UTTRI

Research Report

Final Project Report

Report 6, iCity SOUTH

Ahmadreza Faghih-Imani, James Vaughan, Bilal Yusuf, Eric J. Miller

December 2018

iCITY-SOUTH:
Urban Informatics for Sustainable Metropolitan Growth in Latin America

REPORT 6:
FINAL PROJECT REPORT

A report to CAF, the Development Bank of Latin America.

BANCO DE DESARROLLO DE AMÉRICA LATINA

Más oportunidades, un mejor futuro.

By:

Ahmadreza Faghih-Imani, Ph.D.
NSERC Post-Doctoral Fellow

James Vaughan, h.BSc.

Bilal Yusuf, B.Eng., M.Eng.CEM

Eric J. Miller, Ph.D.
Professor, Department of Civil & Mineral Engineering
Director, UTTRI

December 22, 2018

UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING
Transportation Research Institute

UTTRI

EXECUTIVE SUMMARY

Urban regions within Latin America face enormous challenges in terms of the provision of transportation infrastructure and services to meet the travel needs of their growing population in a cost-effective, equitable and sustainable manner. High quality, comprehensive information concerning travel behaviour and transportation system performance is a fundamental prerequisite for successful urban transportation planning and decision-making to address these pressing, first-order needs. In recognition of this need, CAF established the Urban Mobility Observatory (OMU, *Observatorio de Movilidad Urbana*) to assemble and utilize standardized transportation-related data for Latin American cities. As one component of CAF's strategy for promoting its urban sustainable mobility objectives, it has partnered with the University of Toronto Transportation Research Institute (UTTRI) to create the *iCity-South* research program to develop and apply advanced urban informatics vision and capabilities in Latin American cities.

This report presents the results of one *iCity-South* project. The objective of this project is to investigate traditional and new data collection methods in Montevideo, Uruguay for use in travel demand analysis and modelling. This report is the sixth and final in a series of reports documenting the project's results. This report has three main purposes. First, it summarizes the work and findings of the project that have been presented in the previous five project reports. Second, it presents in the results from the final analysis of Antel cellphone trace and Intendencia de Montevideo public transit smartcard transaction data that represents the culmination of the project's work and which has not been previously reported. The products of this work are, first, a new dataset of detailed trips by origin, destination, purpose and mode generated by fusing data together from the individual datasets, and, second, the data fusion procedure, which can be applied to future data to similarly generate detailed representations of travel behaviour within the Montevideo region.

Third, and finally, the report discusses possible next steps that could build on this study's results to develop and implement an agent-based microsimulation travel demand model for Montevideo using the full range of data that have been examined in this study.

ACKNOWLEDGEMENTS

This study was funded by CAF. The unstinting support and patience of CAF, and, in particular, Nicolas Estupiñan and Andres Alcala has been greatly appreciated. The collaboration with Diego Hernandez, Universidad Católica del Uruguay, and Antonio Mauttone, Universidad de la República de Uruguay, has been most welcome and helpful. The generous support of Antel in providing access to the cellular data records and of the Intendencia de Montevideo for access to the transit transaction data analyzed in the report is gratefully acknowledged. The assistance of Juan Pablo Pignataro and Juan Andrés Negreira Marin from Antel and Verónica Orellano Chiazzaro and Leonardo Goday from the Intendencia has been much appreciated.

TABLE OF CONTENTS

		Page No.
Executive Summary		1
Acknowledgements		1
Table of Contents		2
List of Figures		3
List of Tables		3
1.	STUDY PURPOSE & MOTIVATION	4
2.	SUMMARY OF PREVIOUS WORK & FINDINGS	6
2.1	Introduction	6
2.2	2016 Montevideo Home Mobility Survey (MHMS)	8
2.3	Analysis of Public Transit Transaction Data	9
2.4	Preliminary Analysis of Antel Cellphone Trace Data	10
3.	DESCRIPTION OF DATA	11
3.1	Introduction	11
3.2	Antel Data	11
3.3	Public Transit Fare Transaction Data	20
3.4	2016 MHMS Data	24
3.5	Network Data	25
3.6	Other Data	26
4.	METHODS	27
4.1	Introduction	27
4.2	Imputing Home & Work Locations	27
	4.2.1 Home Locations	27
	4.2.2 Work Locations	28
	4.2.3 Origin-Destination (O-D) Trip Matrices	28
4.3	Imputing Trip Mode	28
5.	RESULTS	32
5.1	Introduction	32
5.2	Home & Work Locations	32
5.3	Cellphone Trip Mode	37
	5.3.1 Introduction	37
	5.3.2 Neural Net Training Results	37
	5.3.3 Cellphone Imputed Trip Mode Results	37
6.	SUMMARY & POSSIBLE NEXT STEPS	41
6.1	Summary of Results	41
6.2	Possible Next Steps: Data Collection & Fusion Methods	41
6.3	Possible Next Steps: Modelling	43

REFERENCES 44
APPENDIX I: CELLPHONE TRIP MODE SHARES 46

LIST OF FIGURES

Page No.

3.1: Antel Data Zone System 12
3.2: Hourly Trips by Day 13
3.3: Average Hourly Trips within the Sample, Weekdays & Weekends 14
3.4: Trips Origins and Destinations, Weekday AM Peak Period 15
3.5: Trips Origins and Destinations, Weekday Mid-Day Period 16
3.6: Trips Origins and Destinations, Weekday PM Peak Period 17
3.7: Trips Origins and Destinations, Weekday Evening Period 18
3.8: Cellphone vs. MHMS Trip Start Times 19
3.9: Cellphone vs. MHMS Trip End Times 19
3.8: Number of Smartcard Transactions per Card, May 2-29, 2018 19
3.10: Number of Smartcard Transactions per Card, May 2-29, 2018 20
3.11: Transit Boardings by Time of Day & Day, May 2-29, 2018 21
3.12: Comparison of Transit Boardings & Cellphone Trips by Time of Day & Day 21
3.13: Smartcard User Home Locations 22
3.14: Smartcard User Work Locations 22
3.15: Smartcard Weekday Trips to/from Home by Time of Day 23
3.16: Smartcard Weekday Trips to/from Work by Time of Day 23
3.17: Census Segments in the Montevideo Region 24
3.18 Montevideo Road Network 25
3:19: Montevideo Bus Network 26
4.1: Analysis Approach 29
4.2: Trip Trace Representation 30
4.3: Inputs & Outputs from a Single Neuron 30
5.1: Imputed Distribution of Cellphone Users Residential Locations 33
5.2: Imputed Distribution of Cellphone Users Job Locations 34
5.3: Cellphone vs. MHMS Trip Start Times, Home as Origin 35
5.4: Cellphone vs. MHMS Trip Arrival Times, Home as Destination 35
5.5 Cellphone vs. MHMS Trip Start Times, Work as Origin 35
5.6: Cellphone vs. MHMS Trip Arrival Times, Work as Destination 36
5.7: Cellphone vs. MHMS Home-to-Work Trip Start Times 36
5.8: Cellphone vs. MHMS Work-to-Home Trip Start Times 36
5.9: MHMS & Predicted Cellphone Trip Modes by Trip Distance 38
5.10: Observed MHMS & Predicted Cellphone Mode Shares by Trip Start Time 39
5.11: Transit Boardings & Weighted Cellphone Transit Trips by Trip Start Time 40

LIST OF TABLES

Page No.

