Measuring Travel Equity and Representativeness: Opportunities and Challenges of Using Smartphone-based Travel Survey in Dar es Salaam, Tanzania

By

Menghan Li

B.E., Urban Planning and B.A., Economics
Peking University
Beijing, China (2014)

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

Master in City Planning
and
Master of Science in Transportation
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 2017

Signature redacted

Author____
Department of Urban Studies and Planning
Department of Civil and Environmental Engineering
May 24, 2017

Signature redacted

Certified by ____
Associate Professor P. Christopher Zegras
Department of Urban Studies and Planning
Thesis Supervisor

Signature redacted

Accepted by____
Associate Professor P. Christopher Zegras
Chair, MCP Committee
Department of Urban Studies and Planning

Signature redacted

Accepted by____
Professor Jesse Kroll
Chair, Graduate Program Committee
Department of Civil and Environmental Engineering

Measuring Travel Equity and Representativeness: Opportunities and Challenges of Using Smartphone-based Travel Survey in Dar es Salaam, Tanzania

By

Menghan Li

Submitted to the Department of Urban Studies and Planning and the Department of Civil and Environmental Engineering on May 24, 2017 in partial fulfillment of the requirements of the degrees of Master in City Planning and Master of Science in Transportation at the Massachusetts Institute of Technology

Abstract

Dar es Salaam is one of numerous large cities in the developing countries facing tremendous development challenges: growing population, rapid urbanization and motorization, sprawled informal settlements, problematic informal transit operations, and poor and unequal infrastructure service. One major problem Dar faces in overcoming its mobility challenges is lack of quality data. This thesis capitalizes upon a unique opportunity to demonstrate the potential value that mobile telephony presents for transportation planning purposes in Dar and other cities of the Global South. Based on the smartphone-based travel survey data collected by the World Bank from 581 Dar residents for a 4-week period in November-December 2015, this thesis identifies travel patterns depicted by the data, analyzes the representativeness of participants, and demonstrates how the data can be used to help evaluate travel equity in the city. Although the pilot implementation reveals several technical and social challenges in adopting this data collection technology in the African context, it demonstrates promising opportunities in such technology. First, the high-resolution data suggest more precise representation of travel behaviors, compared to traditional paper-based surveys; the smartphone-based survey records higher trip rates, more work-related and private leisure trip-making, and more short-distance travel. Second, little observable bias exists in the likelihood of participation among different socio-demographic groups; except for people with fewer education years and who work for government, no significantly lower likelihood of participation is found among other socio-demographic attributes. Third, the richness of smartphone-based data makes it possible to operationalize new ways to systematically evaluate travel equity in the African city. The ellipse-measured monthly activity space and network-based daily activity space both suggest high travel inequity among participants in Dar. However, although accessibility is unequal (equality of welfare), it is the differences in willingness and ability to pay, rather than the differences in returns on travel welfare investment (equality of effectiveness), that lead to high inequity of travel welfare. Overall, this thesis can help guide future mobility technology-based data collection methods and facilitate the development of evidence-based smart and sustainable planning decisions for Dar and other rapidly developing regions.

Thesis Supervisor: P. Christopher Zegras
Tittle: Associate Professor, Transportation & Urban Planning
Department of Urban Studies and Planning, Massachusetts Institute of Technology

Thesis Reader: Jinhua Zhao
Tittle: Edward H. and Joyce Linde Assistant Professor
Department of Urban Studies and Planning, Massachusetts Institute of Technology

Acknowledgements

The research presented in this thesis was supported by the National Research Foundation Singapore through the Singapore MIT Alliance for Research and Technology's Future Urban Mobility IRG research program.

My experience at MIT has been so valuable. It would not happen without the guidance, support, and help from many people.

First, I would like to thank my thesis advisor Chris Zegras. He provided me the great opportunity to work on this smartphone-based project in Dar es Salaam as well as a previous uncertainty research project in Boston. He gave me enough trust and autonomy to learn, think, work, and grow, while also provided me great guidance and support whenever I needed. He always offered timely feedback, insightful comments, and critical advice to my research update, and helped me explore useful resources to overcome research challenges. He was very patient in editing my writing; he gave sentence-level comments and edits to all my thesis chapters, and I have learned greatly from his editing. He also helped me build connections both in academic and in career and gave me opportunities to present my work to different audience. These opportunities not only promote the influence of my work, but also train my professional skills that are valuable for my future career.

Second, I would like to thank my research advisor and academic advisor Jinhua Zhao. He guided me to grow from an inexperienced researcher to a qualified master candidate on the project of policy knowledge. He always provided inspiring ideas and spotted insightful interpretation behind the data, which taught me how to be a better storyteller. He also gave me detailed instruction on research norms and model building. Beyond the research, he gave me advice on study, life, and career; most importantly, his optimistic attitude exemplifies how happiness can be a personal choice. It motivates me to cultivate a positive and malleable mindset that will bring me long-lasting benefits.

I give thanks to Ajinkya Ghorpade, my research teammate in this future mobility survey (FMS) project. He extracted from raw sensor data the day complexity indicators that I use in this thesis. He also provided many constructive suggestions to my research.

I am thankful to the support and help from the World Bank team: Talip Kilic, Nancy Lozano-Gracia, Marco Tiberti, Ana I. Aguilera, Yonas E. Mchomvu, and Cecilia Escalante. They generously shared the smartphone-based data and baseline survey data they collected and provided useful comments during the process of the research.

I am thankful to African Centre of Excellence for Studies in Public and Non-motorised Transport (ACET) and Japan International Cooperation Agency (JICA), especially Professor Roger Behrens, Professor Estomohi Masaoe, Francis C.M. Maganga, Kazuki Shimada, and Flavia Manyanga. They shared with me their paper-based survey results.

I give thanks to the people I interviewed in Dar es Salaam: Swalehe Nyenye (from Dar es Salaam City Council), planning department of Ialala municipality, Gabriel Ramadhan Anthony (from SUMATRA), Flavia Manyanga (from JICA), Rauxen Zedriga (from The Chartered Institute of Logistics and Transport), Ramadhani S. Sawaka (from The Chartered Institute of Logistics and Transport), Faraji Said (from Dar Rapid Transit), Professor Estomohi Masaoe (from University of Dar es Salaam), Dr. Hannibal J. Bwire (from University of Dar es Salaam), Professor Gordian Wilfred (from Ardhi University), a Yutong vehicle dealer, an inter-city bus owner, and 15 Dar residents. They provided profound understanding of Dar's context in transportation and city development.

I give thanks to many people who offered help to my thesis. I thank Professor Marta C. González for guiding my analysis on describing travel patterns in Dar. I thank Professor Moshe Ben-Akiva and Fang Zhao for their critical comments on FMS analysis. I thank Professor Aaron Golub (from Portland State University) and Jeffrey Rosenblum in helping me frame the travel equity question. I thank Professor Sarah Williams, Mike Foster, Shan Jiang, Hongmou Zhang, Daniel Sheehan and other Rotch GIS lab staffs in helping me solve GIS operationalization issues. I thank Professor Siqi Zheng for providing suggestions on housing cost estimation and Adriano Costa on remote sensing data processing.

I give thanks for JTL members. They gave me useful comments to my research presentations and paper drafts. Especially thanks for Shan Jiang, Joanna Moody, Jeffrey Rosenblum, Anson Stewart, Jungwoo Chun, Jake Gao, Zelin Li, Xuenan Ni, and Zhan Zhao.

Finally, I would like to thank my parents and Haoyang for their unconditional love and support.

Table of Contents

Table of Figures

Table of Tables

Chapter 1 Introduction

Accurate and comprehensive data on individuals' urban travel patterns constitute fundamental inputs into evidence-based and coordinated infrastructure, land use and transportation policies. However, in developing countries, the lack of high-quality travel data is usually a big challenge for sustainable transportation planning. In the case of Dar es Salaam, the economic and environmental costs associated with traffic congestion are rapidly increasing, eroding the benefits entailed by the urban growth. In 2012, Dar had a population of 4.36 million (Tanzania National Bureau of Statistics 2012). From 1990 to 2010, Dar had an annual population growth rate of 4.67% (Nkurunziza 2013) and was ranked the third fastest growing urban area in Africa (Hill and Lindner 2010). The annual growth rate of vehicles reached 10 percent in 2012 (Kiunsi, 2013a). This rapidly motorizing era not only represents a critical moment for transport interventions, but also to more generally redistribute benefits of transport across the society to improve equity.

Transportation infrastructure is unevenly distributed across the city. Five main radial roads have been surfaced and widened since 1990, but local roads are rarely paved and are in increasingly poor condition (Olvera et al. 2003). More than 70 percent of Dar's population lives in the informal settlements scattered across the city (Behrens, McCormick, and Mfinanga 2016). These informal settlements mostly suffer from poor basic infrastructure and services. Olvera et al. (2003) show that unplanned wards have lower access to road, water, food, education, and health service, limited travel ability, and greater burden from transportation costs than affluent planned wards.

Travelling for Dar residents is difficult. The poor level of mobility is caused by rapid population growth, poverty, uncontrolled urban growth, and insufficient provision of transport services and infrastructure (Mkalawa and Haixiao 2014). As of the 2015, Dar's public transit system consists of buses, parantransit (daladalas), and bajajis (motor tricycle taxis). Privately owned daladalas are the key transportation mode for residents in

Dar, followed by walking. JICA (2007; technical report 6, p.1-24) estimated that 59.7% trips were made by daladala and 25.9% by walking based on a household interview survey of 10,000 households; Mkalawa and Haixiao (2014) estimated the mode share of daladala at 68%, of walking at 17%, and of private vehicle at 12% (although the sample size and survey procedure is not well-documented). Daladalas face minimum regulations on market entry, level of service and safety (Transport Research Laboratory 2002). Out of 253 daladala routes identified by Logit and Inter-consultant (2006), only 181 have been authorized.

In order to fundamentally improve the public transit system, the city of Dar es Salaam decided to adopt a bus rapid transit (BRT) system, or Dar es Salaam Rapid Transit (DART). The first phrase of DART was provisionally introduced in April 2015 and entered into full operations in May 2016. The long-awaited DART took more than ten years to formulate the plan, negotiate with stakeholders, and finally construct and operate the first phrase. Questions remain whether DART brings significant improvement to Dar's transit system and whether the resources are fairly distributed across different socio-demographic groups.

Despite the importance of travel data for investing in and evaluating an efficient and fair urban mobility system, little contemporaneous evidence on Dar's rapidly changing spatial and temporal variation of travel patterns exists. In 2007, a household travel survey was conducted as part of a transport master plan with the support from the Japanese government. More recent surveys were conducted by researchers in around 2010 (Nkurunziza 2013; Masaoe et al. 2011). These data may not be recent enough to capture the changes in the rapidly motorizing city. Moreover, traditional travel surveys reveal shortcomings, such as underestimating short-distance trips (Wolf, Oliveira, and Thompson 2003; Ettema, Timmermans, and van Veghel 1996), limited sample size, and imprecision in reported trip start and end times (C. Chen et al. 2010).

This thesis aims to demonstrate the potential value that mobile telephony presents for transportation planning purposes in cities of the Global South. Mobile phone usage in Africa has proliferated in recent years, which increases the potential for leveraging phones as a new travel data collection method. Mobile device penetration in Africa is estimated to be 46%-63% according to different sources, with approximately 10% annual growth (GSMA Intelligent 2016; Mramba et al. 2014). Smartphone adoption has also grown fast, from 4% in 2010 to 23% in 2015 (GSMA Intelligent 2016). In Tanzania, the mobile phone penetration reached 67% and the number of mobile subscriptions increased by 24% in 2015 (Corporate Digest 2017; TanzaniaInvest 2016). Although the average smartphone ownership rate is only 11% for the whole country (Poushter 2016), the rate can be much higher in the city of Dar. The Google Analytics-based estimation of relative smartphone penetration in Dar almost triples that in Mwanza, Tanzania's second largest city (TETEA 2015). This growing mobile penetration provides opportunity to collect travel data via a more precise and cost-effective platform.

The thesis draws from the first implementation of the Future Mobility Sensing (FMS) platform in a developing city. For a 4-week period in November-December 2015, a World Bank project, *Sensors*, deployed FMS to collect high-resolution and high-frequency data on intra-city movements of approximately 550 adult members from 300 households in Dar es Salaam. The subjects were selected randomly from 2,400 households previously interviewed by the Measuring Living Standards within Dar es Salaam Survey (MLDS) in January-February 2015.

1.1 Research Questions and Hypotheses

I aim to answer two basic research questions:

1. Do observable individual characteristics influence the likelihood of participating in smartphone-based travel survey data collection in Dar, thus potentially influencing the representativeness of the travel data collected?
2. What do the data collected via a smartphone-based travel survey indicate regarding transportation equity in Dar?

For the first question, the willingness to cooperate and the familiarity with smartphone may be lower than in developed countries. The lower coverage of cell towers, Wi-Fi access, and detailed contextual data makes the travel inference less accurate. Therefore, understanding the social and technical challenges in the African city enables us to evaluate the feasibility of scaling up this state-of-art data collection approach in the Global South. I build statistical models to analyze who provide more effective travel data based on smartphone-based surveys and understand the structural reasons behind ineffective participation. *Hypothesis 1*: certain individuals, day-activity related, and verification call-related characteristics might bias participation in smartphone-based data collection. Specifically, men, full-time workers, low-educated individuals, wealthy individuals, and people with more complex days will participate less. Men might be less patient in answering verification calls; full-time workers may have less time to cooperate; low-educated individuals may be less familiar with the technology; wealthy individuals' high opportunity cost to may dissuade them from participation; people with more complex days may find it too burdensome to verify data. By testing this hypothesis my research should provide some lessons for how similar the trip data collection methods can be effectively deployed elsewhere in the Global South.

For the second question, I establish a travel equity framework derived from past literature to systematically evaluate the travel equity situation in Dar. I measure people's travel welfare via the revealed individual accessibility and operationalize it by two approaches to calculate activity space, daily and monthly. I then evaluate the society-wide equity index based on two principles: the equality of welfare and the equality of effectiveness and I examine the relationship between travel welfare, on the one side, and socio-demographic attributes and built-environment characteristics, one the other side. The results identify transport-disadvantage groups, locations, and neighborhoods, and provides indicative policy guidance for future transport intervention in Dar es Salaam. *Hypothesis 2*: I expect the two principles (i.e. equality of welfare, equality of effectiveness) and two accessibility measures (i.e. monthly and daily activity space) yield different results. For accessibility measures, I expect that monthly activity space can diminish the randomness in daily activity space, but potentially introduce representativeness problems due to unverified records. For equity principles, I expect to find strong differences across types of individuals in terms of equality of welfare. I also expect to find that the factors influencing the differences in equality of welfare will largely go away when comparing the equality of effectiveness. That is, people in Dar don't have the same level of accessibility, but they get what they pay for.

Chapter 2 Literature Review

2.1 Travel Data Collection

Travel surveys are critical in providing data for transportation modelers and planners to develop evidence-based planning. Mobile-based travel data collection has attracted growing attention for its efficiency, accuracy, convenience, low cost and high coverage. Traditionally, travel behavior data have been collected through paper, phone or interview assisted surveys to reconstruct individual travel behavior for one or more specific days. The self-reporting approach often underestimates short-distance trips (Wolf, Oliveira, and Thompson 2003; Ettema, Timmermans, and van Veghel 1996) and leaves the data quality and quantity questionable. Conducting these traditional surveys is also time, labor and cost intensive, which can be a significant burden, especially for the developing world. With the technological improvement in global positioning system (GPS), new methods have been developed to process the large amount of data collected from GPS devices to infer stops and modes of travel. Although these new models provide a promising alternative to traditional surveys (Bricka et al. 2012; Stopher, Prasad, and Zhang 2010; Bohte and Maat 2009; Greaves et al. 2010; Auld et al. 2010; Li and Shalaby 2008), they face difficulties due to the logistics involved, as the agency conducting the survey is required to purchase and distribute the devices and the participants themselves are required to carry an additional device that they can easily forget (Ghorpade et al. 2015).

In recent years, smartphones have emerged as increasingly pervasive devices with a variety of sensors and sufficient computing power that amount to a new data collection technology. A recent trend in modern survey technologies is to use smartphones for data collection due to their wide spread use (Fan et al. 2012; Nitsche et al. 2014; Zhao, Pereira, et al. 2015). Smartphone-based travel surveys can cover most of the limitations of traditional methods and handheld GPS-assisted methods. Smartphones promise the possibility to economically and frequently enable travel surveys, thus improving the travel data scarcity situation, especially acute in developing countries. Various sensors (e.g. WiFi, accelerometer, GPS, Bluetooth, etc.) built into the smartphone along with mobile-based questionnaires can also increase the detection accuracy of activity and travel mode as well as provide the richness to link transportation to research in other fields such as public health, environment, and psychology.

Despite the increasing popularity of smartphone-based travel survey methods, limited implementations exist in African cities or other developing regions where data availability and quality are almost always a challenge. To begin to bridge this gap, the World Bank deployed a smartphone-based travel survey among 581 residents in Dar es Salaam for a 4-week period in November-December 2015. This thesis analyzes these data and aims to evaluate the quality and feasibility of scaling up smartphone-based data collection in the Global South.

2.2 Travel Equity Literature Review

Travel equity analysis is important and unavoidable in planning. Equity, or social justice, is typically considered as one of three meta-objectives in sustainable development, along with economic growth and environmental preservation. For transportation, no single mature way exists to define and measure travel equity. Similar to equity more generally, the concept of travel equity is very broad and any attempt at "universal evaluation" faces at least five difficulties: the multiple and sometimes contradictory philosophical principles of distributive equity, the difficulty in deciding what, exactly, to distribute in transport and how, travel equity's diverse scope, the range of ways to measure transport well-being, and the variation in how disadvantaged groups can be defined in different contexts.

2.2.1 Political Philosophy of Distributive Justice and Equity

Pereira, Schwanen, and Banister (2016) review five key theories of justice from the perspective of political philosophy and evaluate how they can be applied to the transport domain. The five key theories of justice are utilitarianism, libertarianism, intuitionism, Rawls' egalitarianism, and Capability Approach (CA). Utilitarianism assumes human well-being is the core of justice concerns. Utilitarians are not interested in the distribution of a consequence, but purely in how it can maximize well-being (Kymlicka 2002; Pereira, et. al., 2016). Libertarianism values the idea of self-ownership. It recognizes that all individuals should have the right and freedom to choose and lead their own lives and the free market is just because it does not have interference by the state (Kymlicka 2002; Pereira, et. al., 2016). Intuitionism refers to the perspectives shared by authors for whom basic moral knowledge is acquired through intuition (Rawls 2009). Intuitionists believe that real-life situations are so complex that the proper justice does not have a universal shape; justice is context-dependent (Barry 1965; D. Miller 1999). Rawls' Egalitarianism consists of two overarching principles. The first principle holds that basic rights and liberties of individuals ought to be distributed as equal as possible as long as this does not infringe others' freedom (Rawls 2009; Pereira, et. al., 2016). The second principle applies to primary goods, including income and wealth, powers and prerogatives of authority, and the social bases of self-respect. The second principle requires the least advantaged groups be able to pursue their life plan (Rawls 2009; Pereira, et. al., 2016). The capabilities approach (CA) attaches importance to human capabilities, which are "sets of freedoms and opportunities available for individuals to choose and act", resulting from "a combination of personal abilities and the political, social and economic environment." (Nussbaum 2011; Pereira, et. al., 2016) CA considers the diversity of individuals' preferences, values, abilities, and needs as well as social structures; those characteristics affect individuals' capabilities to convert resources and opportunities into welfare. CA promotes not only equality of opportunities, but also the minimum guarantee of basic capabilities (Sen 1979; Pereira, et. al., 2016).

2.2.2 Dimensions of Travel Equity Evaluation

In the practice of travel equity evaluation, there are two major dimensions: how to define the travel utility of individuals, and how to distribute the utility. With regard to the first dimension, the two contrasting perspectives are welfare and resources (Martens and Golub 2011). Welfare is defined as the "satisfaction" people derive from their actual travel in transportation (Martens and Golub 2011). Equality of welfare implies that people should experience the same level of "satisfaction" from their travel. Criticism of defining utility as welfare comes from questioning the legitimacy of tastes. If the satisfaction of a person's travel discriminates against that of other people, or if people have different level of satisfaction from the same travel, it is unfair to equally distribute their "satisfaction". Equality of resources suggests that people's potential mobility ability should be the same and therefore the initial distribution of transport-related resources is the only thing deserving attention. A problem with this perspective is that not everyone has the same ability to convert resources into welfare. Martens and Golub also introduce the concept of midfare from Cohen (1993) as the middle ground of welfare and resources. Midfare is defined as the extent to which a person is able to translate transport resources into the possibility to participate in activities. Equality of midfare implies that people have the same level of potential opportunity to access services. The problem with midfare definition is that not all the potential opportunities are relevant to the individuals' needs. For example, constrained by an activity schedule, an individual cannot always convert place-based accessibility into utility. According to Sen's (1979) capabilities approach, individuals' needs, abilities, and preferences should be considered in evaluating distributive justice. In addition, the measurement of midfare has limitations in practice, which is discussed in more detail in section 2.2.4.1.

How to distribute welfare (utility) relates to distributive justice. Distribution has two contrasting perspectives: horizontal equity and vertical equity (Litman 2006). Horizontal equity (fairness or egalitarianism) is concerned with providing equal resources to individuals or groups considered equal in ability. It avoids favoring one individual or group over another, aiming to provide services equally regardless of need or ability. Vertical equity (social justice, environmental justice or social inclusion) is concerned with distributing resources between individuals of different abilities and needs. Vertical equity favors groups based on social class or specific needs in order to make up for overall societal inequalities.

The principle of horizontal equity is often applied to assess pricing and allocation of costs. Strict application requires that public policies treat people equally (e.g., face the same costs) (Litman 2006). For example, (D. Chen 1996) suggested that highway funding in

Georgia state was not inequitable; cities received less per capita transport funding than rural districts given the fact that the fund spending favors automobile-oriented facilities. Vertical equity requires the identification of disadvantaged groups and evaluates specific transport plans, projects, services, fees, etc. based on their incidence across different groups (Ng 2005). For example, Casas et. al (2009) evaluate transport-based exclusion of children aged between 5 and 18 years old according to traditional deprivation index, cumulative accessibility, and space-time measures. Jiao and Dillivan (2013) measure the gap between transit service supply and demand for transit dependent people (defined based on age, income, and disability criteria).

2.2.3 The Scope of Travel Equity

The scope of travel equity should extend beyond transport service supply and include broader social and economic dimensions, including land uses, service access, job access, etc. (Lucas 2012). Table 2-1 reviews seven categories of transport-related inequity: physical inequity, transport access inequity, facility access inequity, economic inequity, time-based inequity, fear-based inequity, and space-based inequity; those categories are adapted from the conceptual framework of transport exclusion summarized by Church, et.al (2000). Existing studies often only focus on one or two categories of travel equity depending on the suitability of project context for each analysis. In the evaluation of land use and transportation strategies, transport access equity and facility access equity are commonly considered (Geurs and Van Wee 2004). Time-based equity considers the influence of individuals' time budget or activities' required time commitment in evaluating individual utility. Physical equity and space-based equity are suitable for formulating guidelines in design and space management. Economic equity covers transport affordability. Fear-based equity focuses on the safety on transport mode and reflects the level of city crime. The primary goal of evaluating Dar's travel equity in this thesis is to provide suggestions for planning and investment in land use and transportation, instead of evaluating social issues (i.e. fear-based inequity, economic equity) or space design (i.e. physical equity, space-based inequity). Therefore, the discussion below follows the categories of transport access equity, facility access equity, and time-based equity.

Table 2-1 Different Scope of Transport-related Inequity And The Corresponding Study Example (adapted and extended from Church, et.al (2000))

Category	Meaning	Literature example	Methods	Study places
Physical inequity	Physical barriers, such as vehicle design, lack of facilities or information, inhibit the accessibility of transport services	(Lucas 2011)	In-depth discussion with focus group	Tshwane Region of South Africa
Transport access inequity	Where a person lives can prevent them from accessing transport services, such as in rural areas or on peripheral urban estates	(Currie 2010)	Public transport supply measures based on access distance and level of service	Melbourne
Facilities access inequity	The distance of key facilities such as shops, schools, health care or leisure services from where a person lives prevents their access	(Delbosc and Currie 2011b)	Subjective measure of transport difficulties based on household interview survey	Melbourne
Economic inequity	The high monetary costs of travel can prevent or limit access to facilities or employment and thus impact on incomes	(Olvera, Plat, and Pochet 2003)	Comparison of transport budget and daily travel complexity between neighborhoods based on 1993 Human Resources Development Survey	Dar es Salaam
Time-based inequity	Other demands on time, such as combined work, household and child care duties, reduces the time available for travel (often referred to as time-poverty in the literature);	(Casas, Horner, and Weber 2009)	Space-time prisma	Erie and Niagara Counties, New York
Fear-based inequity	Where fears for personal safety preclude the use of public spaces and/or transport services	(Scott 2003)	Random digit dialing telephone interviews on fear	Canada
Space inequity	Where security or space management prevent certain groups access to public spaces, e.g. gated communities or first class waiting rooms at stations	(Kokoreff 1995)		Paris

2.2.4 Measurement of Travel Welfare

Accessibility, defined as "the potential to reach spatially distributed opportunities", is an often-used method to measure travel welfare in transport studies (Páez, et. al. 2012;

Martens 2016). However, just as equity can be defined and operationalized in many ways, there are many ways to categorize accessibility measures, such as: cost of travel vs. distribution of opportunities, normative measures vs. positive measures based on actual behaviors, location-based vs. person-based (Páez, et. 2012; Martens 2016). This section reviews accessibility measures using three categories: infrastructure-based accessibility, zonal accessibility, and individual accessibility.

2.2.4.1 Three Types of Accessibility Measures: infrastructure-based accessibility, zonal accessibility, and individual accessibility

Infrastructure-based accessibility, zonal accessibility, and individual accessibility reflect the principles of distributing resources, mid-fare, and welfare, respectively. With regard to the scope of travel equity, infrastructure-based accessibility discusses transport access equity; zonal accessibility relates to transport access equity and facility access equity; individual accessibility adds time-based equity. Zonal accessibility is a popular measurement method. It measures the proximity of job or service opportunities to other sets locations from one location or zone. Cumulative opportunities (Martens 2016) and gravity measures (Liu and Zhu 2004; Luo and Wang 2003) are two commonly used methods of zonal accessibility. However, critiques about these measures are that they represent the accessibility of places rather than people, that they tend to produce similar spatial patterns, and that they only measure potentials rather than people's actual needs or their ability to convert potentials into welfare (Kwan 1998). In the equity aspect, zonal accessibility may not capture enough variation between two adjacent neighborhoods that have different level of transport infrastructure as long as they are in the same traffic analysis zone (TAZ). In the case of Dar es Salaam, the road network within neighborhoods can be very poor (Olvera, Plat, and Pochet 2003), and consequently the level of true accessibility varies based on the distance to inter-neighborhood roads. However, depending on the TAZ structure, zonal accessibility measures may not represent well the differences in intra-TAZ travel difficulty.

Individual accessibility is typically based on revealed travel (and/or activity) behaviors. Typical measures of individual accessibility include the space-time prism (H. J. Miller 2006) and the utility-based measure (H. J. Miller 1999). Space-time measures evaluate accessibility in terms of an individual's ability to reach activity locations given the person's daily activity program and spatial-temporal constraints (Landau, Prashker, and Alpern 1982). Utility-based accessibility measures (Neuburger 1971; Ben-Akiva and Lerman 1979) use the expected maximum utility (EMU) from a random utility model. EMU represents the expected ''worth'' to an individual of a set of travel alternatives (Dong et al. 2006). Individual accessibility is estimated based upon the unique situation of each individual. Several challenges to this approach exist. First, the revealed travel behavior may include travel obligations, which may not necessarily increase the travel welfare. Second, revealed travel behavior can underestimate the welfare of people who have the ability to access opportunities without travelling (e.g. people who can work from home and rich people who can call services to home).

Infrastructure-based accessibility is the most straightforward measure because it focuses primarily on service supply. However, infrastructure-based accessibility offers incomplete picture of travel well-being; it only reflects the level of throughput, but does not include any information on the opportunities available to travel (Zegras 2011). Indicators for infrastructure-based accessibility include service coverage, frequency, quality of service, potential mobility efficiency (Martens 2016), level of congestion, average travel speed (Geurs and Van Wee 2004), or a composite index of them (Delbosc and Currie 2011a).

In this thesis, I use space-time prism-based approaches, a measure of individual accessibility, to measure travel welfare.

2.2.4.2 Operationalization of Potential Path Area (PPA)

All space-time (ST) measures are developed based upon Hägerstrand's (1970) time-geographic framework. As opposed to zonal accessibility or place-based accessibility, ST measures consider time constraints in individuals' physical mobility. Individuals cannot always freely travel around to take advantage of all potential opportunities and those opportunities cannot be translated to everyone's travel welfare. People are constrained by their schedules (such as work, school, doctor's appointment, etc.). ST measures evaluate accessibility in terms of an individual's ability to reach activity locations given the person's daily activity program and spatio-temporal constraints (Landau, Prashker, and Alpern 1982). This space-time accessibility can be depicted by a three-dimensional space-time prism (Figure 2-1; H. J. Miller 2006). The x- and y-axis represent spatial coordinates and the z-axis is the time dimension. For fixed activities (such as work and school), the space-time illustration is a line; for time between fixed activities, individuals have the flexibility to explore around the space and this flexibility is depicted by a three-dimensional potential path space (PPS). The projection of the PPS on the two-dimension geographic space is called the potential path area (PPA).

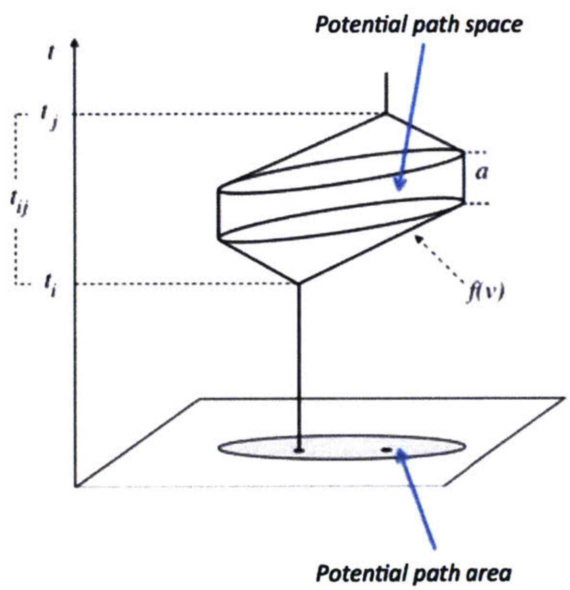


Figure 2-1 A Space-Time Prism (H. J. Miller 2006)

Because the space-time prism and potential path space are very abstract concepts, it is hard to operationalize them with real data. Scholars typically measure PPA with travel diaries, although observed travel behaviors do not always reveal true flexibility in time budgets or activity constrains/flexibility.

Table 2-2 summarizes the methods to calculate PPA used in past studies. These methods can be categorized by two dimensions: the period of time of the travel diary and the approach to derive accessible area. The daily potential path area (DPPA) measures the reachable area estimated within one day. DPPA is the union of PPAs between all pairs of fixed activities in one-day travel diary. Monthly potential path area (MPPA), on the other hand, still measures personal accessible area, but does not follow strictly the concept of a *daily* space-time prism. MPPA estimates the potential area that individuals are exposed to according to the diary of a longer time period. The regularly visited places are given more weight than occasionally visited ones. MPPA and DPPA have different advantages and limitations. MPPA can better represent an individual's travel preference and need by measuring behaviors in the longer term. This mitigates possible randomness from an individual's one-day diary. For example, a regular commuter may work from home for one particular day. If a one-day travel diary only records this work-from-home day, it will misrepresent the individual's typical travel behavior. MPPA often measures behaviors from several weeks and therefore can, for example, take both weekdays and weekends into consideration and depict a richer overall picture of people's travel patterns. However, DPPA has advantages on conceptual rigorousness and data requirement. In fact, MPPA does not consider time constraint in the calculation, while DPPA follows the original space-time concept more closely. In terms of data requirement, DPPA only needs a one-day complete travel diary but MPPA requires weeks of complete diary.

With regard to the approach to deriving geospatial area, geometric approximations are usually more convenient to measure, while network-based calculations have higher data, effort, and computing abilities requirements. A spatial ellipse is the typical geometric approximation (e.g. Lenntorp 1976) but is criticized for not being able to best represent reality. Kernel density estimation can mitigate the weakness by having more fine-grained grid cells, but the results can be sensitive to the assumption of grid cell resolution and kernel function (Schönfelder and Axhausen 2003). Network-based calculation requires data on the detailed road network, segment-specific travel speeds, ideally congestion levels during the time of day, and operation hours of opportunities (e.g. stores, restaurants). These data are not always available. These different geospatial approaches can generate different results; Schönfelder and Axhausen (2003) compare confidence ellipse, kernel density, and shortest paths network (SPN) method based on MPPA schema and report that kernel density and SPN give similar results while results from confidence ellipse are different.

Table 2-2 Methods Summary of Potential Path Area (PPA) Operationalization

	Geometric approximation	Network-based calculation
Monthly Potential Path Area	• <u>**Confidence ellipse**</u>: draw 90%-95% confidence ellipses that are weighted by most frequently visited places (e.g. Schönfelder and Axhausen 2003) • **Kernel density estimation**: interpolate potential travel density based on revealed activities (Schönfelder and Axhausen 2003)	• **Shortest paths networks**: create buffer area of most possibly used roads (e.g. Schönfelder and Axhausen 2003)
Daily Potential Path Area	• **Straight-line based geometric area**: geometric shaped area between fix activities derived from straight-line distance, often in the shape of ellipse (Lenntorp and Lenntorp 1976)	• <u>**Accessible area**</u>: area that is reachable between two fixed activities within the time available (Weber 2003) • **Cognitive feasible opportunity area**: calculate all the potentially accessible area between one-day "fixed activities" based on road network (e.g. Miller 1991), overlay the area with individuals' familiar area, and exclude the aversive area (e.g. Kwan and Hong 1998) • **Available time budget:** determine one-day time budget for discretionary activities, by excluding estimated travel time from the available time between fix activities; the estimated travel time is based on origin-destination cost matrices between fine-grained grid cells (e.g. Fransen and Farber 2017)

Note: The underlined methods are operationalized in this research.

2.2.5 Disadvantaged Groups

The last issue is the definition of disadvantaged groups. With the principle of vertical equity, one needs to define the groups most in need of accessible and affordable transport options. The difficulty in defining disadvantaged group is that the criteria are subjective and highly dependent on the local context, despite a few shared attributes. Litman (2006) summarizes individual characteristics that contribute to transportation-disadvantaged status: low income, non-driver, disability, language barriers, isolation (i.e low geographic access), caregivers (who are responsible for a dependent child or disabled adult), and obligations (e.g., frequent medical treatments, attending school or work). Murray and Davis (2001) define young people (under 16 years old), senior citizens, the low income, migrants and overseas visitors, people with large households, and people with disability as transport disadvantaged groups when evaluating service provision equity in the southeast Queensland region of Australia. Manaugh et. al (2012) use indicators of low median household income, foreign-born/immigrant status, low education, transit-dependency, poor accessibility to job opportunities they are seeking (e.g. less educated people will tend to seek low-skill jobs) to define socially disadvantaged neighborhoods in the city of Quebec, Canada. In developing countries, women often are transport disadvantaged, because they are more vulnerable to violence and crime and often carry out more household obligation trips (e.g. collecting water) (Abidemi 2002; Olvera, Plat, and Pochet 2003).

2.2.6 Travel Equity Studies in Africa

In the African context, existing research reports inequity for females, low-income individuals, and people in rural places, based on differences in travel characteristics and road conditions. On gender, Abidemi (2002) examined intra-urban travel behavior of 232 households in Ibadan, Nigeria and finds that women have significantly more weekly trips of childcare, shopping, fetching water, and getting rid of household waste. Bryceson and Howe (1993) find that rural men only allocate 35% of the time that women do for transport activities and yet benefit more from the rise of transport technology. On the urban-rural analysis, Porter (2002) summarizes difficulties and challenges of living in off-road settlements in sub-Saharan Africa. (C. Venter, Vokolkova, and Michalek 2007) analyze gender differences in trip types, mode use, and travel expenditures in Durban, South Africa as they interact with spatial factors; they report that households in peri-urban and peripheral localities suffer the highest travel burdens and that women bear the greatest burden in most distant, rural localities. In terms of poverty, Behrens (2004) reports lower trip generation and less motorized mode usage for low-income groups based on an activity-based travel survey in South Africa. Lucas (2011) interviewed a focus group of low-income people in South Africa and concluded that they have much lower access to both private and public transportation. Olvera, et. al (2003) report poor access to water and food markets, to education and health services, and to roads in unplanned wards of Dar es Salaam; people who live there also suffer from unaffordable transport cost.

2.3 Proposed Travel Equity Framework and Approach

2.3.1 Principles

Drawing from the insights of justice theories, I use two principles to evaluate society-wide travel equity in Dar: equality of welfare and equality of effectiveness. Equality of welfare requires that everyone has a similar level of welfare. More precisely, the equality of welfare approach holds that a distributional scheme treats people as equals when it distributes or transfers resources among them until "no further transfer would leave them more equal in welfare" (Martens 2016). This principle draws from Rawls' Egalitarianism. Travel welfare is comparable to liberties or opportunities and only the equal distribution is fair according to Rawls' first principle. Although Rawls' second principle also mentions maximizing the prospects of the least advanced groups, it applies to situations when equal distribution is not present.

While equality of welfare may offer one ideal, I introduce the equality of effectiveness, which recognizes that people have diverse preferences and abilities and thus considers the cost of travel welfare to evaluate equity. The principle assumes that when two individuals pay the same price, they should have the same level of welfare. More precisely, given the generalized cost of housing and transport, people should have experienced the same level of welfare. This draws insight from the capabilities approach (CA): we need to consider individuals' capabilities to convert transport and housing resources into travel welfare. I consider individuals' economic abilities to pay rent, travel fare, and car cost as well as their preferences and needs to devote the time to travel (time cost). For example, if a person is capable and willing to pay higher rent to increase accessibility, or if the person prefers to travel further for a better opportunity, this person ought to enjoy a higher level of travel welfare. The generalized cost of housing and transport is not the perfect measure of capabilities and preferences, but it is pragmatic and covers important aspects regarding travel decisions. With the principle of equality of effectiveness, I measure whether the increase of travel welfare is in proportion to the cost of travel. People who are under the median cost-welfare line are in a condition of effectiveness "deficiency." Again, following CA, all individuals should have minimum basic capabilities to do things essential for survival and development. In this study, I use the sufficient generalized cost of housing and transport as an indicator for the minimum level of basic capability. People who are below the minimum level need more government attention.

Inspired by intuitionism, a third way to examine travel equity is not a universal principle, but a context-dependent approach to understand what factors influence the level of travel welfare. To explore this approach I regress travel welfare on socio-economic and built-environment variables to test which factors are associated with lower levels of travel welfare in Dar's context.

Table 2-3 Five Key Theories of Justice and The Adapted Principles in This Research (Table Adapted from Pereira, et. al., 2016)

Theories of justice	Distribution of what?	The fairest distribution	Adaption in this research
Utilitarianism	Welfare, well-being, utility	Whatever distribution that maximizes aggregate welfare	/
Libertarianism	Basic rights and liberties	Absolute equality	/
Rawls' Egalitarianism	Basic liberties	Equal distribution	**Equality of welfare:** Travel welfare is the geospatial opportunity for everyone; people should have the equal level of travel welfare.
	Opportunities	Equal distribution	
	Primary goods (rights and prerogatives of authority, income, and wealth)	Maximin criterion: The distribution that maximises, subject to constraints, the prospects of the least advantaged groups	
Capabilities approach	Opportunities	Equal distribution	**Equality of effectiveness:** - Consider the cost of travel welfare. When two pay the same price, they should have the same level of welfare. - Everyone should have the minimum cost affordability to guarantee basic travel welfare
	Central/basic capabilities	All should get above a minimum basic level	
Intuitionism	Different "whats", for example, resources (food, money, etc.), services (health, education, etc.)	No clear distribution pattern	**Regression: Analyze the influence of those socio-economic factors on welfare**

2.3.2 What and How to Distribute

Equality of welfare is comparable to horizontal equity. It holds that all individuals should have equal geospatial opportunities and the same level of travel welfare. Equality of effectiveness argues that inequality of travel welfare is legitimate as long as (a) people are above the minimum basic capabilities and (b) individuals' levels of welfare are proportional to their cost. This overlaps with vertical equity but is different in application procedure. Similar to vertical equity, equality of effectiveness considers differences in individuals' abilities and needs, but does not focus on all groups, rather than only the disadvantaged groups. Disadvantaged groups are not pre-defined as in vertical equity practices, but are derived from the analysis.

In this thesis, I define travel utility as the observed accessible opportunities derived from the individuals' actual travel. This is adapted from Marten and Golub's welfare concept, but is not based on "satisfaction" or "desire." This approach addresses part of the criticism of the capricious nature of satisfaction, but still leaves unresolved the issue of the legitimacy of only valuing observed behaviors. Chapter 4 presents the calculation methods in more detail, but the main reasons for using revealed behaviors are data availability and their representativeness of individuals' situation. For the rest of the thesis, I use "travel welfare" or "travel well-being" to refer to "travel utility", or the "what" to be distributed in transport, to avoid confusion with utility-based accessibility measures.

In this thesis, I do not define the disadvantaged groups or evaluate the travel welfare for these disadvantaged groups. Instead, I regress the travel welfare on socio-economic and built-environment variables and let the results indicate the transport-disadvantaged groups. In the context of Dar es Salaam, gender, income, education, access to health and school, access to water, access to major road, access to vendors can be possible influence on people's travel welfare (Olvera, Plat, and Pochet 2003). Those variables are included in the regression model.

Ultimately, evaluating travel equity can be challenging because it depends on the foundational principles of justice, the various ways of defining the scope of travel equity, how to measure travel welfare, and how to define the disadvantaged groups. Table 2-4 summarizes my framework, tailored to Dar's context, based on evidence from existing research in Africa and from past literature.

Table 2-4 Summary of Travel Equity Evaluation in Previous Studies and in This Research

	Principles of distributive equity		The scope of travel equity	Measurement of travel welfare	Potential disadvantaged groups
	Philosophy principles	**Applied evaluation dimensions**			
Literature review	• Utilitarianism • Libertarianism • Rawls' Egalitarianism • Capabilities approach • Intuitionism	1. What to distribute • Distribution of satisfaction/welfare • Distribution of mid-fare • Distribution of resources 2. How to distribute • Vertical equity • Horizontal equity	• Physical inequity • Transport access inequity • Facility access inequity • Economic inequity • Time-based inequity • Fear-based inequity • Space inequity	1. Zonal accessibility • Cumulative opportunities • Gravity measures 2. Individual accessibility • Space-time measures • Utility-based measures 3. Infrastructure-based accessibility	Pre-defined • Low income people • Non-driver/car-less people • People with disability • People who live in in inaccessible location • Young people (under 16) • Senior citizens • People with large households • Migrants or oversea visitors • Caregivers • People who have too many obligational activities • ...
In this study	• Rawls' Egalitarianism	1. What to distribute: actual need/welfare	• Physical equity • Geographical equity	• Individual accessibility: Space-	Does not discuss the welfare of certain pre-defined disadvantaged groups, but

	• Capabilities approach	2. How to distribute: • Equality of welfare • Equality of effectiveness	• Equity of facilities, Time-based equity	time measures	analyze factors that can influence welfare in Dar. Factors include: • Gender • Wealth • Education • Age • Total working hours • If having driving/motorcycle license • Access to infrastructure service (or living in formal settlement) • Access to major road • Access to market

Chapter 3 Context: Dar Transport Challenges and Causes

3.1 City Overview

Dar es Salaam (Dar) is the largest and fastest growing city in Tanzania, the nation's main port, political center, and economic hub. In 2012, Dar had 4.36 million persons, accounting for 10% of total population of Tanzania Mainland (Tanzania population and housing census, 2012). From 1990 to 2010, the city's population grew by nearly 5% per year, making it Africa's third most rapidly growing urban area (Mkalawa and Haixiao 2014). Although Dar es Salaam lost its capital city status to Dodoma in 1974, Dar retains the function of actual political center because many central government activities are still located there. Dar is the economic center of Tanzania, contributing 17.3% of national GDP in 2013 (Tanzania NBS, 2014). The city's estimated income per capita, US$1,200 in 2013, is 1.7 times higher than the rest of the nation (Tanzania NBS, 2014). Dar has the busiest port in East Africa and the busiest airport in Tanzania; the Tanzania-Zambia Railway Authority has its northern and main terminal in Dar (Y. Venter et al. 2015).

Dar es Salaam is among the 30 regions of Tanzania (DES master plan draft 2015, p.29). The city was divided into three municipalities (Ilala, Kinondoni, and Temeke) and has recently adjusted the division to be five municipalities (Ilala, Kinondoni, Ubungo, Temeke, and Kiigamboni). Local municipalities or municipal councils are responsible for providing social and economic services to residents, while the City Council performs a coordinating role of these municipalities. Municipalities are further subdivided into wards. There were 90 wards in 2015 before the adjustment of municipal jurisdictions.

