovationgrowth.com/single-post/2015/02/10/Of-early-adopters-Scurves-and-the-business-life-cycle`.

[41] Jérôme Massiani and Andreas Gohs. The choice of bass model coefficients to forecast diffusion for innovative products: An empirical investigation for new automotive technologies. *Research in transportation economics*, 50:17–28, 2015.

[42] Deepa Chandrasekaran and Gerard J Tellis. A critical review of marketing research on diffusion of new products. In *Review of marketing research*, pages 39–80. Emerald Group Publishing Limited, 2007.

[43] Fareena Sultan, John U Farley, and Donald R Lehmann. A meta-analysis of applications of diffusion models. *Journal of marketing research*, pages 70–77, 1990.

[44] Thomas A Becker, Ikhlaq Sidhu, and Burghardt Tenderich. Electric vehicles in the united states: a new model with forecasts to 2030. *Center for Entrepreneurship and Technology, University of California, Berkeley*, 24, 2009.

[45] M Davidson, D Cross-Call, M Craig, and A Bharatkumar. Assessing options for accommodating electric vehicles in santa delano valley. usaee case competition 2013. tt a. p. *Group*, 49:38, 2013.

[46] Walter McManus and Richard Senter Jr. Market models for predicting phev adoption and diffusion. 2009.

[47] Sang Yong Park, Jong Wook Kim, and Duk Hee Lee. Development of a market penetration forecasting model for hydrogen fuel cell vehicles considering infrastructure and cost reduction effects. *Energy Policy*, 39(6):3307–3315, 2011.

[48] Hirokazu Takada and Dipak Jain. Cross-national analysis of diffusion of consumer durable goods in pacific rim countries. *The journal of marketing*, pages 48–54, 1991.

[49] PJ Lamberson. The diffusion of hybrid electric vehicles. future research directions in sustainable mobility and accessibility. *Sustainable mobility accessibility research and transformation (SMART) at the University of Michigan center for advancing research and solutions for society (CARSS)*, 2008.

[50] Austin Louis Oehlerking. *StreetSmart: modeling vehicle fuel consumption with mobile phone sensor data through a participatory sensing framework.* PhD thesis, Massachusetts Institute of Technology, 2011.

[51] Fangyu Wu, Raphael Stern, Miles Churchill, Maria Laura Delle Monache, Ke Han, Benedetto Piccoli, and Daniel Work. Measuring trajectories and fuel consumption in oscillatory traffic: Experimental results. In *Transportation Research Board 96th Annual Meeting*, page 14p, 2017.

[52] Stacey Davis, Susan William, and Robert Boundy. *4. Light Vehicles and Characteristics*, page 4.30. U.S. Department of Energy, 2016.

[53] Simeon Kerr. Saudi arabia looks to reform energy subsidy programme. *Financial Times*, Nov 2015. URL `https://www.ft.com/content/b9e1d072-893d-11e5-90de-f44762bf9896`.

[54] Oxford Business Group. Saudi arabia sets out strategy to reform subsidies and reduce domestic energy usage. *Oxford Business Group*, Nov 2016. URL `https://www.oxfordbusinessgroup.com/analysis/subsidy-reform-government-sets-out-strategy-reduce-domestic-energy-usage`.

[55] Serdar Çolak, Lauren P Alexander, Bernardo G Alvim, Shomik R Mehndiratta, and Marta C González. Analyzing cell phone location data for urban travel: current methods, limitations, and opportunities. *Transportation Research Record: Journal of the Transportation Research Board*, (2526):126–135, 2015.

[56] Philip S Chodrow, Zeyad Al-Awwad, Shan Jiang, and Marta C González. Demand and congestion in multiplex transportation networks. *PloS one*, 11(9):e0161738, 2016.

[57] Shan Jiang, Gaston A Fiore, Yingxiang Yang, Joseph Ferreira Jr, Emilio Frazzoli, and Marta C González. A review of urban computing for mobile phone traces: current methods, challenges and opportunities. In *Proceedings of the 2nd ACM SIGKDD international workshop on Urban Computing*, page 2. ACM, 2013.

[58] Mohammad Javad Abdolhosseini Qomi, Arash Noshadravan, Jake M Sobstyl, Jameson Toole, Joseph Ferreira, Roland J-M Pellenq, Franz-Josef Ulm, and Marta C Gonzalez. Data analytics for simplifying thermal efficiency planning in cities. *Journal of The Royal Society Interface*, 13(117):20150971, 2016.

[59] Nurhan Cetin, Adrian Burri, and Kai Nagel. A large-scale agent-based traffic microsimulation based on queue model. In *IN PROCEEDINGS OF SWISS TRANSPORT RESEARCH CONFERENCE (STRC), MONTE VERITA, CH.* Citeseer, 2003.

[60] Ben Hubbard. Saudi arabia agrees to let women drive, Sep 2017. URL https://www.nytimes.com/2017/09/26/world/middleeast/saudi-arabia-women-drive.html.

[61] United Nations Development Program. Kingdom of saudi arabia millenium development goals. Technical report, Ministry of Economy and Planing, 2013.

[62] Abdullah Almaatouq Fahad Ahasoun Anas Alfaris Marta C. Gonzalez Riyadh Alnasser, Faisal Aleissa. Blue vs. pink: Gender based characterization of communication behavior. An extended abstract is available at http://netmob.org/www15/assets/img/netmob15_book_of_abstracts_posters.pdf.

[63] Federal Highway Administration U.S. Department of Transportation. 2009 national household travel survey. URL http://nhts.ornl.gov.

[64] L. Pappalardo, D. Pedreschi, Z. Smoreda, and F. Giannotti. Using big data to study the link between human mobility and socio-economic development. In *2015 IEEE International Conference on Big Data (Big Data)*, pages 871–878, Oct 2015. doi: 10.1109/BigData.2015.7363835.

[65] Austin Louis Oehlerking. Streetsmart: Modeling vehicle fuel consumption with mobile phone sensor data through a participatory sensing framework. Master's thesis, Massachusetts Institute of Technology, Department of Mechanical Engineering, 2011.

[66] Khalid Alkhathlan and Muhammad Javid. Carbon emissions and oil consumption in saudi arabia. *Renewable and Sustainable Energy Reviews*, 48:105–111, 2015.

[67] Zuhaimy Ismail and Noratikah Abu. A study on new product demand forecasting based on bass diffusion model. *Journal of Mathematics and Statistics*, 9(2):84, 2013.

[68] Alison L Hill, David G Rand, Martin A Nowak, and Nicholas A Christakis. Emotions as infectious diseases in a large social network: the sisa model. *Proceedings of the Royal Society of London B: Biological Sciences*, 277(1701):3827–3835, 2010.

[69] Yu Xiao, Jing Han, Zhouping Li, and Ziyi Wang. A fast method for agent-based model fitting of aggregate-level diffusion data. 2017.