3.1: Mode Share of Trips in MHMS Dataset 25
5.1: Home Location Imputation Results 32
5.2: Work Location Imputation Results 33
5.3: Neural Net Validation Using the MHMS Sample 37
5.4: Predicted Cellphone vs. Observed MHMS Aggregate Mode Shares 38

Chapter 1
Study Purpose & Motivation

Urban regions within Latin America (and elsewhere) face enormous challenges in terms of the provision of transportation infrastructure and services to meet the travel needs of their growing population in a cost-effective, equitable and sustainable manner. High quality, comprehensive information concerning travel behaviour and transportation system performance is a fundamental prerequisite for successful urban transportation planning and decision-making to address these pressing, first-order needs.

In recognition of this need, CAF established the Urban Mobility Observatory (OMU, *Observatorio de Movilidad Urbana*)[1] to assemble and utilize standardized transportation-related data for Latin American cities. 29 cities are currently members of OMU. Collecting consistent, time-series data for these cities, however, is a difficult and costly task for CAF and its partner cities.

At the same time, exciting, new transportation data collection sources are emerging to complement or even replace the traditional methods used to collect the OMU data. These include:

- The pervasive penetration of cellphone and smartphone technology within urban populations.
- The widespread adoption of smartcard systems by public transit agencies in many cities.
- Extensive deployment of many types of sensors (video, thermal, Bluetooth, etc.) for monitoring travel flows.
- Increasing availability of very large (typically crowd-sourced) datasets collected in a variety of ways by private sector companies (Google, Waze, Inrix, etc.) that can provide travel information.
- Web-based survey methods to complement/replace traditional survey methods such as home-interviews, telephone interviews, etc.

In 2015, the University of Toronto Transportation Research Institute (UTTRI) launched the *iCity* research program, which is dedicated to applying modern *urban informatics* (the combination of data collection, data science, modelling, visualization and high-performance computing methods) to the promotion of sustainable metropolitan growth. As one component of CAF's strategy for promoting its urban sustainable mobility objectives, it has partnered with UTTRI to create the *iCity-South* research program to apply the *iCity* urban informatics vision and capabilities in Latin American cities.

Two initial projects were chosen to launch the *iCity-South* research program. One involved the demonstration of agent-based microsimulation methods for modelling urban travel demand in terms of developing a prototype microsimulation model for Asunción, Paraguay.[2] The second is the focus of this report. The objective of this project tis to investigate traditional and new data

---

[1] https://www.caf.com/es/temas/o/observatorio-de-movilidad-urbana/

[2] This project was completed in April, 2017. See Miller, et al., (2017a,b) for the results of this project.

collection methods in Montevideo, Uruguay. This report is the sixth and final in a series of reports documenting the Montevideo project results.

This report has three main purposes. First, it summarizes in Chapter 2 the overall work and findings of the project, most of which have been presented in the previous five project reports.

Second, it presents in the results from the final analysis of Antel cellphone trace and Intendencia de Montevideo public transit smartcard transaction data that represents the culmination of the project's work and which has not been previously reported. Chapter 3 describes the data used in the analysis. Chapter 4 presents the methods used in the analysis, while Chapter 5 discusses the analysis results.

Third, and finally, Chapter 6 discusses possible next steps that could build on this study's results to develop and implement an agent-based microsimulation travel demand model for Montevideo using the full range of data that have been examined in this study.

CHAPTER 2
SUMMARY OF PREVIOUS WORK & FINDINGS

2.1 INTRODUCTION

The starting point for this project was a comprehensive review of the current state of the art of urban travel demand data collection methods (*Project Report 1*, Miller & Habib, 2017). Key findings of this review include:[3]

- The field is currently in a significantly dynamic state, in which the data requirements to support urban transportation planning and modelling are changing in response to increasingly complex issues and correspondingly challenging analysis needs.
- Traditional methods of data collection – principally various forms of home interview surveys – are increasingly expensive and difficult to undertake, for a variety of reasons.
- At the same time, new Information Technology (IT) based data sources and data collection methods provide promise of supplementing, or even replacing, traditional data collection methods with new data sources that may be more comprehensive, cost-effective and applicable to current planning issues than traditional survey methods.
- It is very unlikely that any one data collection method will be able to address all planning analysis and modelling needs. Each method possesses strengths and weaknesses that make it better suited to some applications than others.
- Further, most new, IT-based datasets, such as cellphone traces, transit smartcard transactions, etc., while providing massive amounts of information (so-called "big data") concerning trips, generally lack any information concerning characteristics of the trip-makers (age, gender, income, etc.). They also typically lack information concerning key trip attributes, such as trip purpose and mode.
- Given the previous two points, a *multi-instrument* approach to data collection design is required, in which multiple data collection methods (or sources of data) are jointly used in a coordinated *data collection program*. This data collection program should be explicitly designed to maximize the usefulness of each source of data and to combine the various datasets so as to provide as comprehensive an overall dataset as possible – one in which the "whole is more than the sum of its parts".
- To build this comprehensive dataset from its individual parts requires the use of advanced *data fusion* methods, such as *machine learning* methods, to generate statistically reliable, robust datasets.

These travel data collection issues are particularly challenging in the Latin American context, given:

- The heterogeneity in socio-economic and living conditions existing in many urban regions.
- The presence of informal and/or privatized transit services in many cities
- City-to-city and country-to-country differences in government structures and capacities.

---

[3] See also, Miller, et al. (2012).

At the same time, however, many Latin American cities are aggressively pursuing new IT technologies, such as public transit smartcards, which offer the potential to support the sort of new, multi-frame, comprehensive data collection program envisioned above. Montevideo, in particular, is very well positioned to be a leader in this regard, for several reasons:

- With a metropolitan regional population of 1.9 million persons, it is "large enough" to present interesting challenges with respect to data collection and transportation planning analysis, but it is also not so large that these challenges might overwhelm the developing and testing of new methods.
- Montevideo is a well-governed urban region with a progressive approach to transportation planning.
- Montevideo is blessed with several very good sources of travel-related data to support the development and testing of new methods.

In particular, three major sources of travel-related data are available in Montevideo:

- A recent (2016) traditional home-interview travel survey (the Montevideo Home Mobility Survey – MHMS).
- A very comprehensive and high quality database of public transit usage, including both smartcard and cash-fare transactions.
- Cellphone trace data, generously made available to the project by Antel, the largest cellphone provider in Montevideo.

Given this rich data environment, Montevideo has provided an excellent case study region to investigate the questions raised above concerning the usefulness of new data sets and how they might best be utilized in Latin American planning and modelling applications. To do so, this project proceeded in two stages. The first stage involved analysing each of the three data sets listed above (MHMS survey data; transit transaction data; Antel cellphone traces) independently to access their strengths and weaknesses. The results of these analyses are presented in detail in *Reports 2-5* of this project's report series and are also briefly summarized in Sections 2.2, 2.3 and 2.4, respectively, below.