Previous jurisdiction of three municipalities (2015) Current jurisdiction of five municipalities (2016)

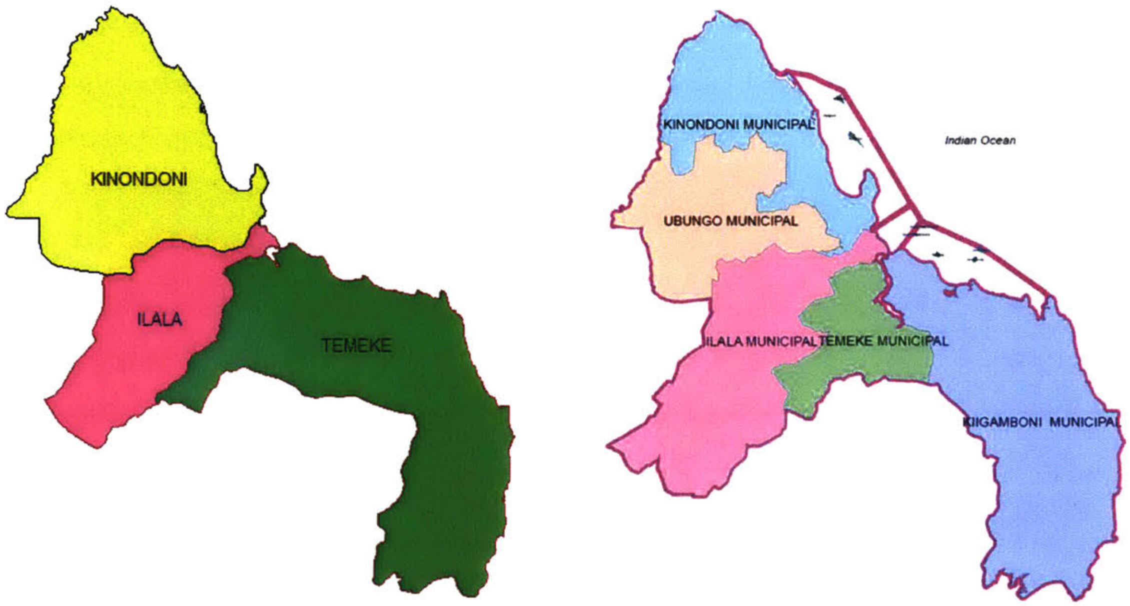


Figure 3-1 Boundary of Dar es Salaam municipalities (InfoBridge Consultants Limited 2009, master plan draft 2015)

Dar es Salaam is relatively monocentric and has been characterized as "low rise" and "sprawling" (Kiunsi 2013; p.96) with an average population density of 31.3 persons per hectare in 2012 (Tanzania NBS, 2014). The Central Business District (CBD) has higher density and some areas exceed 300 persons per hectare (Kiunsi 2013b). The urban development along the arterial roads are called "ribbon development" and the farthest developed area along the corridors reaches 30km away from the city center in 2007 (JICA 2008; technical report 1, p. 1-17).

Informal settlements accommodate 70 percent of Dar's population and occupy 75 percent of the land area (JICA 2008; technical report 1, p.1-17; DES master plan draft 2015, p.25). Informal settlements are unplanned and most have poor access to basic infrastructure services including water, sewerage, electricity, roads, hospitals, and schools. Informal jobs in these neighborhoods are common; approximately 60 percent of Dar residents work in the informal sector (Kiunsi, 2013b). Low-income households are spread across Dar; the percentage of households below the basic-need poverty line are 14% in Kinondoni, 16% in Ilala, and 29% in Temeke (REPOA 2005).

3.2 Dar Transport Challenges

Travelling in Dar is notoriously challenging. Dar residents suffer from severe congestion during peak hours, limited access to activities and opportunities, poor transit service, and potential travel inequity.

3.2.1 Congestion

Severe traffic congestion exists along radial roads, particularly during peak hours (JICA 2008; technical report 2, p. 1-5). Melbye et al. (2015) estimate average speed of four radial roads at 12 km/hr during peak hours. In comparison, the average free flow speed is 50-60 km/hr during off-peak hours. Dar es Salaam Rapid Transit (DART) reported four billion Tshs (or USD 1.79 million) economic and environmental loss per day caused by urban transport challenges (Kiunsi 2013a). The economic loss due to wasted time in congestion is estimated at USD 200 million per year; the estimated wasted fuel reaches per annum USD 92 million (Katala 2012).

3.2.2 Limited access to activities and good transit service

Dar residents suffer from limited mobility. By one estimate, the average gross trip rate (including walk trip) is only 1.2 trips per day (JICA 2008; technical report 6, p.1-20). On the one hand, social, economic, and recreational services are inadequate and residents thus have a limited destination choice set. Key activities concentrate in the city center, which is far away from large neighborhoods located in the suburban areas. On the other hand, residents lack the reliable and efficient transportation means to go to their destinations. Most people rely on daladala (a privately operated minibus) for long-distance travel and on walking for short-distance travel (JICA 2008, technical report 6, p.1-24; Mkalawa and Haixiao 2014). Despite the strengths in demand responsiveness, self-financing, and job creation (Mfinanga et al., 2015), these loosely regulated semiformal daladalas fail to provide safe and high quality transit services. Unsafe behaviors of daladalas include the lack of regular vehicle safety check, dangerous overloading, on-street boarding, speeding, jumping red lights, driving on road shoulders or sidewalks, and other behaviors that violate road rules; the low level of services are reflected in poor vehicle maintenance, overcharging, bad attitudes of conductors, poor route design, and unreliable schedules and/or long waiting times. JICA (2008) estimated the travel time per daladala trip at 77 minutes and waiting time at 33 minutes. JICA's bus passenger survey found high level of dissatisfaction towards daladalas; 80% of surveyed passengers judged the daladala system as unacceptable and 90% were unsatisfied with the routing and comfort level of daladata. Nonetheless, 98% of daladala passengers are captive to this mode because they do not have cars or driving licenses (JICA 2008; technical report 2, p.8-6).

3.2.3 Potential travel inequity

Little research exists evaluating travel equity in Dar es Salaam. However, given the residents' different access to transport modes and the unbalanced infrastructures and services in planned and unplanned neighborhoods, one can imagine that travel welfare and transport resources are likely to be unequally distributed among the population. Olvera et.al (2003) compare the access to water, food, road, school, and hospitals among affluent, planned, and unplanned wards based on a 1993 Human Resources Development Survey. They find people in unplanned wards face longer travel times to water and hospital, fewer schools in the accessible distance, and unaffordable transport costs, although access to informal public transit are similar among the three types of wards. They suggest potential gender inequity because women are largely excluded from the workforce and carry out most house chores, such as collecting water; in 75% households, women travel to obtain water. JICA (2008) estimates that women only have 0.9 trips per day on average, compared to 1.5 trips generated by men.

3.3 Dar Urban Transport Stakeholders

The landscape of Dar urban transport stakeholders is complex. The transport-related responsibilities are spread among a range of institutes and individuals (Figure 3-2, Table 3-1), making coordination difficult. The organization of government has dual principles in setting responsibilities: it concedes authority to national ministries for local issues under a centralized approach; and it devolves citywide or regional functions to local municipalities and the private sector under decentralization and privatization policy (JICA, 2008; technical report 3, pp. 3-3). On the one hand, national agencies have direct control over local functions, as part of the colonial legacy: many key responsibilities related to urban transportation, like planning and supervision, remain at the national-level ministries; national-level Executive Agencies are in charge of most of the transportation regulation and construction. On the other hand, local governments (municipal councils) and the private sector take the main responsibilities for land use enforcement, road construction, and transport services provision, which undermines the integration and coordination power of Dar City Council at the city/regional level. Transport service providers in Dar are almost all from the private sector under the loose regulation of the Surface and Marine Transport Regulatory Authority (SUMATRA). The Bus Rapid Transit (BRT) phase I has been in full operation, by UDA-RT, a private operator, since May 2016 and is currently the largest transit service managed by the government.

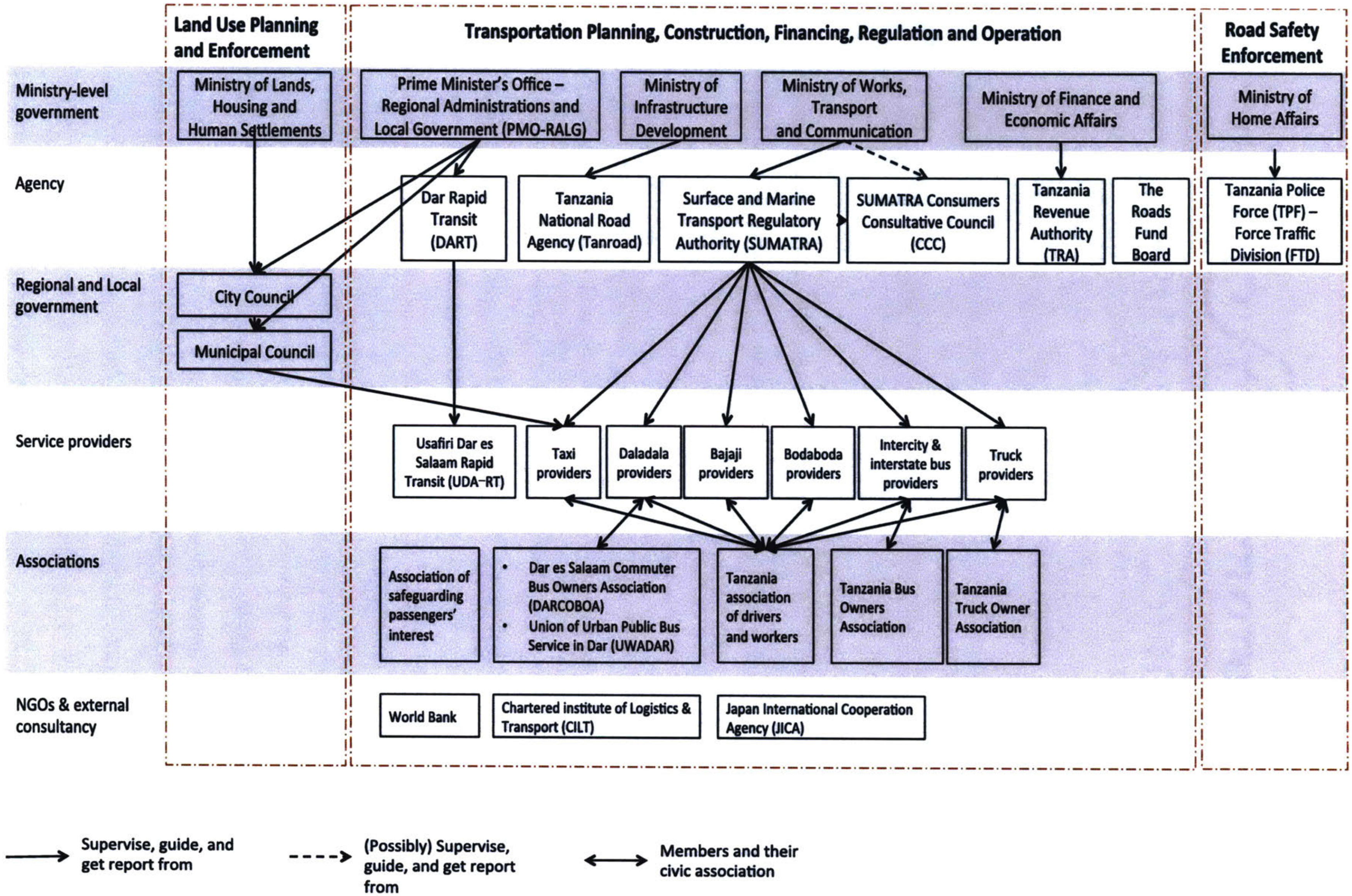


Figure 3-2 Dar Urban Transport Stakeholders and Their Relationships

Table 3-1 Qualitative Evaluation of the Responsibilities of Dar Urban Transport Stakeholders[1]

Level of institute	Institute	Land Use Responsibilities			Transportation Responsibilities						
		Planning	Enforcement	Supervision	Planning	Construction	Supervision	Regulation	Operation	Financing	Road Safety
Ministry-level government	Ministry of Lands, Housing and Human Settlements	+++	++								
	Prime Minister's Office – Regional Administrations and Local Government (PMO-RALG)			$\underline{++}$	$\underline{+}$		$\underline{+}$	$\underline{+}$			
	Ministry of Infrastructure Development					$\underline{+}$	$\underline{++}$				
	Ministry of Works, Transport and Communication				$\underline{++}$		++	++			
	Ministry of Home Affairs						++				+
	Ministry of Finance and Economic Affairs						++			+	
Executive Agency of ministry	Dar Rapid Transit (DART)				$\underline{+}$	$\underline{++}$		++			
	Tanzania National Road Agency (TANROADS)					+++					
	Surface and Marine Transport Regulatory Authority (SUMATRA)							+++		++	
	SUMATRA Consumers Consultative Council (CCC)							+			
	Tanzania Police Force (TPF) – Force Traffic Division (FTD)										+++
	Tanzania Revenue Authority (TRA)									++	
	The Roads Fund Board									+++	
Regional government	City Council	+	+		$\underline{++}$	$\underline{+}$				++	
Local government	Municipal Council	++	+++		$\underline{+}$	$\underline{++}$	$\underline{+}$	+		++	
Service provider	Usafiri Dar es Salaam Rapid Transit (UDA-RT)								+++		
	Daladala providers								+++		
	Bajaji providers								+++		
	Bodaboda providers								+++		
	Taxi provider								++		
	Intercity & interstate bus providers								+++		
	Truck providers								+++		

[1] Note: 1) +++ represents strong responsibility, ++ is moderate, + is limited, blank is none; 2) Signs with underlines represent that the author is not 100% sure about the evaluation

3.3.1 Transportation Responsibilities

3.3.1.1 Transport Planning

The most recently published transport master plan in Dar was conducted in 2008, with technical support from the Japanese government through the Japan International Cooperation Agency (JICA) (JICA 2008). JICA is currently working on a new master plan in collaboration with the Ministry of Transport, Dar es Salaam City Council and the municipal council. During transport planning, the Ministry of Transport reviews the policies and strategies of transport development and compiles and maintains sub-sector data. The Dar city council coordinates with municipalities and other stakeholders and is supposed to be the implementing agency of the study; the municipal council participates as the local stakeholder and provides support.

Bus rapid transit (BRT) has been the most ambitious recent public transport intervention proposed for Dar. The BRT plan has involved in a variety of international agencies and consultants, including the Institute for Transportation and Development Policy (ITDP), the World Bank, and the African Development Bank. Those organizations have been working with the Dar Rapid Transit (DART) on the implemented Phase I BRT and the planning of the rest route network (JICA, technical no.2, 9-3).

3.3.1.2 Transport Construction

Transport infrastructure provision is largely inadequate in Dar. The Tanzania National Road Agency (TANROADS), under the supervision of Ministry of Infrastructure Development, takes the responsibility to design, build, repair and maintain the national roads or trunk roads. As a road administration body, TANROADS also prepares guidelines, standards and specifications for road works for development and maintenance (JICA, 2008; technical report 4, pp.2-12). Municipal councils manage the local roads. Out of 3118.61 km of road in Dar in 2013, TANROADS supervised 18.8% and municipalities supervised 81.2% (Tanzania NBS, 2014). Dar City Council coordinates cross-border major road development and maintenance. However, only a few arterial roads and ring roads are well maintained and 79% roads in Dar are unpaved (Tanzania NBS, 2014). For other transport infrastructure, Dar City Council is responsible for inter-city bus terminals across Dar, but the current two bus terminals are fewer than what are needed; the local municipalities are responsible for taxi stations and parking space, which are also under-provided (The World Bank 2011).

3.3.1.3 Transport Regulation

The Surface and Marine Transport Regulatory Authority (SUMATRA) is the main regulatory body for urban transportation in Dar. SUMATRA is responsible for regulating, promoting, and facilitating the surface and marine transport services (except for BRT and taxi). Nevertheless, the current situation does not deliver an efficient and high-quality urban public transit system. SUMATRA issues permits to intercity buses (operating between Dar and other cities) and daladalas to authorize them to operate on specific routes with specific vehicle capacity. However, SUMATRA does not plan the routes nor require operators to bid for routes. Permit issuing is on a first-come-first-served basis, with private operators deciding on their routes and then applying for the permits. The extent of SUMATRA's regulation in this respect is to ask a proponent to choose another route if the route already has too many operators on it[2]. SUMATRA also formulates codes (e.g., for safety) and monitors transit services (The World Bank 2011), but regulation of, e.g., vehicle maintenance and driver and operator training, is lacking[3].

SUMATRA oversees road transport in collaboration with Ministry of Transport and Tanzania Police Force. The Ministry of Transport also takes responsibility for some aspects of road transport regulation (The World Bank 2011). For taxi permitting and operations, municipal councils are responsible for them. The Prime Minister's Office – Regional Administrations and Local Government (PMO-RALG) monitors the activities of municipal administrations.

---

[2] The information comes from an interview with a SUMATRA officer

[3] The information comes from an interview with two officers from The Chartered Institute of Logistics and Transport

The SUMATRA Consumer Consultative Council (CCC) used to be one division of SUMATRA, but later became independent in an attempt to more objectively represent the interests of consumers[4]. CCC aims to guarantee that travel is affordable for residents. It consults with industry, government, and other consumer groups on matters of interest to consumers; CCC also makes submissions and provides views and information to the Authority, Minister and Sector Ministers (The World Bank 2011).

Dar Rapid Transit (DART) was formed under the Prime Minister's Office (PMO-RALG) in 2006 to manage BRT operations. DART supervises only BRT operation; the regulation of the rest of surface transport services belongs to SUMATRA. DART holds an 8%[5] share of the BRT operator (UDA-RT) and therefore also helps with the BRT business model and risk management. As the next step of BRT development, DART plans to work with other stakeholders to expand BRT routes, to attract business opportunities to BRT stations for transit-oriented development (TOD), and to establish a control center for BRT operation[6].

3.3.1.4 Transport Service Providers

Public transport plays an important role in Dar es Salaam, for the majority of residents rely on it to travel around the city. The public transport modes include Daladala (semi-informal minibus), Bajaji (three-wheeler auto-richshaws), Bodaboda (two-wheeler public motorcycles), roving saloon taxis, BRT, and intercity bus[7]. The estimated mode share of daladala ranges from 59.7% to 68% (JICA 2008, technical report 6, p.1-24; Mkalawa and Haixiao 2014). Privately owned, Daladalas usually operate on inter-neighborhood roads and are the key mode for Dar residents' city-wide travel. Most of the intra-neighborhood roads are too narrow and rough for daladalas. Bajajis and Bodabodas operate mainly within neighborhoods, serving as the last-mile solution. But the fares for

---

[4] The information is given by a SUMATRA officer in an interview

[5] The percentage is from the interview with a DART officer

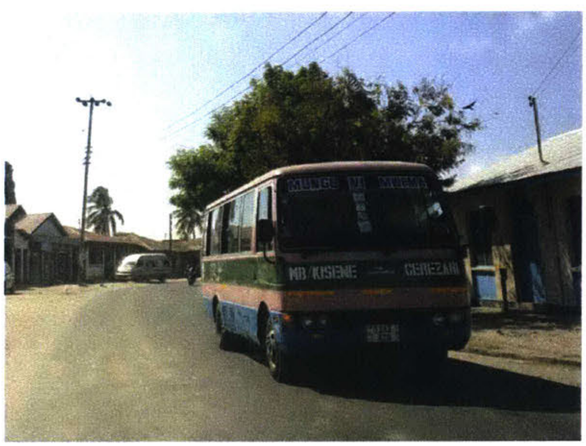
[6] Based on the interview with a DART officer and a transport specialist in the World Bank Dar es Salaam office

[7] Private operators run intercity bus between Dar es Salaam and other cities in Tanzania. SUMATRA regulates this service.

Bajajis and Bodabodas (typically 1000 Tshs or 0.45 USD) are higher than for Daladalas (typically 400 Tshs or 0.18 USD), making them difficult for most people to afford to use regularly. Many neighborhoods are large and it can take half an hour to walk to the daladala station on an inter-neighborhood road from the home of residents.

Dar es Salaam Commuter Bus Owners Association (DARCOBOA) and Union of Urban Public Bus Service in Dar (UWADAR) are the most important daladala associations in Dar. They represent the interests of daladala owners and drivers. The government often communicates with daladala operators through DARCOBOA and UWADAR.

Usafiri Dar es Salaam Rapid Transit (UDA-RT) is the operator of the BRT under the management of DART. The private sector holds 92%[8] of UDA-RT's shares and the rest belongs to the government. UDA-RT currently generates positive net operating income and reportedly performs much more efficiently and effectively than its former company Usafiri Dar es Salaam (UDA)[9]. UDA was the previous formal public bus company, but declined as the competition rose from the individual daladala players.



Daladala (informal minibus)

Bajaji (three-wheeler auto-rickshaws)

[8] The percentage is from the interview with a DART officer
[9] The information is obtained from the interview with a DART officer

Bodaboda (two-wheeler public motorcycles)

BRT

Roving saloon taxis

Intercity Bus

Figure 3-3 Public Transport Modes in Dar

3.3.1.5 Transport Financing

Revenues are generated from various sources. SUMATRA collects fees from issuing operating permits to daladalas; it also imposes fines on the operators that violate the permits. The Tanzania Revenue Authority (TRA) collects revenues on behalf of the government, including income tax from bus operators and drivers and vehicle licensing revenues. The Road Fund Board collects fuel taxes, which is the major source for road repair and maintenance. Municipal councils collect local taxes and fees from taxi permitting and car parking fees and fines (The World Bank 2011). DART shares part of the BRT revenue.

External sources of funding are also important to Dar. In the fiscal year 2005/06, external debt accounted for 41% of the total income source of the national budget (JICA 2008; technical report 4, p.3-1). A large portion of the funds to construct BRT comes from external sources, including the World Bank, the African Development Bank, and the African Growing Together Fund (AGTF)[10].

3.3.2 Land Use Responsibilities

Land use planning in Dar may be characterized as inadequate (DES master plan draft 2015; p.13) and reactive. The last master plan of Dar was in 1979, although it was not really followed and is largely outdated due to the city's rapid population growth. The most recent land use plan was the Sustainable Dar es Salaam Project (SUDP) conducted in 1992, which is also not well enforced (DES master plan draft 2015; p.12). For most parts of the city, the government does not have zoning plans[11]. When people want to develop a piece of land, they apply for a land development permit from municipal councils. Municipal councils then prepare the plot plan partially based on SUDP and submit the plan to the planning committee consisting of counselors from wards. Sometimes people who want to develop the land make the plot plan themselves and municipal councils only review the plan. After municipal councils prepare the plan, the city council will sub-approve it and send it to the Ministry of Lands, Housing and Human Settlements for the final approval. However, even though the ministry has the right of final approval, municipalities are more intensely involved in evaluating the land development and doing the actual groundwork. After the plot plan is approved, municipal

[10] the loan from African Growing Together Fund (AGTF) is mentioned in this BRT General Procurement Notice: https://www.afdb.org/fileadmin/uploads/afdb/Documents/Procurement/Project-related-Procurement/GPN_%E2%80%93_Tanzania_-_Dar_Es_Salaam_bus_rapid_transit_system_project.pdf

[11] This information is from an interview with an officer from the planning department of Ilala municipality and Professor Hannibal J. Bwire from the University of Dar es Salaam

councils issue permits to land users and are supposed to monitor the development process. Municipal councils also deal with complaints from residents on land use issues[12].

The Ministry of Lands, Housing and Human Settlements is in charge of developing a new master plan for Dar with external technical assistantship from a consortium of four planning and design consultants—Dodi Moss Ltd (an Italian architecture firm), Buro Happold Ltd (a British engineering firm), Afri-Arch Associates (a Tanzanian architecture firm), and Q Consult Ltd (a Tanzanian infrastructure and real estate firm). Dar City Council and municipal councils provide suggestions and data for the master plan. Dar City Council also coordinates with municipalities and other stakeholders on this master plan.

3.4 Sources of Dar's Transport Challenges

The urban mobility challenges in Dar es Salaam are enormous, difficult, and complex. The sources are various: rapid population growth and city expansion, poor road infrastructure, monocentric land use structure and sprawled informal settlements, problematic semiformal transit operations, and institutional weaknesses.

3.4.1 Population Growth and Travel Demand Increase

Growing at over 5% per year for more than one decade, Dar's population has increased 75 percent from 1.9 million in 2002 to 2.5 million in 2012 (Tanzania population and housing census 2012, vol.1). Figure 3-4 shows the expansion of built-up area from 1947-2001. The rapidly growing population and expanding urban area increase travel demand and traffic volumes, which produce huge pressures on the transport system.

To worsen the transport burden, the city's motorization growth rate is even higher than the population growth and continues accelerating. Melbye et al (2015) report 600,000 private cars in 2012 compared to only 220,000 in 2000, suggesting 8.7% annual

[12] The land use approval and enforcement process is summarized based on the interview with an officer from the planning department of Ilala municipality

automobile growth on average; the annual growth of cars has exceeded 50,000 since 2009, an approximate 12% annual growth rate. Other data suggest lower motorization rate; Masaoe et al (2011) report, for example, that 90% of households have zero automobiles. But in any case, at the relatively early stage of motorization, the rapid growth in car ownership and usage poses a great challenge for Dar's road system.

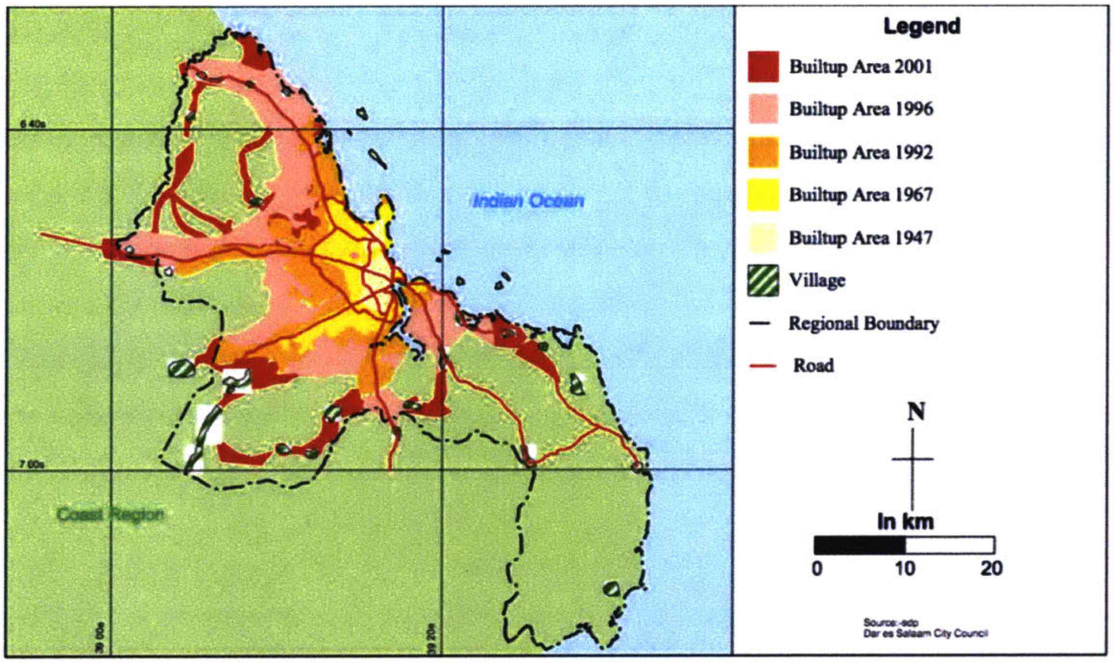


Figure 3-4 Urban expansion in Dar es Salaam from 1947 to 2001 (JICA 2008, technical report 1, p.1-16)

3.4.2 Road Infrastructure Quality

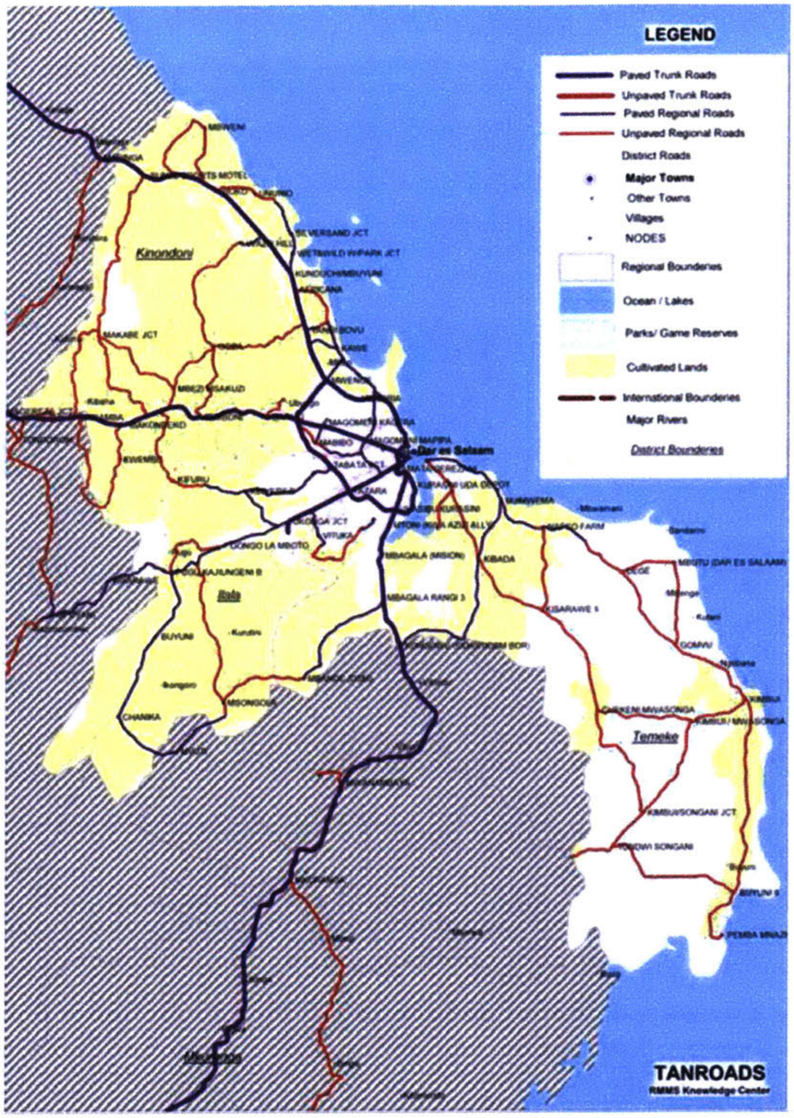
The lack of well-functioning roads worsens the gap between travel demand and transport network capacity. The road density is only 1.91 km/km^2(Tanzania NBS, 2014). For comparison, road density in New York City is 12.79 km/km^2 and in London is 9.42 km/km^2 (Di 2013). There are only a few trunk roads in Dar that are well surfaced and maintained. The quality of other inter-neighborhood roads varies and most of them are poorly maintained. In 2013, the road length was 3118.61km in Dar; only 21% are paved, 35% have gravel surface, and 44% have earth surface (Tanzania NBS, 2014). The intra-neighborhood roads are generally in even worse conditions: narrow, rough, and without proper drainage. The poor non-trunk roads make it unusually difficult to drive and walk, thus pushing the overwhelming traffic volumes to the trunk roads, worsening congestion.

Figure 3-5 shows the existing road infrastructure in Dar es Salaam. Paved and well-maintained four-lane roads only exist in four radial corridors (Bagamoyo, Morogoro, Nyerere and Kilwa Roads) and two ring roads (Nelson Mandela/Sam Nujoma and Kawawa Roads). High traffic volumes on these trunk roads leads to frequently severe congestion, especially at the intersections and during peak hours. Other two-lane regional roads are paved, but poorly maintained (often cracked and/or with potholes) in the good cases, and many are even not paved. But even those unpaved regional roads are considered as "better roads" in Dar[13], because they are at least wide and flat enough for daladalas to operate on. The daladala route map[14] in Figure 3-6 represents the proxy indicator of the roads that are at least tolerable for vehicles to drive on. However, even with this broader categorization of "good roads," the density of the road network is still low.

Intra-neighborhood roads are in the worst condition. Those unpaved roads are rough and narrow and can be flooded by a heavy rain (Figure 3-7). They are usually too tough for daladalas to operate on, so often only bajajis and bodabodas can operate there. In most cases, walking is the main mode of transport on those roads, although it can take a very long time to walk to the trunk and regional roads where daladala stations and many informal markets are located.

[13] Based on the author's field trip observation

[14] Ramani Huria (a team from the World Bank) organized a group of mapping volunteers to track daladala routes by a smartphone app or GPS device. Ramani Huria and Ally (a German company) worked together to process the data and import all tracked routes into OpenStreetMap (http://ramanihuria.org/putting-dar-es-salaam-dala-dalas-map/)



Bagamoyo road (radial road)

Sam Nujoma road (ring road)

Kimweri Ave (paved regional road)

Figure 3-5 Existing road infrastructure map (DES master plan draft 2015) and example situations of trunk roads

Inter-neighborhood road - Makumbusho

Inter-neighborhood road - Buguruni

Figure 3-6 Existing daladala routes (as of Feburary 2017, Source: Open Street Map) and the road infrastructure examples

Main neighborhood road - Tegata

Inter-neighborhood road - Mwananyamala

Intra-neighborhood road (a wide one)- Hanna Nassif

Intra-neighborhood road - Kimara

Figure 3-7 Unpaved inter- and intra-neighborhood road examples

3.4.3 Land Use Issues

Land use patterns determine trip generation and distribution. In the case of Dar es Salaam, the land use structure increases the complexity of the transport challenges.

3.4.3.1 Mono-centrality of key services and scattered informal business

Dar es Salaam has both mono-centrality and dispersion features. Most key services and formal businesses, such as central and local government agencies, banks, and private business activities, locate at the Central Business District (CBD) in the city center. Some companies and shopping centers have gradually emerged along four radial roads, but the CBD contains the highest concentration of jobs in the entire city (DES master plan draft 2015; p.140). The CBD is the densest area in Dar; density exceeds 100 people per hectare within 10km of the CBD (Y. Venter et al. 2015), more than three times the city average (Tanzania NBS 2014; p.19). People employed in the formal economy often commute to the CBD, thus creating high inbound demand during the morning peak hour and high outbound demand in the evening. JICA estimated that the city center attracted 1.15 million trip ends in 2008, accounting for 40% of total trips (2.87 million) made in the city (JICA 2008; technical report 6, p.6-16 + p.1-20). Other than these commuting trips, work-related trips in the formal economy can also end at the CBD. As online services are not very common in Dar, work-related trips are one important component of total travel demand.

On the other side, informal markets are scattered around the city. The informal economy plays a crucial role in Dar, with 60 percent of residents having jobs in informal businesses (Kiunsi, 2013b). The most common informal jobs are vendors and drivers. Vendors have stands alongside the major roads, near daladala stations, or in their neighborhoods. Those stands and the informal markets can be found across the whole city, providing accessible shopping opportunities to many residents. However, although these stands increase mix of residential and commercial use, they cannot satisfy all demands. Many informal markets are small and the goods are not diverse (DES master plan draft 2015). Residents often need to go to a few big markets once or twice a week. The largest market is Kariakoo, located in the city center close to the CBD. Overall, the best shopping services are still in the city center, attracting additional traffic beyond commuting trips, but the scattered small markets do provide alternatives for people with limited access to good transport and alleviate the pressure caused by the mono-centric land use structure to some degree.

CBD (formal business) Street market (informal business)

Figure 3-8 CBD and street market in Dar es Salaam

3.4.3.2 Sprawled informal settlements

More than 70 percent of the population in Dar lives in informal settlements or unplanned areas with poor infrastructure services (JICA 2008, technical report 1, p.1-17; Figure 10, Table 3-2). These neighborhoods have limited access to water, roads, health care, and schools. For example, residents in unplanned neighborhoods often rely on vendors to buy bucket water, while planned neighborhoods have several water kiosks for residents to collect water.

Informal settlements not only create tough living situations for residents, but also produce difficulties in urban transportation planning and development. First, the residents typically build their houses themselves without applying for development permits or following basic planning guidance. The areas have sparse and irregular road networks and travel demand often exceeds road capacity, worsening connectivity of the whole city. Once the areas are occupied, expanding the road network can be difficult and expensive. In addition, most of the informal houses are one-story. The fast expansion of these low-rise settlements exacerbates urban sprawl, increases travel time and distance, and makes it more difficult to operate mass transit, which requires compactness for adequate demand.

A complex set of reasons have led to the sprawled informal settlements. First, early urban migrants were allowed to build houses outside of the urban boundaries where building regulations were not enforced (DES master plan draft 2015, p. 27). Second, some villages outside of the urban built-up area decades ago began to expand themselves. While not originally included in the previous master plan, they are now part of the functional city area. Finally, demands for housing and development have simply outpaced government planning and enforcement capabilities.

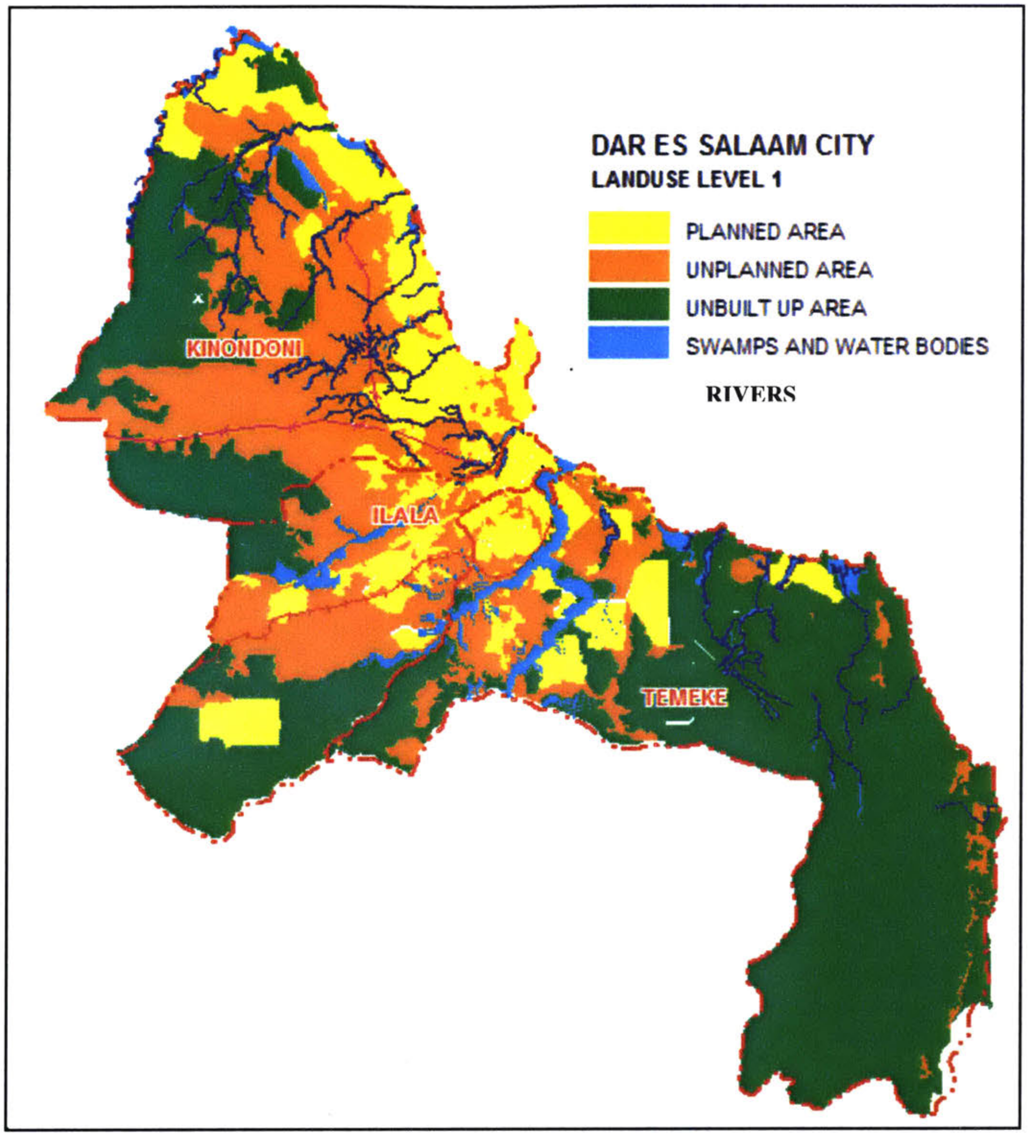


Figure 10 Planned, unplanned (informal settlements), and unbuilt-up area of Dar es Salaam (InfoBridge, 2009)

Table 3-2 Land Coverage by municipality (square km): Planned, Unplanned, Unbuilt-up, and Swamps and Water bodies (InfoBridge, 2009)

Municipality	Planned	Unplanned	Unbuilt-up	Swamps and Water bodies	Total
Ilala	41.4	121.2	126.1	20.7	309.4
Kinondoni	89.4	240	109.7	24.8	463.9
Temeke	34	85.7	424.7	64.2	608.6
Dar Total	**164.8**	**446.9**	**660.5**	**109.7**	**1381.9**
Dar Total (percentage)	11.9%	32.3%	47.8%	7.9%	

3.4.4 Problems in Informal Transit Operation

Dar's public transit system today is the result of the decline of the formal bus sector and a public policy of liberalization and privatization (JICA 2008; technical report 2, p.8-2), consistent with experiences in many other parts of the global south. A privately owned British company, the Dar es Salaam Motor Transport Company (DMT) started providing bus service since the colonial age; U safiri Dar-es-Salaam (UDA), a publicly owned company took over operations of public buses, under government control, starting in 1974. However, UDA service rapidly deteriorated and private operators moved into the market. In 1983, the national government issued a decree that officially allowed the operation of daladalas (Sohail, Maunder, and Miles 2004). While daladalas provide crucial public transport services to Dar residents, the nature of the business and its regulation pose problems in the levels of road safety and service quality.

3.4.4.1 Individual operators: maximizing profit, not service quality

Dar's privately operated public transport services are not directly subsidized and the market is competitive (JICA 2008; technical report 3, p. 2-8). When an individual decides to join the daladala business, he buys the 17-seat minibus with roughly 40,000 – 50,000

USD[15], applies for an operating permit from SUMATRA and pays SUMATRA 40,000 – 80,000 Tsh (18-36 USD) per year to operate on one specific route. If he does not drive the daladala himself, he hires a driver and a conductor for one daladala and requires them to hand in a fix amount of money every day (10,000Tsh, roughly 45 USD). In order to earn more money, the driver often speeds, competes on the road with other vehicles to load more passengers, and overloads the daladala. Owners have little financial incentive to maintain or replace their vehicles. Routes are selected based on demand. In short, the market dominates and the supply of daladala is low in suburban, poor, and inconveniently accessed places where public transport services are in most need. A similar dynamic exists for Bajaji and bodaboda owners, who often drive the vehicles themselves; these require less up-front investment but generate lower daily revenue as well.

3.4.4.2 Regulatory Shortcomings

Since the private sector is the major actor providing public transit service, the importance of effective regulation cannot be overstated. As described in section 3.3.1.3, SUMATRA is the main regulatory body for daladalas. But, the market tends to dominate route planning and enforcement. Problems include: permit allocation on a first-come-first-serve basis; lack of professional training (and even licensing) for drivers, conductors, and owners; minimal rules for vehicle maintenance and operating life; and an overall fragmented ownership structure which hampers monitoring and enforcement. Logit and Inter-consult (2006) estimated that the number of daladala vehicles reached 47,000 in 2006. Out of 255 daladala routes they identified at that time, only 181 routes were operated with permits.

3.4.5 Transportation planning, governance, and finance

Institutional shortcomings are largely at the root of the problems of poor road infrastructure, land use issues, and informal transit operation. I have described some of

---

[15] The price information comes from interviews with a bus dealer and a bus owner

the institutional challenges to land use planning above. This section summarizes some additional concerns.

The first and also the most recent transportation master plan was only done in 2008, as part of the feasibility and integration study for BRT. Prior to this, the National Transport Policy was released in 2003 as an urban and national transport guide at a very high level; most other more concrete transport studies focused on road improvement and development (Maganga 2014, p.40).

Some of the problems rest in fragmented responsibility and lack of coordination. Urban transportation responsibilities are spread vertically and horizontally among multiple national, regional, and local governments and agencies. This enlarges the complexity of enforcement and coordination and creates a gap between concrete needs and actual development (JICA, 2008; technical report 3, pp. 1-1). On land use administration, the fragmented planning process (plot preparation, planning, review, approval, etc) across local, regional, and national level of government creates loopholes on zoning enforcement. On traffic management, the Ministry of Transport, SUMATRA, PMO-RALG, DART, and local municipalities are all involved and the scopes of responsibilities are complicated and sometimes overlap with each other. Accountability can therefore be vague. For example, TANROADS is responsible for national or trunk roads and local municipalities are responsible for local roads (JICA, 2008; technical report 3, pp. 2-4). However, the categorization of trunk roads and local roads is not always clear and upgrading local roads is not the direct responsibility for either of them.

Intra-agency communication and coordination can also be problematic. In fact, different government units and agencies often have their own strategies and plans, without clear coordination, much less integration. To consolidate the fragmented responsibility, JICA proposed, in the 2008 transport master plan, to establish Dar es Salaam Urban Transport Authority (DUTA). DUTA was expected to act as an intermediate and coordinating authority between the national policy and the agencies responsible for design, management and control of the city's transport development (JICA 2008; summary, p.26). However, this authority has not yet been established.

The lack of funding poses another important constraint. TANROADS was only able to execute 35% of total required maintenance work for the fiscal year 2007/2008 (JICA, 2008; technical report 4, pp.2-14). Local municipalities are in an even more severe shortage in their transportation budgets. Local tax revenues are not high and the budget allocation to transportation was only around 0.2%-4% in fiscal year 2005/2006-2006/2007 (JICA, 2008; technical report 4, pp.2-8).

3.5 Summary and connection to research questions

The urban mobility challenges in Dar es Salaam are enormous, difficult, and complex. A variety of factors play a role: rapid population growth and city expansion, poor road infrastructure, monocentric land use structure and sprawled informal settlements, problematic informal transit operation, and institutional shortcomings. No easy solution exists. Fortunately, we have seen some endeavors towards transport improvement from both within the government and external assistance. Dar City Council has launched a program of Consensus for Urban Transport and Policy Improvement in Dar es Salaam (CUPID), as an attempt to integrate plans and policies across government entities. The Ministry of Lands, Housing and Human Settlements is developing a new master plan, the first since 1979. BRT phase I has been in full operation since May 2016 and more BRT lines are expected to be built and managed under DART with support from international organizations. SUMATRA has been working to to encourage individual daladala operators to form an operation company to formalize and standardize transit services.

As authorities work to improve Dar's mobility conditions, better data will almost certainly have to play a role. This thesis aims to provide a small step in that direction, enabling travel behavior analysis which, if scaled up, might contribute to more evidence-based planning decisions for Dar and other rapidly developing regions.

Chapter 4 Data and Methods

4.1 Future Mobility Survey (FMS) App: Data Collection Platform

FMS is a smartphone-based prompted-recall travel survey platform consisting of three components: the smartphone app, the backend server, and the web-based user interface for verification (Figure 4-1) (Zhao, Ghorpade, et al. 2015). FMS is intended to identify a user's trip origins (starts), destinations (stops), non-travel activities, modes, and routes. The smartphone app is non-intrusive, running in the background of the phone, collecting sensor data (from GPS, WiFi, GSM and accelerometer) without user intervention to minimize the app's influence on participants during their daily activities. The application is designed to be lightweight (in terms of memory use), easy to use, and energy efficient. Various approaches like time-phased data collection of, e.g., GPS signal were employed to minimize battery consumption (Nawaranthne et al. 2014), a major concern for location-based applications and which naturally poses a tradeoff with locational precision.

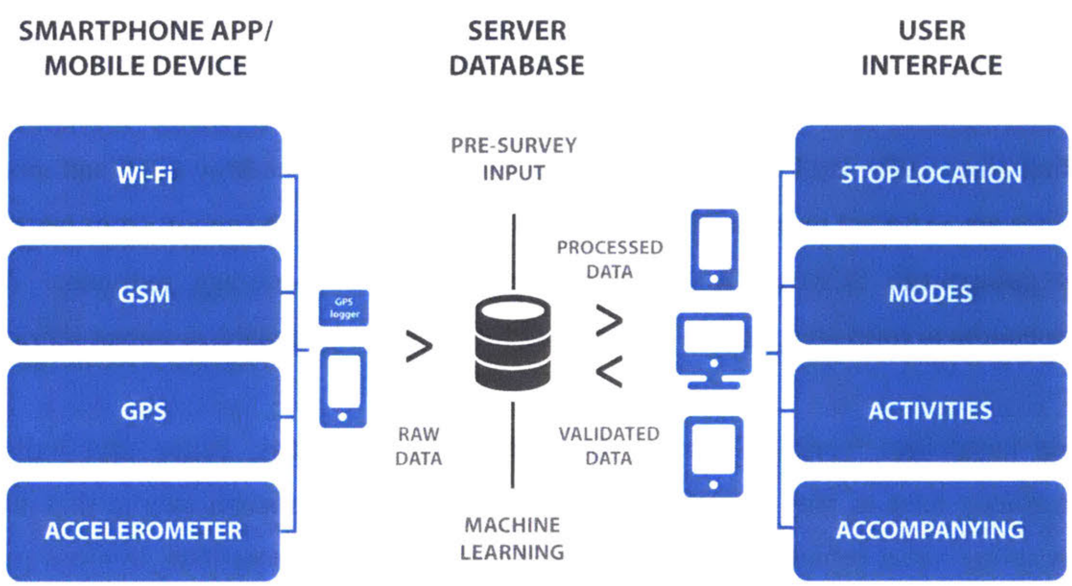


Figure 4-1 FMS Architecture

The sensor data collected on the phone are transferred to the back-end server through either the cellular network or the WiFi. The server includes the database as well as the data processing and learning algorithms for inferring stops, stop duration, stop location, travel modes, etc. The server-based inference is designed to minimize the user interaction burden during the verification stage (a main objective being to minimize false stops presented since users tend to be reluctant to change system-generated stops; Zhao, Ghorpade, et al., 2015). Detected stops play a key role in identifying activities and travel modes; the latter because detected stops also impact mode detection. FMS defines a stop based on how long a user spends time in a place performing a travel behavior-relevant activity; detecting a change of modes is relevant, for example, while delays due to traffic are not. In practice, stop detection is a challenge due to the inherent data gaps arising from the battery-saving phased sampling approach as well as other causes of poor locational information (such as densely built up areas and underground areas).

In FMS, the first round of stop detection is based on location data, while GSM, Wi-Fi and accelerometer information is used to merge stops that would otherwise be interpreted as a distinct stop. Travel modes are detected based on GPS and accelerometer features, as well as available public transit network information. Non-travel activities (e.g. home, work, change mode/transfer) are also detected based on previous validations by the user, POI data and other contextual information. When a GPS reading is unavailable, location data will come from cell tower information and WiFi (if on) which adds further noise. All of these factors impact the availability and quality of the location data collected and, ultimately, the inferences made and presented for verification.

FMS participants are presented the output of FMS inference on a website in form of an activity diary timeline. The activity diary is shown in Figure 4-2. The participants can access this diary by logging into their account on FMS website. Once logged in, the users can add or delete a stop (referred to as Non-Travel Activity on the interface), edit the stop location, edit the stop times, answer additional survey questions pertinent to the stops (eg., amount paid for parking) and travel modes (eg., number of accompanying passengers), change the mode of travel, and add or change the activity purposes for each stop. The participant can perform these operations to change the presented activity pattern if the inferences presented to them are inaccurate. If the inferences from FMS are accurate then the participants are required to confirm them by marking the stop as verified. Once all the stops and trips are verified by the user, the system will mark the day as verified.

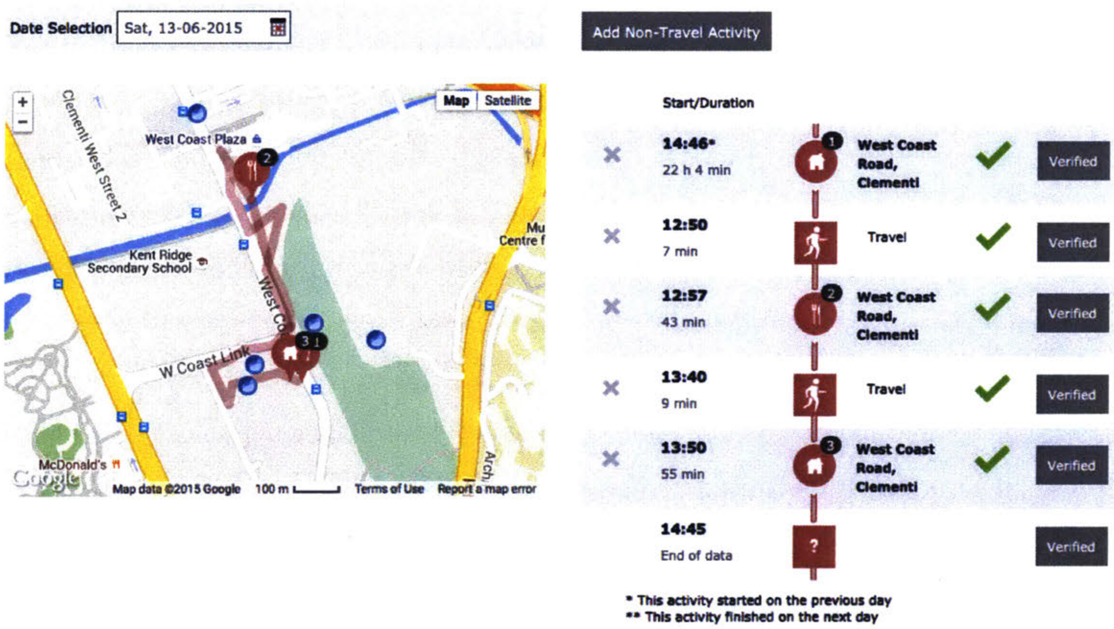


Figure 4-2 FMS activity diary timeline as presented to the participant on the FMS website for verification

FMS was originally developed in Singapore and field-tested there between October 2012 and September 2013 together with the Singapore government's traditional household travel survey (HITS 2012). That pilot demonstrated the viability of the platform and the potential richness of the data obtained, showing better accuracy in terms of reported time of day of travel, less under-reporting of trips, more accurate travel times, and very low marginal cost of additional days of data collection, revealing the relative heterogeneity of most people's travel and their tendency to self-report "typical" days (Zhao, Pereira, et al., 2015). FMS has subsequently been deployed for official household travel surveys in cities in the USA, Israel, and Singapore (2016, ongoing).

FMS applications to date reveal a number of important implications for minimizing user burden and maximizing the quality of the information generated and verified. Stop detection accuracy plays a critical role; a few erroneous stops increase verification burden, with false positives generally preferred over false negatives (Zhao, Ghorpade et al, 2015). Various problems make verification cumbersome and error-prone: long times indoors can generate signal jumps (false short trips), since inactive GPS will prompt the app to find cell tower-based location readings; capturing public transport alighting can be especially challenging, as can very short trip purposes (such as pick-up/drop-off stops); and mode detection can be tough in especially congested situations where signals among different modes are difficult to distinguish. These challenges can be overcome through learning from user validation (e.g., to learn a user's "home" signal) and by incorporating contextual information such as on transit routes and stops and POIs (Zhao, Ghorpade et al, 2015). Since to date all FMS applications have been in industrialized cities, implementation in a data-sparse setting poses a range of unique challenges.

4.2 Data

4.2.1 FMS and Baseline Surveys

The main data used in the thesis are from a World Bank project, *Sensors*. The data include the smartphone-based travel diary (581 candidates), the associated pre-survey, and citywide living standard survey with a larger sample size (around 2,400 households). The smartphone-based travel diary was collected by the World Bank project over a 4-week period in November-December 2015. The project deployed FMS among 581 adult members from 300 households selected at random from a larger sampling frame (a sample of 2,400 households previously interviewed by the Measuring Living Standards within Dar es Salaam Survey (MLDS) in January-February 2015. MLDS follows two-stage cluster sampling procedure: census is used to define primary sampling units, and then clusters are stratified at geographical level and at rural/urban clusters (Gollina, Kirchberger, and Lagakos 2015)). For each selected household, up to 2 smartphones were deployed among household members, aiming to strike a gender balance. Each selected respondent answered a short multi-topic questionnaire with individual-specific MLDS modules, and was given an Android smartphone with free data bundles and equipped with the FMS application.

Table 4-1 compares socio-demographic characteristics of the FMS candidates, FMS participants, the citywide MLDS respondents, and the Tanzania national population. The socio-demographic variables include gender, age, education, home location, job type, marriage status, driving license, index for wealth per household member, depression score, total working hours per day, and monthly sales for the family business. All these variables are used in models to measure representativeness and travel equity. The household wealth index is the composite assessment of durable goods ownership (e.g. computer, air conditioner, refrigerator, etc.), small items ownership (e.g. radio, stove, fan, etc), and housing conditions (e.g. wall materials, roof materials, water source, etc.). Principle component analysis is used to extract the score for wealth index[16]. Index for wealth per household member is the household wealth index divided by household size. The depression score is the test score of 20 questions regarding emotional depression status from the questionnaire. The MLDS provides the best estimate available of citywide statistics in DES. Compared to MLDS, FMS candidates do not include < 18 year-olds, but the age difference of adults is small (~0.6 year, significant at 10% level); FMS candidates have significantly more married individuals and self-employed individuals (also consistent with the >18 year-old constraint); FMS candidates have slightly higher years of education (~1 year), shorter distance to CBD (~0.6km), and lower individual wealth index (1%). Sample distributions of other socio-demographics in FMS are similar to MLDS.

Relative to prior and other ongoing FMS implementations, the Dar case had one key distinction. Rather than rely on the individual to carry out her own verification via the web-interface, verification was done via third party enumerators, who called the FMS users approximately every three days for a maximum 15 minutes per interview. The enumerators inquired about the travel mode, activity purpose and location, self-reported duration of trips and activities, and costs and challenges associated with trips, as

[16] I am grateful that the World Bank team provides me the wealth index they processed

displayed on the web interface. Under strict supervision and respondent corroboration, the enumerators would delete "phantom" trips/stops that were wrongfully deduced by FMS. The enumerators were asked not to add additional trips/stops that were not deduced by FMS. There were three major reasons for choosing this approach in this setting. First, not all users would be familiar with interacting with a web interface, which would imply major training efforts and risks of introducing too many errors in the case of user self-verification. Second, some users may not have access to the internet. Third, using enumerators can avoid selection bias in users choosing the simpler days to validate.

Table 4-1 Socio-demographic Statistics of FMS Candidates, FMS Participants, Citywide Survey Respondents (MLDS), and Tanzania (Mainland) National Population

		All FMS candidate		**FMS participants (who generated at least one validated record)**		**MLDS**		**Tanzania Natioal Census - Tanzania Mainland (2012)**	**Difference in mean or proportion (FMS candidates vs. MLDS)**
		mean/ proportion	*sd*	*mean/ proportion*	*sd*	*mean/ proportion*	*sd*	*mean/ proportion*	*T-test (mean)/ Z-test (proportion)*
Categorical variables	Urban	100%	-	100%	-	100%	-	29%	-
	Rural	0%	-	0%	-	0%	-	71%	-
	Gender: female	51%	-	51%	-	52%	-	51%	0.70
	Gender: male	49%	-	49%	-	48%	-	49%	0.70
	Type of primary job: Government	4%	-	3%	-	4%	-	-	0.01
	Type of primary job: Private company	15%	-	16%	-	17%	-	-	0.18
	Type of primary job: Private individual(s)	14%	-	14%	-	16%	-	-	0.27
	Type of primary job: Public company	3%	-	2%	-	2%	-	-	0.91
	Type of primary job: self-employed	64%	-	65%	-	45%	-	-	0.03
	Type of primary job: others	0%	-	0%	-	1%	-	-	0.00
	Type of primary job: no job	24%	-	20%	-	16%	-	-	0.43
	Marriage: Married (monogamous)	64%	-	66%	-	50%	-	51%	0.00
	Marriage: Married (polygamous)	2%	-	2%	-		-		
	Marriage: Living together with partner, but not married	5%	-	5%	-	-	-	7%	-
	Marriage: Divorced	2%	-	1%	-	1%	-	3%	0.17
	Marriage: Separated	3%	-	2%	-	3%	-	1%	0.42
	Marriage: Widowed	3%	-	3%	-	4%	-	3%	0.26
	Marriage: never married	22%	-	21%	-	42%	-	36%	0.00

	Driving license - car	11%	-	11%	-	-	-	-	-
	Driving license - motorbike	2%	-	2%	-	-	-	-	-
	Driving license - none	86%	-	87%	-	-	-	-	-
Continuous variables	Age: <18 (proportion)	0%		0%		38%		50%	-
	Age: >=18 (mean and sd)	35.83	10.53	36.06	10.59	35.19	13.30	-	0.07
	Education years	11.02	3.45	11.11	3.47	10.01	4.29		0.00
	Distance from home to CBD	13.14	8.36	13.16	8.18	13.77	8.46	-	0.09
	Index for wealth per household member	-0.03	0.37	-0.02	0.36	0.01	0.31	-	0.03
	Depression score	36.41	6.98	36.23	6.95	-	-	-	-
	Total working hours per day	7.82	5.47	8.10	5.40	8.11	4.54	-	0.23
	Monthly sales of the family business	226900	734075	204200	503156	335900	3974462	-	0.10

4.2.2 Paper-based Surveys

I compare the FMS data with two travel surveys collected in a traditional paper-based approach. The results are discussed in chapter 6.

4.2.2.1 2007 Household Interview Survey (HIS)

JICA conducted a household interview survey (HIS) in 2007 when they prepared the transportation master plan for Dar (JICA 2008; technical report 6). The sample size of HIS was approximately 50,000 individuals in about 10,000 households. The sample households were selected randomly based on the National Master Sampling Framework used for the National Census. The dedicated survey team collected travel information on one weekday per respondent during May-July 2007 (45 effective survey days in total). The survey is trip-based and paper-based. Questions include origin place (address, landmark, building name) and departure time, destination place and arrival time, trip purpose, transport modes, access/egress time, waiting time, travel cost, travel time and transfer point, car parking place and parking fees.

4.2.2.2 2010 Travel Survey (TS)

Behrens and Masaoe conducted a place-based travel survey in 2010 (Masaoe, Del Mistro, and Makajuma 2011; Maganga 2014). The sample size is 2008 households. The survey used stratified sampling based on pre-determined wards. The survey was paper-based, and was carried out on Tuesdays through Saturdays after 4pm from August – September in 2010. The survey team asked each respondent to complete a travel diary of the previous day. Every respondent has one weekday's travel information in the database. Behrens and Masaoe (2011) found that this place-based travel diary format yielded higher rates of trip recall and lower rates of measurement error than trip-based or activity-based surveys in Dar es Salaam and Cape Town.

Table 4-2 Data collection basics comparison among 2015 future mobility survey (FMS), 2010 travel survey (TS), and 2007 household interview survey (HIS)

	2015 FMS	2010 TS	2007 HIS
Data collection technology	Smartphone-based	Paper-based	Paper-based
Travel diary format	Activity-based	Place-based	Trip-based
Sample size	300 households 581 individuals	2008 households	10,000 households 50,000 individuals
Sample days per individual	4 weeks	One weekday	One weekday
Sample period	November-December, 2015	August-September, 2010	May-July, 2007

4.2.3 Spatial Opportunity Data

Accessibility is associated with spatial distributions of activity opportunities. In my accessibility calculations, I use two spatial opportunities to weight activity space (details in section 4.5.1): weighted population and land use information.

4.2.3.1 Worldpop

The first spatial opportunity is the average day and night population estimated by Worldpop (www.worldpop.org.uk). Worldpop is an open-access demographic archive of spatial demographic datasets for Central and South America, Africa and Asia. The project was initiated in 2013 to support development, disaster response and health applications. Worldpop was prepared by a core team in collaboration with multiple academic institutes, government units, and non-governmental funders around the world.

Worldpop provides high-resolution estimates of population at 100m * 100m level in year 2010 and 2015 in Tanzania. It uses areal interpolation to redistribute sub-national level census data to the grid cell. The weighting surface is modeled by random forest technique and inputs include various remotely-sensed and geospatial datasets (e.g. settlement locations, settlement extents, land cover, roads, building maps, health facility locations, satellite nightlights, vegetation, topography, refugee camps) (Stevens et al. 2015). Where census data are outdated or unreliable, Worldpop estimates population distribution through a combination of satellite-derived feature extractions and household surveys. From its methods, it should contain both daytime foot traffic and nighttime residential population and can be a reasonable indicator of attractions.

4.2.3.2 Remote-sensed land use data

The second spatial opportunity is land use coverage inferred from remote sensing data in 2010. The World Bank classified the remote-sensing images and I base my analysis on their land use output. Table 4-3 shows the classification accuracy of the land use data. The land use resolution is around 15m * 15m. I use regular residential areas, commercial/industrial areas and paved roads as indicators of "good" spatial opportunities in Dar. In the travel equity analysis, I treat commercial/industrial area and paved road area as proxies for activity opportunities and regular residential area as the built-environment variable at home location.

Table 4-3 Classification Accuracy of Remote-Sensed Land Use Data (provided by the World Bank)

Dar es Salaam c. 2010		Regular Residential	Irregular Residential	Commercial / Industrial	Vegetation	Bare Ground	Water	Paved Roads	Clouds/ Shadows	Total	Producer's Accuracy	User's Accuracy
		Reference Data										
Classified Data	Regular Residential	43	1	1	2	3	0	0	0	50	71.67%	86.00%
	Irregular Residential	2	47	0	1	0	0	0	0	50	94.00%	94.00%
	Commercial/ Industrial	0	1	40	0	4	0	5	0	50	93.02%	80.00%
	Vegetation	3	0	0	46	1	0	0	0	50	83.64%	92.00%
	Bare Ground	1	0	0	3	45	0	1	0	50	77.59%	90.00%
	Water	0	0	0	1	4	45	0	0	50	100.00%	90.00%
	Paved Roads	11	1	2	2	1	0	33	0	50	84.62%	66.00%
	Clouds/ Shadows	0	0	0	0	0	0	0	50	50	100.00%	100.00%
	Total	60	50	43	55	58	45	39	50	400		
	Overall Accuracy	0.873										
	Kappa	0.854										

4.3 Identify Travel Patterns

4.3.1 Basic Trip Statistics

I calculate trip rate, trip purpose, trip distance and duration, and mode shares from the 2015 FMS and compare the results with two paper-based surveys described in Table 4-1.

4.3.2 Temporal Analysis

Temporal analysis focuses on the travel patterns within one day. It includes density of trips throughout the day, trip length distribution, and travel time by mode.

- *Time of Day of Travel*: Figures plotting the percentage of users travelling for Work, Home, and other different activity types in every hour of the day are generated, to understand the travel patterns within a day.
- *Trip Time Distribution*: Compared to traditional travel surveys, a smartphone-based travel survey can provide more precise estimation of trip duration. This analysis shows the probability of trips against travel time, suggesting people's tolerance of trip duration for each activity type.

4.3.3 Spatial Analysis

Spatial analysis focuses on how people's travel behaviors interact with the city. It visualizes activity trajectory and activity density.

- *Activities Trajectory*: all the verified activities in the dataset are visualized by an animation. The visualization can help build intuition of the spatial travel patterns, identify the most congested places, and provide foundations for transport planning. It is also a powerful tool to communicate the data value of FMS and to advocate for corresponding policies by demonstrating the evidence vividly.
- *Activity Density:* this analysis maps the activity intensity in the city, to identify the most popular urban areas. Those places are indicated by the destination locations of all verified trips.

4.3.4 Daily mobility motif

Daily mobility motif represents the number of visited places per day and the order of these visits. Motifs are networks consisting of nodes and direct links, representing the locations where individuals perform activities and the trips between these locations

(Schneider et al. 2013; Jiang, et al., 2015). Figure 4-3 shows the examples of individual activity pattern structure (daily tours) commonly observed in a travel survey and represented in transportation modeling language (Jiang, et al., 2015). I use daily mobility motif as one of the indicators to depict travel patterns in Dar es Salaam. The level of complexity of the daily mobility motif may be correlated with the level of convenience of traveling in the city. Therefore, the distribution of daily motifs can represent some level of travel equity. However, daily motif is not the perfect indicator of travel welfare because it does not consider the spatial distribution of activities. A more refined approach to analyze travel equity is presented in section 4.5.

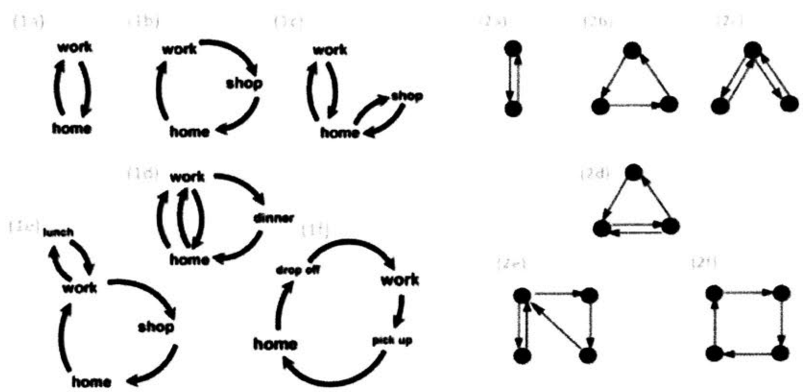


Figure 4-3 Examples of daily mobility network (Jiang, et al., 2015)

4.4 Measuring Participant Representativeness

To assess the basic representativeness of the participants in the FMS data collection pilot, I model a recruited individual's likelihood to participate at all (binary logit model). To analyze the relative degree of participation effectiveness and the role of individual- and household-level factors, I estimate a count model of an individual's fully verified days (zero-inflated Poisson model). Finally, to examine those factors, including the role of relative travel burden, influencing the likelihood of verification status at the day level, I estimate a binary logit model.

4.4.1 Measures of Travel Behavior

As discussed above, the raw sensor data collected via FMS are inherently noisy. These imperfections in the sensor data result in inaccurate inferences by the FMS algorithms, thus increasing verification burden. The complexity of the travel activity pattern can also be associated with increased verification burden. Hanson and Hanson (1981) identified seven groups of travel-activity pattern by principle component analysis (PCA): 1) frequency of travel, 2) dispersion of destination visited, 3) shopping, variety, and multi-stop travel, 4) travel to work, 5) social travel, 6) travel to recreation, and 7) overall distance traveled. Due to the challenges in travel mode and activity detection without verification, I do not use measures related to particular activity purpose (indicators 3-6). I use number of stops detected (a measure of frequency of travel), total distance traveled, distance from home to activities centroid and standard radius of activities (measures of dispersion of destination visited) as four indicators of travel burden[17].

4.4.2 Mixed logit model

I use the mixed logit technique (Hensher and Greene 2003) to estimate two binary choice models: to model who are non-participants in FMS and to model which days will be fully verified. The fixed effects are for attributes of households, individuals, and/or days. The models include random effects, for unobserved factors correlated within households and individuals. The random effects framework follows statistical theory dealing with panel data (Cameron and Trivedi 2005; Greene 2003; Wooldridge 2010; Dinardo, Johnston, and Johnston 1997). I use the glmmADMB package (Fournier et al. 2012; Skaug et al. 2012) in R to implement the mixed logit model. GlmmADMB uses Laplace approximation to fit the model. The model formulation for participating or not at the individual level is

$$y_{ih} = S_{ih} + S_h + \alpha_h + \varepsilon_{ih} \tag{1}$$

[17] I am grateful to my colleague Ajinkya Ghorpade for generating these indicators from the FMS inference algorithms (non-verified results).

where y_{ih} is the binary choice whether the individual i from household h effectively participated in FMS; S_{ih} and S_h are attributes associated with the individual i and household h respectively; α_h is the random effects capturing correlated error term within household; ε_{ih} is the independently and identically distributed (i.i.d.) noise. The model formulation for whether an individual's given day is fully verified is

$$y_{ihck} = S_h + S_{ih} + S_{ihc} + \lambda(T_k) + \alpha_h + \alpha_{ih} + \varepsilon_{ihk} \quad (2)$$

where y_{ihck} is the binary variable whether day k of individual i from household h is fully verified by enumerator call c; S_h, S_{ih}, S_{ihc} are attributes of household h, individual i, and call c; α_h, α_{ih} are random effects for household and individual; and, ε_{ihk} is i.i.d. noise.

4.4.3 Zero-inflated Poisson (ZIP) model with random effects

I use a ZIP model to estimate the number of fully verified days every participant generated. ZIP is used to model count data that have an excess of zero counts. In a ZIP model, a count response variable is assumed to be distributed as a mixture of a Poisson (λ) distribution and a distribution with point mass of one at zero, with mixing probability p (Hall 2000; Lambert 1992). I incorporate random effects to accommodate within-household correlation and between-household heterogeneity. I implement the model in Mplus 7 (Muthén and Muthén 2007). The model formulation for the number of fully verified days is

$$N_{ih} = S_{ih} + S_h + \alpha_h + \varepsilon_{ih} \quad (3)$$

where N_{ih} is the number of completely verified days of person i in household h; S_h, S_{ih} are attributes of household h and individual I; α_h are random effects for household; and, ε_{ihk} is i.i.d. noise.

4.4.4 Spatial autocorrelation test: Moran's I

Except for possible unobserved factors correlated within households and individuals, spatial autocorrelation (lag or error) may exist. As it adds considerable complexity to incorporate both spatial and non-spatial random effects into generalized linear mixed models, I conduct post-estimate test for spatial autocorrelation. I test for spatial autocorrelation using Moran's I (Moran 1950) of the model residuals. The value of Moran's I can range from -1 to +1. Negative values indicate negative spatial autocorrelation and positive values indicate positive spatial autocorrelation.

4.5 Measuring Travel Equity

4.5.1 Individual Travel Welfare Measure: Potential Path Area

I use the concept of Hägerstrand's space-time (ST) prism, as described in Chapter 2, to measure an individual's accessibility (ability to reach locations given daily activity program and spatial-temporal constraints. Kwan and Hong (1998) formulate the space-time prism or potential path space (PPS) as:

$$PPS_{sa} = \left\{(k,t)\middle|t_i + \frac{d_{ik}}{v} \leq t \leq t_j - \frac{d_{jk}}{v}\right\} \tag{4}$$

where (k,t) is all the possible activity locations for flexible activity set sa within reach in PPS; t_i is the earliest ending time of the previous fixed activity at location i; t_j is the latest starting time of the fixed activity at location j after activity a; v is the average travel speed on the transport network; d_{ik} is the distance from the first fixed location i to location k; d_{jk} is the distance from the next fixed location j to location k.

The potential path area (PPA) is the projection of PPS on two-dimensional geographic space (H. J. Miller 2006). PPA is the area of all possible activity locations k, which can be defined as:

$$PPA_{sa} = \{Area(k)|(k,t) \in PPS_{sa}\} \tag{5}$$

To incorporate opportunity density into consideration, the weighted potential path area (WPPA) can be described as:

$$WPPA_{sa} = \{w * Area(k) | (k, t) \in PPS_{sa}\} \qquad (6)$$

where w is the weight to show the relative importance of the area k. Weights in previous studies include the number of opportunities, area-weighted sum of opportunities, weighted area of opportunities only for travel conducted between 9am and 6pm, and length of road segments in the PPA (Weber 2003; Kwan 1999). In this thesis, I use the average day and night population and land use mix (commercial area, formal settlement area and paved road area within 1km buffer) as the spatial opportunity indicators to calculate WPPA.

While it is relatively easy to describe PPA conceptually, it is not straightforward to operationalize this abstract concept in a data sparse setting such as Dar es Salaam. Without information on the detailed transit network, segment-specific traffic speeds, and congested speeds at different time of the day, calculating network-based measures can be inaccurate. In light of this data challenge, I choose a geometry-approximation method to evaluate monthly potential path area (MPPA) and create a speed prediction model to calculate daily potential path area (DPPA).

4.5.1.1 Operationalization (A): Confidence Ellipse of Monthly Visited Places

The first approach is to measure the spatial distribution of the locations a traveller has visited according to the travel diary over a month. The geometry and size of the MPPA is determined by the location of home, regular activities, and subsequent travel (Golledge 1997; Schönfelder and Axhausen 2003). In this approach, the spatial distribution of locations is not calculated as accessible area between fixed activities as described by equation (2), but is approximated by all the visited places over several weeks. The confidence ellipses drawn to cover these trip destinations represent the MPPA.

The confidence ellipses are analogous to the confidence interval of univariate distribution as the smallest possible area in which the true value of the population should be found with a certain probability (Schönfelder and Axhausen 2003). In the context of spatial behavior, the confidence ellipse describes the dispersion of activity locations or the potential accessible area of an individual. The size of the ellipse or the opportunities within the ellipse represents PPA or weighted PPA.

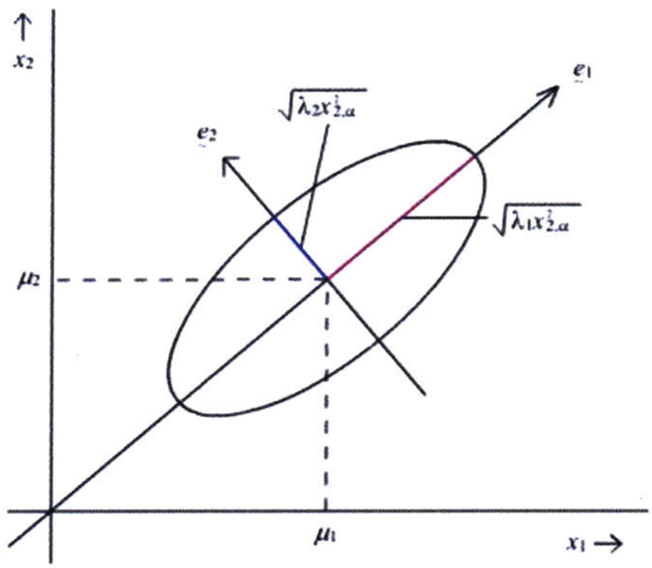


Figure 4-4 Illustration of the prediction ellipse of the bivariate normal distribution (PennState Eberly College of Science 2017)

The stop location distribution is assumed to be a bivariate normal distribution. In a bivariate normal distribution, the ellipsoid of the $(1-\alpha)$ confidence region is at the means (μ_1, μ_2). The longer axis of the ellipse is in the direction of the first eigenvector e_1 and its length is $\sqrt{\lambda_1 \chi^2_{2,\alpha}}$, where λ_1 is the first eigenvalue and α is the significance level; the shorter axis is perpendicular to the longer axis and is at the direction of the second eigenvector e_2 with the length of $\sqrt{\lambda_2 \chi^2_{2,\alpha}}$. The area of the two-dimensional ellipse is $\pi \chi^2_{2,\alpha} |\Sigma|^{1/2}$, where Σ is the covariance matrix (Chew 1966; PennState Eberly College of Science 2017).

Figure 4-5 shows an example of one FMS participant's all verified trip destinations within four weeks and the inferred confidence ellipse. In the final calculation, I use a 95% confidence ellipse based on all verified stops. This approach has at least three limitations. First, an ellipse may not be a realistic representation of the actual travel behavior. In reality, the PPA can be an irregular shape and an ellipse can over-simplify the complexity in trip destinations. The unrealistic shape can overestimate the PPA for individuals who have to travel further to reach their destinations but do not have many chances to explore opportunities around the destinations. Second, the ellipse only considers the spatial distribution of activities but does not explicitly consider the individual's time constraints. Finally, missing data will influence the results. For example, a small resulting PPA may be due to actually low accessibility or to missing stops. This may be problematic in the analysis of Dar, where the FMS implementation resulted in a fair portion of not fully verified days (details in Chapter 5).

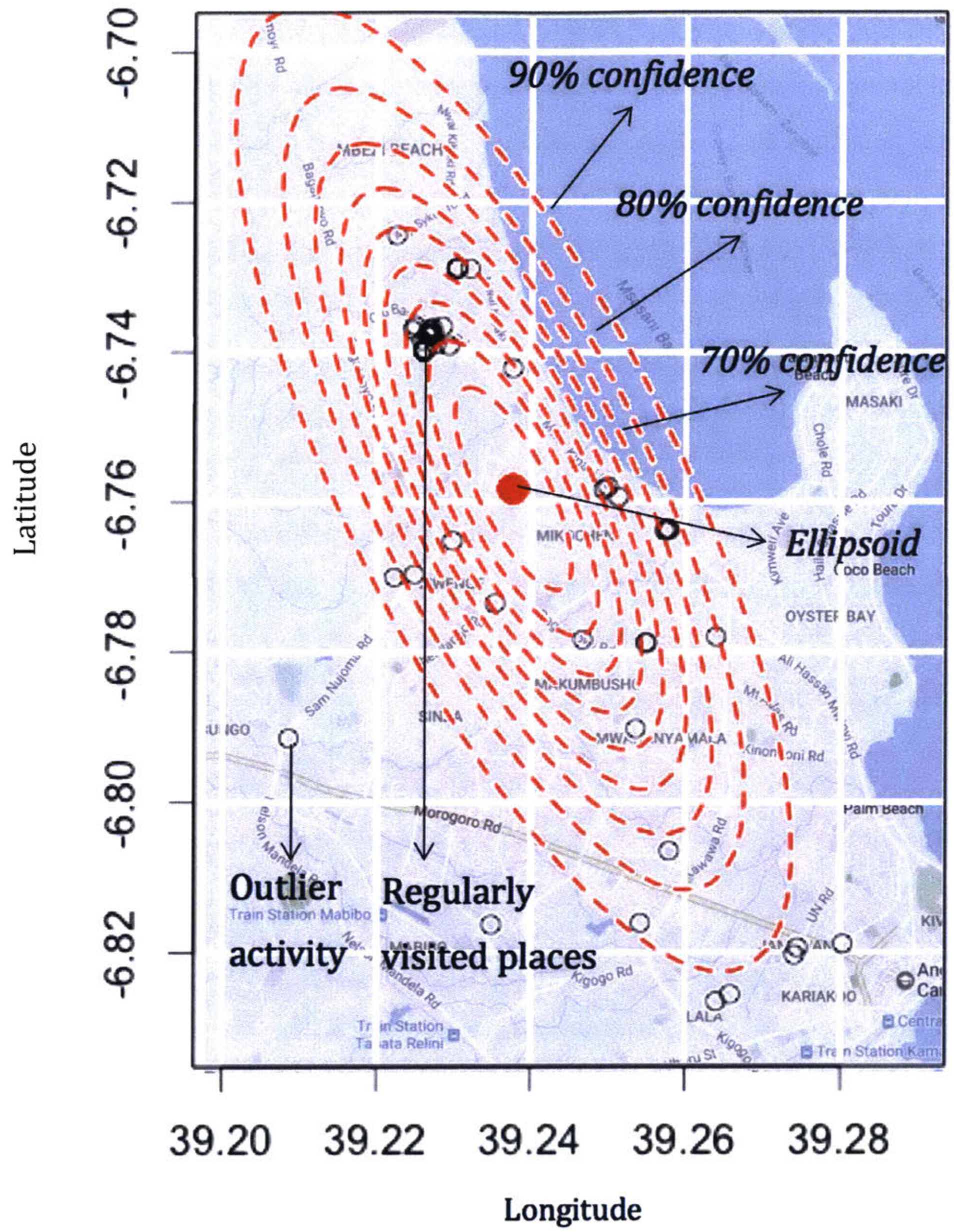


Figure 4-5 Confidence Ellipse Example of One Participant

4.5.1.2 Operationalization (B): Network-Based Daily Potential Path Area

The second approach uses a road network-based calculation to measure the area that an individual can access between pairs of fixed locations given time constraints within one day. Some daily activities are fixed (e.g. home, work school, and doctor appointment) while others are flexible (e.g. grocery, shopping, entertainment, restaurant, etc.). The area that an individual can reach between one day's fixed activities is the DPPA. It depicts the potential opportunities one can access given the spatio-temporal constraint. Deriving the DPPA requires four steps:

1. select fixed locations of one fully verified weekday for as many participants as possible;
2. prepare the road network with segment-specific speeds;
3. generate PPAs between each pair of fixed locations (equation 5);
4. take the union of all the PPAs for one day to obtain DPPA and weight DPPA by spatial opportunities (equation 6).

For step one, I randomly select a one-day travel diary from participants who had at least one fully verified weekday and prepared the fixed locations. Although FMS was conducted for a four-week period, not all participants fully verified a weekday, for various reasons described in Chapter 5. Only 333 people have at least one fully verified weekday out of 581 candidates. As DPPA has to be calculated based on the precise travel diary, having a completely verified day is important. For those people (188) with multiple fully verified weekdays, I randomly selected the weekday from the final week of the FMS implementation, when people were most familiar with FMS. I defined fixed activities as home, work, work-related, school, and religious activity. For the 333 fully verified weekdays, I extracted 827 fixed locations as the anchors for calculating DPPA.

For step two, I use the road network from Open Street Map (OSM). Dar's OSM does not have any speed information, so I use the observed (from FMS) median travel speed by different mode (Table 4-4, second column). This simplifies the actual travel speeds which may greatly vary, depending on road types, modes, locations, and time of the day.[18] I calculated travel speed for each mode using data from the FMS travel diary, which contains the coordinates and starting and ending times of each stop. Using the coordinates, I estimated the distance between each stop via shortest path, via ArcGIS Network

[18] I developed and tested a speed estimation model, but the explanatory power was weak and risked introducing too much noise to the calculation. I documented the speed model in the Appendix.

Analysis. This distance estimate, combined with the FMS-based travel time estimate, provided estimated travel speeds by mode (Table 4-4).[19]

Table 4-4 Estimated Travel Speed, Travel Duration, And Travel Distance By Mode

	obs	travel speed (kph)		travel duration (min)		shortest-path travel distance (meter)		straight-line distance (meter)	
		mean	median	mean	median	mean	median	mean	median
bicycle	8	11.1	8.0	27.9	27.6	3908	3277	1964	1080
bus (daladala)	243	13.7	11.6	49.0	37.4	8356	6911	6675	5629
foot	372	6.7	5.3	14.6	10.3	1360	909	870	517
motorcycle	71	20.0	18.0	16.7	12.1	4931	3914	3742	2793
taxi	123	16.1	13.5	22.7	10.6	4078	1975	3014	1272
vehicle	75	13.0	11.3	26.3	14.4	4670	2704	3668	2298

For step three, I generate PPAs for flexible activities between each pair of fixed locations, or identify the network area that an individual can access given observed constraints of time, origin, and destination. This step operationalizes equations (4)-(5). As current geospatial platforms do not provide easy solutions for this purpose, I derive the PPAs by iteratively intersecting isochrones between origin i (the previous fixed location) and destination j (the next fixed location) (Kwan and Hong 1998). More specifically, the time between the ending activity at i and starting activity j are the upper-bound time budget with which an individual could theoretically explore other spatial opportunities. The time budget is divided into various combinations $(I_n + J_n)$ of travel time from i to the k, the location of the flexible activity, and from k to j, with a constant increment. For an example of 30-minute budget, the travel time combinations are 5min+25min, 10min+20min, 15min+15min, 20min+10min, and 25min+10min. For each of the travel time combinations $(I_m + J_m)$, an I_m-minute isochrones of i, or A_{I_m}, represents all the

[19] I did this calculation based on a random selection of 956 stops. There are 14408 verified trips in total, but calculating shortest paths for all of them via ArcGIS took too long to run. In addition, occasional snapping problems can crash the running; I had to manually remove those points and continue the script execution. Generating trip distances for these 956 stops takes approximately three days. Calculating distances for all 14408 trips would have been infeasible for this thesis.

area that can be reached within I_n minutes starting from i. Similarly, A_{J_m}represents the area that can be reached within J_m minutes ending at j. The intersection area of A_{I_m} and A_{J_m}, denoted by $Inter_{I_m+J_m}$, is the area that an individual can access given the origin and destination (i,j) constraint for a particular time combination $(I_m + J_m)$. Finally, the union of all $Inter_{I_m+J_m}$, or U_{ij}, is the PPA between fixed locations (i,j) given a time constraint (Figure 4-6). These isochrones are generated by python geospatial packages (Shapely and Fiona) and Mapzen's isochrone function.[20]

Finally, I take the union of all the PPAs of one person's day, visualize and calculate the size of the DPPA, and weight the DPPA by spatial opportunities (equation 6). ArcGIS can easily handle these functions. The results, DPPA size and weighted opportunities, are used to evaluate society-wide travel equity.

Figure 4-7 illustrates the derived DPPA of one participant with seven stops: home – first shopping – first workplace – second workplace – first workplace – second shopping – home. This individual's fixed activities are home, first workplace, second workplace, and home. I calculated the four PPAs for consecutive fixed stops and took the union of these PPAs to obtain DPPA.

This second approach addresses the two limitation of the confidence ellipse approach. However, in my application this approach is still not perfect. First, the sample size is small. The one weekday diary may not be representative enough of the person's typical travel pattern and the results may contain noise from the randomly selected weekday. Second, the starting and ending time of fixed activities have some flexibility to change in real life, but this flexibility is not represented in this approach. This can result in underestimating DPPA. For example, if an individual is relaxing at home for the whole day, the person's DPPA is zero. But the time of the person's home activities are in fact not fixed for most time of the day. Considering the actual flexibility in time budget would increase the DPPA size. Ideally, we would collect the time flexibility of fixed activities

---

[20] Mapzen is an open-source routing platform (https://mapzen.com/documentation/mobility/isochrone/api-reference/).

through a survey, but the Dar implementation of FMS did not collect such information. Third, it simplifies the transportation system performance. For example, I did not have the daladala network data, so daladala route limitations were not included in the calculation (daladalas were assumed to simply follow the road network, not a specific route). This likely overestimates the PPAs for daladala users.