The second, and final, stage in the project involved bringing the three datasets together through a data fusion process to create a more holistic and comprehensive dataset characterizing travel behaviour in the Montevideo urban region. In particular, this work involves using MHMS and transit transaction data to impute trip purposes and travel modes for trips observed as Antel cellphone traces. As part of this analysis process, trip purposes for the transit transaction data are also imputed. The final product of this analysis is a much richer and comprehensive "snapshot" of travel behaviour in the region than that which can be provided by any one of the datasets individually. As discussed further in Chapter 6, it is hoped that this new, fused dataset will be of on-going use to Montevideo for a variety of transportation planning purposes, including, possibly, the development of a comprehensive model of travel within the region.

Chapters 3, 4 and 5 document this second stage work, which has not been previously reported within this project. Chapter 3 provides an overview of the datasets used in the analysis. Chapter 4 describes the analysis methods used. Finally, Chapter 5 presents the results obtained.

2.2 2016 Montevideo Home Mobility Survey (MHMS)

The 2016 Montevideo Home Mobility Survey (MHMS) was designed and executed by the municipal governments of the Metropolitan Area of Montevideo (AMM) and the Universidad de la República (Udelar) under funding from CAF. It is a classic home interview survey in which trained interviewers survey randomly selected households in their homes. The survey was conducted during the period of August-October 2016. The survey study area consisted of the entire AMM. In total, 2,230 households and 5,946 persons within these households were interviewed, representing a 0.34% sample of the approximately 656,000 households (1,807,000 persons) within the survey study area (based on 2011 Census data). The survey and key results are extensively documented in the July, 2017 report, *Encuesta de movilidad en el Área Metropolitana de Montevideo 2016, Principales resultados e indicadores.*

The UTTRI iCity-South project team reviewed the draft design of the survey questionnaire in December, 2015. We found the MHMS to be a very well-designed survey, but we were also able to provide some suggestions to the MHMS design team concerning possible minor changes in the wording of several questions, as well as the general layout of the questionnaire. The iCity-South team also had an opportunity to further discuss the survey design with the MHMS design team while it was in pilot testing during our first project visit to Montevideo in early June, 2016. These discussions confirmed the quality of the survey design and care with which the survey was being implemented in the field.

Once the survey had been completed and the cleaned dataset was available for analysis, the iCity-South team reviewed the survey data with respect to:

- Definition of traffic zones.
- The spatial distribution of the respondents.
- The socio-economic representativeness of the sample.
- The trip attributes collected.
- Implications of the sample size/rate for travel behaviour analysis and modelling.

The analysis and findings of this review are documented in *Report 2* of the project report series (Miller, et al., 2017c). In general, it was found that the spatial distribution of respondents, the socio-economic representativeness of the sample and the trip attributes collected were all satisfactory, confirming the overall quality of the survey and the usefulness and representativeness of the survey data for travel analysis purposes.

The only weakness identified in the survey stems from its small sample size (0.34%), which means that origin-destination (O-D) trip matrices can only be reliably constructed using a very aggregate traffic zone system which is too gross for detailed travel demand modelling purposes. Thus, while the disaggregate trip records are individually valid and useful, they should ideally be augmented by larger-sample data that would permit more spatially (and temporally) detailed O-D travel patterns to be constructed to support more detailed travel demand model development.

Given this key finding, a primary purpose of the work reported in this document is to explore how the MHMS data can be best combined with such other larger travel-related datasets to address this need.

2.3 ANALYSIS OF PUBLIC TRANSIT TRANSACTION DATA

A major task in this project was the detailed analysis of Montevideo public transit transaction data. Montevideo has an excellent database to support this analysis, which consists of every smartcard transaction, plus an electronic recording of every cash transaction as well. Hence essentially a 100% record of every transit boarding is available.

The transaction data is a "tap-on" system, in which every boarding is recorded, but no alighting information is recorded. Thus, a major task involved in translating these transaction data into trip data is to impute the destination for each trip. For smartcard transactions, the "tap-on" location for subsequent trips for a card-holder for a given day can be used to impute the alighting location for previous trips, while the tap-on location of the first trip of the day can generally be used to impute the alighting location of the last trip of the day. Transfers can also be tracked for smartcard users, since the tap-on for the second (and subsequent) transit line used in a given trip is also observed. Multi-day/week data significantly enhance the statistical reliability and usefulness of the data.

Non-smartcard boardings are also recorded, but transfers and trip sequences for trip-makers paying by cash cannot be directly imputed. These boardings, however, are still useful to provide information concerning total daily transit usage by line and boarding location.

The Intendencia de Montevideo provided a one-week sample of data for Montevideo's integrated public transit system STM (Sistema de Transporte Metropolitano), composed of buses from 4 different operators that serve the City of Montevideo and surrounding areas (Coetc, Comesa, Cutcsa, Ucot). This system has 144 bus lines with 107 different destinations, and 4,835 stops.[4] The data consists of 4 main components:

1. Boarding records: 7 consecutive days of passenger boarding records, including the five weekdays and a weekend from August 15th to August 21st, 2016. These records belong to smartcard (STM card) and no-card passengers recorded by the system.
2. Lines and branches: Information about bus routes including the direction and order of stops. Each bus run or trajectory in one direction, is labeled with a unique identification number that can be paired with this data to obtain the run's line and branch.
3. Stops: Number, coordinates, and description of the closest intersection from the stop.
4. Automatic Vehicle Location (AVL): Position and speed of 295 bus runs without timestamps.

The passenger boarding records correspond to smartcard and non-smartcard users during a complete week (Monday-Sunday). The total boarding records for smartcards are 5,077,674 and for no cards 2,371,815, representing a 68% to 32% split.

The analysis of these data was undertaken in two phases: a preliminary analysis (documented in *Project Report 3*, Parada and Miller (2017) and a second, more detailed analysis, documented in *Project Report 5* (Parada and Miller, 2018). More detailed results are also available in Parada's MASc thesis (Parada, 2018). Key issues investigated include:

---

[4] http://www.montevideo.gub.uy/transito-y-transporte/stm-sistema-de-transporte-metropolitano/el-sistema

- Descriptive analysis of smartcard trips and non-smartcard trips to identify similarities and differences between the two user groups.
- Detailed analysis of transit users' travel patterns.
- Application of a method to estimate trip alighting locations for smartcard users.
- Estimation of trip origins and destinations (zone-based) from stop boarding and alighting locations within the transaction data.

Overall, this analysis clearly demonstrated the usefulness of transit transaction data (especially smartcard data) for characterizing transit travel behaviour in Montevideo (and, by extension, elsewhere). In particular, although considerable analysis is required, it is possible to impute robust estimates of transit trip origin-destination (O-D) travel from transit stop-based transaction data for a very large percentage of trips.

Subsequent to this analysis, the Intendencia de Montevideo has developed their own transaction data processing methods to generate transit O-D trips by time of day. The new transaction data described in Chapter 3 that are used in the data fusion exercise described in Chapters 4 and 5 are based on the Intendencia's methods.

2.4 Preliminary Analysis of Antel Cellphone Trace Data

As a first step in the investigation of the use of cellphone trace data for travel demand analysis purposes, a one-day sample of cellphone traces by 5,000 Antel customers was obtained from Antel. The trace data were aggregated in 15-minute time intervals and a 74-zone system for the Montevideo region, resulting in a dataset of approximately 480,000 data records to process.

Project Report 4 (Faghih-Imani and Miller, 2018) presents the results of this analysis, which include:

- A literature review of usage to date of cellphone trace data, as well as the analysis/modelling methods used
- Cellular trace data clearly possess potential for describing trip-making in an urban region.
- **BUT**:
  - Finer temporal and spatial resolution are both required for the trace data to be truly useful. 5-minute, rather than 15-minute time intervals are required, as is a finer zone system if trips are to be characterized in both time and space at a level of disaggregation that is suitable for useful travel demand modelling.
  - Multiple days/weeks of observations for individual trip-makers is required for statistically robust estimations of their travel behaviour.
  - A much larger sample of cellphone users is required to generate statistically robust representations of travel by origin, destination and time of day. A key strength of cellphone trace data is the potentially very large size of the dataset. Exploiting the strength of such "big data" is essential in order to maximize its usefulness.

CHAPTER 3
DESCRIPTION OF DATA

3.1 INTRODUCTION

This chapter provides overall descriptions of the various data used in the data fusion exercise undertaken in this study. These consist of:

- A very large sample of Antel cellphone traces, consisting of 40% of all such traces for the time period May 2-29, 2018.
- All public transit fare transaction records for the same time period, provided by the Intendencia de Montevideo.
- 2016 MHMS records.
- Road and transit network data, as well as other ancillary datasets used in the analysis.

Sections 3.2 – 3.5 present descriptions of each of these datasets, respectively.

3.2 ANTEL DATA

Antel, the primary cellular telecommunications company in Montevideo, provided four weeks of cellphone traces for analysis in this project, consisting of a random sample of 40% of mobile phones using their service within Montevideo and the surrounding metropolitan area. The data are for the time period May 2nd to May 29th, 2018. Overall, the dataset includes more than 117,862,000 cellphone traces for about 948,600 unique cellphones.

These cellphone traces were preprocessed from raw data by Antel to eliminate as much noise in the data as possible and to preserve user anonymity. The trace data are temporally reported in minutes, thereby providing excellent temporal precision. They are also spatially aggregated to 135 zones in the Montevideo region (see Figure 3.1). The average area of the zones in the study area is about 37.1 km^2, with the smallest zone having an area of 0.38 km^2 and the largest zone about 1,020 km^2.

Figure 3.1: Antel Data Zone System

In order to do a preliminary analysis of the data and to identify the potential of the data for travel analysis, an assumption is made that any time a cellphone user stays more than 30 minutes in a zone is due to participation in an activity at that location. With this assumption, 32,417,002 activities are estimated for the four weeks of data. Further, trips are defined when a user moves between two activities with first activity location identified as the trip origin and the second activity location as the trip destination. More than 14 million of trips are estimated in the Montevideo region within the dataset. The average estimated daily trip rate per person is 1.99, with a minimum of 1 and maximum of 13.54. This daily trip rate is slightly lower than that observed in the 2016 MHMS of 2.11 (see Section 3.4, below).

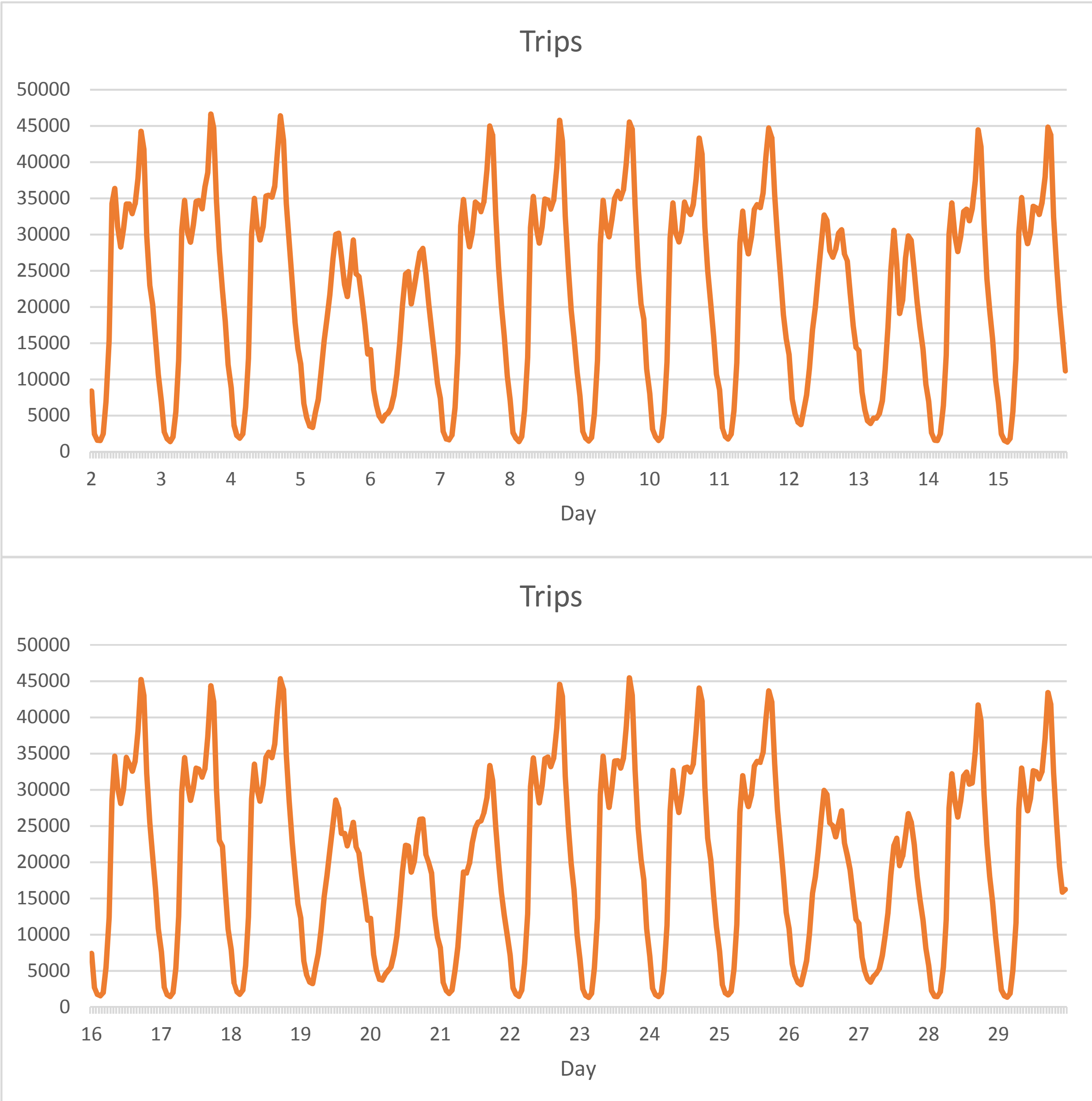

Figure 3.2: Hourly Trips by Day

Figure 3.2 presents the hourly trips in the Montevideo region for each day in the sample. Figure 3 also presents the hourly trips for a typical weekday and weekend day. These two figures clearly show that the temporal pattern of trip-making in Montevideo can be identified using cellphone location data.