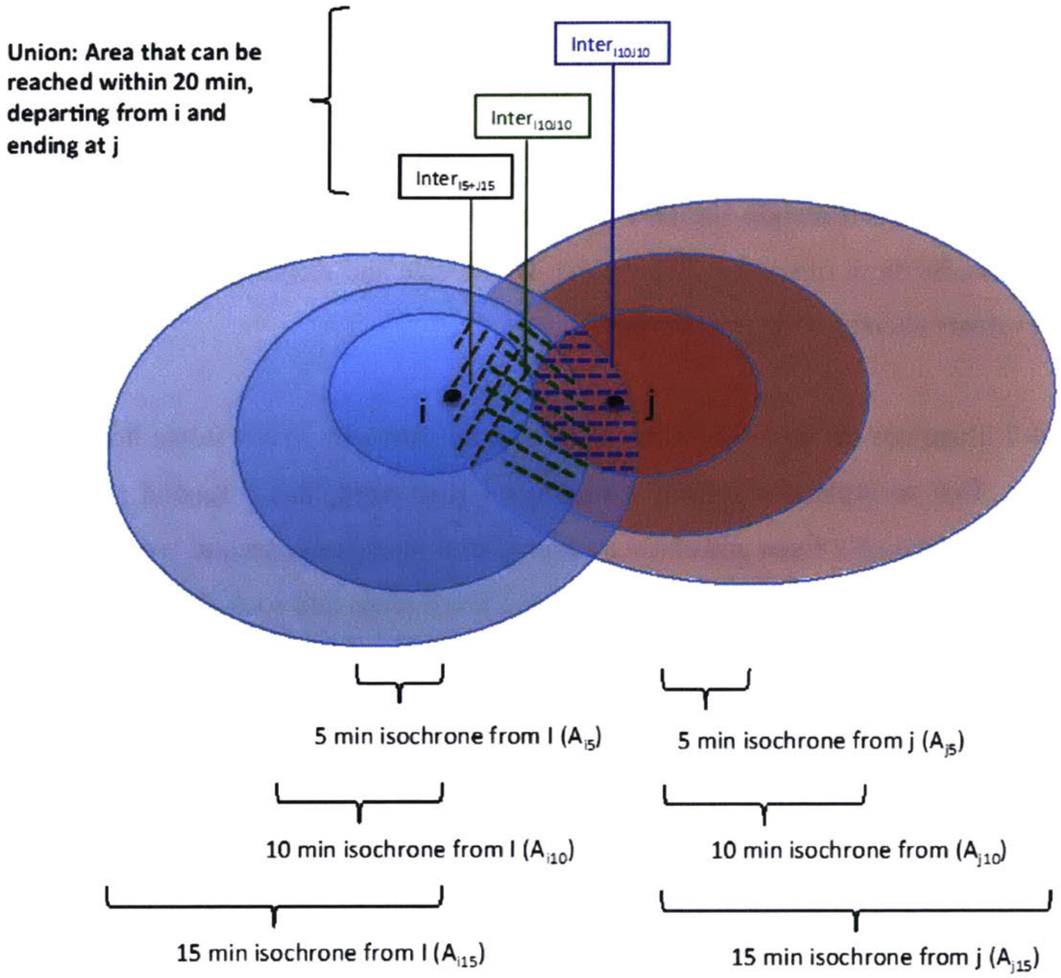
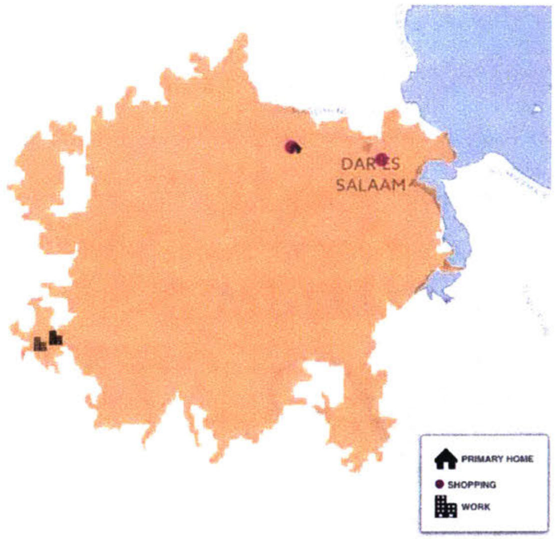
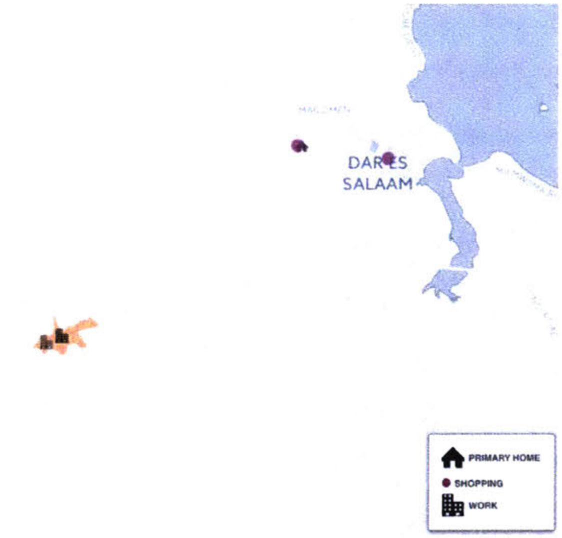


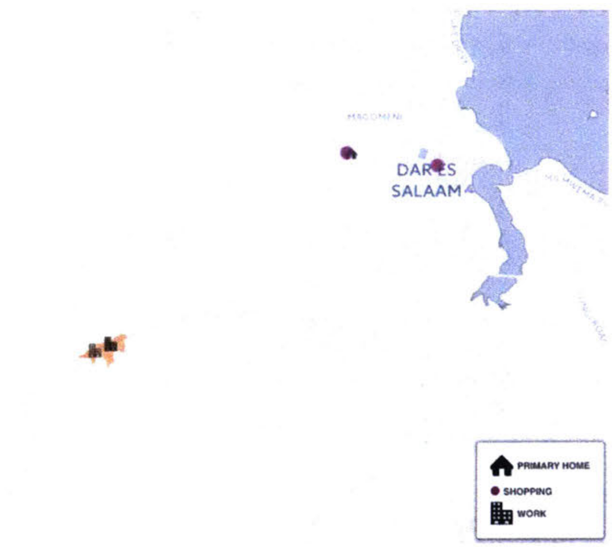
Figure 4-6 Illustration of generating the PPA between two fixed locations (i and j)



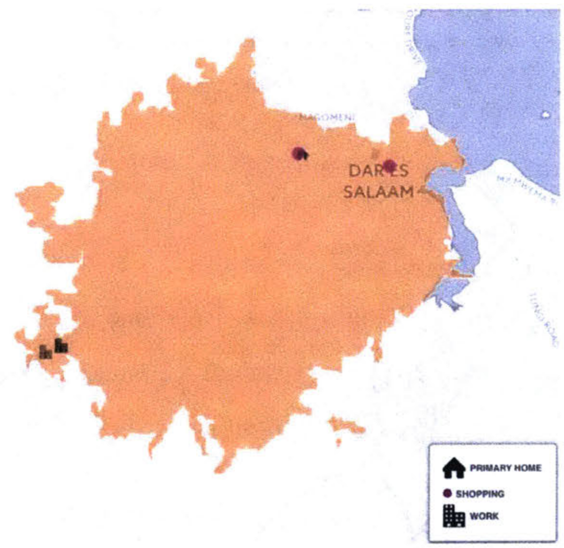
1st PPA: between home and the first workplace



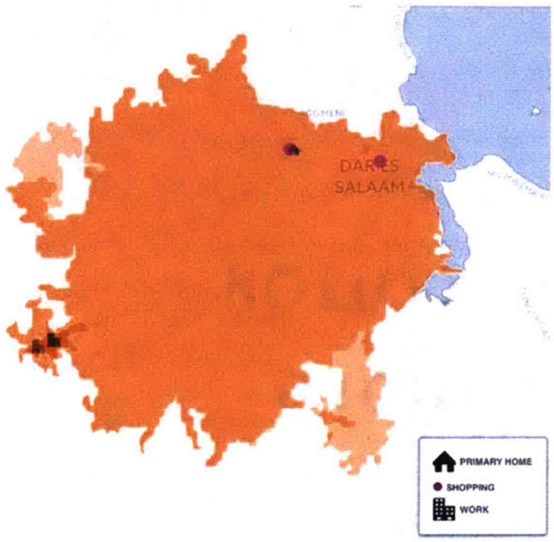
2nd PPA: between the first workplace and the second workplace



3rd PPA: between the second workplace and the first workplace



4th PPA: between the first workplace and home



DPPA: union of all PPAs

Figure 4-7 Illustration of the DPPA of one participant (travel diary: home – shopping –first workplace – second workplace – first workplace - shopping – home)

4.5.2 Society-Wide Equity

As discussed in Chapter 2 this study adopts two principles to evaluate society-wide equity: equality of welfare and equality of effectiveness. Equality of welfare holds that all individuals have equal geospatial opportunities and the same level of travel welfare. Equality of effectiveness argues that inequality of travel welfare is legitimate as long as (a) people are above the minimum basic capabilities and (b) the welfare of an individual is proportional to its cost. For the first principle, I use a Gini coefficient and Lorenz Curve to evaluate the disparity of travel welfare. For the second principle, I use housing cost and the generalized transport cost to adjust travel welfare. I first derive a LOESS curve to define the "sufficient" welfare effectiveness line. People below the line presumably suffer from deficient travel welfare compared to their cost. I then define a sufficient cost affordability line by the median cost. This represents the minimum level of capability an individual should have to pursue his or her well-being (Sen 1979; Pereira, et. al., 2016). Finally, I use the severity index of travel welfare deficiency to evaluate society-wide equity. The sub-sections below discuss the evaluation methods in more detail.

Table 4-5 Summary of Society-Wide Equity Principles and Evaluation Methods

Equity principles	Root justice theory	Distribution of what?	The fairest distribution	Evaluation methods
Equality of welfare	Egalitarianism	Travel welfare, measured by individual accessibility	Equal distribution of travel welfare	Gini Coefficient And Lorenz Curve
Equality of effectiveness	Capabilities approach	Cost-adjusted travel welfare. The cost includes housing cost and generalized transport cost	1) People who pay the same cost should enjoy the same level of welfare 2) Everyone should have the minimum cost affordability to guarantee basic travel welfare	• Sufficient welfare effectiveness line: LOESS curve • Sufficient cost affordability line: median cost • The Severity of Travel Welfare Deficiency

4.5.2.1 Equality of Welfare: Gini Coefficient And Lorenz Curve

Lorenz curves are a graphical representation of the cumulative distribution function of wealth across the population (Lorenz 1905). The Gini coefficient is a single mathematical metric to represent the overall degree of inequity based on that cumulative distribution function (Gini 1912). Lorenz curves and Gini coefficients are commonly used to show the income disparity and have been adapted to multiple fields including transport (Delbosc and Currie 2011a). I adapt Lorenz curves and Gini coefficients to measure the inequity of travel welfare in Dar es Salaam. Specifically, the Gini coefficient is given by:

$$G_1 = 1 - \sum_{k=1}^{n} (X_k - X_{k-1})(Y_k + Y_{k-1}) \tag{7}$$

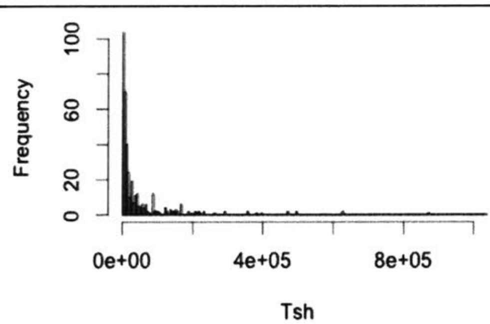
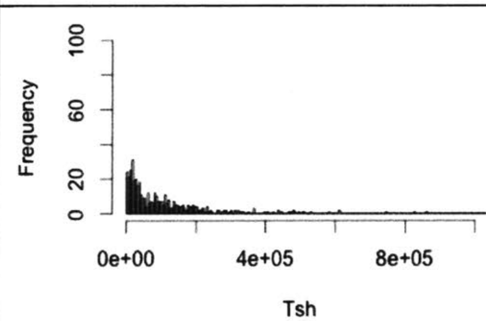
where X_k is the cumulated proportion of the population variable, for k=0,...,n, with $X_0 = 1, X_n = 1$. Y_k is the cumulated proportion of the travel welfare variable defined by activity space, for k=0,...,n, with $Y_0 = 1, Y_n = 1$.

4.5.2.2 Equality of Effectiveness: Travel Welfare Adjusted by Housing and Generalized Transport Cost

Equality of effectiveness considers individuals' different capabilities and preferences to convert spatial opportunities into welfare. Instead of focusing only on travel welfare, this principle adjusts welfare by the price people pay to gain it. The price includes transport-related economic cost. More precisely, I use estimated housing and generalized transport costs. Housing located at a place with higher accessibility often costs more and this trade-off is often at the discretion of a household. Generalized transport cost includes financial (out-of-pocket) costs (e.g. transit fare, gas cost, parking fees) and the monetized time cost.

To estimate housing costs, I standardize it as the monthly rent per person adjusted by rooftop area. First, for house owners, I amortize the market values of their houses into the monthly rents, using 12%, the benchmark interest rate in Tanzania (Trading Economics 2017b). The amortization term is 15 years, the maximum mortgage term of Bank of Africa (Bank of Africa 2017). Second, I adjust the rent so that every person has 25-square-meters of roof-top area. Both house attributes (e.g.. square footage, number of rooms, house structure) and location characteristics (e.g. distance to bus stops, distance to city center) influence housing price, but house attributes are irrelevant in the discussion of travel welfare. To decompose the cost for housing location or accessibility, I control the impact of housing size by standardizing the rent using the median roof-top area per person (25m^2). The adjusted rent for each individual is $\frac{household\ rent}{house\ rooftop\ area * household\ size} * 25\ (m^2)$. Ideally, I should use square footage instead of roof-top area and standardize the effect of other housing attributes, but those data were not available.

For generalized travel cost, the first component is financial cost, including transit fare, amortized car price, gas cost and parking fees. The amortized car price is 50 Tsh per minute, based on the assumption of purchasing price of 40 million Tsh, durability at 400,000km, and the average speed of 30km/h[21]. The gas cost per minute is estimated to be 49 Tsh per minute, with the gas price at 1959 Tsh/liter (Global Petrol Prices 2017), fuel efficiency at 5 liter/100km, and the average speed at 30km/h[22]. The parking price and transit fare come from those reported in the FMS travel diary. The second part of generalized travel cost is the monetized time value. There are debates on the fair methods to evaluate time value for individuals (Galvez and Jara-Diaz 1998); I simply use JICA's estimate of the value of time by vehicle type based on equivalent income. The time value of people on foot, bicycle, or bus is 602 Tsh per hour and that of people use car, taxi, motorcycle, and bajaji is 1846 Tsh per hour in 2007 (JICA 2008; technical report 5, p3-3). I inflate those values to 2015 using a 5.2% inflation rate (Trading Economics 2017a). Figure 4-8 shows histograms of the estimated housing and generalized transportation cost.

	Housing cost	**Generalized transporation cost**
Total	 Histogram of adjusted monthly housing cost (equivalent monthly rent per person adjusted for roof-top area)	 Histogram of generalized monthly transportation cost (including transit fare cost, car cost, and time value)

[21] The parameter assumptions come from Numbeo website – Dar es Salaam (https://www.numbeo.com/gas-prices/in/Dar-Es-Salaam). Numbeo contains user-contributed data on living conditions in cities worldwide.
[22] Source: Numbeo – Dar es Salaam (https://www.numbeo.com/gas-prices/in/Dar-Es-Salaam)

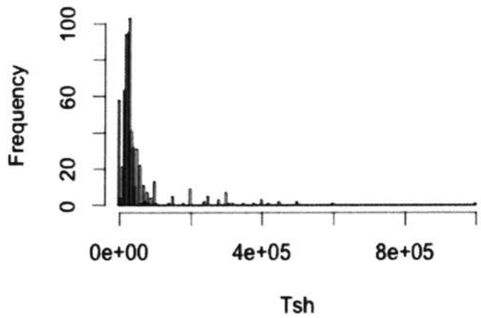
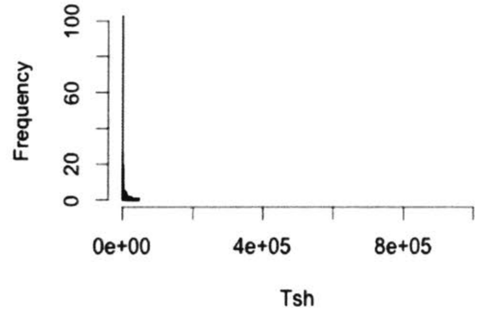
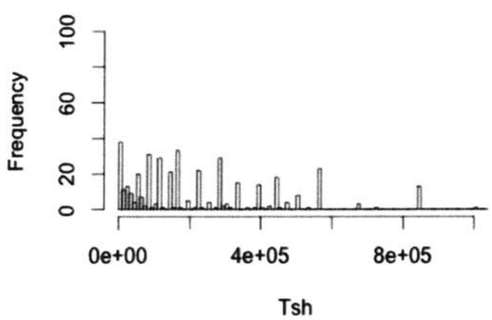
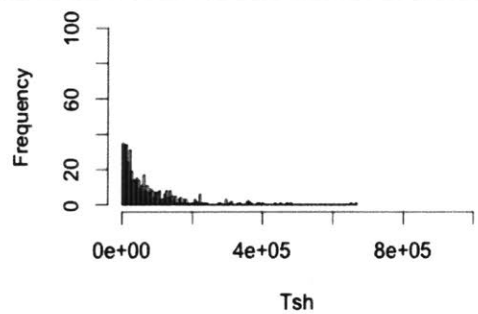
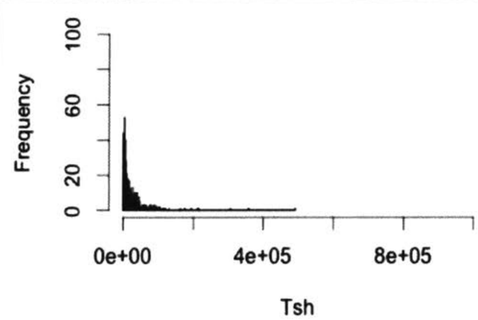
Component (A)	 Histogram of monthly rent per household (for house renters)	 Histogram of monthly transit fare cost
Component (B)	 Histogram of monthly amortization per household (for house owners)	 Histogram of monthly car cost (including parking fees, gas cost, and car amortization)
Component (C)		 Histogram of monthly time value during travel

Figure 4-8 Histogram of Housing and Generalized Transportation Cost

4.5.2.3 Equality of Effectiveness: Local Regression (LOESS) Curve

Local regression (LOESS) is a non-parametric smoothing technique to depict relationships between variables (Jacoby 2000). It combines multiple regression and knearest neighbor to present the trend underlying data. The weight function (Cleveland and Loader 1996) is

$$w(x) = (1 - |x|^3)^3 I[|x| < 1] \quad (8)$$

In the two-dimensional scatterplot of travel welfare and cost, I derive a LOESS curve to define the sufficient welfare effectiveness line. The curve is analogous to median travel welfare at a given cost. This approach to determine the "sufficient" line is adapted from Martens's (2016) method to define sufficient accessibility as the median accessibility of the city. In the transportation field, more generally, it is difficult to draw a travel poverty line because of the variety in city contexts as well as individuals' preferences and needs. Martens (2016) uses median accessibility to define sufficient accessibility to account for differences in context. People who are below the line are suffering, relatively, from deficient travel welfare compared to their cost.

4.5.2.4 Equality of Effectiveness: The Severity of Travel Welfare Deficiency

Martens (2016) also introduces the severity of travel welfare deficiency, adapting this measure from a poverty measure proposed by Foster, Greer, and Thorbecke (1984). The poverty measure gives a weight to each poor person of a group based on the size of the income shortfall. I modify the equation slightly to evaluate the travel fairness in Dar using sample data from FMS:

$$P = \frac{1}{N}\sum_{i=1}^{q}(\frac{z - y_i}{z})^2 \quad (9)$$

where P is the severity of travel welfare deficiency, q is the number of people with travel welfare no greater than the sufficiency line z, y_i is the travel welfare of person i who is below the sufficiency line z, and N is the total number of people. The value of P ranges from 0 to 1, with 0 representing no "poverty" in travel welfare and 1 indicating the entire population suffers from an insufficient travel welfare level. In this study, the sufficiency line z is defined by the LOESS curve between travel welfare and its cost.

4.5.3 Factors Influencing the Level of Travel Welfare: Linear Mixed-effects Regression

I use a linear mixed-effects model to identify factors that contribute to low level of travel welfare measured by PPAs or weighted PPAs. The MPPA and DPPA, or weighted MPPA and weighted DPPA are dependent variables in separate regressions. Independent variables include socio-economic characteristics (gender, wealth, age, education, total working hours, driving license ownership) and built environment features (living in formal settlement, access to major road, and access to the market). Those independent variables are fixed effect. The sample has two members per household. I use random effect for the household to capture potential within-household correlation in the error term.

Chapter 5 Travel Behavior and Patterns

5.1 Challenges in Data Collection

Despite the usage of third-party verification to minimize possible user errors, the Dar implementation still reveals many unexpected social and technical challenges associated with smartphone-based travel survey data collection in the global south. I define the degree of participation effectiveness as the number of fully verified days, in order to reveal individuals' complete day travel and activity patterns. Verification quality can be assessed in two different ways in the Dar implementation: FMS diary-based, simply using the FMS diary and considering a "complete" day to be one in which every trip and stop has been verified; or enumerator-based, using the enumerator's evaluation of the day verification status. The latter is arguably a less rigorous definition of verified days since enumerators may tolerate minor incompleteness and still label a given day as fully verified in the accompanying interview notes available. Out of 581 candidates who completed the pre-survey and received the free smartphones, 482 people provided at least one verified trip/stop, 354 people had at least one fully verified day using the enumerator-based criterion, and 329 had at least one fully verified day according to the FMS diary-based criterion. Out of 13,944 person-days (581 candidates times 24 active survey days), the FMS diary-based criterion indicates 1087 fully verified days, or 2.1 days per candidate; the enumerator-based criterion indicates 1574 fully verified days, or 2.83 days per candidate (Figure 5-1; Table 5-1).

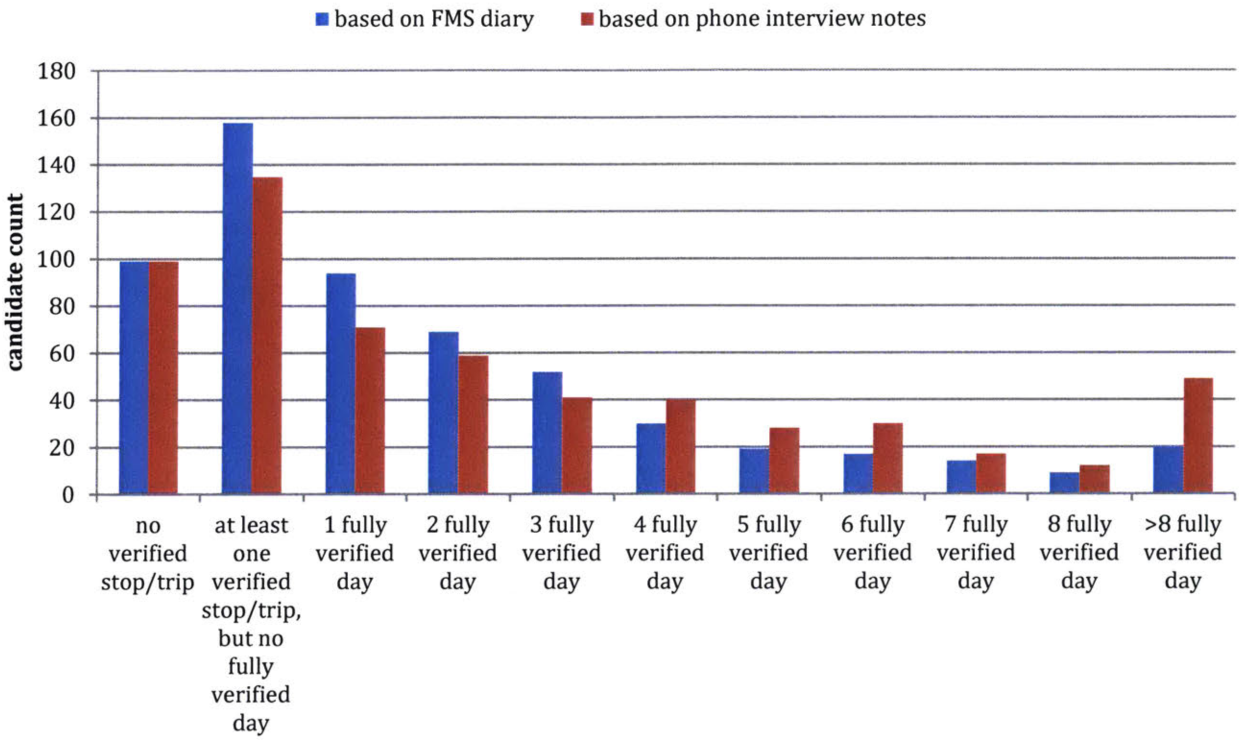


Figure 5-1 Distribution of FMS candidates according to the degree of participation

Table 5-1 Statistics on the number of fully verified days per person

	FMS diary-based		**Enumerator-based**	
	All participants	**Participants with > 0 verified days**	**All participants**	**Participants with > 0 verified days**
Mean	2.1	3.3	2.8	4.4
Median	1	2	2	3.5
Variance	6.73	6.59	11.76	11.28

Table 5-2 describes examples of social and technical challenges encountered during the Dar FMS implementation, including deliberate non-cooperation, limited ability to effectively contribute, and forced low degree of participation. Examples of deliberate non-cooperation include selling the phone after deployment, constantly removing the SIM card, selectively refusing to answer the phone interview, and being unwilling to consult with the troubleshooting team for technical issues. Examples of limited ability to effectively contribute include turning on airplane mode due to unfamiliarity with a smartphone, forgetting to carry the phone, carrying the wrong phone (e.g., of another household member), poorly recalling activities (especially common for people employed as vendors or drivers), and insufficient battery charge due to low access to electricity. Reasons for forced low degree of participation include both social and technical aspects. Several candidates had their phones stolen; some damaged their phones by accident. The biggest technical challenge related to the sparse cell tower coverage, resulting in large data gaps. Other technical issues are random problems from the smartphone or app itself, including wrong system time, internet connection problems, battery problems, etc. These challenges resulted in a relatively low number of fully verified days (Figure 5-1) and, overall, reduce the efficacy of the data collection, and even worse, may skew the representativeness of the data collected. For example, certain socio-demographic groups may tend to have a higher degree of participation and thus the collected data would be biased towards them.

Table 5-2 Example user problems identified in Dar FMS implementation

Type of challenge	Reasons	Count
Deliberate non-cooperation (18)	Refusing to meet with the fixing team for technical issues	6
	Refusing to answer the phone interview	4
	Refusing to participate (e.g. tired of carrying the phone, disappearing)	4
	Selling the phone after deployment	3
	Constantly removing the SIM card	1
Limited ability to effectively contribute (32)	Mixed up username	14
	Poorly recalling activities (e.g., for people employed as vendors or drivers)	12
	Always not carrying the phone	3
	Insufficient battery charge (e.g. due to low awareness or low access to electricity)	3
Forced low degree of participation (22)	Phone being stolen	8
	Technical problems (e.g. battery problem, not showing data, internet connection problem, wrong phone time)	7
	Phone damage (e.g. dropped in the water)	4
	Poor Tigo network	3

5.2 Comparison of Basic Trip Statistics Between FMS and paper-based travel survey

This section compares basic trip statistics (trip rate, trip purpose, trip distance and duration, and mode share) from three data sources when available. For FMS data, I also compare the results from all verified trips and those from only fully verified days to understand possible problems in using the partially verified data in other analyses.

5.2.1 Trip rates

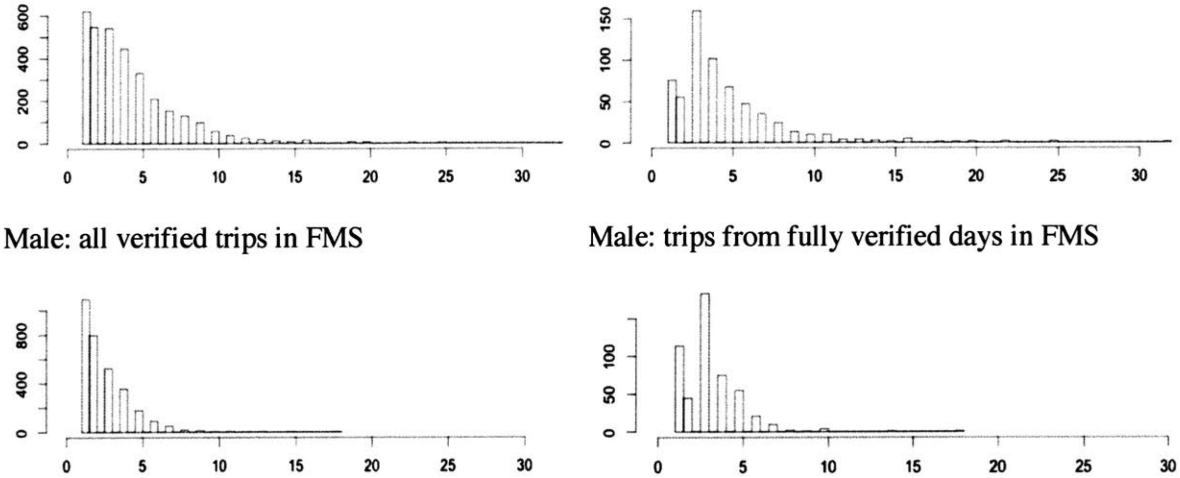
Table 5-3 summarizes trip rates by gender. Both 2007 HIS and 2015 FMS suggest a notably lower trip rate for females, with a similar gender difference (females about 70% lower). This reflects the fact that women are relatively excluded from the workforce in Tanzania (Rwebangira 1996) and that women tend to live seclusively or only travel in the immediate vicinity of home (Olvera, Plat, and Pochet 2003).

The total number of trips per person per day is the highest in 2015 FMS, followed by 2010 TS and then 2007 HIS. While some of this difference may be due to increasing trip rates over time, at least some is likely attributable to the difference in data collection methods. A smartphone-based travel survey has an advantage in capturing short-distance travel that is often underestimated in paper-based surveys (Wolf, Oliveira, and Thompson 2003; Ettema, Timmermans, and van Veghel 1996). Furthermore, Behrens and Masaoe (2011) found that their place-based travel diary format yielded higher rates of trip recall than trip-based survey, the latter of which was used in 2007 HIS.

Table 5-3 Trip rates by gender from 2015 FMS, 2010 TS, and 2007 HIS

	FMS-all verified trips (2015)	FMS – data from fully verified days (2015)	Behrens and Masaoe's TS (2010)	JICA's HIS (2007)
Male	3.13	3.66	-	1.5
Female	1.52	2.17	-	0.9
Total (mean)	2.39	2.99	2.06	1.2
Total (median)	2.00	2.00		

The trip rate estimated based on all verified trips as opposed to all verified days in FMS is lower because there are missing stops in the partially verified days. The result, on average, is 0.5 fewer trips per person per day. The trip rate distribution (Figure 5-1) suggests that this is mostly due to the large number of incompletely verified days that have one or two stops.



Male: all verified trips in FMS

Male: trips from fully verified days in FMS

Female: all verified trips in FMS

Female: trips from fully verified days in FMS

Figure 5-2 Trip rate distribution by gender from FMS data

5.2.2 Trip purposes

Compared to HIS, FMS recorded much higher shares of work-related trips and private leisure trips. This implies that the smartphone-based data collection more precisely depicts travel, especially for non-commuting trips that tend to be underreported in paper-based travel surveys. The FMS sample did not include people under 18 years old and therefore educational trips are not as high as JICA's HIS estimate. For FMS, the trip purpose composition does not vary greatly between the use of all verified trips or fully verified days.

A gender difference is notable in the percentage of work trips, work-related trips, and shopping trips. This is consistent with women's lower rate of participation in work and their tendency to undertake more house chore-related travels (such as grocery shopping) in Dar (Olvera, Plat, and Pochet 2003). The percentage of other personal and leisure trips is similar between women and men. However, because women generate only about 60% of the trips as men, the absolute number of those trips is lower for females.

Table 5-4 Percentage of trips by activity purpose, from 2015 FMS and 2007 HIS

	FMS (2015) – trips from fully verified days			FMS (2015) - all verified trips			JICA's HIS (2007)
	Female	Male	Total	Female	Male	Total	Total
to home	41%	28%	32%	39%	21%	26%	48.7%
to work	10%	19%	16%	11%	19%	17%	24.6%
work-related trips	5%	21%	16%	6%	24%	19%	3.7%
educational/religious trips	1%	2%	2%	2%	2%	2%	11.0%
shopping trips	11%	4%	6%	12%	4%	6%	12.2%
personal errands	8%	5%	6%	7%	6%	6%	
social/recreational/entertainment trips	16%	12%	13%	16%	13%	14%	
Eat meals out	1%	2%	2%	1%	2%	2%	
others	6%	8%	7%	7%	8%	8%	
Total trips per day	2.17	3.66	2.99	1.52	3.13	2.39	1.2

5.2.3 Trip distance and duration

Table 5-5 summarizes straight-line trip distance and trip duration for unlinked trips from FMS. An unlinked trip is defined as a separate trip segment each time the person changes modes or vehicles (Reddy et al. 2009; Utsunomiya, Attanucci, and Wilson 2006), as opposed to an entire journey the person takes from the origin to destination (linked trip). The FMS data set contains lat/lon coordinates of origins and destinations but does not contain the real trip distance. Therefore, I report the Euclidean travel distance instead of trip distance. The straight-line trip distance distribution using all verified trips is similar to that of trips from fully verified days. The actual trip distance can be estimated as the straight-line distance times a detour index. As a benchmark, Boscoe, Henry, and Zdeb (2012) estimate the detour index in the United States at 1.417 nationwide. The estimated detour index in Dar, based on 956 randomly selected verified FMS trips, is 1.65[23]. Using 1.65 as the average detour index in Dar, the crude estimation of actual travel distance is

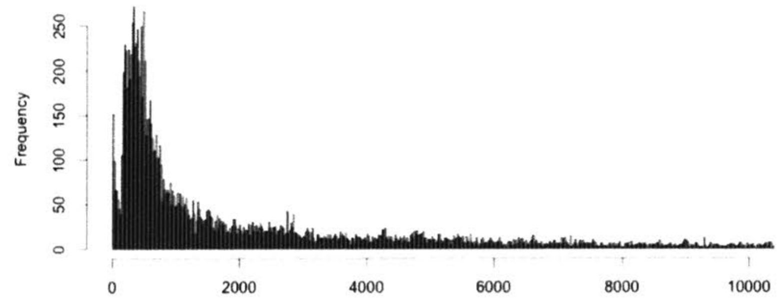
[23] Detour index estimation: shortest path length suggested by ArcGIS divided by straight-line distance. More details about shortest path length calculation can be found in Chapter 4.

approximately 6.4 km per unlinked trip. The average time per unlinked trip is about 33 minutes. For comparison, the average trip distance from 2007 HIS is 10.6km per linked trip and the average travel time is 76.5 minutes per linked trip (JICA 2008; technical report 6). The likely higher total trip distance and duration in 2007 HIS versus 2015 FMS again provides evidence that the paper-based HIS underestimates short-distance travel.

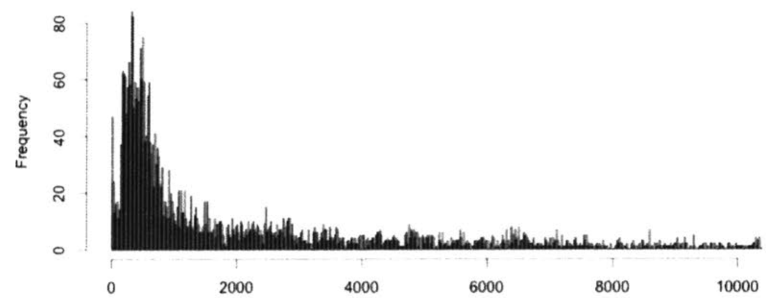
Comparing trip lengths and durations by gender again reflects women's lower work participation and more seclusive lifestyle (Olvera et. al 2003) (Table 5-5). Trip durations estimated for men calculated from all verified trips in FMS are underestimated by around 10 minutes compared to that from fully verified days; trip distance and the female's trip duration is similar across the two trip pools.

Table 5-5 Straight-line trip distance and trip duration (unlinked trips) from FMS

		Female	Male	Total
Straight-line trip distance (meters)	All verified trips	3428	4089	3888
	Fully verified days	3596	3997	3869
Trip duration (minutes)	All verified trips	28.6	34.7	32.8
	Fully verified days	32.6	46.1	41.8



Straight-line travel distance (meters): all verified (unlinked) trips



Straight-line travel distance (meters): unlinked trips from fully verified days

Figure 5-3 Distribution of straight-line travel distance from FMS trips

5.2.4 Mode shares

Mode shares from 2015 FMS suggests another gender difference in travel characteristics: males use vehicles more and females ride buses and walk more. The mode share composition from all verified records is similar to that from trips in fully versified days. Compared to 2007 HIS, the share of bus (daladala) in FMS' unlinked trips is lower and the shares of taxi, motorcycle and vehicles are higher. This is consistent with Dar's rapid motorization, which raises growing transportation challenges in the city.

Table 5-6 Mode share from 2015 FMS (unlinked trips), 2010 TS (linked trips), and 2007 HIS (unlinked trips)

(% of unlinked trips)	FMS - all verified trips (2015)			FMS – trips from fully verified days (2015)			Behrens and Masaoe's TS (2010), linked trips	JICA's HIS (2007)
	Female	Male	Total	Female	Male	Total	Total	Total
Foot	60.9%	47.8%	51.9%	59.6%	47.2%	51.3%	85.4%	56.0%
Bus	26.2%	21.2%	22.7%	27.4%	24.9%	25.7%	10.8%	41.0%
Taxi (Car, Bike, Motorbike, Bajaji)	6.1%	6.6%	6.5%	6.4%	7.3%	7.0%		
Motorcycle (Personal)	1.0%	9.5%	6.8%	0.9%	8.2%	5.8%		
Bicycle (Personal)	0.1%	1.5%	1.1%	0.1%	1.7%	1.2%		
Other	0.5%	0.9%	0.8%	0.5%	0.6%	0.5%	3.8%	3.0%
Vehicle	5.0%	12.5%	10.1%	5.0%	10.0%	8.4%		
Total trips	4734	10420	15154	1507	3065	4572	-	-

5.3 Temporal Travel Characteristics From FMS Data

The section above suggests that travel characteristics of all verified trips and of trips from only fully verified days are similar. To utilize the larger pool of observations, this section uses all verified trips to further explore temporal travel characteristics in Dar.

5.3.1 Trip generation by time of day

Trip generation in this section displays the number of trips according to the departure time. Figure 5-4, and Figure 5-5, show trip generation by day of the week and by activity with x-axes being the time of the day. Figure 5-6 shows total trip generation by activity on different dates; each date shows the sum of verified trips. Note that two factors may influence the graph interpretation: the number of verified observations by date and time of the day, and the travel behaviors. I base the following interpretation mostly on travel behavior, but underlying verification bias may potentially impact the interpretation. The departure time pattern from Monday to Thursday is similar, where 7-9am is the morning peak and 5-7pm is the evening peak. The bumps on Fridays are less obvious than other weekdays, implying more flexible commuting trips on Fridays. On the weekends, the observed trip generation is more stable throughout the day (Figure 5-4). Verified trip generation from November 20, 2015 to December 10, 2015 increases gradually, potentially reflecting respondents' greater familiarity with the smartphone-based travel survey and increasing comfort with verifying trips (Figure 5-6). The significant drop after December 10 likely reflects the fact that respondents finished the 4-week survey in succession. The composition of different activities verified does not vary much across the day, even for the work and work-related trips (Figure 5-6). Work and work-related trips account for similar percentages from Monday to Saturday and only decrease on Sunday; the magnitude of the drop is not very large.

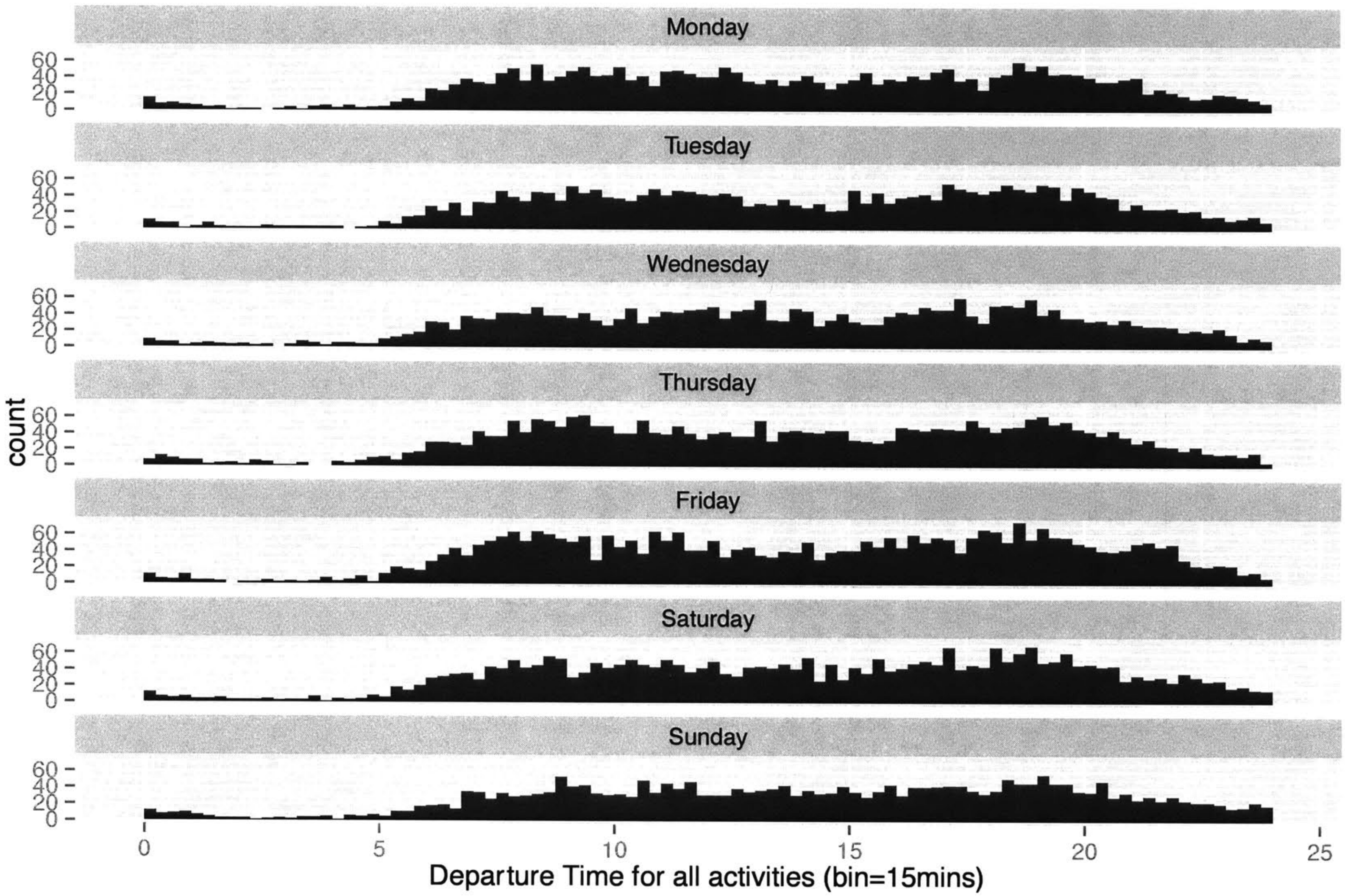


Figure 5-4 Departure time by weekday for all types of activities, 2015-11-20 to 2015-12-10 (21 days, 3 weeks)

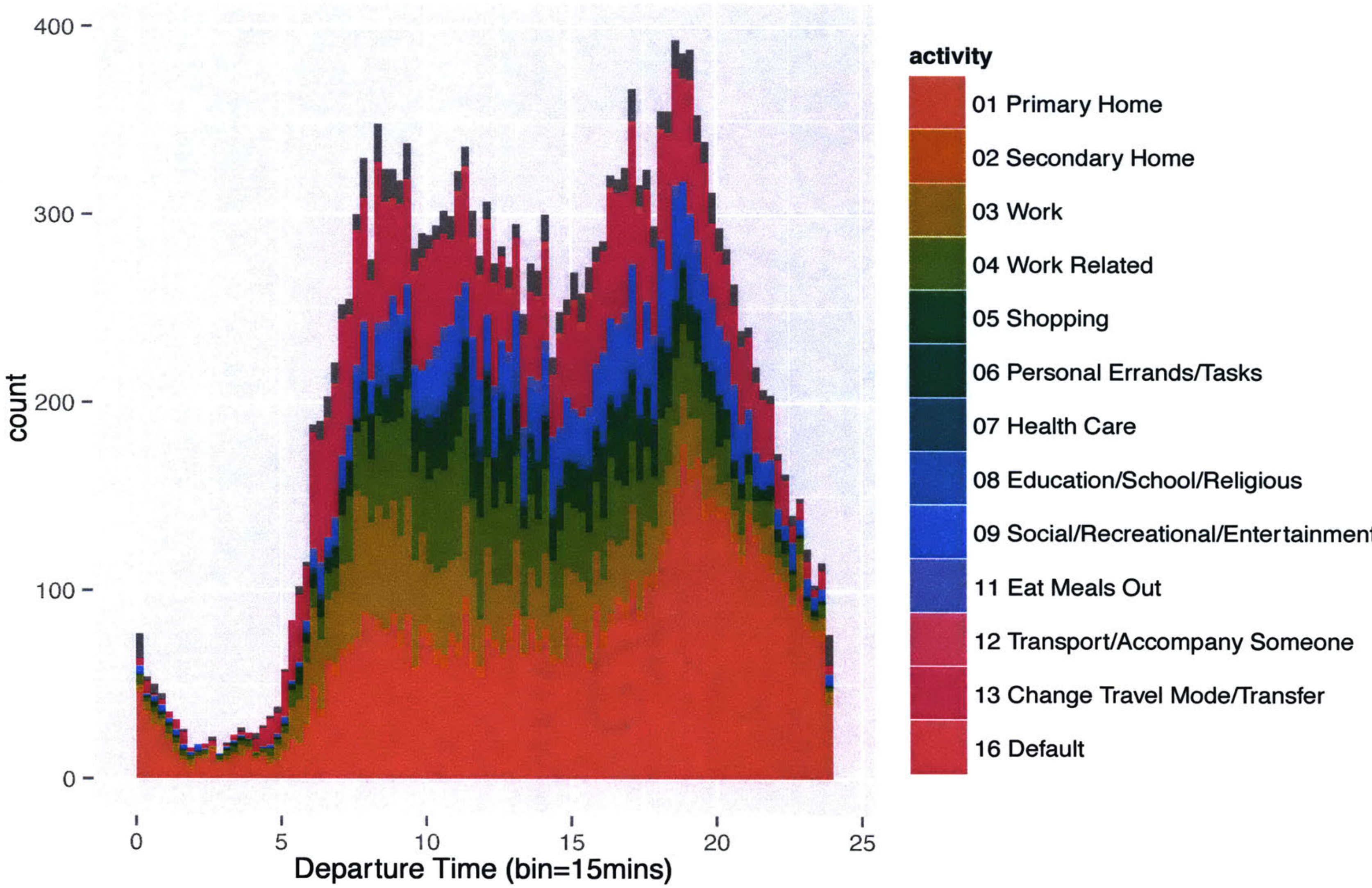


Figure 5-5 Departure time by time of the day and activity

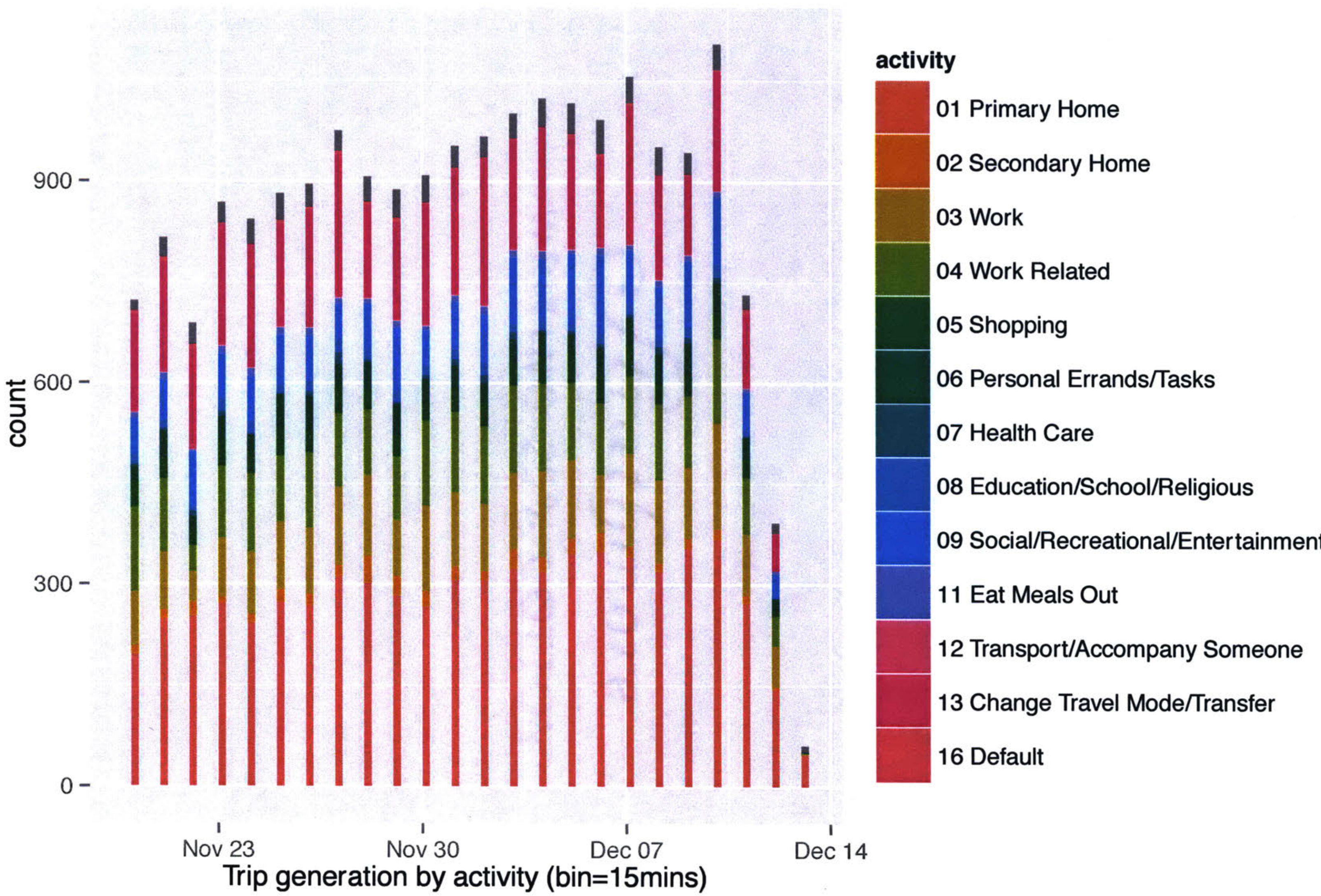


Figure 5-6 Trip generation by activity and date (2015-11-20, Friday to 2015-12-13, Sunday)

5.3.2 Traveling to activities

Traveling to activities represents the number of travelers at different times of day by activity purpose. Different from trip generation where the count is based on departure time point, traveling to activities counts the travelers as long as they are in the trip. For example, if one travels 20 minutes to work, the person is a "traveler" during this 20-minute period. Travelling to activities (Figure 5-7) reflects the travel demand on the road by time of the day. Several interesting travel patterns emerge in Dar es Salaam. First, the number of travelers to home in the evening peak hour is around twice the size of travelers to work in the morning peak hour. For comparison, FMS in Singapore reveals a more symmetrical pattern of traveling to home in evening and traveling to work in morning (Zhao, Pereira, et al. 2015). At least two factors may explain the observed differences between Dar and Singapore: a) the share of commuting trips in Dar is less than that in Singapore; b) the Dar sample is biased towards unemployed population. The following chapter suggests that people who are employed tend to have fewer fully verified days.