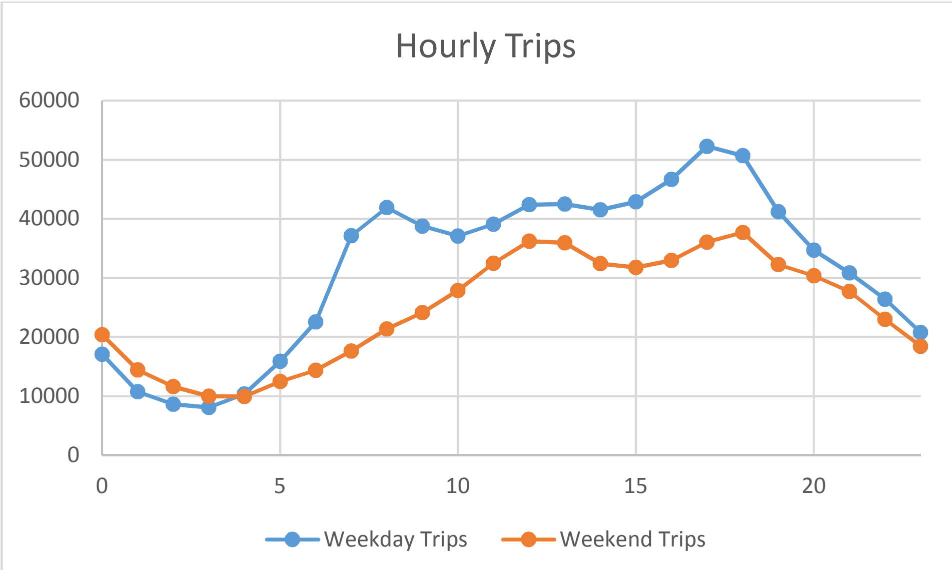

Figure 1 Hourly trips for a typical weekday or weekend

Figure 3.3: Average Hourly Trips within the Sample, Weekdays & Weekends

Each identified trip has its own start and end time and an origin and destination based on the detected activities which are connected by the trip. Thus, it is possible to create Origin-Destination (O-D) matrices for each time period by accumulating the trips by time period. Figures 3.4 to 3.7 depict the average number of weekday trips generated by origin and destination, aggregated to four periods in a day. Based on trip start time, the four time periods are: AM (6:00-9:00), Midday (9:00-15:00), PM (15:00-19:00), and Evening (19:00-24:00). Note that these maps display the total number of trip origins or destinations for a given zone during the given time period, not an hourly average. Again, these maps demonstrate that the spatial pattern of trips within Montevideo can be obtained by processing cellphone location traces at a level of spatial detail (i.e., traffic zones), which is useful for transportation planning and modelling purposes.