Second, the number of verified work-related trips is almost as high as commuting trips. As online services are not very common in Dar, work-related trips are indispensable for many professional activities. The peak hour (7:30am-1pm) for work-related trips lasts longer than work trips (7-9am), and the peak volume is almost the same as for work trips. Third, verified social/recreational/entertainment trips are observed throughout the day. The volume of social trips is high from 11am to 8:30pm with bumps from 11am-1pm and 5-6:30pm.

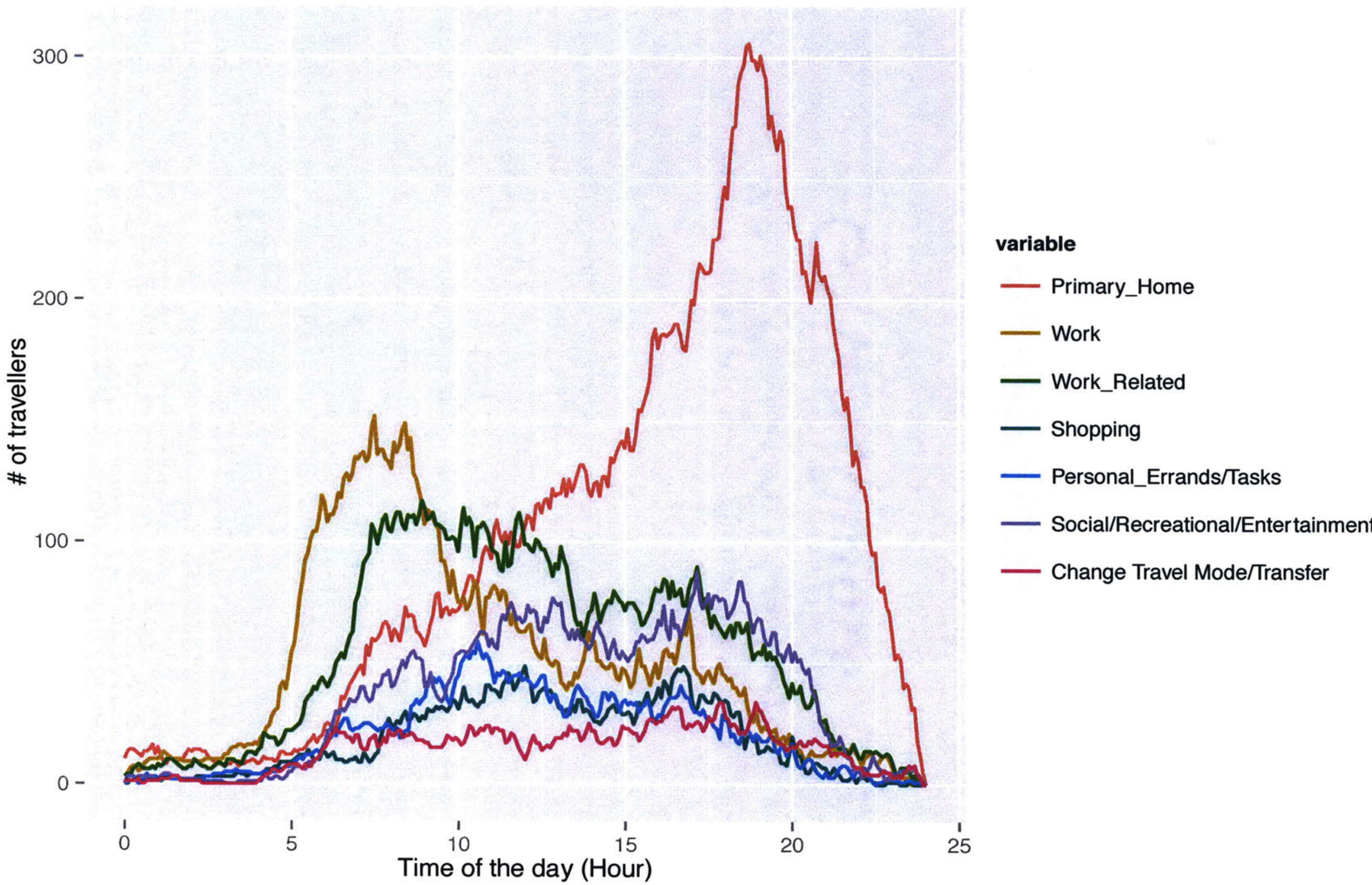


Figure 5-7 Number of travelers to different activities by the time of the day

5.3.3 Trip time distribution

Trip time distribution (TTD) reveals the distribution of travel time for different trip purposes, suggesting people's tolerance level of trip duration for each activity type (Figure 5-8). In developed cities, such as metropolitan areas in the United States, TTD for commuting trips typically skews to the right more than for other trip purposes, because people can often tolerate (or must tolerate) long commuting travels (National Cooperative Highway Research Program et al. 2012). The verified FMS trips for Dar es Salaam suggest that TTDs for all trip purposes follow almost the same trend: the frequency of trips peaks at 10 minutes but with a long tail. The results suggest that Dar residents mostly conduct activities close to their homes or neighborhoods with occasional longer-distance travels. But when travelling beyond their neighborhoods, they take very long trips (sometimes up to 120 minutes). On the one hand, the poor transportation infrastructure can be one constraint to people's travel behaviors. Motorized travelling in Dar is challenging and walking is the most common mode, constraining the activity space of Dar residents and making travel outside neighborhood burdensome. On the other hand, informal businesses within neighborhoods provide shopping and working opportunities close to home. Although those opportunities may not be as attractive as those in the city center, they serve as compromise alternatives to guarantee basic daily needs for Dar residents.

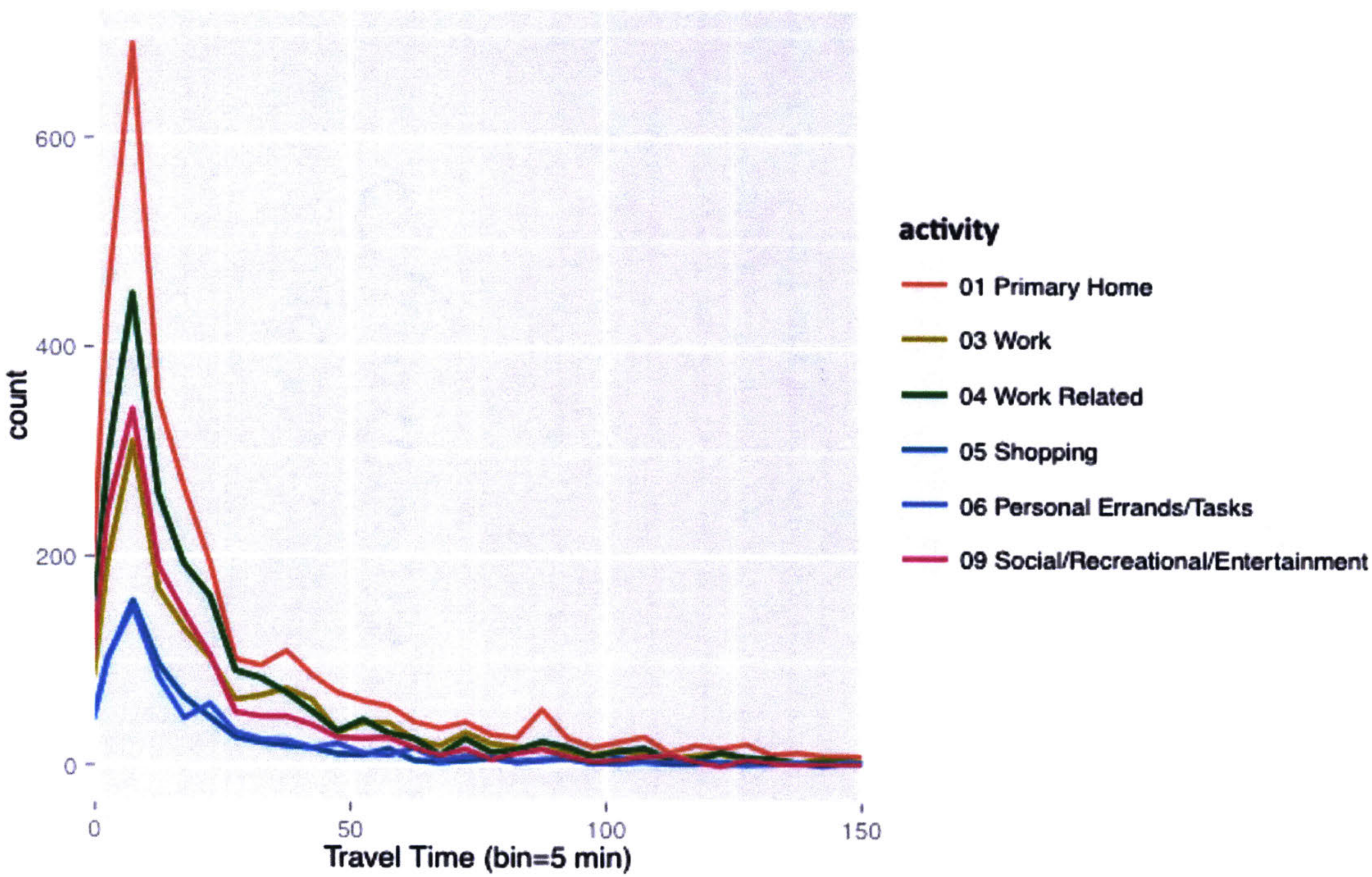
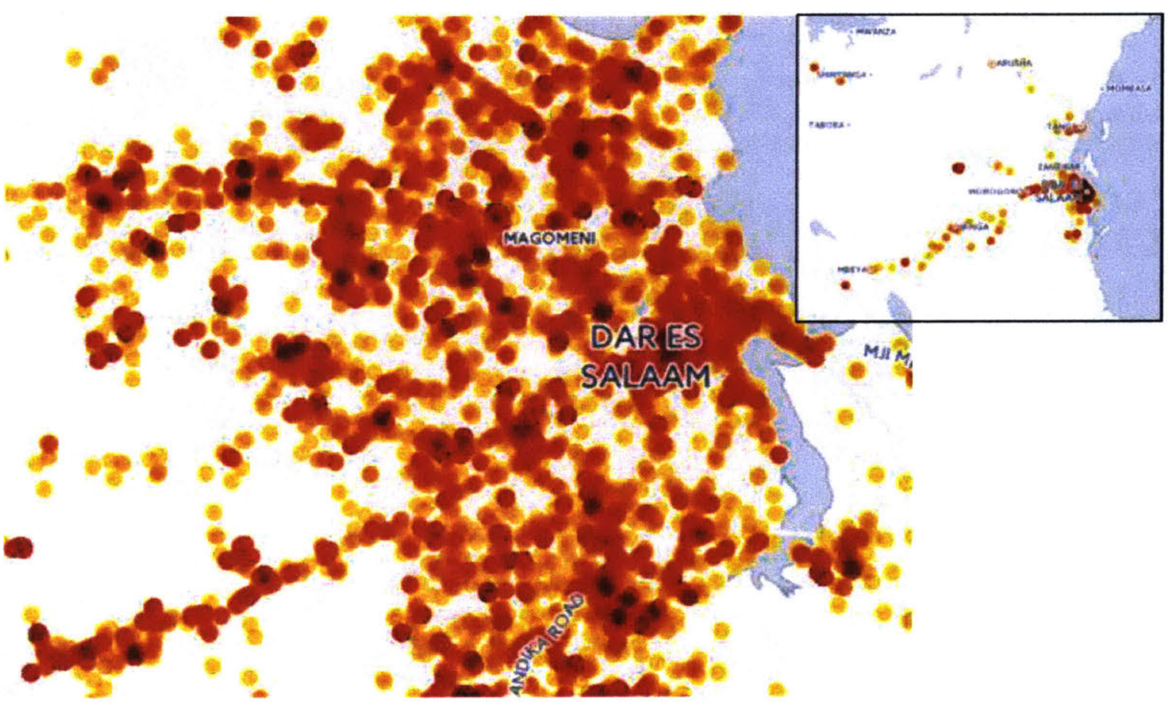


Figure 5-8 Distribution of travel time by activity type

5.4 Spatial Travel Characteristics From FMS Data

Figure 5-9 shows the spatial intensity and spatial distribution of all verified activities.[24] Although Dar is a monocentric city, the FMS verified activities spread out across the city to areas within the second ring road. Government and big companies typically locate in the city center and generate more work and work-related activities, but the intensity for all activities is not visibly higher than in other places. For areas other than the city center, no clear separation of residential areas and workplaces appears and opportunities seem diverse. Within each neighborhood, activities tend to concentrate. Based on my field observation, activities, especially shopping activities, typically cluster around daladala stops. The observed spatial activity pattern is associated with short-time trips dominating. Two mixed and inter-related effects may contribute to these two travel patterns: the constraint due to poor transport infrastructure and services, and the popularity of informal businesses served by self-employed residents in their own neighborhoods. Of course, these patterns may also be explained by the nature of the FMS sample and unknown factors underlying the verification rates.



[24] The video of activity in time and space can be found here: https://goo.gl/IpWxAo.

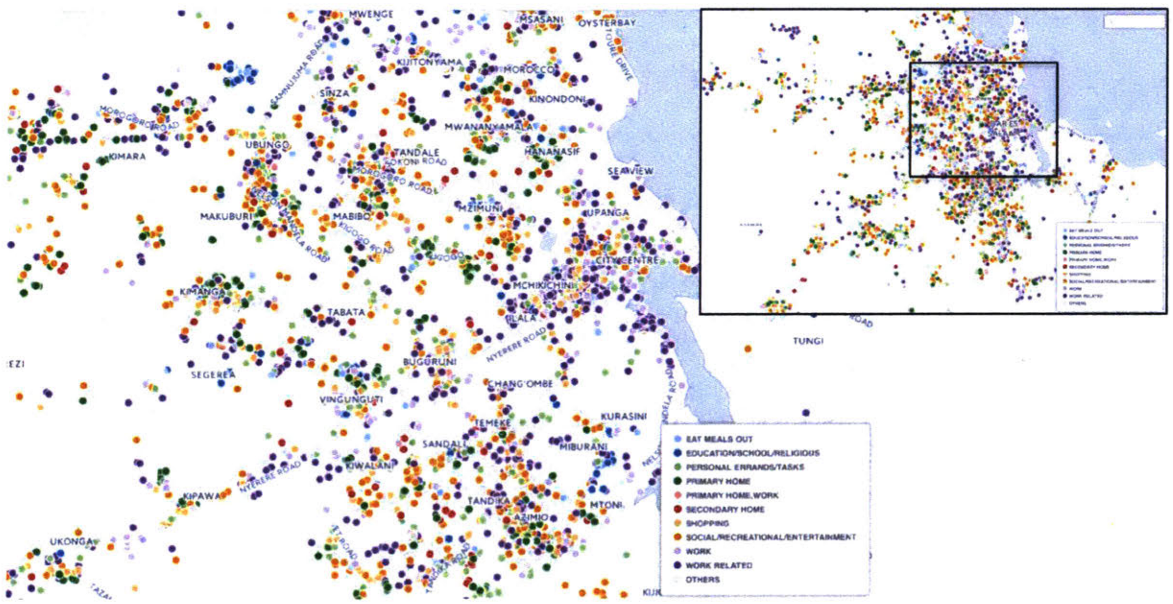
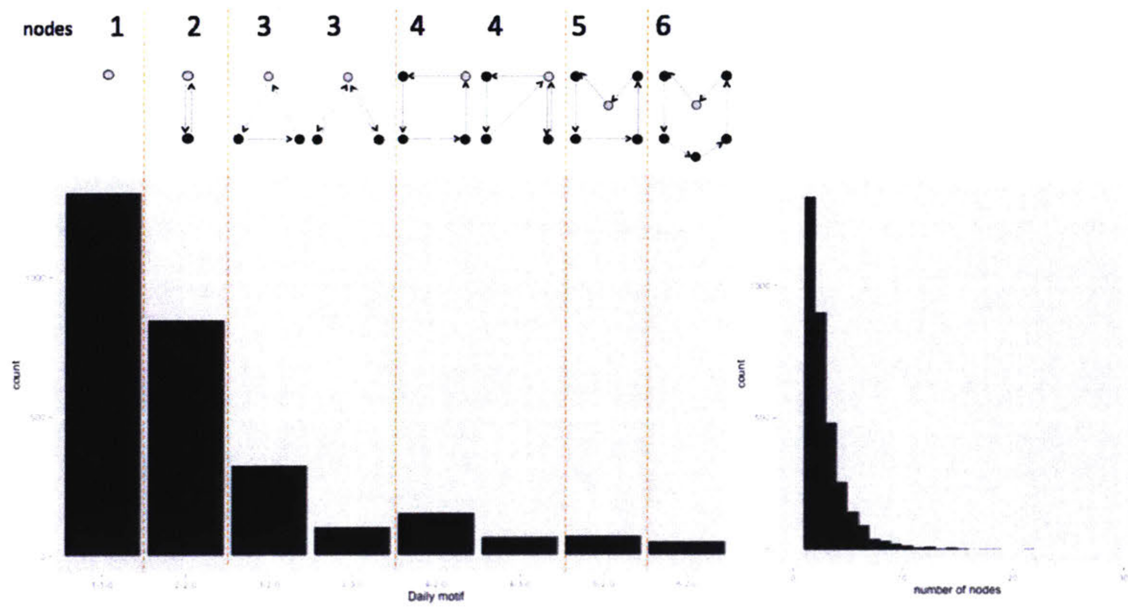


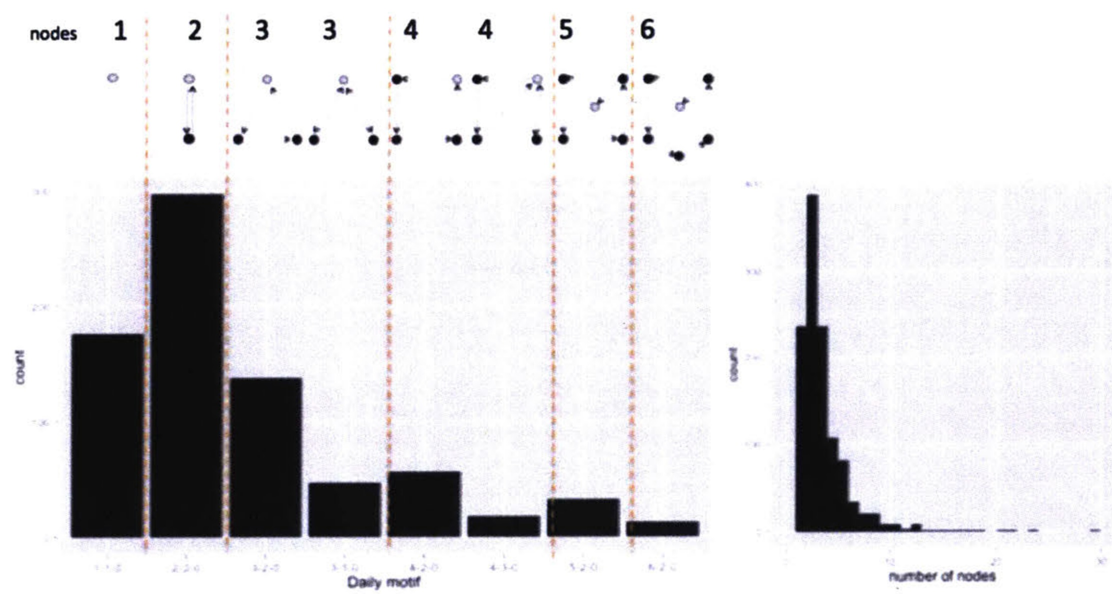
Figure 5-9 Spatial intensity (top) and spatial distribution (bottom) of all activities

5.5 Daily Mobility Motif from FMS Data

The analysis of daily motif needs to be based on travel from fully verified days. Most days have two nodes (i.e. home – one other place), followed by staying at home (one node) or visiting two out-of-home places (three nodes) (Figure 5-10). For the days with more than two nodes, the majority of people do not go back to home in the middle of the day. The average number of daily visited places (nodes) provides one indicator of an individual's mobility ability. More nodes or a more complex daily motif is potentially related to more convenience in traveling around.



Daily motif based on all verified records (overestimateing the motif of staying at home)



Daily motif based on records from fully verified days

Figure 5-10 Daily mobility motif and number of nodes visited in one day

Figure 5-11 visualizes the number of nodes on a map. The dots show the home location of the respondents, and the darkness of color represents the average number of nodes they visited per day – a simplified way to represent travel equity. A more refined approach to analyze travel equity is presented in chapter 7. Figure 5-11 shows no clear spatial pattern of the complexity in daily motif. Blue lines represent the planned BRT phases while the green line (phase I) has been in operation since May 2016. None of these BRT lines were in operation at the time FMS was being implemented. Although the visual map does not suggest an unfair BRT allocation, the limitation in sample representativeness and in the simplicity of number of nodes indicator do not generate a conclusive finding.

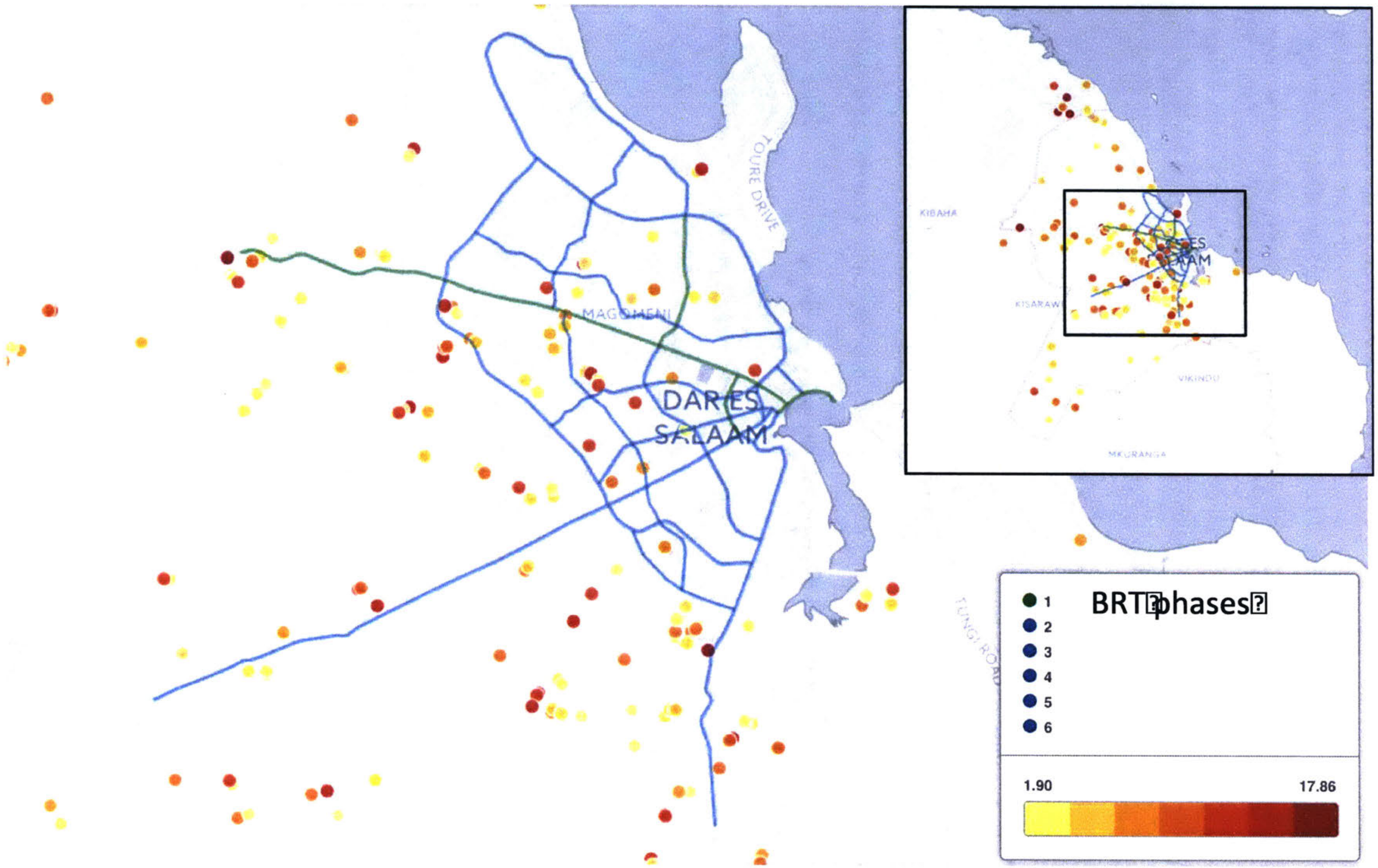


Figure 5-11 Average number of nodes visited by the home locations of respondents (based on fully verified days)

5.6 Summary

This chapter describes basic travel characteristics of Dar residents at the aggregate level. I compare trip rate, trip purpose, trip distance, and mode share using FMS and, when possible, two other survey sources, 2010 TS and 2007 HIS. I analyze gender differences and temporal and spatial travel patterns in Dar and derive daily mobility motif as one indicator to represent individual's mobility ability. Key findings from this chapter are as follows.

- Notable gender differences exist in travel characteristics. Males generate more trips per day and travel longer. Women have fewer work trips, work-related trips, and personal trips while conducting more household chore trips. Females have higher mode shares for walking and bus.
- The partially verified trips in FMS do not show bias based on measures of trip distance, trip duration, and mode share; those statistics are similar to those calculated from trips in fully verified days. However, the number of stops per day is underestimated if not using fully verified days.
- The comparison between FMS, TS and HIS suggests more precise recall of travel using smartphone-based methods than paper-based methods. FMS records more trips, among which short-distance trips and non-commuting trips appear to be underreported in paper-based surveys.
- The evidence suggests an increase in private vehicle mode share and a decline of daladala share from 2007 to 2015.
- The verified FMS unlinked trips suggest that the amount of travelers to home in the evening peak hour is around twice the number of travelers to work in the morning peak hour. Work-related trips are another major composition of all trips. Social/recreational/entertainment trips happen throughout the day.
- Dar's formal professional functions are located in the city center and result in the highest generation of work and work-related trips. However, many urban activities are spread across the entire built area and have the tendency to cluster.

Based on field trip observation, the clusters are typically informal opportunities around local daladala stops.

- The verified unlinked FMS trips suggest a tendency to travel short-distances around neighborhoods. Two factors potentially contribute to this travel behavior: poor transportation infrastructure and services and the popularity of informal businesses served by self-employed residents in their own neighborhoods.
- The complexity of daily mobility motif can be an indicator of traveling accessibility to urban opportunities. Based on the daily motif inferred from travels in fully verified days, no clear spatial pattern of travel complexity emerges among the FMS sample.

A few limitations need highlighting. First, the comparison between 2015 FMS, 2010 TS, and 2007 HIS is only based on basic statistics at the aggregate level. Future studies can dive deeper into the comparison between paper-based travel survey and smartphone-based travel survey, for example, by different socio-demographic attributes, by time of day, and by origin-destination pair. Second, the calculation of trip distance, trip duration, and mode share is based on unlinked trips for simplicity. However, linked trips are better representations of the complete travel behaviors. I did not calculate linked trips in this thesis because of time limitation. Third, the travel characteristics analyzed in this chapter are not representative so should only be viewed as indicative, at best. As discussed in the following chapter, a smartphone-based travel survey can potentially introduce sample bias and the travel patterns derived from FMS should ideally correct the bias by applying proper weights. Future studies and smartphone-based data applications should be aware of this limitation.

Chapter 6 Representativeness of FMS Participants

As seen in the previous Chapter, FMS provides high-resolution data on travel behavior, with some apparent advantages over more traditional data collection methods. Nonetheless, given the nature of the data collection technology and context, questions emerge about the representativeness of the individuals participating. Recall that of the 581 candidates who received the free smartphones, only 482 people provided at least one verified trip/stop, and only 329 or 354 people had at least one fully verified day (using the FMS diary-based or enumerator-based verification criterion, respectively). In this Chapter I present the estimation results from various models to examine any factors which apparently influence individual participation in FMS.

6.1 Model 1: Factors influencing individual non-participation

In the first model, I estimate how socio-demographic factors play a role in influencing participation, to approximate the general representativeness of FMS participants. I use three different definitions of non-participants: (1) individuals who do not generate any verified record (99 observations); (2) individuals who do not have any diary-based fully verified days (227 observations); (3) individuals who do not have any enumerator-based fully verified days (252 observations). Independent variables include basic socio-demographic variables – gender, education, age, job type, marriage status, index for wealth per household member – from MLDS and individual-specific MLDS modules and a relative locational variable – distance from home to CBD. Another set of independent variables, from a short multi-topic questionnaire administrated with FMS, represents those conceptually likely to influence the level of participation, including an individual's depression scores, vehicle ownership, total working hours per day, and monthly sales of the family business. More than half of people in DES work in a family business; monthly sales is used to capture the business size.

The model results (Table 6-1) reveal little bias among socio-demographic groups, irrespective of how I define non-participation. I exclude non-significant variables from the final model results. Participation status is statistically the same among people with different gender, age, home location, marriage status, wealth, depression scores, working hours, and monthly sales of the family business. The only characteristics of potential concern are education and certain job types. With all three definitions, more years of education are correlated with a higher likelihood of participating, perhaps implying that lack of familiarity with technology could potentially be a barrier to scale up smartphone-based method among less-educated individuals. On the other hand, people employed in the public sector (government job or public company) are more likely to be non-participants, an interesting finding, perhaps representing some fear public employees may have of being 'tracked'. This result warrants further investigation. The Moran's I test of the residuals reveal no significant autocorrelation (non-participant by definition 1) or marginally significant autocorrelation (definitions 2 and 3).

Table 6-1 Results of Binary Logit Model Showing the Socio-demographic Factors Influencing the Likelihood of Non-participation

	Percent	Non-participant Definition (1): individuals do not have any verified record			Non-participant Definition (2): individuals do not have any fully verified day (based on diary)			Non-participant Definition (3): individuals do not have any fully verified day (based on notes)		
		Estimate	Pr(>lzl)		Estimate	Pr(>lzl)		Estimate	Pr(>lzl)	
(Intercept)	-	-3.493	0.000	***	0.111	0.700		0.511	0.149	
Female	49%							-0.212	0.094	.
Male (base)	51%									
Education years	-	-0.123	0.029	*	-0.037	0.083	.	-0.043	0.028	*
ln(distance from home to CBD)	-							-0.106	0.216	
Type of primary job: Government	3%	2.131	0.038	*	1.059	0.019	*	0.768	0.039	*
Type of primary job: Private company	12%	0.676	0.334		0.218	0.480		0.016	0.946	
Type of primary job: Private individual(s)	11%	1.108	0.101		0.556	0.076	.	0.021	0.930	
Type of primary job: Public company	2%	1.853	0.047	*	0.265	0.598		0.241	0.568	
Type of primary job: self-employed	52%	0.589	0.301		0.261	0.295		-0.144	0.415	
Type of primary job: no job (base)	20%									
Total working hours	-				-0.032	0.068	.			
Sample size		581			581			581		
Sigma (household random effect)		5.236			0.626			0.362		
log-likelihood		-174.941			-343.649			-336.264		
AIC		365.900			705.300			692.500		
Moran's I of the residuals		-0.004	0.924		-0.045	0.101		-0.054	0.047	*

6.2 Model 2: Factors influencing an individual's number of fully verified days

To further analyze the degree of participation, I estimate a ZIP model predicting each individual's number of completely verified days, using the same set of independent variables as in the previous model. I estimate models on two versions of the dependent variable: the FMS diary-based number of fully verified days; and the enumerator-based number of fully verified days. The latter represents an arguably less conservative definition of verified day, as enumerators may tolerate some incompleteness in the verification process (see, also, Table 5-1 in Chapter 5).

The results (Table 6-2) indicate some biases in the representativeness of the verified days, with some differences depending on the definition of day verification (diary-based versus enumerator-based). The probability of having zero fully verified days does not differ much across different socio-demographic groups, consistent with the participation model above. However, the number of verified days for the non-zero group are influenced by several socio-demographic factors. With both definitions, the number of verified days goes down for married and widowed individuals, individuals with car driving license, and people who work for government, private individuals and self-employed; the number goes up with age. Married people may have other priority matters (although interestingly this effect is not statistically significant for polygamous cases); the widowed may have more burden for housework. People with a car driver's license may have more complex trip chains, which dissuade verification. People who work in government may be concerned about being 'tracked" and people self-employed may be preoccupied with their own business. Older people may have more time to or interest in verifying days. Other results are less consistent across the two definitions of the dependent variable: gender; working for public company; living together with partners but not married; wealth per household member; monthly sales of family business. Women may have less time to or interest in verifying days; un-married people living together with partners may have similar reasons as married people for not verifying more full days; people from households with more wealth per person may have smart device themselves and are thus more familiar with this data collection methods, or they have fewer children to take care of and thus have more energy to cooperate with FMS program. The fact that these results vary depending on the definition of the dependent variable means these results are less robust than those that are consistent across two definitions.

It also suggests a possible enumerator random effect. However, for the most part, the differences are marginal in terms of statistical significance. For enumerator-based fully verified days, the small significant negative Moran's I indicates slightly negative spatial autocorrelation, but the close to zero value should have little impact on the model results.

Table 6-2 Results of Zero-inflated Poisson Model Showing the Socio-demographic Factors Influencing the Number of Fully Verified Days

	Number of fully validated days (FMS diary-based)			Number of fully validated days (enumerator-based)		
	Estimate	Pr(>|z|)		Estimate	Pr(>|z|)	
Count part for non-zero group						
Female	-0.243	0.062	`	-0.088	0.408	
Male (base)						
Age	0.01	0.042	*	0.011	0.025	*
ln(distance to CBD)	0.091	0.314		0.114	0.14	
Type of primary job: Government	-1.059	0	***	-0.41	0.073	`
Type job: Private company	-0.176	0.383		-0.032	0.835	
Type job: Private individual(s)	-0.339	0.094	`	-0.458	0.004	**
Type of primary job: Public company	0.016	0.958		0.307	0.049	*
Type of primary job: self-employed	-0.305	0.034	*	-0.269	0.017	*
Type of primary job: no job (base)						
Married (monogamous)	-0.388	0.003	**	-0.273	0.039	*
Married (polygamous)	-0.352	0.361		0.013	0.961	
Living with partner, not married	-0.404	0.146		-0.582	0.006	**
Divorced	0.364	0.177		0.099	0.742	
Separated	-0.054	0.88		-0.07	0.828	
Widowed	-0.643	0.082	`	-0.421	0.038	*
Never married (base)						
Wealth per HH member	0.149	0.239		0.23	0.038	*
Depression scores	0.011	0.112				
Driving license - car	-0.232	0.097	`	-0.42	0.005	**
Driving license - motorbike	0.284	0.268		0.083	0.767	
Driving license - none (base)						
Monthly sales of family business (0 if the person is not self-employed)	0.001	0.083	`			
Probability to be in zero group						
Female				-0.353	0.088	`
Male (base)						
Education years				-0.068	0.025	*
ln(distance from home to CBD)	-0.217	0.12		-0.141	0.312	
Type of primary job: Government				1.127	0.062	`
Type job: Private company				0.014	0.969	
Type job: Private individual(s)				-0.113	0.767	
Type of job: Public company				0.317	0.629	
Type of job: self-employed				-0.292	0.289	
No job (base)						
Wealth per HH member	-0.366	0.183				
Intercepts						
Probablity to be in zero group	-0.037	0.913		0.775	0.159	
Count part to non-zero group	0.818	0.033	*	1.31	0	***
Sample size	581			581		
Log-likelihood	-1014.724			-1184.589		
AIC	2075.448			2423.179		
BIC	2173.374			2538.136		
Moran's I of the residuals	-0.019	0.525		-0.065	0.017	*

6.3 Model 3: Factors influencing the verification of a day

Finally, I turn to the analysis of factors influencing the likelihood of a given user day being verified. Beyond attributes of the individual and the household, the models also include four indicators of day complexity, the day of the week, and characteristics of the verification calling (whether the enumerator's verification call gets through and last call duration; the latter assumes that a prior long call will dissuade the likelihood of an individual participating in the call for verification).

Although I include day complexity indicators in the model, these may suffer from endogeneity. Several possible sources of endogeneity exist. Most basically, an individual's low level of interest in participation can hypothetically influence both the day verification status and the number of stops detected by FMS. For example, if a person is not very interested in the program and often leaves the phone at home or often fails to charge the phone, it can result in low number of detected stops; this low level of interest may simultaneously lead to non-cooperation in verifying the records. Unfortunately, no simple way to correct for endogeneity exists in this case. As a simplified method to examine the potential for endogeneity, I report model results with and without the day complexity indicators as independent variables; those models generate similar results and imply little endogeneity. More sophisticated techniques such as two step control-function and multiple indicator solution are needed to formally test the endogeneity issue (Guevara 2015), but they require instrumental variables that are hard to obtain in this case. In total, there are 7,285 days from all candidates that have an FMS sensor data record, of which 1,087 are fully verified according to the FMS diary-based definition and 1,574 are fully verified according to the enumerator-based definition.

The results point to several findings (Table 6-3). First, counter-intuitively, whether the verification call gets through does not significantly influence the likelihood of day being fully verified. Two related factors might influence the day verification status: the participants' willingness to verify, and the experimental design that enumerators were not allowed to add stops. If stops were not detected by FMS, enumerators would leave the activity diary intact and label the record as not verified. The result implies that this enumerator effect dominates the participant's effect. In other words, the day verification status is influenced more by the severity of the data gap than by participants' willingness to cooperate. Consistent with this implication, a long duration of the previous call does not significantly increase people's impatience to verify the data either. Second, the number of detected stops highly significantly increases the likelihood of the day being fully verified. If not caused by endogeneity, this suggests that the more complete travel diary detected by FMS for a given day, the less likely that day will be labeled as not verified by enumerators because of missing stops. Third, the negative influence of distance from home to activities centroid may imply that participants have more chances to charge their phone (and/or do not use their phone battery as much) if their destinations are close by. Fourth, the results for the socio-economic and demographic variables are similar to those relating to the number of days validated as estimated in the ZIP count model. Once more there is significant negative effect for government employees (for both the FMS-diary and enumerator-based definitions of dependent variable) and significant negative effect for private individuals and self-employed (for enumerator-based definition only). A consistently negative effect is also revealed for people with a car driver's license. Again, this may be due to relative travel complexity of car drivers. Those working long hours, on the other hand, have a higher likelihood of verifying a given day, perhaps representing relatively mundane (easy to verify) days, with fewer total trips. No apparent relationship emerges based on the day of the week. Other variables I tried, but that were non-significant include gender, index for wealth per household member, depression scores, and monthly sales of family business. Finally, the intercept is very significant, indicating that many unobserved factors influence the day verification.

Table 6-3 Results of Binary Logit Model of Likelihood of A Day Being Fully Verified

		%	Day being fully verified (based on FMS diary)				Day being fully verified (based on phone notes)			
			Full model		Basic Model (w/o potentially endogenous variables)		Full model		Basic Model (w/o potentially endogenous variables)	
			Coef.	Pr(>|z|)	Coef.	Pr(>|z|)	Coef.	Pr(>|z|)	Coef.	Pr(>|z|)
	(Intercept)	-	-1.710	1.30E-10 ***	-1.610	1.60E-09 ***	-1.620	1.80E-08 ***	-1.560	6.70E-08 ***
day mobility indicators (*potentially endogenous variables*)	The number of detected stops		0.0242	3.90E-08 ***			0.012	0.005 **		
	Total distance traveled		-0.0007	0.223			-0.001	0.171		
	Distance from home to activities centroid		-0.0013	0.055 .			-0.001	0.039 *		
	standard radius of activity circle		0.0026	0.068 .			0.000	0.838		
day characteristics	Day of week: Tuesday	16%					-0.071	0.319	-0.072	0.312
	Day of week: Wednesday	16%					-0.077	0.288	-0.080	0.267
	Day of week: Thursday	16%					0.066	0.349	0.064	0.362
	Day of week: Friday	12%					0.159	0.023 *	0.150	0.033 *
	Day of week: Saturday	12%					0.067	0.346	0.063	0.375
	Day of week: Sunday	12%					0.065	0.364	0.056	0.432
	Day of week: Monday (base)	15%								
call characteristics	Last call duration	-	0.000	0.225	0.000	0.218	0.000	0.558		
	Whether the call gets through		0.051	0.438	0.035	0.589	0.052	0.365		
socio-economic variables	Education years	-	0.019	0.107	0.017	0.143	0.018	0.145	0.017	0.180
	Age	-	0.008	0.049 *	0.009	0.029 *	0.009	0.049 *	0.009	0.049 *
	ln(distance from home to CBD)	-	0.089	0.104	0.091	0.096 .	0.110	0.076 .	0.106	0.089 .
	Type of primary job: Government	3%	-0.664	0.006 **	-0.557	0.020 *	-0.623	0.021 *	-0.589	0.029 *
	Type of primary job: Private company	13%	-0.161	0.204	-0.115	0.366	-0.305	0.064 .	-0.308	0.062 .
	Type of primary job: Private individual(s)	11%	-0.197	0.154	-0.153	0.267	-0.364	0.033 *	-0.358	0.036 *
	Type of primary job: Public company	2%	0.039	0.876	0.191	0.453	0.030	0.917	0.083	0.771
	Type of primary job: self-employed	54%	-0.097	0.342	-0.069	0.501	-0.275	0.043 *	-0.275	0.043 *
	Type of primary job: no job (base)	19%								
	Marriage: Married (monogamous)	66%	-0.285	0.007 **	-0.301	0.004 **	-0.213	0.066 .	-0.218	0.060 .
	Marriage: Married (polygamous)	2%	-0.187	0.486	-0.216	0.422	-0.166	0.552	-0.181	0.516
	Marriage: Living together with partner, but not married	5%	-0.257	0.193	-0.227	0.249	-0.261	0.230	-0.256	0.241
	Marriage: Divorced	1%	0.364	0.232	0.307	0.315	0.440	0.176	0.425	0.192
	Marriage: Separated	2%	-0.239	0.369	-0.194	0.467	-0.061	0.816	-0.045	0.863
	Marriage: Widowed	3%	-0.417	0.073 .	-0.446	0.055 .	0.092	0.692	0.085	0.715
	Marriage: never married (base)	21%								
	Driving license - car	11%	-0.206	0.074 .	-0.136	0.236	-0.343	0.006 **	-0.324	0.009 **
	Driving license - motorbike	2%	0.326	0.173	0.409	0.089 .	-0.139	0.601	-0.133	0.618
	Driving license - none (base)	87%								
	Total working hours	-					0.019	0.052 .	0.020	0.036 *

goodness-of-fit	**Sample size**		7285	7285	7285	7285
	sigma (random effect for household)		0.298	0.290	0.441	0.441
	sigma (random effect for individual)		0.425	0.435	0.462	0.467
	log-likelihood		-1945.6	-2402.6	-2989.4	-3186.2
	AIC		4816.1	4847.1	6419.0	6428.3
	Moran's I of the residuals		N/A because one person has multiple observations in the same location			

6.4 Summary

This chapter analyzes results from the first attempt to implement a smartphone-based travel survey in an African city. I identify potential social and technical challenges of implementing such a survey, utilizing the example of FMS deployment in Dar es Salaam. The findings show some promise for the use of such technologies for travel behavior data collection, revealing little observable biases in the likelihood of an individual to participate. That said, the results suggest that, conditional on participating, the degree of participation (the number of fully verified days) is associated with an individual's marriage status, having a driving license, age, and the monthly sales of family business. Finally, the verification of a given day suggests the apparent negative impact due to one defect of the experimental design: not allowing enumerators to add stops. The verification process is supposed to enable the completion of the activity diary generated by the FMS backend. If enumerators cannot add missing stops, however, this has a larger influence on the data quality than on participants' willingness to participate. Nevertheless, the low observable bias in participation demonstrates the promise of smartphone-based survey to provide representative data. The differences in the degree of participation require future endeavors to overcome. Possible solutions include enhancing mobile phone training of participants, improving enumerators' verification procedure, and estimating weights to adjust the impacts from different degree of participation.