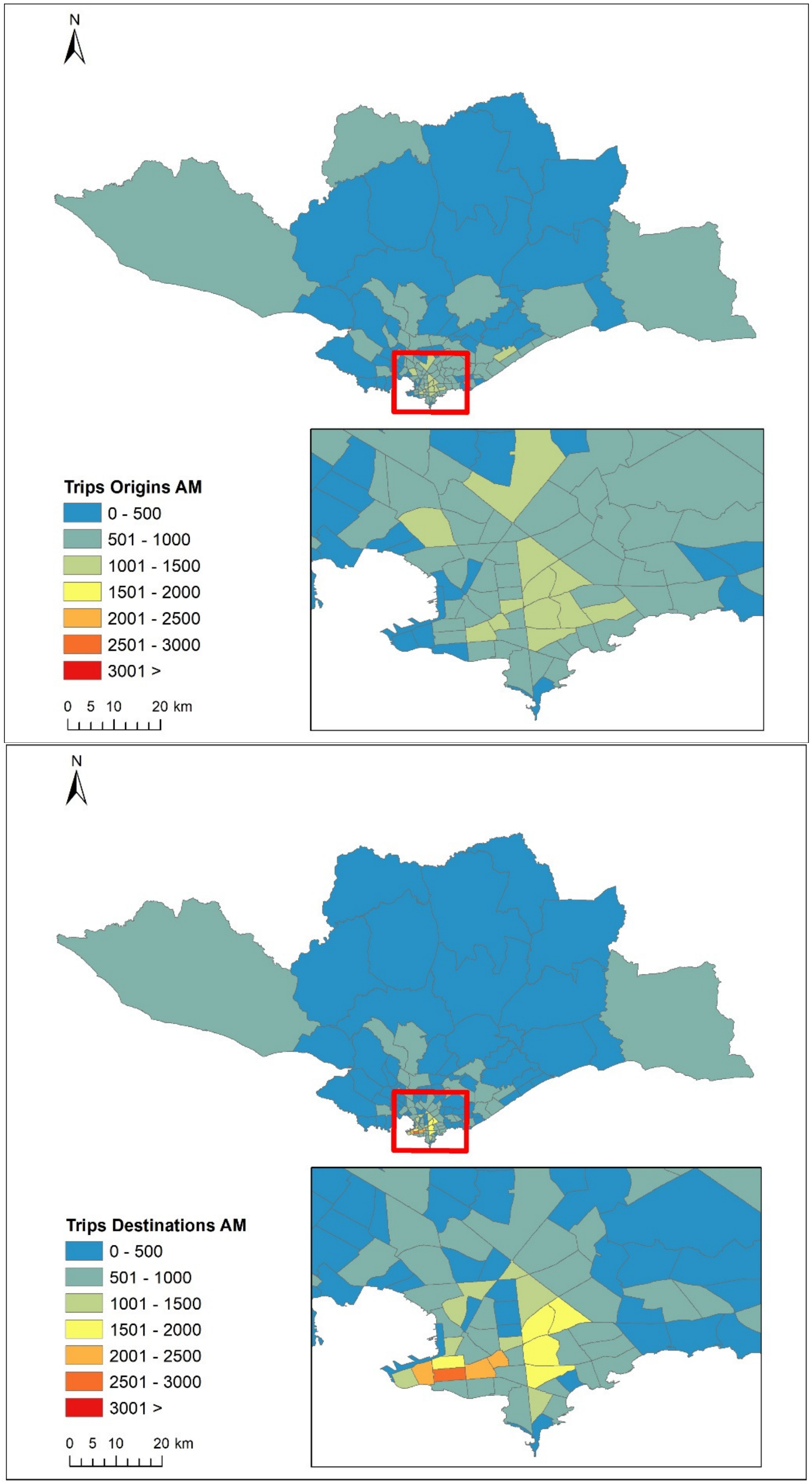

Figure 3.2: Trip Origins and Destinations, Weekday AM Peak Period

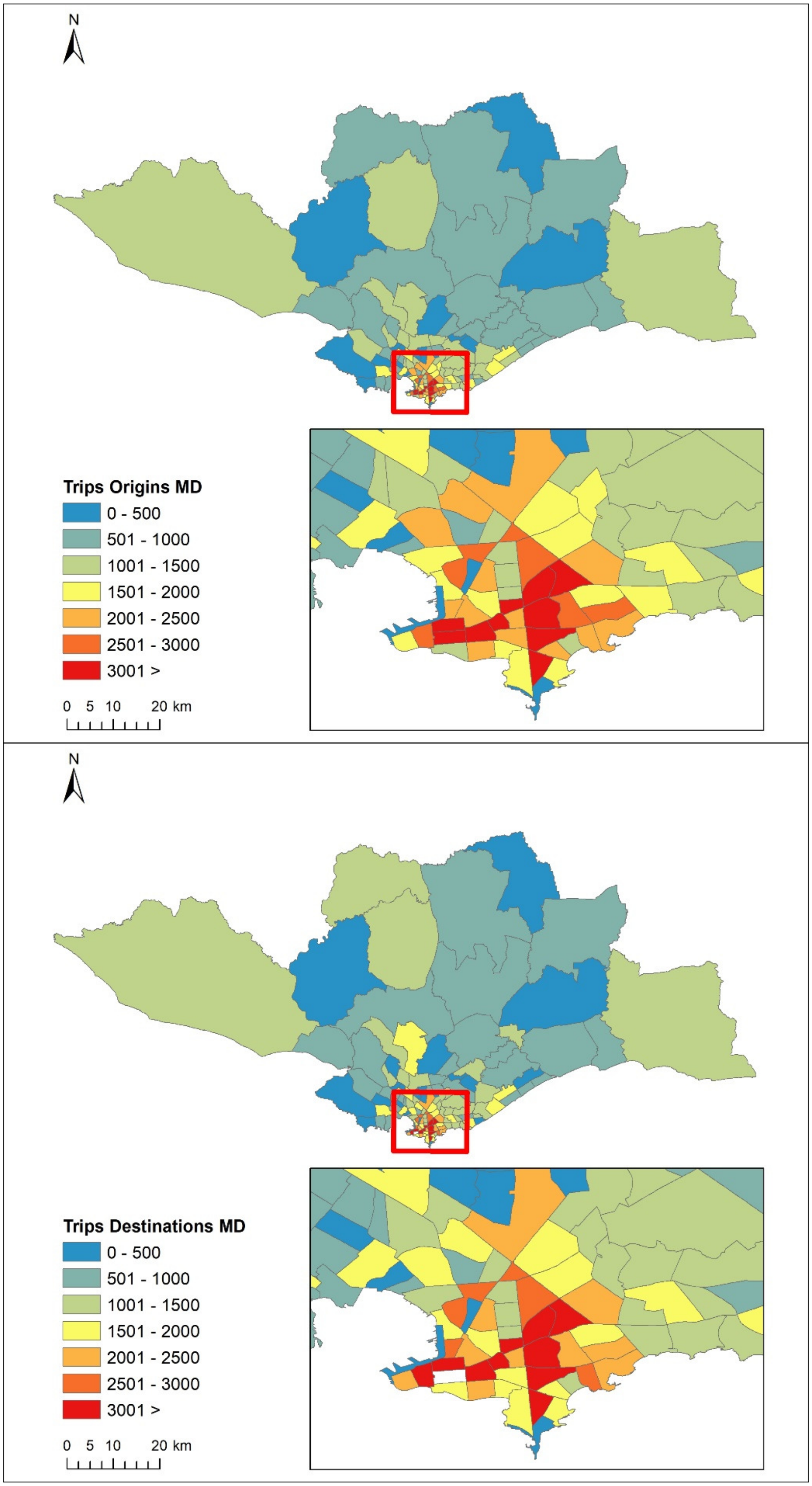

Figure 3.5: Trip Origins and Destinations, Weekday Mid-Day Period

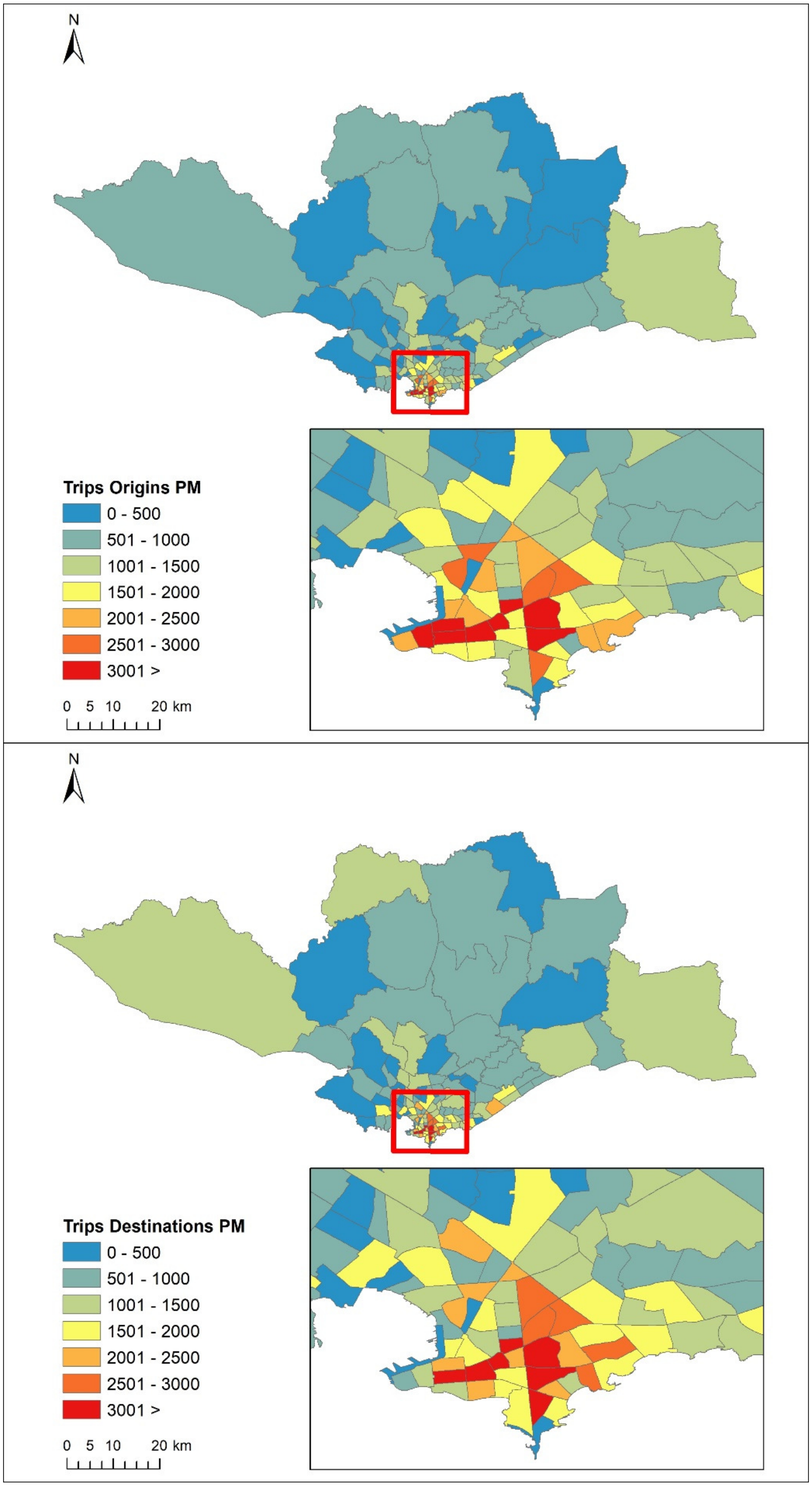

Figure 3.6: Trip Origins and Destinations, Weekday PM Peak Period

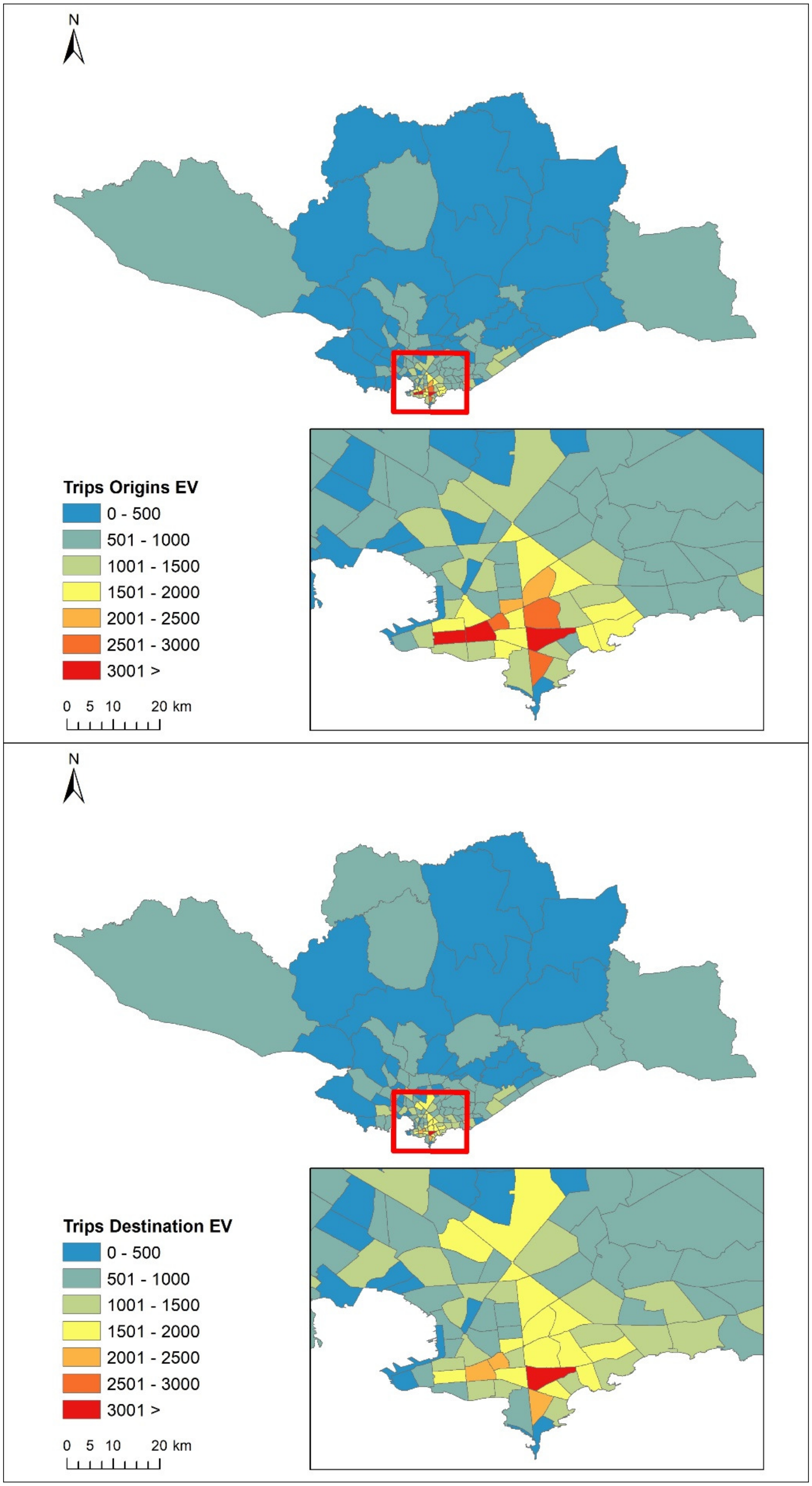

Figure 3.7: Trip Origins and Destinations, Weekday Evening Period

Figures 3.8 and 3.9 compare Antel cellphone trace and MHMS trip start and end times, respectively, by time of day. As can be seen, the cellphone distributions generally compare well with the survey data, except in the morning peak period, where relatively fewer trips are observed in the cellphone trace data. The reason for this discrepancy is not clear and requires further investigation. It may reflect the greater number of mid-day trips observed in the cellphone data, which is an expected result, since it is likely that non-home-based, mid-day trips are under-reported in the survey. It may also reflect loss of short trips due deletion of intra-zonal trips in the analysis, which possibly particularly occur during the morning peak period.