This research not only supports the feasibility of smartphone-based data collection methods in the African context, but also provides insights into adapting FMS and similar such approaches for data collection in developing regions. Techniques of travel activity inference should, for example, account for possible poor phone network and the lack of contextual data (e.g. land use, building type), and battery constraints arising from limited access to electricity. Operationally, if enumerators are used, they need to be better trained to, and allowed to, add additional stops not captured by the app. Because of the imperfect contextual information mentioned above, inferred activities might not be accurate, requiring auxiliary inputs from users. Other means for ensuring more effective participation should also be examined, such as calling more frequently for users with a complex travel diary, requiring a deposit to avoid users from selling the phone, and training users on basic usage of the phone.

Chapter 7 Travel Equity

The previous chapter showed the data collected in the FMS sample show little signs of observable bias in terms of the representativeness. This chapter demonstrates the potential these data have in suggesting the relative equity of travel – as measured via accessibility indicators – among participants in the FMS pilot survey. I operationalize two basic types of space-time measures to evaluate equity using an accessibility-based travel welfare indicator, following the theoretical foundations presented in Chapter 2 and the methodological approaches described in Chapter 4.

7.1 Society-wide Distribution

7.1.1 Equality of Welfare

By the equality of welfare principle, individuals should have the same level of travel welfare or absolute PPAs. Table 7-1 shows the travel welfare inequity in Dar by different measures. First, keep in mind that FMS collected data also on inter-city travel over the course of the one-month of data collection. For MPPA, inter-city travel may not represent "typical" trips for an individual's daily behavior, but will increase the size of PPA considerably. Inter-city trips might occur for business, leisure, or visiting friends and relatives; they tend to be more random than intra-city travel, although some people may travel between Dar and other cities regularly. Somewhat similarly, for DPPA, if an individual stays at home for the selected day, his or her DPPA is zero. However, this person may just happen to be at home for that day, for example choosing not to go out even if s/he does not have to be fixed at home. Both inter-city travel (for MPPA) and stay-at-home days (for DPPA) can greatly influence the estimates, as shown in Table 7-1. For the remainder of this chapter I exclude inter-city travel for MPPA and stay-at-home days for DPPA, unless otherwise specified.

The Gini coefficients of MPPA and DPPA are alike, both ranging from 0.55 to 0.64 and suggesting a relatively high travel inequity. For comparison, the Gini coefficents for distribution of household income in 145 countries range from 0.237 to 0.632, with Slovenia, Ukraine, and Denmark the lowest three and Lesotho, South Africa, and Central African Republic the highest three (Central Intelligence Agency 2017). Note several caveats in interpreting the Gini coefficients of PPAs. First and foremost, unlike in the case of income where more income is essentially always better, the PPA reflects the results of individuals' travel preferences and needs. For example, if a person lives only five minutes from work and does not like entertainment activities, the person can have very low PPA but also may not want a higher PPA. The Gini coefficients may, thus, over-estimate the level of inequity. Second, the results are only estimated for fewer than 500 individuals, who may not be representative of society, more generally. Third, the PPA operationalization has simplified assumptions about travel speed, activity schedule, and the transportation network, for the DPPA, and the ellipse shape of activity area for MPPA. Different approaches to estimation can potentially result in different Gini coefficients.

Table 7-1 Gini Coefficients of MPPAs and DPPAs

		Size of PPA	Weighted population in PPA	Commercial /Industrial area in PPA	Paved road in PPA
MPPA	(including inter-city travels)	0.938	/	/	/
	(excluding inter-city travels)	**0.637**	**0.621**	**0.628**	**0.568**
DPPA	(including non-travel days)	0.792	0.790	0.787	0.795
	(excluding non-travel days)	**0.631**	**0.627**	**0.586**	**0.558**

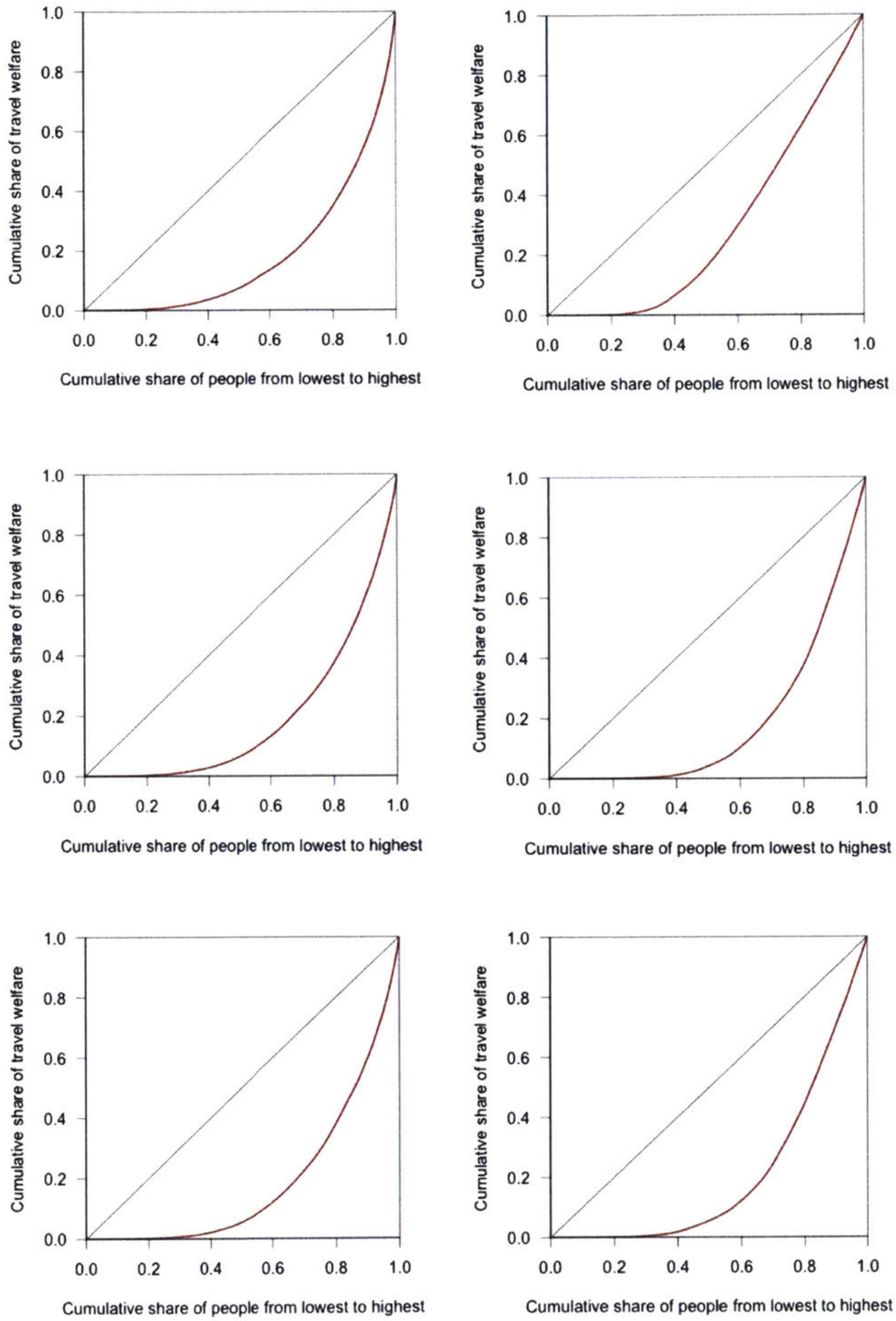
MPPA
(excluding inter-city travels)

DPPA
(excluding non-travel days)

Size of PPA

Weighted population in PPA

Commercial/Industrial area in PPA



Paved road in PPA

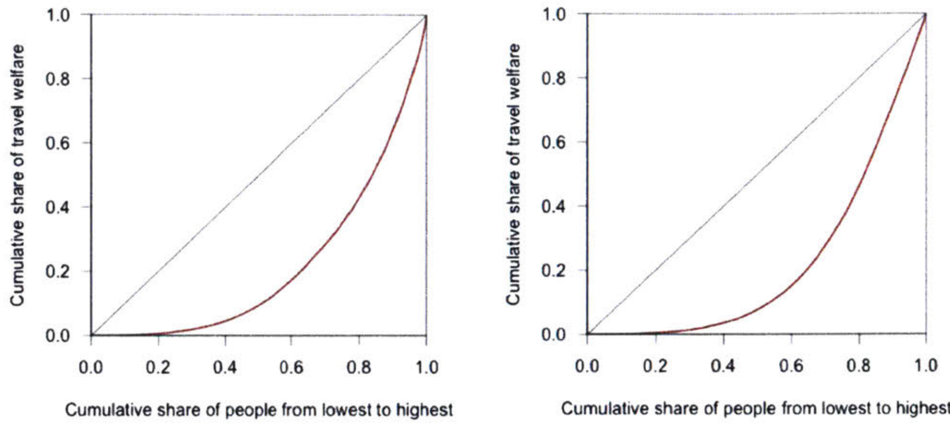


Figure 7-1 Lorenz Curves of MPPAs and DPPAs

7.1.2 Equality of Effectiveness

As described in Chapter 2, applying the equality of effectiveness principle adjusts travel welfare by the price people pay to gain it. As detailed in Chapter 4, in this case the "price" includes estimated housing costs, and an estimated generalized cost of travel, including financial (out-of-pocket) costs (e.g. transit fare, gas cost, parking fees) and monetized time costs. This cost is analogous to one's ability to transform resources into functioning. The generalized cost reflects the investment individuals can spend to gain welfare and is the proxy of travel capability.

In the scatter plot of travel welfare defined by PPA and cost (Figure 7-2), the LOESS curve delineates one estimate of a "sufficient" welfare effectiveness line. People under the line suffer from insufficient travel welfare compared to the estimated cost they pay. However, some rich people may choose to live in expensive houses and own luxurious cars with less concern for cost. To identify those people, I draw a sufficient cost affordability line, using the median cost.[25] People who are both to the left of the line and under the LOESS curve are in the travel welfare effectiveness-disadvantaged group. I

[25] I use the median cost to define the line to simplify the issues and to illustrate the concept. More sophisticated ways could be used to draw the cut-off lines, such as the estimated minimum cost to ensure basic transport needs and an empirically sufficient line of travel capability

apply the severity of travel welfare deficiency index to those disadvantaged groups and evaluate their travel poverty.

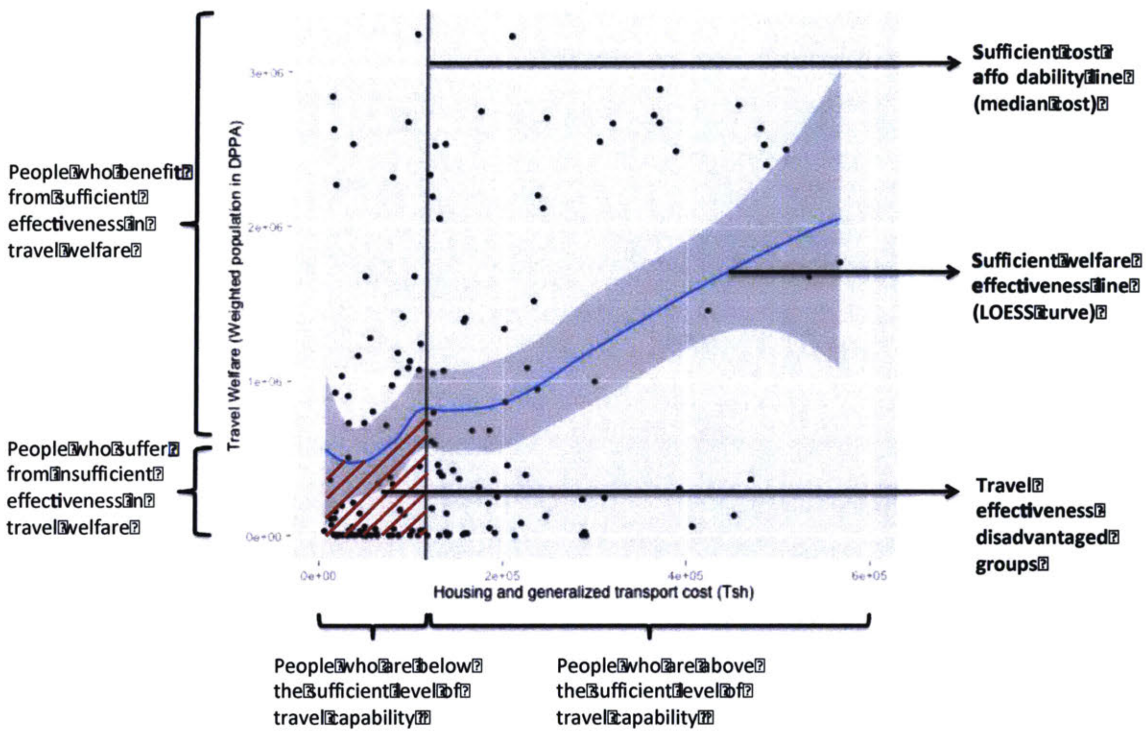


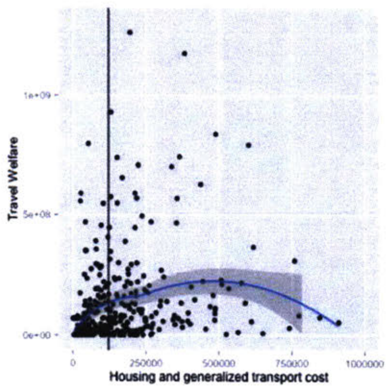
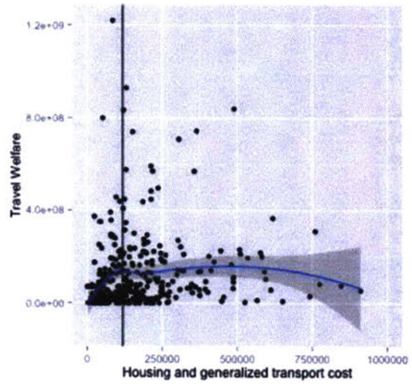
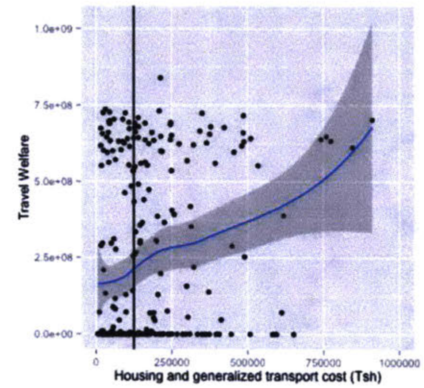
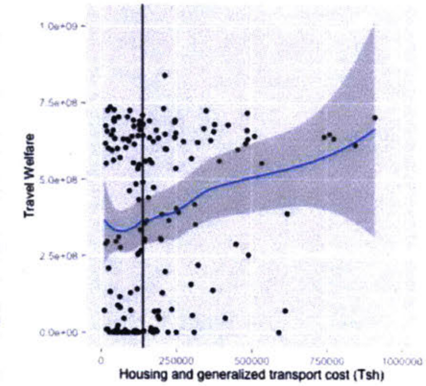
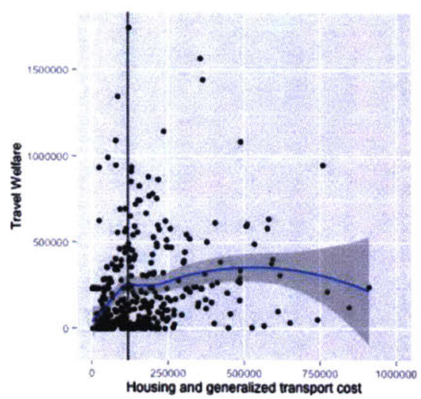
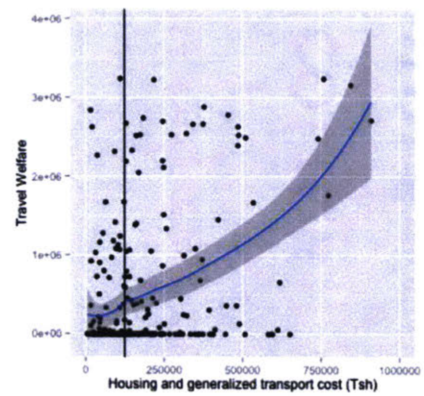
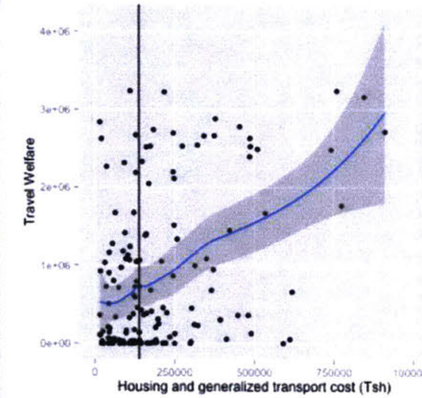
Figure 7-2 Illustration of travel welfare effectiveness-disadvantaged groups

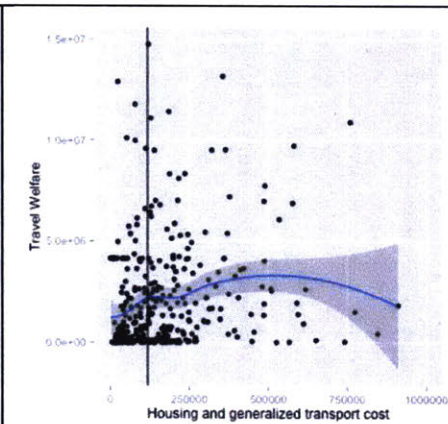
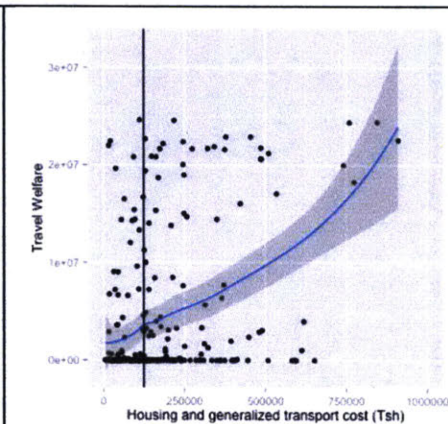
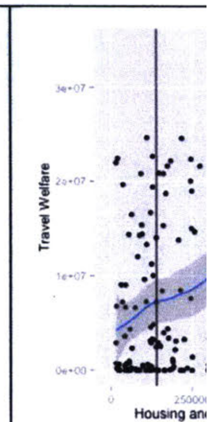
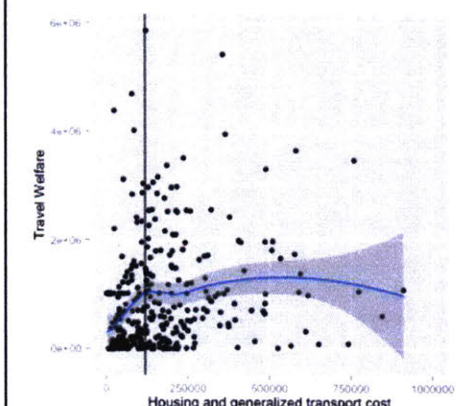
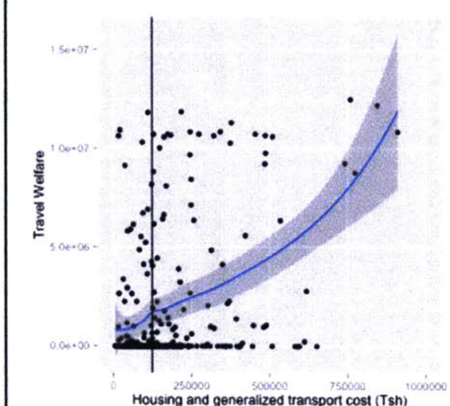
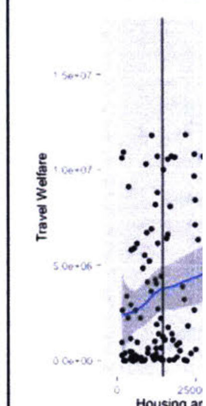
Table 7-2 illustrates the relationship between different PPAs and the corresponding cost and summarizes the severity of travel welfare deficiency index for each measure. Weighted MPPA including inter-city travels are not calculated, because land use information only covers the city of Dar. Despite the variety in individuals' PPA and cost positions, PPA tends to be monotonically increasing with cost before the cost goes too high. This is consistent with the assumption that the more people pay, the higher travel welfare they tend to gain. After cost reaches a high level, DPPA continues to increase with cost, while MPPA decreases with cost. This implies that higher cost tends to always increase one's ability to access more spatial opportunities within one day, but does not necessarily enlarge the area people actually visit over a longer period of time.

The fact that many individuals are far from the LOESS curve shows the complexity of travel behavior, which cannot be simply explained by the cost. The cost dimension here is used only to evaluate travel welfare effectiveness, not to explain or predict PPAs. Compared to MPPA, the dots in the DPPA plot tend to be more extreme – individuals either have very high PPAs or very low PPAs. This is because DPPA only uses one day of travel and contains more randomness. MPPA uses travel data over four weeks, with the results smoothed over a long period.

The severity of travel welfare deficiency is 0.3-0.5 for MPPAs and around 0.5 for DPPAs. This relatively high index indicates that disadvantaged groups suffer fairly severely from insufficient travel welfare effectiveness. The severity of travel welfare deficiency is adapted from the level of poverty measure (Foster, Greer, and Thorbecke 1984); I use sufficient welfare effectiveness (or LOESS curve) as the benchmark line while Foster et. al use poverty line as the benchmark line (equation 9 in Chapter 4). For comparison, Foster, Greer, and Thorbecke (1984) estimate the level of poverty in Nairobi at 0.166 for permanent residents and 0.06 for all residents. It is not an apple-to-apple comparison because the poverty line is supposed to be lower than median income, the latter of which is analogous to sufficient welfare effectiveness. Therefore, the result of the severity of travel welfare deficiency is supposed to be higher than its concept predecessor, but 0.5 still seems to be a relatively high number.

Table 7-2 Scatter Plots of Travel Welfare (Y-axis) and Housing and Generalized Transportation Cost (X-axis) with LOESS curves and The Severity of Travel Welfare Deficiency Index by Multiple PPA measures

	MPPA		DPPA	
	(including inter-city travel)	(excluding inter-city travel)	(including non-travel day)	(excluding non-travel day)
Size of PPA	 The severity of travel welfare deficiency index: 0.561	 The severity of travel welfare deficiency index: 0.491	 The severity of travel welfare deficiency index: 0.612	 The severity of travel welfa deficiency index: 0.511
Weighted population in PPA	/	 The severity of travel welfare deficiency index: 0.497	 The severity of travel welfare deficiency index: 0.609	 The severity of travel welfa deficiency index: 0.507

	 The severity of travel welfare deficiency index: 0.400	 The severity of travel welfare deficiency index: 0.629	 The severity deficiency index
	 The severity of travel welfare deficiency index: 0.323	 The severity of travel welfare deficiency index: 0.614	 The severity deficiency index

7.2 Travel Welfare Spatial Distribution

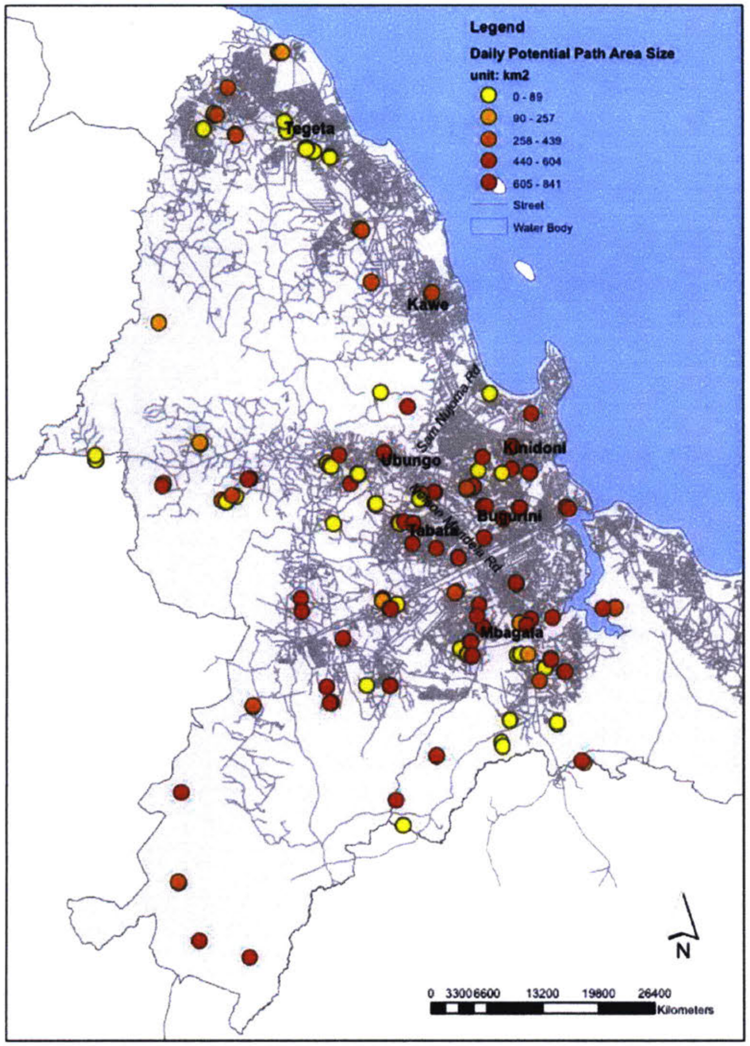
Figure 7-3 visualizes on maps the travel welfare defined by monthly potential path area (MPPA), daily potential path area (DPPA), and MPPA and DPPA weighted by three spatial opportunities: average day and night population (proxy for the level of urban development), commercial/industrial areas, and paved road areas. As mentioned in the context chapter (Chapter 3), the commercial opportunities and paved roads are sparse and precious in Dar. The right column maps demonstrate the PPAs after adjusting for housing and generalized transport cost, reflecting the "equality of effectiveness" principle. The left column maps show the PPAs before cost adjustment, following the "equality of welfare" principle. Each dot represents one individual at his/her home location and darker red denotes higher PPA.

The travel welfare distribution maps reveal several noteworthy patterns. First, comparing DPPA and MPPA shows relatively strong differences, as DPPA's apparent monocentricity is not obviously discernible for MPPA. Recall that DPPA measures one-day accessible area with a spatiotemporal constraint; by this measure, people who have high DPPA tend to be concentrated in homes closer to the central city. Those areas often have better transport infrastructure, which allow people to travel faster and further. In addition, people who live in those areas are often closer to opportunities including workplaces and schools. They may have more time flexibility because of shorter commuting times. Nonetheless, given the nature of the DPPA calculation, people who live further away from core city areas may still have a large DPPA. For example, if they travel far to the market in the city center, all out-of-home activities and travel are considered flexible in time schedule; the resulting DPPA can be large. However, because the majority of individuals from the FMS sample (76%) have jobs, such behaviors are likely not common during weekdays. In contrast, MPPA measures the potentially accessible areas based on monthly activity trajectories; people who live far away from their destinations tend to have a higher ellipse-measured MPPA. However, it is hard to distinguish individuals who have to travel far to reach their destinations (e.g. workplaces) from those who want to travel far to explore interesting activities (e.g. relaxing on the beach). Consequently, no obvious spatial patterns emerge (Figure 7-4): people who live in neighborhoods with good accessibility may travel more often, while people who live in distant neighborhoods may have to travel far; but they may both generate high MPPAs. In that sense, although ellipse-measured MPPA reduces the randomness by considering a month period records, the measure itself does not reveal reasons behind such travel behaviors.

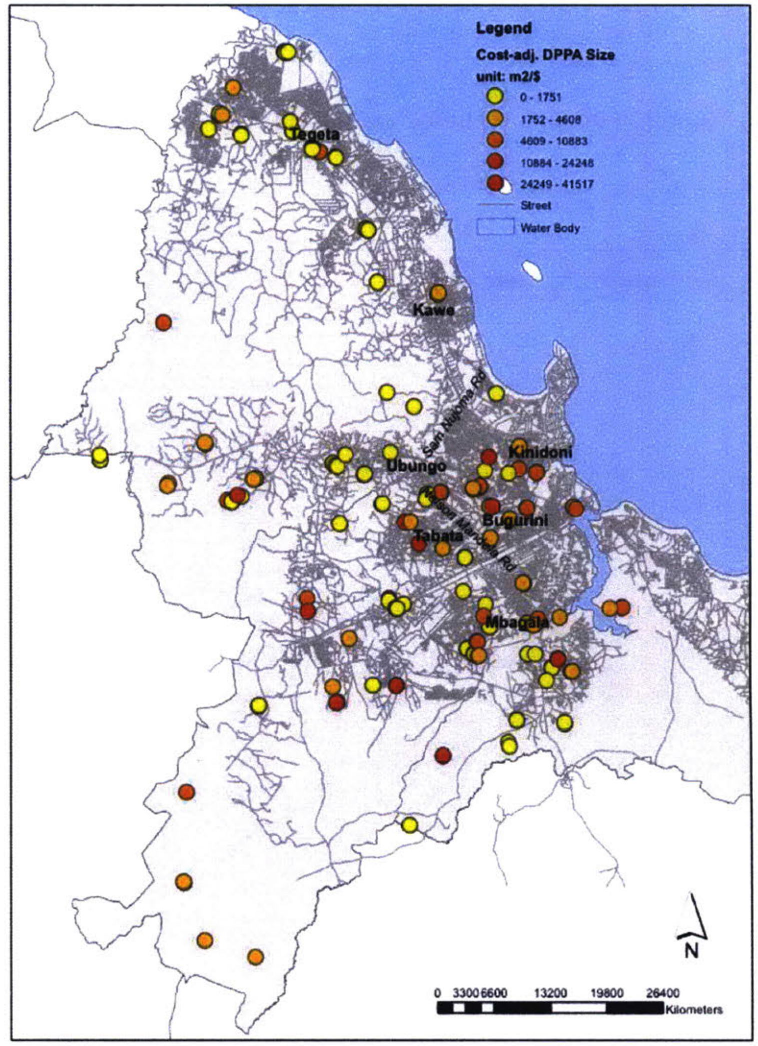
Second, the cost adjustment mitigates the variation of travel welfare, as measured by PPA, among the population. After accounting for an individual's estimated travel and location costs, apparent welfare "advantages" disappear (for both weighted and un-weighted PPAs). In other words, from an "egalitarian" theoretical perspective, direct comparison of welfare suggests inequality among individuals; from a "capabilities" theoretical perspective, comparison of cost-adjusted welfare suggests more equality, in terms of effectiveness. This implies that some of the observed imbalance in travel welfare arises from differences in preferences: in the trade-off between living in places with higher accessibility and paying less for housing and transportation, some people choose the latter in sacrifice of the former.

Third, observing these maps reveals some neighborhoods with good travel welfare defined by multiple measures. The absolute weighted DPPA without cost adjustment indicates that neighborhoods within the central city, especially within the first ring road, tend to have high accessibility; absolute weighted MPPA suggests a more scattered spatial pattern. More specifically, individuals who live in Mbagala and Kinondoni have better weighted PPAs in all measures. Apart from them, neighborhoods with relatively high travel welfare effectiveness by weighted cost-adjusted DPPA measure include Tabata, Kawe, Bugurini, and other neighborhoods within the second ring road (Sam Nujoma Rd and Nelson Mandela Rd); neighborhoods with high effectiveness by weighted cost-adjusted MPPA measure include Tegeta.

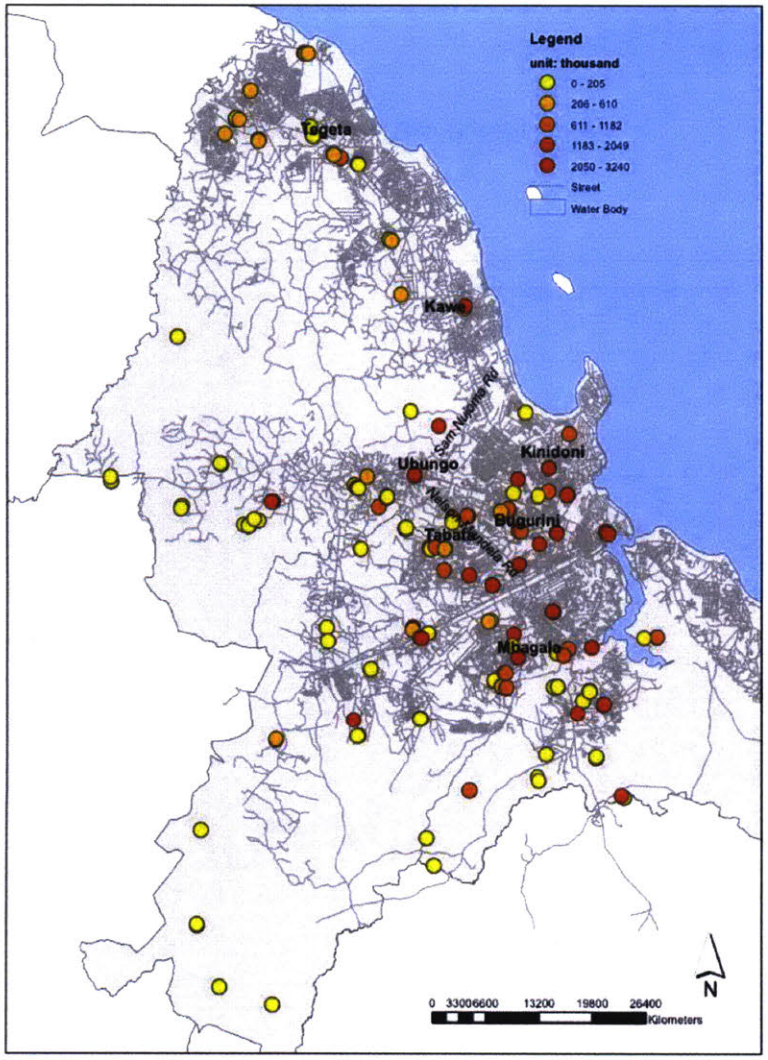
Fourth, both cost-adjusted and non cost-adjusted PPAs concentrate more in the central city after considering opportunities density weights (e,g. unweighted DPPA (1a) versus population-weighted DPPA (2a) in Figure 7-3; unweighted MPPA (1a) versus population weighted MPPA (2b) in Figure 7-4). This is consistent with the monocentric city structure in Dar. More dense opportunities in the city center increase the value of PPA. In other words, for two PPAs with same sizes but different locations, the PPA closer to city center potentially brings more benefits to individuals. The PPAs weighted by three spatial opportunities show very similar patterns (e.g. population-weighted DPPA (2a) versus commercial area-weighted DPPA (3a) versus paved road-weighted DPPA (4a) in Figure 7-3), implying similar distribution of population, commercial areas and paved roads. Arguably, the weighted PPAs provide a better representation of travel welfare, as opportunities within the PPAs are more relevant to an individual's welfare than purely PPA size.



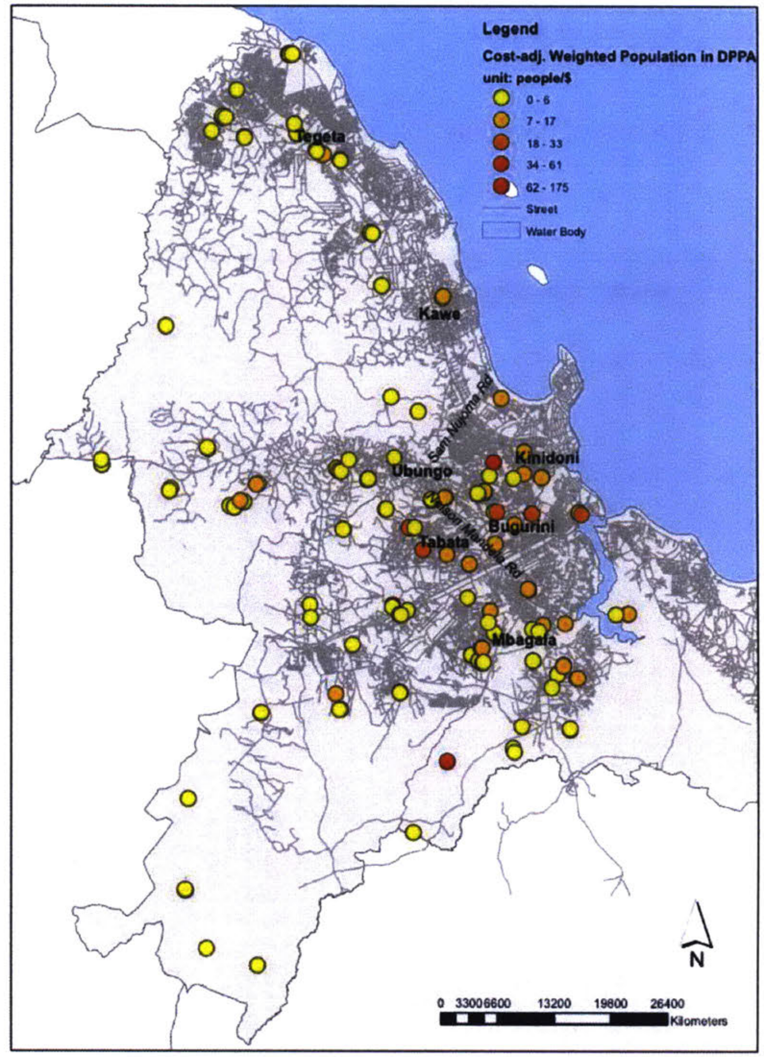
1a. Daily Potential Path Area (DPPA) Size



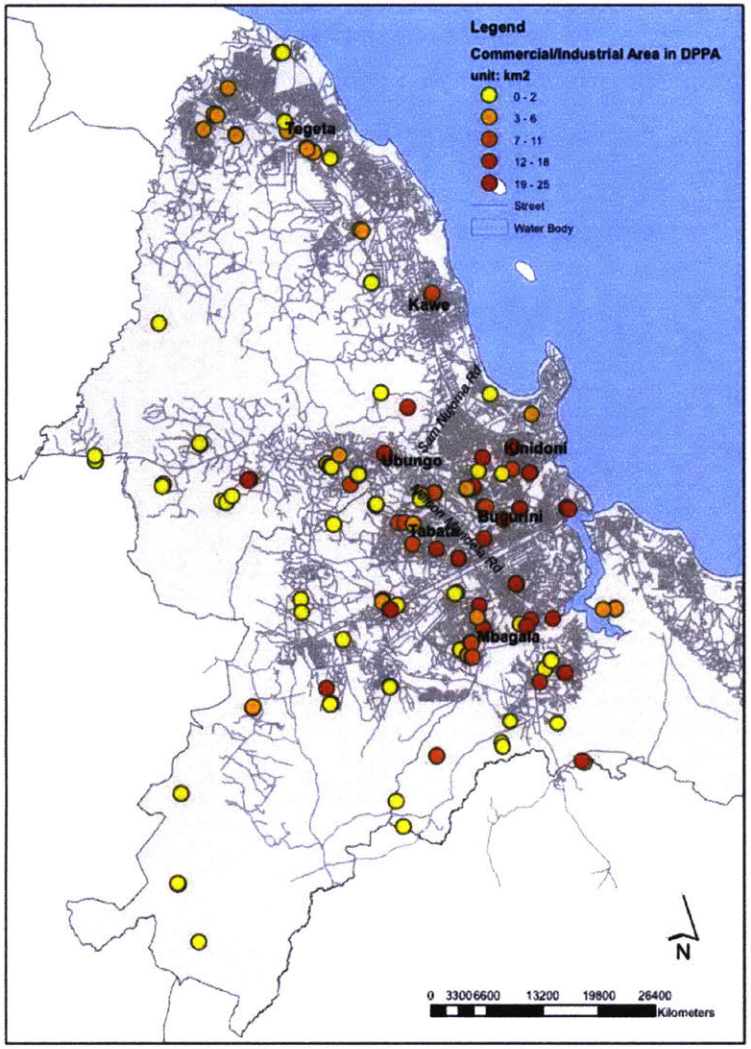
1b. Cost-adjusted Daily Potential Path Area (DPPA) Size



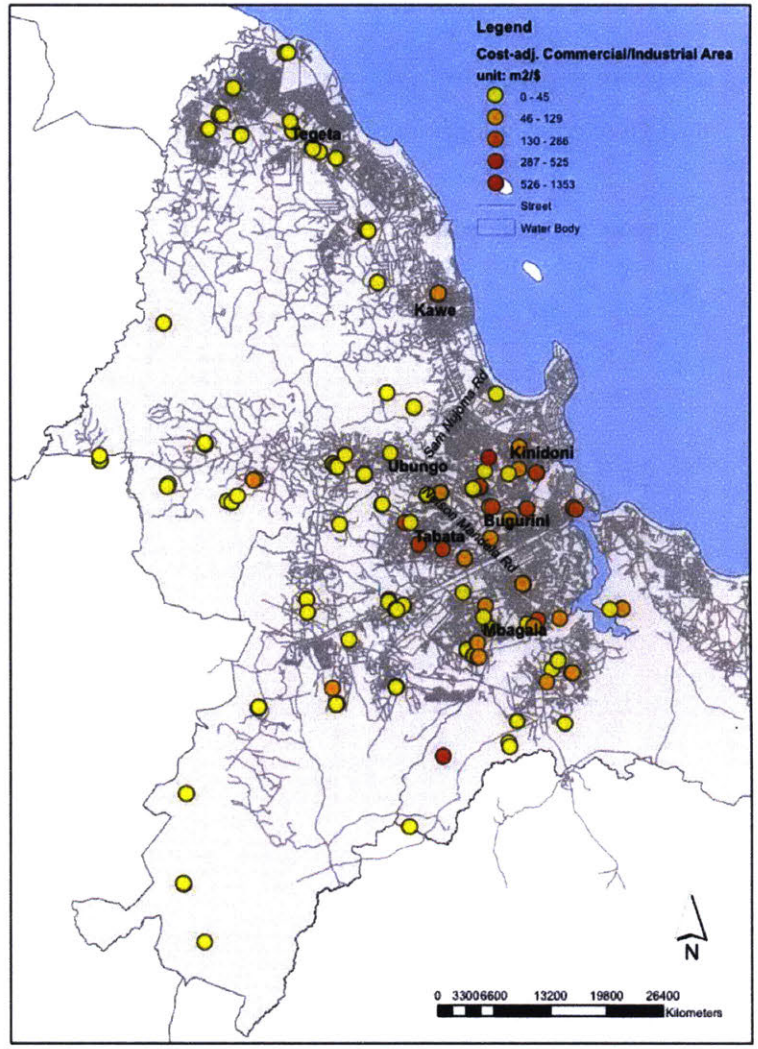
2a. Weighted Population in DPPA



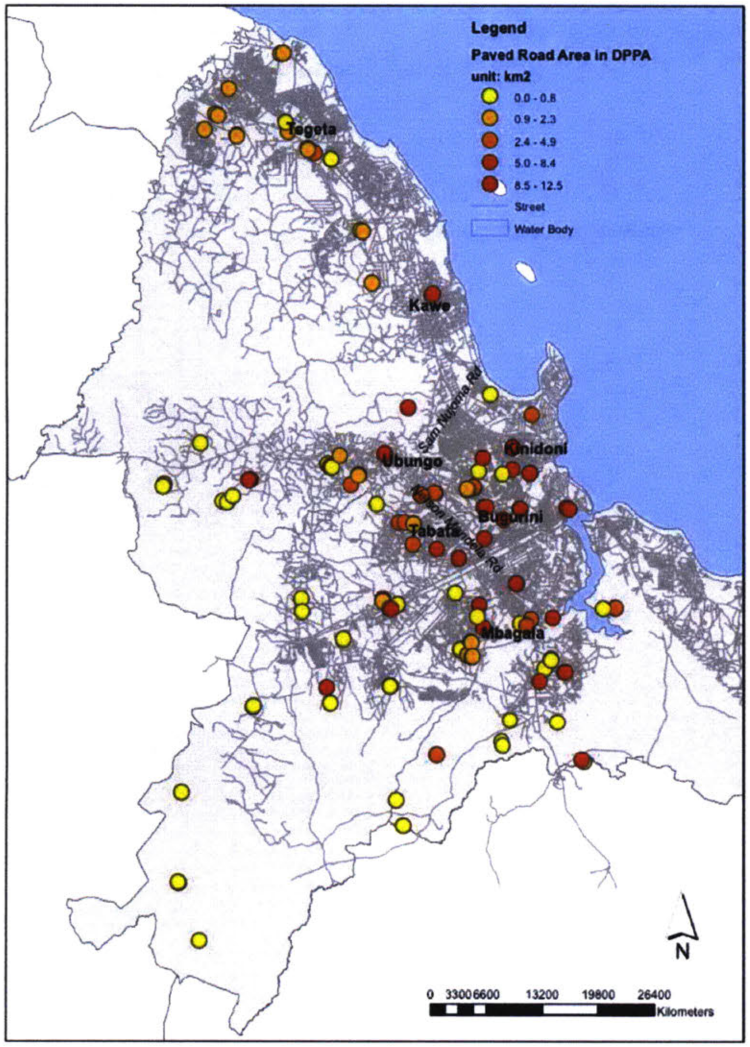
2b. Cost-adjusted Weighted Population in DPPA



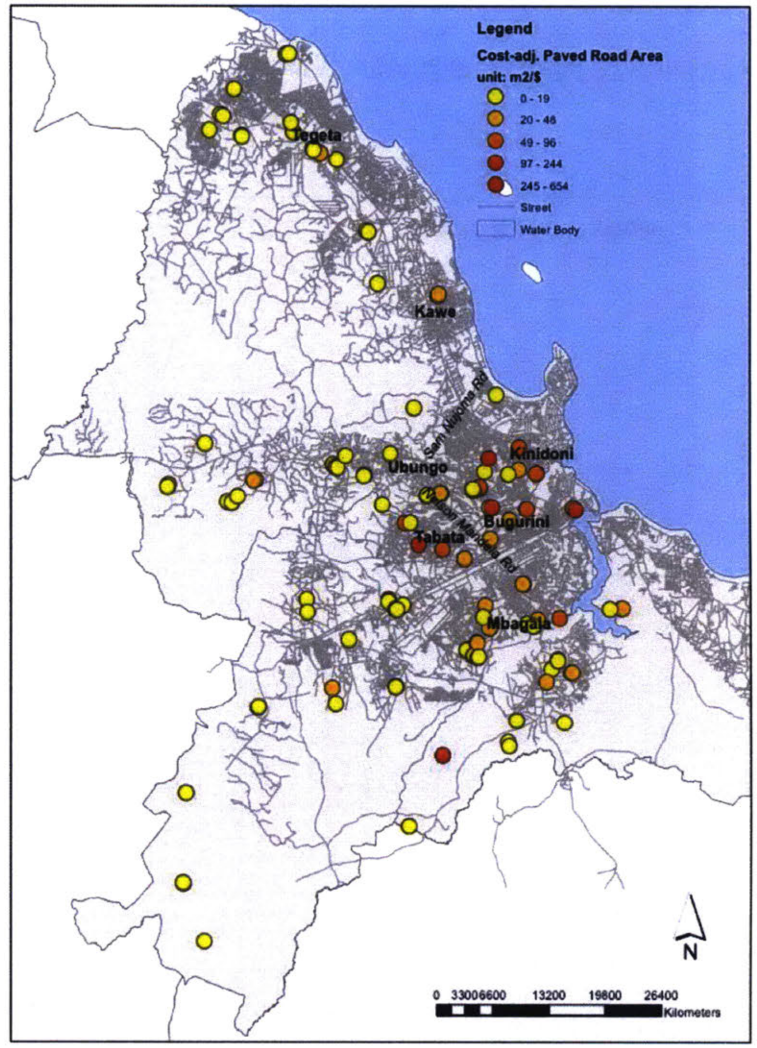
3a. Commercial/Industrial Area in DPPA



3b. Cost-adjusted Commercial/Industrial Area in DPPA

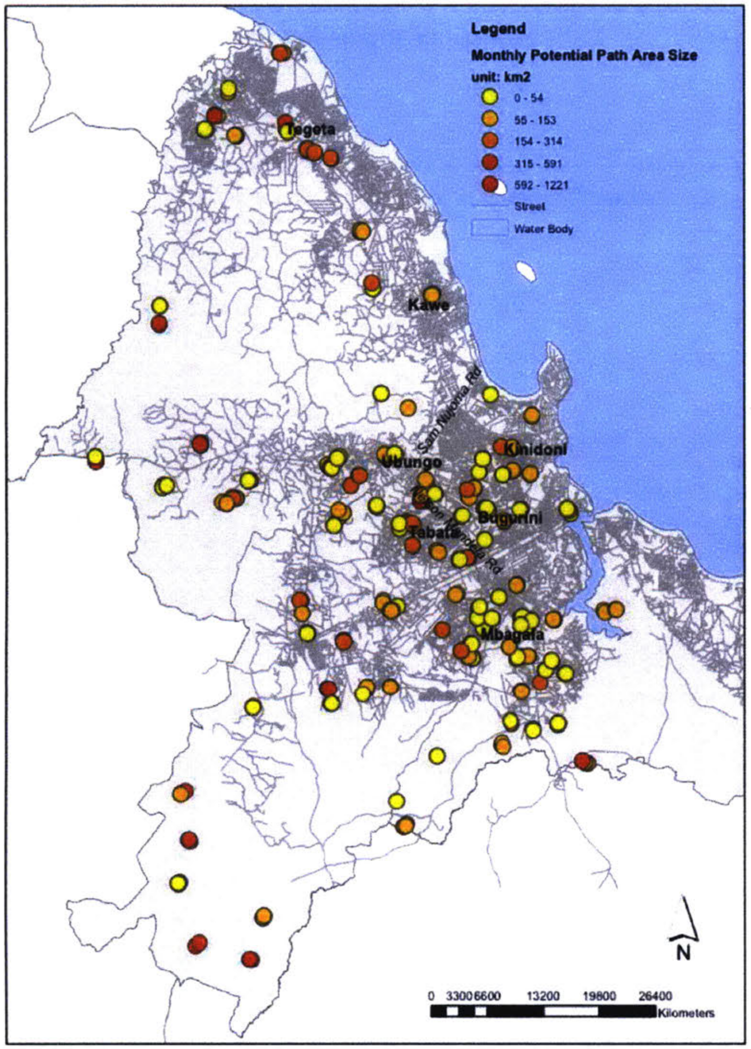


4a. Paved Road Area in DPPA

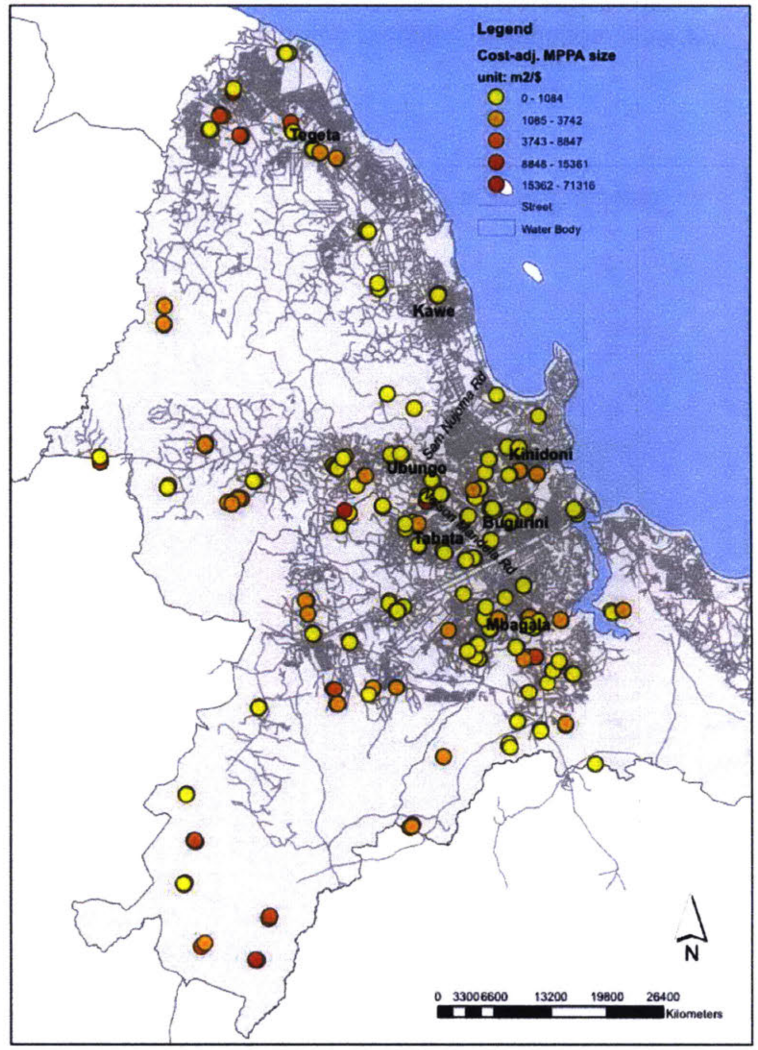


4b. Cost-adjusted Paved Road Area in DPPA

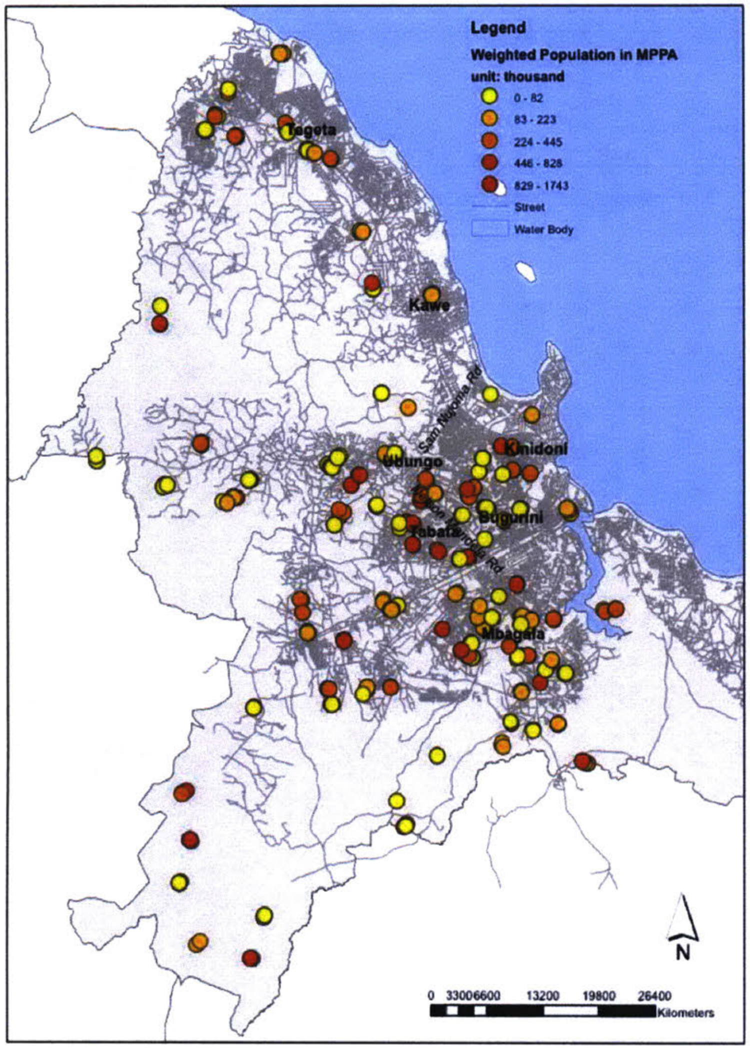
Figure 7-3 Spatial distributions and Histograms of Daily Potential Path Area (DPPA) and Weighted DPPA on Maps (Each dot represents one individual, located at home)



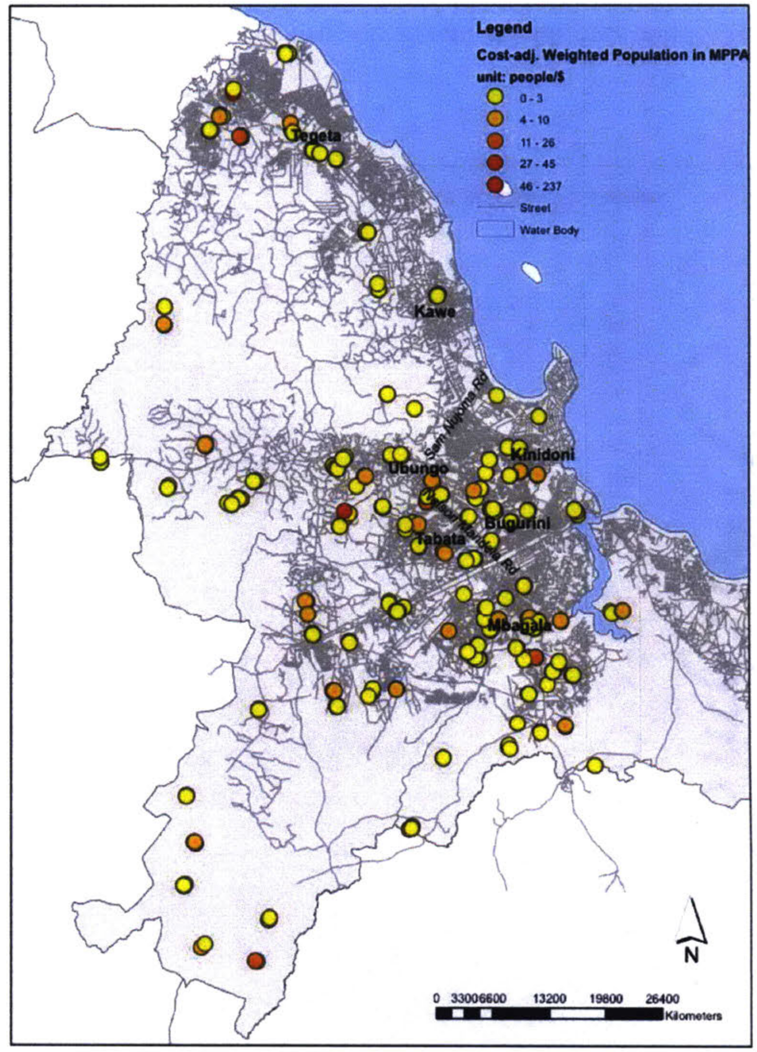
1a. Monthly Potential Path Area (MPPA) Size



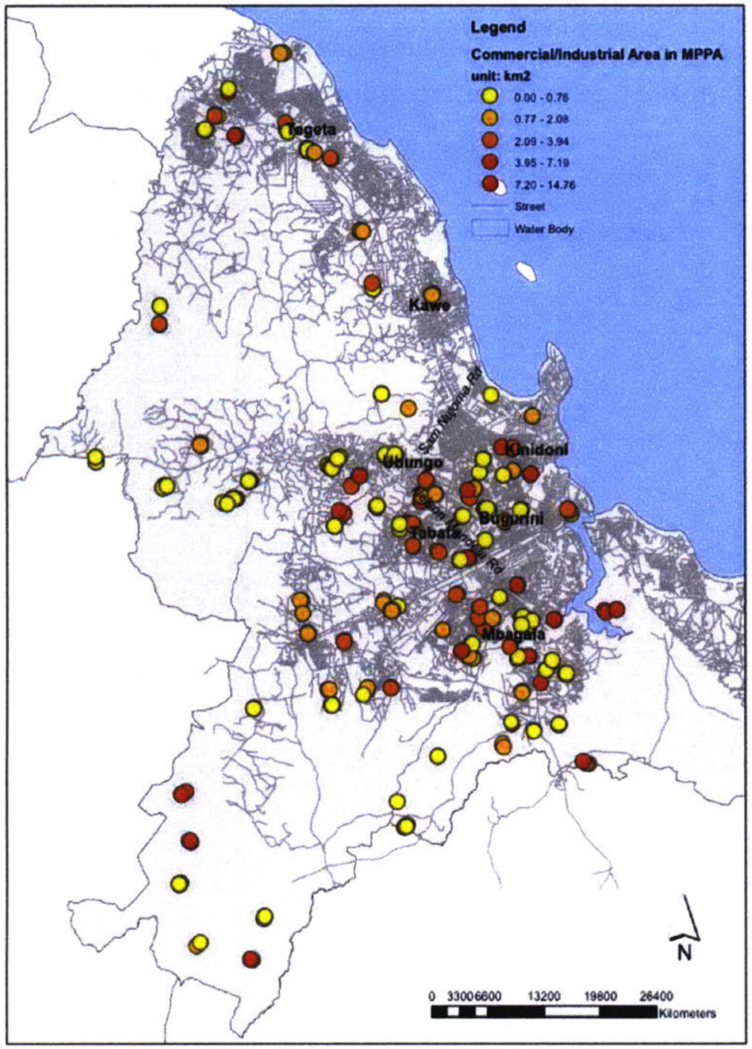
1b. Cost-adjusted Monthly Potential Path Area (MPPA) Size



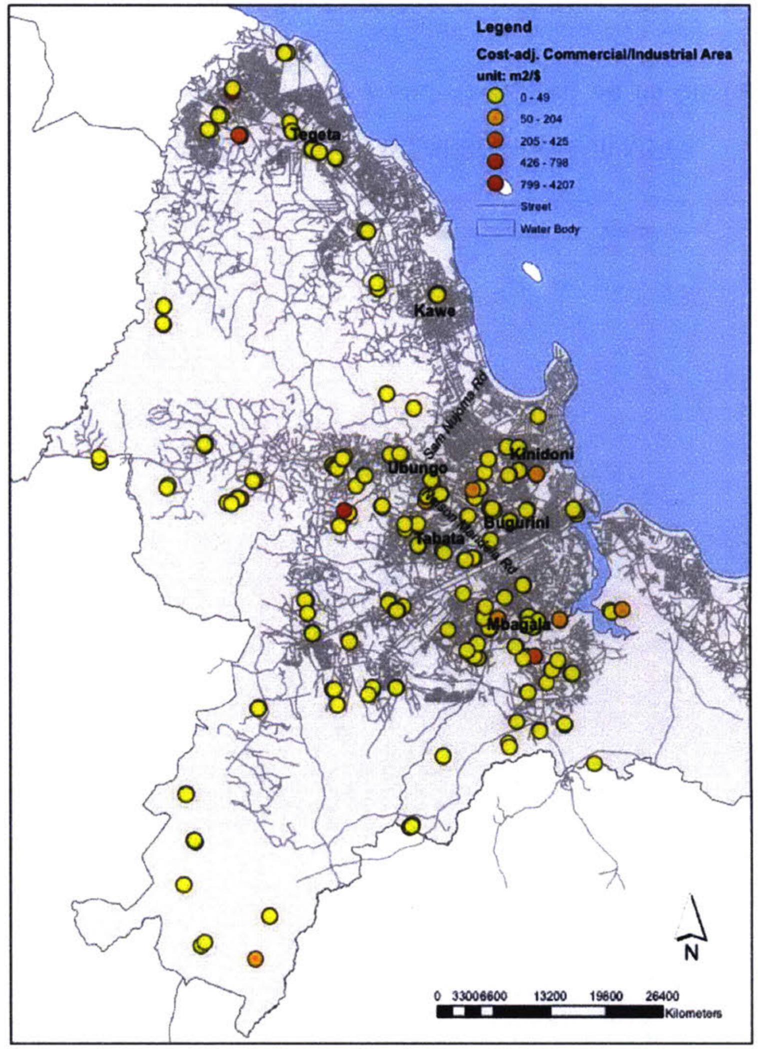
2a. Weighted Population in MPPA



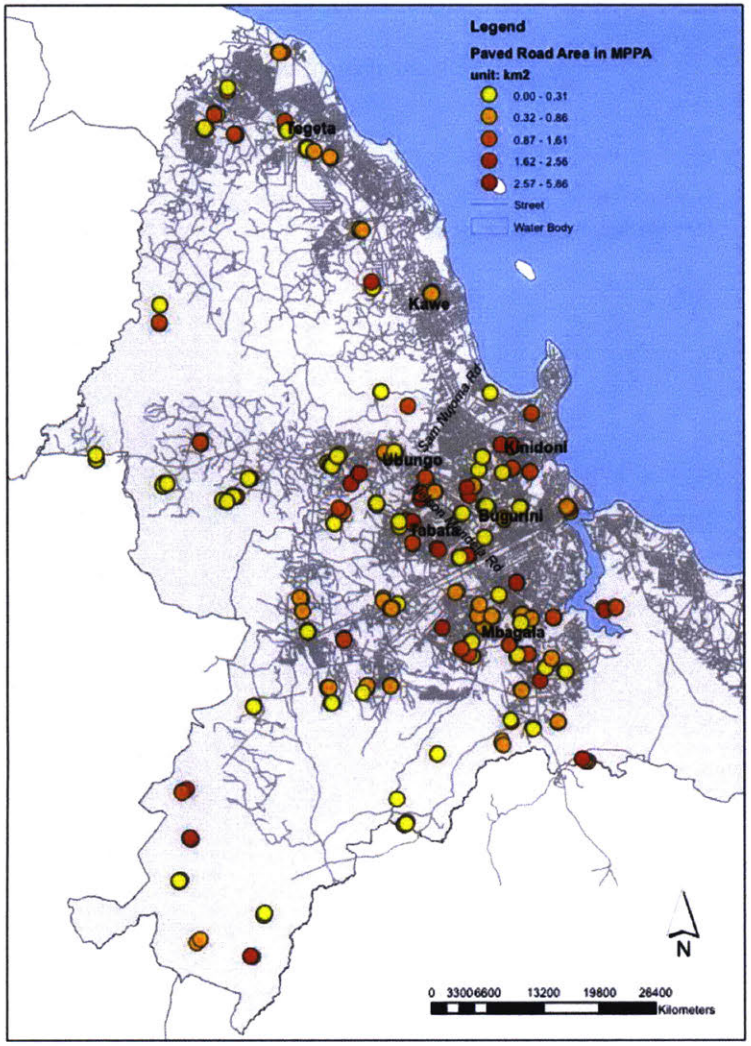
2b. Cost-adjusted Weighted Population in MPPA

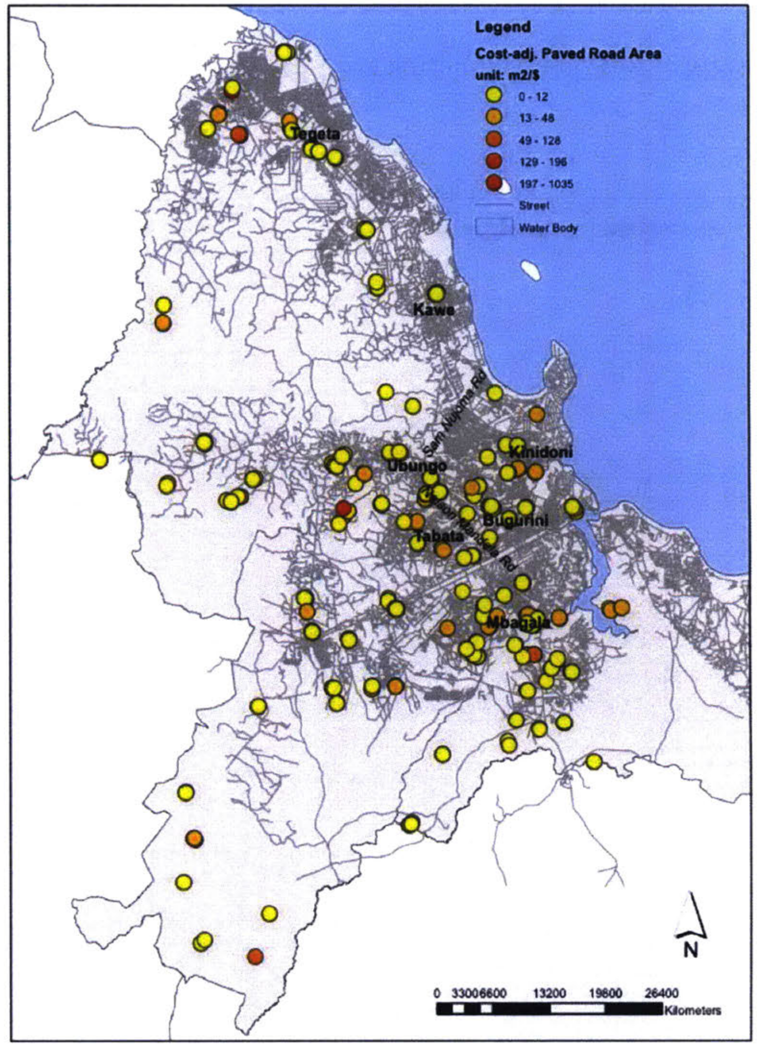


3a. Commercial/Industrial Area in MPPA



3b. Cost-adjusted Commercial/Industrial Area in MPPA





4a. Paved Road Area in MPPA

4b. Cost-adjusted Paved Road Area in MPPA

Figure 7-4 Spatial distributions and Histograms of Monthly Potential Path Area (MPPA) and Weighted on Maps (Each dot represents one individual, located at home)

7.3 Factors Influencing the Level of Travel Welfare

Table 7-3 and Table 7-4 and show the results of socio-economic and built-environment variables influencing PPAs. The dependent variables are weighted and un-weighted, cost-adjusted and non cost-adjusted MPPAs and DPPAs in separate models. The independent variables are socio-economic and built-environment variables. I also tried urban expansion factors (Páez, Scott, and Morency 2012) to represent variation across locations, instead of distance from home to CBD in the model, but the interaction term of expansion factors and individual characteristics are harder to interpret. Therefore, I only report the models using distance from home to CBD as the measure of relative location.

The results for the DPPAs and MPPAs are different, as expected. The similarities between the two sets of models are somewhat general: they both suggest different influences on non cost-adjusted PPAs versus cost-adjusted PPAs; the signs of the coefficients are consistent regardless of the opportunity weighting scheme used; and, the non-significant Moran's I in all models suggests negligible spatial autocorrelation.

For DPPAs, males, middle-age individuals (~ 40 years old)[26], people who have university or higher education, people who have motorcycle-driving license, and people who live closer to CBD tend to have higher absolute weighted DPPAs. However, all individual characteristics, other than location, do not have significant influence on cost-adjusted DPPAs (i.e. travel welfare effectiveness). This is also consistent with the finding that cost adjustment mitigates the variation of travel welfare (section 7.2). This implies that travel welfare inequity of different social groups does exist, but it seems to be caused by the different capabilities to obtain accessibility (by housing location and travel costs) rather than differences in travel welfare returns per cost unit. The latter reflects land use and transportation situation, while the former reveals social issues beyond infrastructure. For example, men have higher weighted DPPA, but the male benefit fades when I adjust it by

---

[26] The positive sign of age and the negative sign of age squared suggest a parabolic curve that opens downwards. With the general form of a quadratic $= ax^2 + bx + c$, the curve reaches vertex when $x = -\frac{b}{2a}$. Based on coefficients in Table 7-3, DPPA peaks when age is close to 40.

capabilities. Therefore, men's higher accessibility is mostly accounted for by the fact that they can pay more for it. This capability difference is more of a social problem; land use and transportation infrastructure does not reinforce that structural gender difference because the travel welfare effectiveness does not significantly differ between men and women. In the end, after adjusting for the cost (equality of effectiveness), the only factor that explains the DPPA (weighted) difference is the locational factors – living in a consolidated neighborhood, or closer to the CBD. All else equal, someone living further from the CBD gets less accessibility per unit invested; those living in a consolidated neighborhood gets more accessibility per unit invested. It suggests unequally distributed good infrastructure in formal communities and poor infrastructure in suburban areas. I tried several other variables in various model estimations, but excluded them from the final models because they do not have significant influence. These variables include total working hours per day, monthly sales of family business, marriage status, population density in the neighborhood, and interaction terms between distance to CBD and other individual characteristics. I dropped the number of paved road pixels within 1km buffer from home because it is highly correlated with that of regular neighborhood pixels (correlation ~ 0.8).

For MPPAs, similarly, middle age individuals (~ 45 years old) have higher absolute MPPAs but not higher MPPA travel welfare effectiveness; males even have lower cost-adjusted MPPAs. Wealth and education do not have significant influence on absolute MPPAs, but have impacts on cost-adjusted MPPAs and their influences depend on the distance from home to CBD. For wealth, the further people live far from CBD, the more negative effects of wealth on cost-adjusted MPPAs; if the individuals live further than 5km from CBD, the impact of wealth becomes negative. For people with university or higher education, their MPPA travel welfare effectiveness is negative near CBD; the further their homes are from the CBD, the less negative the effects of university education; for those who live farther than 15km, their university education begins to have positive effect on MPPA travel welfare effectiveness. In the estimation process, I tested numerous other variables which I left out of the final specification, including: whether individuals have a driving license and the type, monthly sales from family business, marriage status, whether living close to formal settlement, population density, and total working hours. In general, it is harder to explain the logic behind the factors influencing MPPAs because, as mentioned above, two mixed effects can contribute to a large ellipse-measured MPPA: people have to travel further and people who want to explore diverse destinations.

Table 7-3 Factors including weighted and un-weighted, cost-adjusted and non cost-adjusted DPPAs

	Size of DPPA (km2)			Cost-adjusted Size of DPPA (m2/$)			Weighted population in DPPA (thousand)			Cost-adjusted Weighted population in DPPA (people/$)			Commercial/Industrial area in DPPA (km2)			Cost-adjusted Commercial/Industrial area in DPPA (m2/$)			Paved road in DPPA (km2)			Cost-adjusted Paved road in DPPA (m2/$)		
	coef	*p-value*		*coef*	*p-value*		*coef*	*p-value*		*coef*	*p-value*		*coef*	*p-value*		*coef*	*p-value*		*coef*	*p-value*		*coef*	*p-value*	
(Intercept)	9.2	0.97		305.31	0.96		399.3	0.63		11.09	0.50		2.72	0.73		116.81	0.41		1.47	0.71		54.82	0.45	
Male	84.7	0.01	*	-2994.84	0.01	**	314.0	0.02	*	-3.44	0.20		2.56	0.04	*	-30.79	0.18		1.29	0.05	*	-14.92	0.21	
Age	31.3	0.00	**	410.83	0.24		84.7	0.05	*	0.48	0.57		0.84	0.04	*	3.58	0.62		0.35	0.09	`	1.52	0.69	
Age squared	-0.4	0.01	**	-5.21	0.25		-1.0	0.05	`	-0.01	0.54		-0.01	0.04	*	-0.05	0.61		0.00	0.12		-0.02	0.70	
Education: no school or adult school	-163.3	0.43		-3855.90	0.56																			
Education: secondary school	0.8	0.98		-196.66	0.87		238.0	0.12		-0.01	1.00		2.39	0.09	`	2.48	0.92		1.13	0.11		-1.61	0.90	
Education: university or higher	220.2	0.01	*	2772.51	0.25		751.3	0.02	*	6.67	0.25		6.36	0.02	*	55.69	0.26		3.20	0.02	*	24.42	0.32	
Has driving license - car	-43.5	0.42		-950.43	0.54		148.6	0.46		-0.77	0.83		1.14	0.52		-5.86	0.85		0.58	0.52		-1.81	0.91	
Has driving license - motorcycle	70.7	0.55		-930.32	0.78		1035.2	0.02	*	2.01	0.80		7.02	0.06	`	11.85	0.86		5.43	0.01	**	8.93	0.81	
Wealth index per person	-63.5	0.26		-2046.64	0.16		-282.0	0.13		-5.86	0.09	`	-2.56	0.11		-55.66	0.06	`	-1.13	0.17		-25.23	0.09	`
Number of regular neighborhood pixels within 1km buffer from home	0.0	0.89		0.78	0.05	*	0.0	0.17		0.00	0.00	**	0.00	0.19		0.02	0.00	**	0.00	0.22		0.01	0.01	**
log(distance from home to CBD)	-108.3	0.00	***	-849.93	0.25		-620.8	0.00	***	-4.71	0.01	**	-5.82	0.00	***	-45.51	0.00	**	-2.54	0.00	***	-20.30	0.01	**
sigma (random effect)	136.8			6161.64			781.2			14.97			6.72			122.64			3.25			60.58		
AIC	2557.2			3468.47			2856.7			1346.07			1093.29			1857.08			780.66			1495.11		
BIC	2602.1			3512.29			2897.7			1386.04			1133.10			1895.96			819.08			1532.69		
Moran's I: observed	0.0			0.03			0.0			0.00			0.02			-0.01			-0.02			-0.01		
Moran's I: p-value	0.3			0.37			0.7			0.97			0.62			0.83			0.77			0.75		

Table 7-4 Factors including weighted and un-weighted, cost-adjusted and non cost-adjusted MPPAs

	Size of MPPA (km2)			Cost-adjusted Size of MPPA (m2/$)			Weighted population in MPPA (thousand)			Cost-adjusted Weighted population in MPPA (people/$)		
	coef	*p-value*		*coef*	*p-value*		*coef*	*p-value*		*coef*	*p-value*	
(Intercept)	-135.5	0.10		1277.86	0.06	`	-148.6	0.36		4.10	0.06	`
Male				-817.43	0.00	**				-1.49	0.02	*
Age	9.0	0.05	*				16.2	0.06				
Age squared	-0.1	0.09	`				-0.1	0.14				
Education: no school or adult school				1460.44	0.62		97.0	0.71		0.14	0.98	
Education: secondary school				115.09	0.89		55.1	0.08	`	0.76	0.72	
Education: university or higher				-2736.70	0.09	`	79.3	0.18		-7.13	0.07	`
Wealth index per person	69.1	0.00	**	1589.27	0.29		87.5	0.03	*	5.27	0.30	
Distance from home to CBD	4.8	0.00	***	28.49	0.51		-0.3	0.84		-0.05	0.71	
* Wealth index per person[27]				-335.92	0.00	**				-1.17	0.00	***
* Education: secondary school				-4.89	0.93					-0.01	0.97	
* Education: university or higher				209.28	0.07	`				0.52	0.06	`
sigma (random effect)	127.7			2125.42			230.3			4.80		
AIC	5573.3			7536.07			5824.2			3047.76		
BIC	5601.8			7583.51			5864.5			3095.20		
Moran's I: observed	-0.0			0.00			-0.0			0.00		
Moran's I: p-value	0.5			0.93			0.1			0.92		

	Commercial/Industrial area in MPPA (km2)			Cost-adjusted Commercial/Industrial area in MPPA (m2/$)			Paved road in MPPA (km2)			Cost-adjusted Paved road in MPPA (m2/$)		
	coef	*p-value*		*coef*	*p-value*		*coef*	*p-value*		*coef*	*p-value*	
(Intercept)	1.07	0.06	`	45.11	0.24		0.20	0.36		14.06	0.16	
Male												
Age	0.03	0.03	*				0.02	0.00	***			
Age squared												
Education: no school or adult school	-2.52	0.32		4.17	0.96		-0.25	0.79		4.80	0.88	
Education: secondary school	0.21	0.50		-0.66	0.98		0.20	0.09	`	2.60	0.81	
Education: university or higher	0.73	0.21		-157.96	0.00	**	0.48	0.03	*	-46.15	0.02	*
Wealth index per person	0.45	0.27		111.32	0.23		0.25	0.11		37.23	0.11	
Distance from home to CBD	-0.01	0.45		-0.25	0.92		0.00	0.78		-0.06	0.93	
* Wealth index per person[27]				-24.10	0.00	***				-7.02	0.00	***
* Education: secondary school				0.61	0.77					0.04	0.96	
* Education: university or higher				9.61	0.01	*				3.02	0.03	*
sigma (random effect)	2.27			58.17			0.83			21.87		
AIC	1829.11			4797.02			1004.99			3659.86		
BIC	1864.64			4839.61			1039.97			3701.75		
Moran's I: observed	-0.01			0.00			-0.03			0.00		
Moran's I: p-value	0.79			0.89			0.31			0.91		

[27] Note: the three variables with "*" signs in the beginning are interaction terms (with distance from home to CBD)

7.4 Summary

This chapter operationalizes the evaluation of one complicated concept in transportation: travel equity. Travel equity is important in planning and infrastructure investment, yet it is a difficult concept to operationalize in practice. The difficult and complexity are embedded in the various perspectives by which one can define the concept, determine the suitable principles, and, thus, measure it.

As suggested in this chapter, equity evaluation results highly depend on the chosen measure and principle. This chapter uses two space-time measures – monthly potential path area (MPPA) and daily potential path area (DPPA) – to measure travel welfare. Then the following analysis evaluates equity based on two principles: equality of welfare and equality of effectiveness. This chapter discusses empirical differences between the two space-time measures and between the two principles. First, MPPA and DPPA are very different travel welfare measures in nature. MPPA measures potentially passed area based on monthly activity trajectory, while DPPA measures one-day accessible area with spatiotemporal constraints. Their spatial distributions and the regression results on socio-economic and built-environment variables differ. In Dar's analysis, DPPA suggests a monocentric distribution and decreases significantly when distance to CBD increases; MPPA spreads relatively randomly across the whole city, partially because it is hard to distinguish individuals who want to travel far to explore from those who have to travel far. The bias in the number of fully verified days (discussed in Chapter 6) also potentially introduces representative issues to the MPPA measure. Due to these confounding effects, I draw most insights from DPPA in this chapter. Second, equality of welfare and equality of effectiveness suggest different disadvantaged groups. Equality of welfare requires that individuals have the same level of travel welfare or absolute PPAs, overlooking variations in people's travel preferences and needs. Equality of effectiveness provides another perspective: aiming to equalize the "returns" of travel welfare proportional to the cost. In this chapter's analysis, most social characteristics that significantly influence equity from equality of welfare perspective do not matter from the perspective of equality of effectiveness.

The analysis in this Chapter reveals three major findings. First, both MPPA and DPPA suggest relatively high travel inequity in Dar, evaluated by both Gini coefficient and the severity index of travel welfare deficiency. Second, although accessibility is unequal from the perspective of equality of welfare, the benefit is not highly distorted based on what people pay (equality of effectiveness). For DPPA, the differences in travel welfare are due to gender, age, education, and location. However, once one accounts for how much people pay to obtain the accessibility, the significant influence from all individual characteristics disappear; the variation of travel welfare also decreases across the city. Third, location matters for travel welfare. For DPPA, people who live in informal settlements have significantly lower travel welfare effectiveness; people who are distant from city center have both lower absolute travel welfare and lower returns of welfare per cost unit invested.

Two policy implications emerge from the findings. First, although travel inequity is relatively high in Dar, we cannot attribute the entire problem to land use and transportation infrastructure issues. A more fundamental issue may be rooted in the different capabilities to obtain accessibility among different social groups. This may be a broader social issue (e.g. gender difference). Second, different policy goals can lead to different strategies in improving travel equity. If we are concerned with evening out the accessibility levels, we should target (for DPPA): women, younger folks, non-university folks, and those living further from CBD. If, however, we are concerned with increasing the equality of effectiveness (equality in people getting what they pay for), we should focus more on spatial characteristics – improving informal neighborhoods and, again, conditions for those further from CBD.

In the end, the results in this chapter should be viewed as indicative, and more a demonstration of a potential lens of analysis rather than authoritative. The analysis presented here has several assumptions and limitations. Readers should be mindful of them when interpreting the results. First, the sample size is small and missing data can bias the results in unknown ways. This pilot included only 581 FMS candidates of which

484 ultimately participated. Dar has more than 4 million people and the travel equity situation may be more complex than what the sample reveals. In addition, the problem of data verification (a low rate) means that the results may underrepresent people who generate fewer verified data; the lack of verification may well be related to dimensions of relevance for equity. This has greater impact on the MPPA results because only the MPPA uses all verified stops. Under-recorded travel can also result in underestimated transport costs in the travel welfare effectiveness calculation. Second, the geometry-approximating approach to operationalize MPPA understates the complexity of the real travel area shape. The simplified shape can overestimate the PPA for individuals who have to travel further to reach their destinations but do not have many chances to reach opportunities around the destinations. For future studies, kernel density and shortest path network calculations can be used as a more realistic representation. Third, the operationalization of DPPA simplifies the complex reality of the transportation network, time flexibility in fixed activities, and travel speeds. For example, due to the lack of daladala network data, the DPPA calculation uses the road network, with estimated daladala speed, to represent public transit use. This can overestimate the DPPA for daladala users. Also, DPPA operationalization does not consider time flexibility in fixed activities which can underestimate DPPA. The travel speed uses the observed average speed for the individual's primary mode and does not include different levels of congestion by different locations and time of day. Finally, for the principle of equality of effectiveness, a series of assumptions underlie the calculation of cost: the approach to standardize the housing cost, the amortization term and discount rate for housing and car costs, the value of travel time, and the definition of sufficient cost affordability. Future studies can test the sensitivity of the results to different cost assumptions.

Chapter 8 Conclusion and Discussion

Dar es Salaam is one of numerous large cities in the developing world facing tremendous development challenges: growing population, rapid urbanization and motorization, sprawled informal settlements, problematic informal transit operations, and poor and unequal infrastructure service. One major problem Dar faces in overcoming its mobility challenges is lack of quality data, which hampers planning and analysis. My thesis aims to help fill this gap in Dar and similar African cities. Proliferating mobile phone usage in Africa increases the potential for leveraging these devices as a new travel data collection method. My thesis capitalizes upon a unique opportunity to demonstrate the potential value that mobile telephony presents for transportation planning purposes in Dar and other cities of the Global South. More specifically, based on a World Bank project deploying a pilot smartphone-based travel survey (FMS) in Dar es Salaam, I identify travel patterns depicted by the data (chapter 5), analyze the representativeness of FMS participants (chapter 6), and demonstrate how the data can be used to help evaluate travel equity in the city (chapter 7). Hopefully this work can help guide future mobile technology-based data collection methods, inspire more cost-efficient and accurate survey tools, shed new light on methods to assess travel equity, and, overall, facilitate the development of evidence-based smart and sustainable planning decisions for Dar and other rapidly developing regions.

The pilot FMS implementation reveals several challenges in adopting this data collection technology in the African context. First is the technical challenge. The sparse cell tower coverage, poor cellphone signal, and low density of Wi-Fi access points make it difficult to capture high-quality raw sensor data. The gap in sensor data creates difficulties in accurately detecting trips and stops. In addition, the lack of contextual land use data, points of interest (POIs), and public transport network data, adds more difficulty to inferring activity locations, purposes, and travel modes. Second is the social challenge. There are three types of social challenges: deliberate non-cooperation, limited ability to effectively contribute, and forced drop out. Examples of deliberate non-cooperation include selling the phone after deployment, constantly removing the SIM card, selectively refusing to answer the phone interview, and unwillingness to consult with the troubleshooting team for technical issues. Examples of limited ability to effectively contribute include turning on airplane mode due to unfamiliarity with a smartphone, forgetting to carry the phone, carrying the wrong phone (e.g., of another household member), poorly recalling activities (e.g., for people employed as vendors or drivers), and insufficient battery charge (due, e.g., to low access to electricity). The reason for forced drop out is the phone being stolen. Most of these social challenges are likely rarer in the FMS implementations in developed regions, such as Singapore and Boston.

Despite these challenges, analysis of this experimental smartphone-based travel survey demonstrates promising opportunities. Most basically, the high-resolution data collected suggest a better representation of individual travel in the city, compared to traditional paper-based surveys (chapter 5). Among the observations that can be drawn from the analysis, FMS apparently records higher overall trip rates, more work-related and private leisure trip-making, and more short-distance travel. The analysis also shows clear gender differences exist among individual travel characteristics in Dar, with males generating more trips (especially work and work-related trips) and traveling longer. This is associated with women's relatively low participation in work and their traditional lifestyles. Overall, the data indicate that Dar residents tend to undertake short-distance travel around their neighborhoods. Two factors potentially influence this short travel range: poor transport infrastructure and severe congestion and the popularity of informal businesses around neighborhoods.

Analyzing the likelihood of participating in the FMS pilot (chapter 6) shows little observable bias. This answers my first research question about representativeness of FMS participants. The results do not support my hypothesis on lower participation among men, full-time workers, and wealthy individuals. Having fewer years of education and working in government are associated with a lower likelihood of an individual participating, but no biases are found in other socio-economic and demographic attributes. That said, the results suggest that, conditional on participating, the degree of participation (i.e. the number of fully verified days) is associated with an individual's marriage status, having a driving license, age, and the monthly sales of family business. Those factors differ from those specified in my original hypothesis. Beyond participants' willingness to cooperate, the particular experimental design that inhibited enumerators from adding stops during the call-based diary verification was another obstacle to collecting fully verified days. Nevertheless, the low observable bias in participation demonstrates the promise of smartphone-based survey to provide representative data. The differences in the number of fully verified days among social groups require future endeavors to overcome. Possible solutions include enhancing mobile phone training for participants and improving enumerators' verification procedure.

Third, the richness of smartphone-based data makes it possible to operationalize new ways to systematically evaluate travel equity in the African city. Research about travel equity in Dar and other African cities is somewhat scarce, yet urgent city development challenges require immediate evidence-based guidance in decision-making of planning and infrastructure investment. Various ways exist to evaluate travel equity. In Chapter 7, I demonstrate a systematic, accessibility-based, approach in Dar and hope future studies can enrich this topic. I establish a framework for assessing travel equity based on the FMS-generated revealed preferences and operationalize it in Dar. The results answer my second research question about transportation equity. The ellipse-measured monthly activity space and network-based daily activity space both suggest high travel inequity among FMS participants, evaluated by both Gini coefficient and the severity index of travel welfare deficiency. However, although accessibility is unequal (equality of welfare), the benefit is not highly distorted based on what people pay (equality of effectiveness). For daily activity space, the differences in travel welfare are due to gender, age, education, and location. However, once one accounts for how much people pay to obtain the accessibility, the significant influence from all individual characteristics disappear; the variation of travel welfare also decreases across the city. This finding is consistent with my hypothesis. It implies that it is the differences in willingness and ability to pay that lead to high inequity of travel welfare. In addition, location matters for travel welfare. For daily activity space, people who live in informal settlements have significantly lower travel welfare effectiveness; people who are distant from the city center have both lower absolute travel welfare and the lower returns of welfare per cost unit invested. The analysis also reveals empirical differences between monthly potential path area (MPPA) and daily potential path area. Although they result in similar Gini coefficients and the severity index of travel welfare deficiency, their spatial distributions and the regression results on socio-economic and built-environment variables differ. MPPA diminishes the randomness measured by DPPA, but potentially introduces at least two more issues in this thesis: the bias in the number of fully verified days can introduce data representativeness concerns; it is also hard to distinguish individuals who want to travel far to explore from those who have to travel far.

By examining these challenges and opportunities, my thesis suggests the promise of smartphone-based data collection methods in other cities of the global south. Future efforts should learn from the pilot FMS implementation in Dar. First, the implementation requires a multi-disciplinary team. The technology and local context are both complex and require collaborations across multiple institutions. In this FMS implementation, the World Bank (WB) acted as the project manager; their local office provided local transport knowledge; the WB and Oxford Policy Management provided policy management knowledge; a local contractor operated the survey and a local mobile company, TIGO, provided the mobile network; the FMS platform was originally developed at the Singapore-MIT Alliance of Research and Technology (SMART). Second, future researchers can design mechanisms to reduce ineffective participation. Examples can be calling more frequently for users with complex travel diaries, requiring deposits to avoid users selling the phone, training the users on basic usage of the phone, and providing monetary incentives to users to finish verification. Third, techniques of travel activity inference should account for possible poor phone network and the lack of contextual data (e.g. land use, building type), and battery constraints arising from limited access to electricity. Fourth, if enumerators are used in the operation, they need to be better trained and be allowed to add stops that are not captured by the app. In this study, it was difficult to differentiate lack of verification caused by participants' unwillingness to cooperate versus that caused by enumerators' not adding stops. Future studies should minimize this "enumerator effect." Fifth, the application of smartphone-based surveys can benefit from integrating with other data sources, including comprehensive household surveys and passive big data (e.g. call detail records, text messages, social media information, etc). Smartphone-based surveys provide high-resolution behavior data. If combined with massive call data, they may provide ways to effectively scale up from the sample and eventually generate smarter solutions for city mobility issues. The integration with household surveys can extend data applications to broader city problems. Examples include poverty and mobility, housing and transportation, health care and transportation, education and transportation, etc.

Some pending issues still exist for future smartphone-based implementation. The first is the privacy concern. It is not unique to African context and potentially introduces participation bias. For example, the lower participation likelihood of people who work for government potentially reflects their concern of being tracked. Second, the technological complexity poses difficulties for local agencies in developing regions to adopt and manage the technology alone. The back-end computation of FMS is heavy and complex; the local government is not likely to run this survey regularly in the future without external help. Nevertheless, this thesis demonstrates promising opportunities in the smartphone-based data collection technology. If future scaled up implementation can properly address those challenges and pending issues identified in this thesis, it will certainly contribute to more evidence-based planning decisions for Dar and other rapidly developing regions.

Reference

Abidemi, Asiyanbola R. 2002. "Gender Differences in Intra-Urban Travel Behaviour: A Preliminary Survey in Ibadan, Nigeria." *AA Balkema*. https://dev.codatu.org/wp-content/uploads/Gender-differences-in-Intra-Urban-travel-behaviour-a-preliminary-survey-in-Ibadan-Nigeria-A.-ABIDEMI.pdf.

Auld, Joshua, Martina Z. Frignani, Chad Williams, and Abolfazl Kouros Mohammadian. 2010. "Results of the Utracs Internet-Based Prompted Recall Gps Activity-Travel Survey for the Chicago Region." *Lisbon, Portugal, July* 11: 15.

Bank of Africa. 2017. "Mortgage Loan | BOA Tanzania." Accessed April 6. https://www.boatanzania.com/personal-banking/mortgage-loan.

Barry, Brian. 1965. *Political Argument: A Reissue with New Introduction*. 20. Univ of California Press. https://www.google.com/books?hl=en&lr=&id=vr3AiIgj_xEC&oi=fnd&pg=PR2&dq=Barry,+B.+(1965).+Political+argument.+London:+Routledge+%26+Kegan+Paul.&ots=NDXhHlwMkL&sig=j7QqAaKI0Kyrz88q-t39l-6HtHM.

Behrens, Roger, and Roger Behrens. 2004. "Understanding Travel Needs of the Poor: Towards Improved Travel Analysis Practices in South Africa." *Transport Reviews* 24 (3): 317–336.

Behrens, Roger, Dorothy McCormick, and David Mfinanga. 2016. *Paratransit in African Cities : Operations, Regulation and Reform*. London ; New York : Routledge, Taylor & Francis Group, 2016.

Ben-Akiva, Moshe, and Steven R. Lerman. 1979. "Disaggregate Travel and Mobility Choice Models and Measures of Accessibility." *Behavioural Travel Modelling*, 654–679.

Bohte, Wendy, and Kees Maat. 2009. "Deriving and Validating Trip Purposes and Travel Modes for Multi-Day GPS-Based Travel Surveys: A Large-Scale Application in the Netherlands." *Transportation Research Part C: Emerging Technologies* 17 (3): 285–297.

Boscoe, Francis P., Kevin A. Henry, and Michael S. Zdeb. 2012. "A Nationwide Comparison of Driving Distance Versus Straight-Line Distance to Hospitals." *The Professional Geographer : The Journal of the Association of American Geographers* 64 (2). doi:10.1080/00330124.2011.583586.

Bricka, Stacey G., Sudeshna Sen, Rajesh Paleti, and Chandra R. Bhat. 2012. "An Analysis of the Factors Influencing Differences in Survey-Reported and GPS-Recorded Trips." *Transportation Research Part C: Emerging Technologies* 21 (1): 67–88.

Bryceson, Deborah Fahy, and John Howe. 1993. "Rural Household Transport in Africa: Reducing the Burden on Women?" *World Development* 21 (11): 1715–1728.

Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge university press. https://books.google.com/books?hl=en&lr=&id=TdlKAgAAQBAJ&oi=fnd&pg=PP1&dq=Microeconometrics:+Methods+and+Applications&ots=yIkhM2bHto&sig=r6ja0BGz-N8zRNorwJOQPVNyL1g.

Casas, Irene, Mark W. Horner, and Joe Weber. 2009. "A Comparison of Three Methods for Identifying Transport-Based Exclusion: A Case Study of Children's Access to

Urban Opportunities in Erie and Niagara Counties, New York." *International Journal of Sustainable Transportation* 3 (4): 227–45. doi:10.1080/15568310802158761.
Central Intelligence Agency. 2017. "The World Factbook." Accessed April 20. https://www.cia.gov/library/publications/the-world-factbook/rankorder/2172rank.html.
Chen, Cynthia, Hongmian Gong, Catherine Lawson, and Evan Bialostozky. 2010. "Evaluating the Feasibility of a Passive Travel Survey Collection in a Complex Urban Environment: Lessons Learned from the New York City Case Study." *Transportation Research Part A: Policy and Practice* 44 (10): 830–840.
Chen, Don. 1996. "Getting A Fair Share: An Analysis of Federal Transportation Spending." Surface Transportation Policy Project.
Chew, Victor. 1966. "Confidence, Prediction, and Tolerance Regions for the Multivariate Normal Distribution." *Journal of the American Statistical Association* 61 (315): 605–17. doi:10.2307/2282774.
Church, Andrew, Martin Frost, and Keith Sullivan. 2000. "Transport and Social Exclusion in London." *Transport Policy* 7 (3): 195–205.
Cleveland, William S., and Clive Loader. 1996. "Smoothing by Local Regression: Principles and Methods." In *Statistical Theory and Computational Aspects of Smoothing*, 10–49. Springer. http://link.springer.com/chapter/10.1007/978-3-642-48425-4_2.
Cohen, G. A. 1993. *Equality of What? On Welfare, Goods, and Capabilities. I Martha Nussbaum Og Amartya Sen (Red.) The Quality of Life*. Oxford: Clarendon Press.
Consortium of Dodi Moss Srl – Buro Happold Ltd – Afri-Arch Associates – Q Consult Lt. 2015. "Dar Es Salaam City Master Plan: 2012 - 2032 (Draft)."
Corporate Digest. 2017. "Tanzania Has A Mobile Phone Penetration of 67%." Accessed April 19. http://www.corporate-digest.com/index.php/tanzania-has-a-mobile-phone-penetration-of-67-report.
Currie, Graham. 2010. "Quantifying Spatial Gaps in Public Transport Supply Based on Social Needs." *Journal of Transport Geography* 18 (1): 31–41. doi:10.1016/j.jtrangeo.2008.12.002.
Delbosc, Alexa, and Graham Currie. 2011a. "Using Lorenz Curves to Assess Public Transport Equity." *Journal of Transport Geography* 19 (6): 1252–1259.
———. 2011b. "The Spatial Context of Transport Disadvantage, Social Exclusion and Well-Being." *Journal of Transport Geography*, Special section on Alternative Travel futures, 19 (6): 1130–37. doi:10.1016/j.jtrangeo.2011.04.005.
Di, Pan. 2013. "Key Transport Statistics of World Cities." Journeys. https://www.lta.gov.sg/ltaacademy/doc/13Sep105-Pan_KeyTransportStatistics.pdf.
Dinardo, John, Jack Johnston, and John Johnston. 1997. "Econometric Methods." *Forth Edition) McGraw-Hill Companies, Inc*, 204–326.
Dong, Xiaojing, Moshe E. Ben-Akiva, John L. Bowman, and Joan L. Walker. 2006. "Moving from Trip-Based to Activity-Based Measures of Accessibility." *Transportation Research Part A: Policy and Practice* 40 (2): 163–180.

Ettema, Dick, Harry Timmermans, and Leo van Veghel. 1996. "Effects of Data Collection Methods in Travel and Activity Research." https://trid.trb.org/view.aspx?id=573629.
Fan, Yingling, Qian Chen, Chen-Fu Liao, and Frank Douma. 2012. "Smartphone-Based Travel Experience Sampling and Behavior Intervention among Young Adults." http://conservancy.umn.edu/handle/11299/132726.
Foster, James, Joel Greer, and Erik Thorbecke. 1984. "A Class of Decomposable Poverty Measures." *Econometrica: Journal of the Econometric Society*, 761–766.
Fournier, David A., Hans J. Skaug, Johnoel Ancheta, James Ianelli, Arni Magnusson, Mark N. Maunder, Anders Nielsen, and John Sibert. 2012. "AD Model Builder: Using Automatic Differentiation for Statistical Inference of Highly Parameterized Complex Nonlinear Models." *Optimization Methods and Software* 27 (2): 233–249.
Fransen, Koos, and Steven Farber. 2017. "Does Accessibility Lead To Social Inclusion? A Person-Based Accessibility Measure For Modeling Activity Participation." Transportation Research Board (TRB) poster.
Galvez, T. E., and S. R. Jara-Diaz. 1998. "On the Social Valuation of Travel Time Savings." *International Journal of Transport Economics/Rivista Internazionale Di Economia Dei Trasporti*, 205–219.
Geurs, Karst T., and Bert Van Wee. 2004. "Accessibility Evaluation of Land-Use and Transport Strategies: Review and Research Directions." *Journal of Transport Geography* 12 (2): 127–140.
Ghorpade, Ajinkya, Francisco Câmara Pereira, Fang Zhao, Christopher Zegras, and Moshe Ben-Akiva. 2015. "An Integrated Stop-Mode Detection Algorithm for Real World Smartphone-Based Travel Survey." In *Transportation Research Board 94th Annual Meeting*. https://www.researchgate.net/profile/Ajinkya_Ghorpade2/publication/275949958_An_integrated_stop-mode_detection_algorithm_for_real_world_smartphone-based_travel_survey/links/554a281a0cf29f836c95d9c6.pdf.
Gini, C. W. 1912. "Variability and Mutability, Contribution to the Study of Statistical Distribution and Relaitons." *Studi Economico-Giuricici Della R.*
Global Petrol Prices. 2017. "Gasoline Prices in Dar Es Salaam, Tanzania - 03-Apr-2017 | GlobalPetrolPrices.Com." Accessed April 6. http://www.globalpetrolprices.com/Tanzania/Dar_es_Salaam/gasoline_prices/.
Golledge, Reginald G. 1997. *Spatial Behavior: A Geographic Perspective*. Guilford Press. https://books.google.com/books?hl=en&lr=&id=2JPMvpMbLrMC&oi=fnd&pg=PR7&dq=Golledge,+R.G.,+Stimson,+RJ.,+1997.+Spatial+Behavior,+The+Guilford+Press,+New+York.&ots=55KNVuIdi1&sig=-hyuOrK0t1rslcbHNAo800L9mjM.
Gollina, Doug, Martina Kirchberger, and David Lagakos. 2015. "Measuring Living Standards Across Space in the Developing World (Draft)." The World Bank. https://collaboration.worldbank.org/docs/DOC-20611.
Greaves, Stephen, Simon Fifer, Richard Ellison, and George Germanos. 2010. "Development of a Global Positioning System Web-Based Prompted Recall

Solution for Longitudinal Travel Surveys." *Transportation Research Record: Journal of the Transportation Research Board*, no. 2183: 69–77.

Greene, William H. 2003. *Econometric Analysis*. Pearson Education India.

GSMA Intelligent. 2016. "The Mobile Economy Africa 2016." https://www.gsmaintelligence.com/research/?file=3bc21ea879a5b217b64d62fa24c55bdf&download.

Guevara, C. Angelo. 2015. "Critical Assessment of Five Methods to Correct for Endogeneity in Discrete-Choice Models." *Transportation Research Part A: Policy and Practice* 82: 240–254.

Hägerstraand, Torsten. 1970. "What about People in Regional Science?" *Papers in Regional Science* 24 (1): 7–24.

Hall, Daniel B. 2000. "Zero-Inflated Poisson and Binomial Regression with Random Effects: A Case Study." *Biometrics* 56 (4): 1030–1039.

Hanson, Susan, and Perry Hanson. 1981. "The Travel-Activity Patterns of Urban Residents: Dimensions and Relationships to Sociodemographic Characteristics." *Economic Geography* 57 (4): 332–47.

Hensher, David A., and William H. Greene. 2003. "The Mixed Logit Model: The State of Practice." *Transportation* 30 (2): 133–176.

Hill, Alexandra, and Christian Lindner. 2010. "Land-Use Modelling to Support Strategic Urban Planning-the Case of Dar Es Salaam, Tanzania." In *45th ISOCARP Congress*, 1–12. http://isocarp.net/Data/case_studies/1820.pdf.

InfoBridge Consultants Limited. 2009. "Dar Es Salaam City Physical Mapping Project."

Jacoby, William G. 2000. "Loess:: A Nonparametric, Graphical Tool for Depicting Relationships between Variables." *Electoral Studies* 19 (4): 577–613.

Japan International Cooperation Agency (JICA). 2008. "Dar Es Salaam Transport Policy and System Development Master Plan." http://open_jicareport.jica.go.jp/710/710_416.html.

Jiang, Shan, Joseph Ferreira Jr, and Marta C. González. 2015. "Activity-Based Human Mobility Patterns Inferred from Mobile Phone Data: A Case Study of Singapore." In *The 4th International Workshop on Urban Computing*. http://scholar.google.com/scholar?cluster=14616055753181799139&hl=en&oi=scholarr.

Jiao, Junfeng, and Maxwell Dillivan. 2013. "Transit Deserts: The Gap between Demand and Supply." *Journal of Public Transportation* 16 (3): 2.

Katala, J. N. 2012. "The Cost of Traffic Congestion and Accidents to the Economy in Tanzania." *Transportation Planning Manager-Dart Agency Accessed at Www. Iekenya. Org/Download/JumbeN. KatalaThecostoftraff Iccongestion..... DART. Pdf.*

Kiunsi, Robert. 2013a. "A Review of Traffic Congestion in Dar Es Salaam City from the Physical Planning Perspective." *Journal of Sustainable Development* 6 (2): 94.

———. 2013b. "The Constraints on Climate Change Adaptation in a City with a Large Development Deficit: The Case of Dar Es Salaam." *Environment and Urbanization*. http://eau.sagepub.com/content/early/2013/07/08/0956247813489617.abstract.

Kokoreff, Michel. 1995. *Aller à La Défense. Polarisation et Pratiques Urbaines Des Jeunes*. Paris, RATP, Département du Développement Prospective.

Kwan, Mei-Po. 1998. "Space-Time and Integral Measures of Individual Accessibility: A Comparative Analysis Using a Point-Based Framework." *Geographical Analysis* 30 (3): 191–216. doi:10.1111/j.1538-4632.1998.tb00396.x.

———. 1999. "Gender and Individual Access to Urban Opportunities: A Study Using Space–time Measures." *The Professional Geographer* 51 (2): 210–227.

Kwan, Mei-Po, and Xiao-Dong Hong. 1998. "Network-Based Constraints-Oriented Choice Set Formation Using GIS." *Geographical Systems* 5: 139–162.

Kymlicka, Will. 2002. *Contemporary Political Philosophy: An Introduction*. Oxford: Oxford University Press. https://www.google.com/books?hl=en&lr=&id=yYCcAQAAQBAJ&oi=fnd&pg=PR7&dq=Contemporary+political+philosophy+:+an+introduction&ots=BFL_d6ZpDg&sig=kukB4u-OHZtpgW0Snd2N_nmFRvY.

Lambert, Diane. 1992. "Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing." *Technometrics* 34 (1): 1–14.

Landau, Uzi, Joseph N. Prashker, and Bernard Alpern. 1982. "Evaluation of Activity Constrained Choice Sets to Shopping Destination Choice Modelling." *Transportation Research Part A: General* 16 (3): 199–207.

Lenntorp, B., and B. Lenntorp. 1976. "Paths in Time-Space Environments: A Time Geographic Study of Movement Possibilities OfIndividuals. Gleerup." *Lund Studies in Geography Series B: Human Geography, Lund, Sweden*.

Li, Zhigang Jason, and Amer S. Shalaby. 2008. "Web-Based GIS System for Prompted Recall of GPS-Assisted Personal Travel Surveys: System Development and Experimental Study." In *Transportation Research Board 87th Annual Meeting*. https://trid.trb.org/view.aspx?id=848806.

Litman, T. 2006. "Evaluating Transport Equity: Guidance for Incorporating Distributional Impacts in Transportation Planning." *Victoria BC: Victoria Transport Policy Institute*.

Liu, Suxia, and Xuan Zhu. 2004. "Accessibility Analyst: An Integrated GIS Tool for Accessibility Analysis in Urban Transportation Planning." *Environment and Planning B: Planning and Design* 31 (1): 105–24. doi:10.1068/b305.

Logit and Inter-consult. 2006. "Consultancy Services for the Conceptual Design of a Long Term Integrated Dar Es Salaam BRT System and Detailed Design for the Initial Corridor." Draft final report. Logit and Inter-consult Ltd. for Dar es Salaam City Council and United Republic of Tanzania, Dar es Salaam.

Lorenz, Max O. 1905. "Methods of Measuring the Concentration of Wealth." *Publications of the American Statistical Association* 9 (70): 209–219.

Lucas, Karen. 2011. "Making the Connections between Transport Disadvantage and the Social Exclusion of Low Income Populations in the Tshwane Region of South Africa." *Journal of Transport Geography* 19 (6): 1320–1334.

———. 2012. "Transport and Social Exclusion: Where Are We Now?" *Transport Policy*, URBAN TRANSPORT INITIATIVES, 20 (March): 105–13. doi:10.1016/j.tranpol.2012.01.013.

Luo, Wei, and Fahui Wang. 2003. "Measures of Spatial Accessibility to Health Care in a GIS Environment: Synthesis and a Case Study in the Chicago Region." *Environment and Planning B: Planning and Design* 30 (6): 865–84. doi:10.1068/b29120.

Maganga, Francis Chamu Michael. 2014. "Travel Behaviour In Dar Es Salaam City (M.Sc. Thesis)." University of Dar es Salaam.

Manaugh, Kevin, Ahmed El-Geneidy, and others. 2012. "Who Benefits from New Transportation Infrastructure? Using Accessibility Measures to Evaluate Social Equity in Public Transport Provision." *Chapters*, 211–227.

Martens, Karel. 2016. *Transport Justice: Designing Fair Transportation Systems*. Routledge. https://www.google.com/books?hl=en&lr=&id=m0yTDAAAQBAJ&oi=fnd&pg=PP1&dq=transport+justice+karen+martens&ots=RErKF9n3Oy&sig=9kY0PY3Klx6Bt4DqTjxFEioN95c.

Martens, Karel, and Aaron Golub. 2011. "Accessibility Measures from an Equity Perspective." In *Transport Planning Research Colloquium*. http://www.cvs-congres.nl/cvspdfdocs/cvs11_065.pdf.

Masaoe, Estomihi N., Romano F. Del Mistro, and George Makajuma. 2011. "Travel Behaviour in Cape Town, Dar Es Salaam and Nairobi Cities." *SATC 2011*. http://www.repository.up.ac.za/handle/2263/17373.

Mfinanga, David A., Herrie Schalekamp, and Dorothy McCormick. 2015. "The Nature of Paratransit Operations." In . Routledge. http://repository.udsm.ac.tz:8080/xmlui/handle/20.500.11810/3682.

Miller, David. 1999. *Principles of Social Justice*. Harvard University Press. https://www.google.com/books?hl=en&lr=&id=y2wMzJtEZ_8C&oi=fnd&pg=PA1&dq=Miller,+D.+(1999).+Principles+of+social+justice.&ots=5IOUPlwrW8&sig=qiOOpCjiRZSm-9sQ7eBCLpVkBwc.

Miller, Harvey J. 1991. "Modelling Accessibility Using Space-Time Prism Concepts within Geographical Information Systems." *International Journal of Geographical Information System* 5 (3): 287–301.

———. 1999. "Measuring Space-Time Accessibility Benefits within Transportation Networks: Basic Theory and Computational Procedures." *Geographical Analysis* 31 (1): 1–26. doi:10.1111/j.1538-4632.1999.tb00408.x.

———. 2006. "Social Exclusion in Space and Time." In *AXHAUSEN KW Moving through Nets: The Physical and Social Dimensions of Travel. Selected Papers from the 10th International Conference of Travel Behaviour Research H. Elsevier Science Publishing Company*, 353–380. Citeseer. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.80.6098&rep=rep1&type=pdf.

Mkalawa, Charles Cosmas, and Pan Haixiao. 2014. "Dar Es Salaam City Temporal Growth and Its Influence on Transportation." *Urban, Planning and Transport Research* 2 (1): 423–446.

Moran, Patrick AP. 1950. "Notes on Continuous Stochastic Phenomena." *Biometrika* 37 (1/2): 17–23.

Mramba, Nasibu, Erkki Sutinen, Michael Haule, and Peter Msami. 2014. "Survey of Mobile Phone Usage Patterns among Street Vendors in Dar Es Salaam City-Tanzania." *International Journal of Information Technology and Business Management* 8 (1): 1–10.

Murray, Alan T., and Rex Davis. 2001. "Equity in Regional Service Provision." *Journal of Regional Science* 41 (4): 557–600.

Muthén, Linda K., and Bengt O. Muthén. 2007. "Mplus." *Statistical Analysis with Latent Variables. Version* 3. http://www.statmodel.com/virg_nov_course.shtml.

National Bureau of Statistics - Ministry of Finance, and Office of Chief Government Statistician President's Office - Finance, Economy and Development Planning. 2012. "Tanzania - Population and Housing Census 2012." http://catalog.ihsn.org/index.php/catalog/4618.

National Cooperative Highway Research Program, National Research Council (U.S.), American Association of State Highway and Transportation Officials, United States, and Cambridge Systematics, eds. 2012. *Travel Demand Forecasting: Parameters and Techniques*. NCHRP Report 716. Washington, D.C: Transportation Research Board.

Nawaranthne, K., F. Zhao, F. C. Pereira, C. C. Zegras, and M. Ben-Akiva. 2014. "The Impact of GPS Based Outdoor Activity Detection on Smartphone Battery Life." In *10th International Conference on Transport Survey Methods, Australia*.

Neuburger, H. 1971. "User Benefit in the Evaluation of Transport and Land Use Plans." *Journal of Transport Economics and Policy*, 52–75.

Ng, M. 2005. "Delaware Valley Regional Planning Commission: An MPO's Methodology toward Equitable Accessibility And Involvement In Regional Planning." In *Racial Equity In Transportation Workshop, by the Harvard Civil Rights Project and the Brookings Institute*.

Nitsche, Philippe, Peter Widhalm, Simon Breuss, Norbert Brändle, and Peter Maurer. 2014. "Supporting Large-Scale Travel Surveys with Smartphones–A Practical Approach." *Transportation Research Part C: Emerging Technologies* 43: 212–221.

Nkurunziza, A. 2013. "Sustainable Transport in Dar-Es-Salaam: The Potential for BRT and Cycling from a User Perspective." http://doc.utwente.nl/90324/.

Nussbaum, Martha C. 2011. *Creating Capabilities*. Harvard University Press. https://www.google.com/books?hl=en&lr=&id=Gg7Q2V8fi8gC&oi=fnd&pg=PR7&dq=Creating+capabilities:+The+human+development+approach&ots=l10we2TPgl&sig=PuV_HcaxUesBYK-i9_6VSryKNPw.

Olvera, Lourdes Diaz, Didier Plat, and Pascal Pochet. 2003. "Transportation Conditions and Access to Services in a Context of Urban Sprawl and Deregulation. The Case of Dar Es Salaam." *Transport Policy*, Transport and Social Exclusion, 10 (4): 287–98. doi:10.1016/S0967-070X(03)00056-8.

Páez, Antonio, Darren M. Scott, and Catherine Morency. 2012. "Measuring Accessibility: Positive and Normative Implementations of Various Accessibility Indicators." *Journal of Transport Geography* 25: 141–153.

PennState Eberly College of Science. 2017. "Applied Multivariate Statistical Analysis (Chapter 4.6 - Geometry of the Multivariate Normal Distribution)." Accessed April 5. https://onlinecourses.science.psu.edu/stat505/node/36.

Pereira, Rafael HM, Tim Schwanen, and David Banister. 2016. "Distributive Justice and Equity in Transportation." *Transport Reviews*, 1–22.

Porter, Gina. 2002. "Living in a Walking World: Rural Mobility and Social Equity Issues in Sub-Saharan Africa." *World Development* 30 (2): 285–300.

Poushter, Jacob. 2016. "Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies." *Pew Research Center*. http://www.diapoimansi.gr/PDF/pew_research%201.pdf.

Rawls, John. 2009. *A Theory of Justice*. Harvard university press. https://www.google.com/books?hl=en&lr=&id=kvpby7HtAe0C&oi=fnd&pg=PR11&dq=A+theory+of+justice&ots=tigmw4Aa78&sig=QW4iotT-h1KoOrVmITvOuy1c8WE.

Reddy, Alla, Alex Lu, Santosh Kumar, Victor Bashmakov, and Svetlana Rudenko. 2009. "Entry-Only Automated Fare-Collection System Data Used to Infer Ridership, Rider Destinations, Unlinked Trips, and Passenger Miles." *Transportation Research Record: Journal of the Transportation Research Board*, no. 2110: 128–136.

Research on Poverty Alleviation (REPOA). 2005. "Poverty and Human Development Report 2005." https://www.google.com/search?q=repoa&oq=repoa&aqs=chrome..69i57j69i60.1820j0j7&sourceid=chrome&ie=UTF-8#q=repoa+2005+Poverty+and+Human+Development+Report&*.

Rwebangira, Magdalena Kamugisha. 1996. *The Legal Status of Women and Poverty in Tanzania*. Vol. 100. Nordic Africa Institute. https://books.google.com/books?hl=en&lr=&id=m66BqxyBiIgC&oi=fnd&pg=PA5&dq=Rwebangira,+M.+K.,+1996.+The+Legal+Status+of+Women+and+Poverty+in+Tanzania.+Research+Report+n%C2%B0100,+Nordiska+Afrikaninstitutet,+Uppsala.&ots=X0WHmHY_d6&sig=zQ3inJ530OLNkAtIONC4OZWAaQQ.

Schneider, Christian M., Christian Rudloff, Dietmar Bauer, and Marta C. González. 2013. "Daily Travel Behavior: Lessons from a Week-Long Survey for the Extraction of Human Mobility Motifs Related Information." In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*, 3. ACM. http://dl.acm.org/citation.cfm?id=2505829.

Schönfelder, Stefan, and Kay W Axhausen. 2003. "Activity Spaces: Measures of Social Exclusion?" *Transport Policy*, Transport and Social Exclusion, 10 (4): 273–86. doi:10.1016/j.tranpol.2003.07.002.

Scott, Hannah. 2003. "Stranger Danger: Explaining Women's Fear of Crime." *Western Criminology Review* 4 (3): 203–214.

Sen, Amartya. 1979. *Equality of What? Tanner Lecture on Human Values*. *Tanner Lectures*. Stanford university.

Skaug, H., D. Fournier, A. Nielsen, A. Magnusson, and B. Bolker. 2012. "Generalized Linear Mixed Models Using AD Model Builder." *R Package Version 0.7* 2.

Sohail, M., D. A. C. Maunder, and D. W. J. Miles. 2004. "Managing Public Transport in Developing Countries: Stakeholder Perspectives in Dar Es Salaam and Faisalabad." *International Journal of Transport Management* 2 (3): 149–160.

Stevens, Forrest R., Andrea E. Gaughan, Catherine Linard, and Andrew J. Tatem. 2015. "Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data." *PLOS ONE* 10 (2): e0107042. doi:10.1371/journal.pone.0107042.

Stopher, Peter R., Christine Prasad, and Jun Zhang. 2010. *Comparing GPS and Prompted Recall Data Records*. Institute of Transport and Logistics Studies.

http://sydney.edu.au/business/__data/assets/pdf_file/0007/86713/itls-wp-10-24.pdf.

Tanzania National Bureau of Statistics, and Regional Commissioner's Office Dar es Salaam. 2014. "Dar Es Salaam Region Socio-Economic Profile." https://www.google.com/#q=DAR+ES+SALAAM+REGION+SOCIO-ECONOMIC+PROFILE%2C+2014.

TanzaniaInvest. 2016. "Tanzania Mobile Subscriptions Increase by 24% in 2015 and Internet Penetration Reach 34%," April 28. http://www.tanzaniainvest.com/telecoms/tanzania-mobile-subscriptions-reach-39-million-and-internet-penetration-reach-34-percent-in-2015.

TETEA. 2015. "Relative Smartphone Penetration in Tanzania | TETEA Inc." February 10. http://www.tetea.org/relative-smartphone-penetration-in-tanzania/.

The World Bank. 2011. "Fare Collection Systems - Dar Es Salaam, Tanzania." https://www.ssatp.org/sites/ssatp/files/publications/Toolkits/Fares%20Toolkit%20content/case-studies/dar-es-salaam,-tanzania.html.

Trading Economics. 2017a. "Tanzania Inflation Rate | 1999-2017 | Data | Chart | Calendar | Forecast." Accessed April 6. www.tradingeconomics.com/tanzania/inflation-cpi.

———. 2017b. "Tanzania Interest Rate | 1972-2017 | Data | Chart | Calendar | Forecast." Accessed April 6. www.tradingeconomics.com/tanzania/interest-rate.

Transport Research Laboratory. 2002. "Scoping Study." http://regulationbodyofknowledge.org/wp-content/uploads/2013/03/WorldBank_Scoping_Study_Urban.pdf.

Utsunomiya, Mariko, John Attanucci, and Nigel Wilson. 2006. "Potential Uses of Transit Smart Card Registration and Transaction Data to Improve Transit Planning." *Transportation Research Record: Journal of the Transportation Research Board*, no. 1971: 119–126.

Venter, Christoffel, Vera Vokolkova, and Jaroslav Michalek. 2007. "Gender, Residential Location, and Household Travel: Empirical Findings from Low-income Urban Settlements in Durban, South Africa." *Transport Reviews* 27 (6): 653–77. doi:10.1080/01441640701450627.

Venter, Yolandi, Romano Del Mistro, Eric Bruun, and David A. Mfinanga. 2015. "The State of Public Transport Systems in Three Sub-Saharan African Cities." In . Routledge. http://repository.udsm.ac.tz:8080/xmlui/handle/20.500.11810/3684.

Weber, Joe. 2003. "Individual Accessibility and Distance from Major Employment Centers: An Examination Using Space-Time Measures." *Journal of Geographical Systems* 5 (1): 51–70.

Wolf, Jean, Marcelo Oliveira, and Miriam Thompson. 2003. "Impact of Underreporting on Mileage and Travel Time Estimates: Results from Global Positioning System-Enhanced Household Travel Survey." *Transportation Research Record: Journal of the Transportation Research Board*, no. 1854: 189–198.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT press. https://books.google.com/books?hl=en&lr=&id=hSs3AgAAQBAJ&oi=fnd&pg=PP1&dq=Wooldridge,+J.+(2010),+Econometric+Analysis+of+Cross+Section+and+Panel+Data,&ots=VWTPnDW1Rp&sig=a2c7OoLYEgv0Ba31qwL9acVl6Rw.

Zegras, Chris. 2011. "Mainstreaming Sustainable Urban Transport: Putting the Pieces Together." In *Urban Transport in the Developing World: A Handbook of Policy and Practice*, (H. Dimitriou, R. Gakenheimer, eds), 548–88. Edward Elgar Publishing.

Zhao, Fang, Ajinkya Ghorpade, Francisco C Pereira, Chris Zegras, and Moshe Ben-Akiva. 2015. "Quantifying Mobility: Pervasive Technologies for Transport Modeling." In , 1039–44. ACM.

Zhao, Fang, Francisco Câmara Pereira, Rudi Ball, Youngsung Kim, Yafei Han, Christopher Zegras, and Moshe Ben-Akiva. 2015. "Exploratory Analysis of a Smartphone-Based Travel Survey in Singapore." *Transportation Research Record: Journal of the Transportation Research Board* 2 (2494): 45–56.

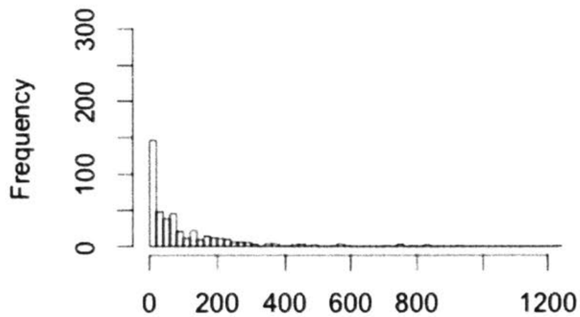
Appendix

1. Average Travel Speed Estimation Model by mode

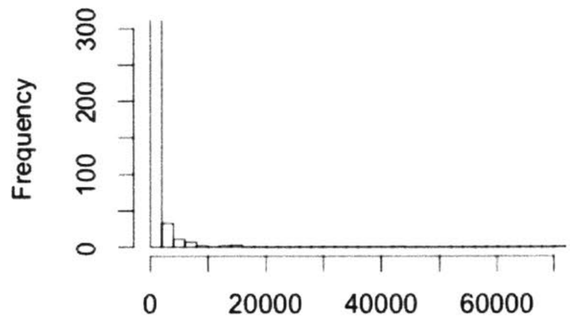
Note: Dependent variable: log-form of speed

	Bike			Bus (Daladala)			Foot			Taxi			Vehicle		
	coef.	Pr(>|t|)		coef.	Pr(>|t|)		coef.	Pr(>|t|)		coef.	Pr(>|t|)		coef.	Pr(>|t|)	
(Intercept)	-845.467	0.009	**	-71.340	0.127		-92.048	0.010	*	2.789	< 2e-16	***	2.245	0.000	***
lat of origin	6.247	0.114		0.632	0.275		0.624	0.215							
lon of origin	22.396	0.008	**	1.993	0.098	.	2.469	0.008	**						
trip duration (minutes)	-0.041	0.004	**	-0.006	< 2e-16	***	-0.015	0.000	***	-0.009	0.000	***	-0.003	0.139	
wealth index (per person)	-0.507	0.367		-0.027	0.766								0.376	0.125	
euclidean distance from the origin to CBD				0.000	0.003	**				0.000	0.394		0.000	0.294	
log(euclidean distance from the origin to CBD)	1.400	0.033	*				0.154	0.005	**						
time of the day: 12-16				0.003	0.977					0.123	0.604		-0.078	0.765	
time of the day: 16-20				-0.062	0.575					-0.051	0.824		-0.052	0.845	
time of the day: 20-23				-0.227	0.099	.				-0.354	0.217		0.295	0.411	
time of the day: 5-7				0.057	0.712								0.390	0.460	
time of the day: 7-10				-0.052	0.659					0.176	0.574		-0.266	0.346	
time of the day: night (23-5)				0.296	0.267					1.058	0.200		0.773	0.080	.
time of the day: 10-12 (base)															
R-square	0.881			0.374			0.115			0.177			0.222		
adj. R-square	0.807			0.344			0.106			0.127			0.114		
obs	14			243			372			123			75		

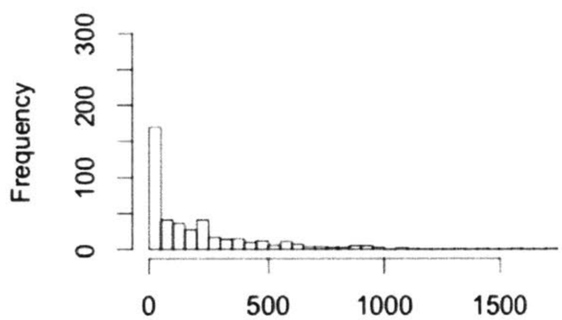
2. Histograms of non cost-adj. and cost-adj, weighted and un-weighted MPPAs and DPPAs



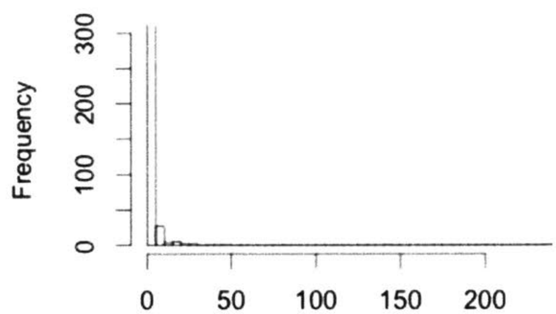
Monthly Potential Path Area (MPPA) Size



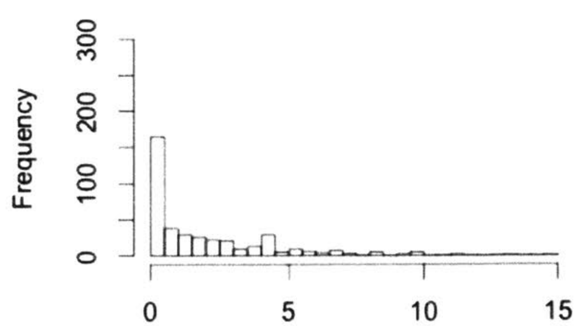
Cost-adjusted Monthly Potential Path Area (MPPA) Size



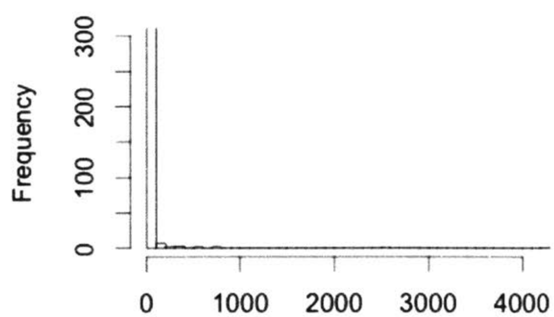
Weighted Population in MPPA



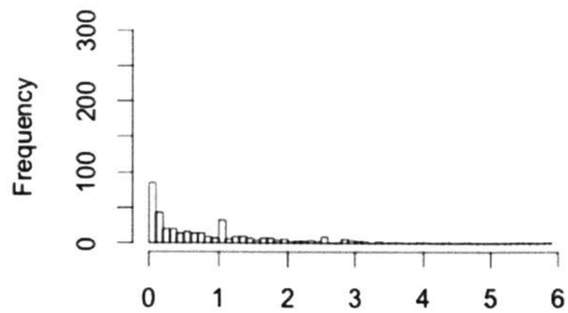
Cost-adjusted Weighted Population in MPPA



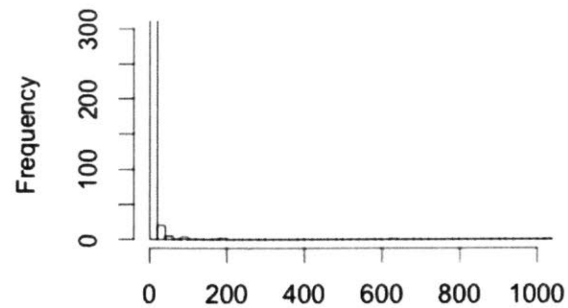
Commercial/Industrial Area in MPPA



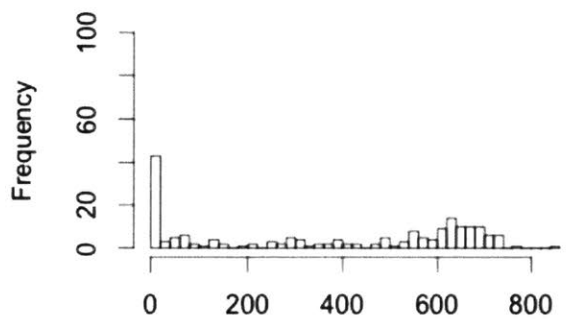
Cost-adjusted Commercial/Industrial Area in MPPA



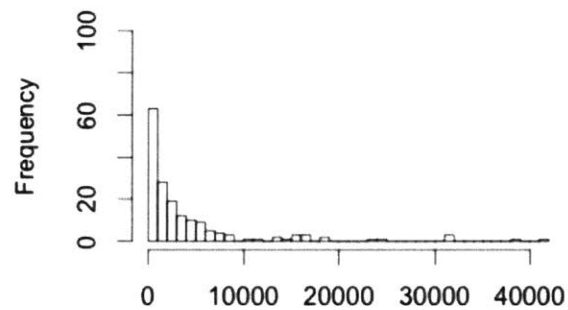
Paved Road Area in MPPA



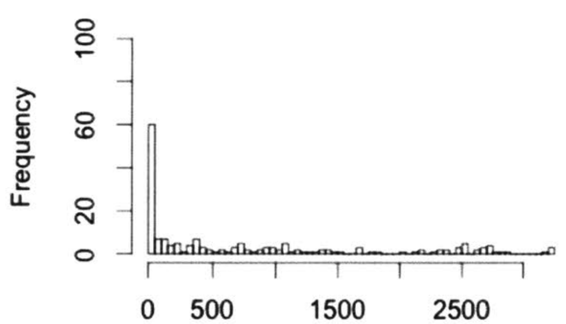
Cost-adjusted Paved Road Area in MPPA



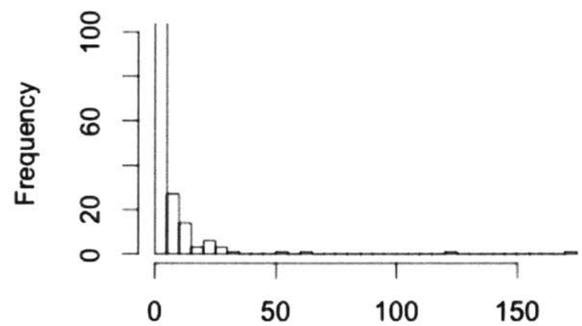
Daily Potential Path Area (DPPA) Size



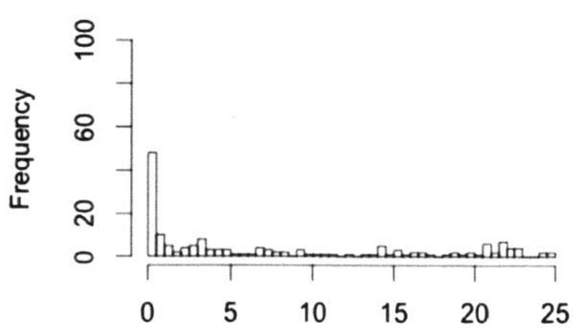
Cost-adjusted Daily Potential Path Area (DPPA) Size

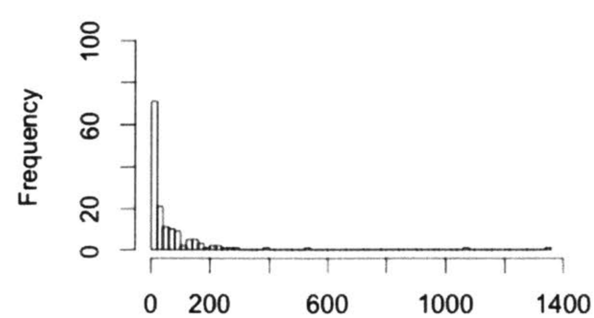


Weighted Population in DPPA



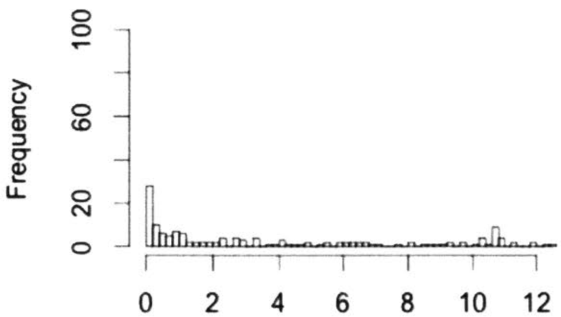
Cost-adjusted Weighted Population in DPPA



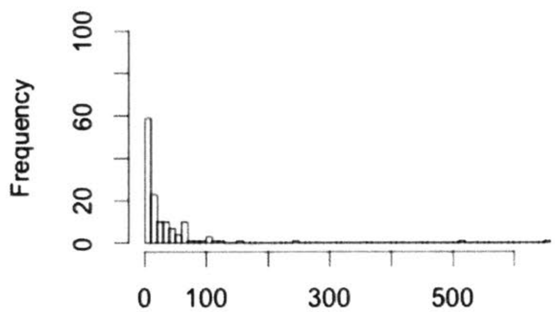


Commercial/Industrial Area in DPPA

Cost-adjusted Commercial/Industrial Area in DPPA



Paved Road Area in DPPA



Cost-adjusted Paved Road Area in DPPA