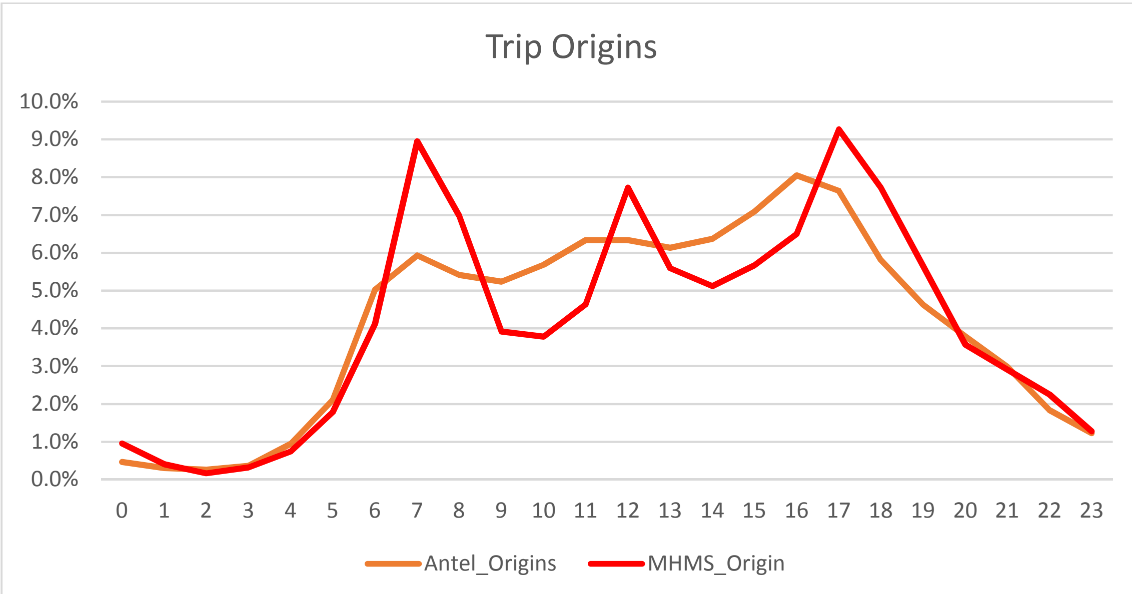

Figure 3.8: Cellphone vs. MHMS Trip Start Times

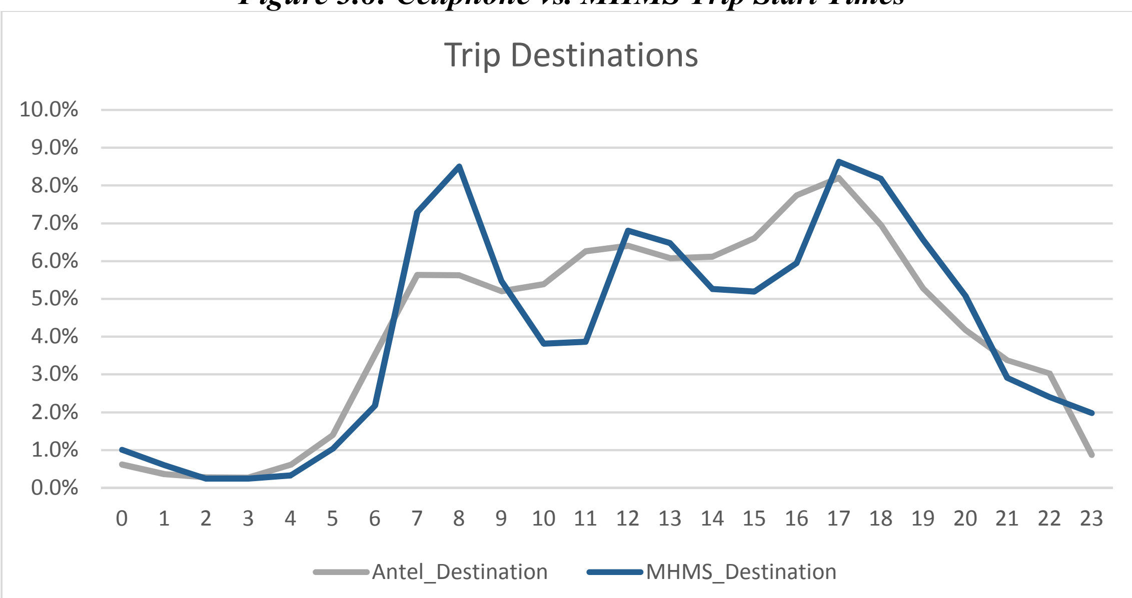

Figure 3.9: Cellphone vs. MHMS Trip End Times

3.3 PUBLIC TRANSIT FARE TRANSACTION DATA

All fare transaction data for the Montevideo transit system were provided by the Intendencia de Montevideo for the same May 2-29, 2018 time period as for the provided Antel data. These consisted of: 29,868,716 recorded transactions, 82.5% of which were made by smartcards from 734,569 unique smartcards. For the smartcard records, the Intendencia used their in-house procedure for identifying trip alighting stops. Their algorithm successfully identified 62% of these alighting locations, which were attached to the transaction records. Unique (but anonymized) identifiers were attached to smartcard records for each smartcard so that the trip-making behaviour of the smartcard owner could be tracked day-to-day within the observation period.

Figure 3.10 shows the frequency distribution of the number of smartcard transactions per user during the sample period. On average, a smartcard was used for 33.6 transactions (SD=29.3).

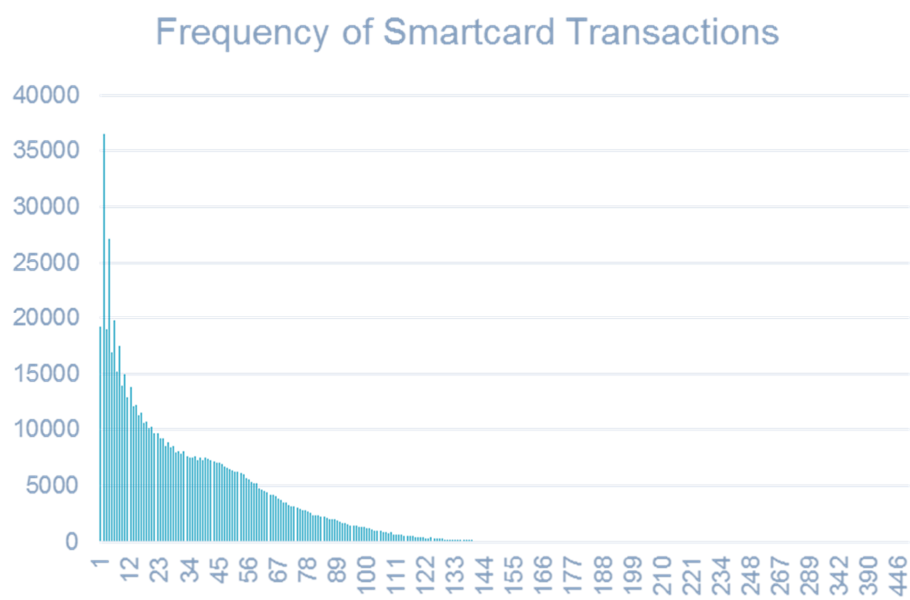

Figure 3.10: Number of Smartcard Transactions per Card, May 2-29, 2018

Figure 3.11 shows the daily distribution of boardings by time of day for 28 days observed in the sample. Figure 3.12 overlays this distribution of transit trips with the same distribution of Antel cellphone trace-based trips. In general, good correspondence between the two distributions is observed.

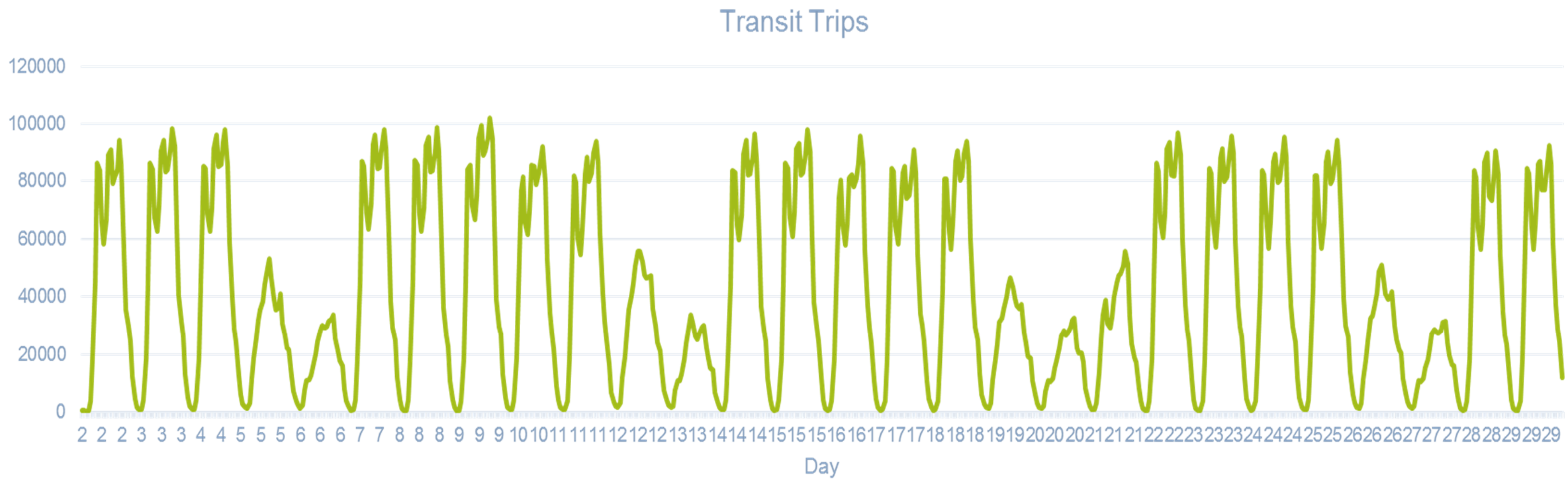

Figure 3.11: Transit Boardings by Time of Day & Day, May 2-29, 2018

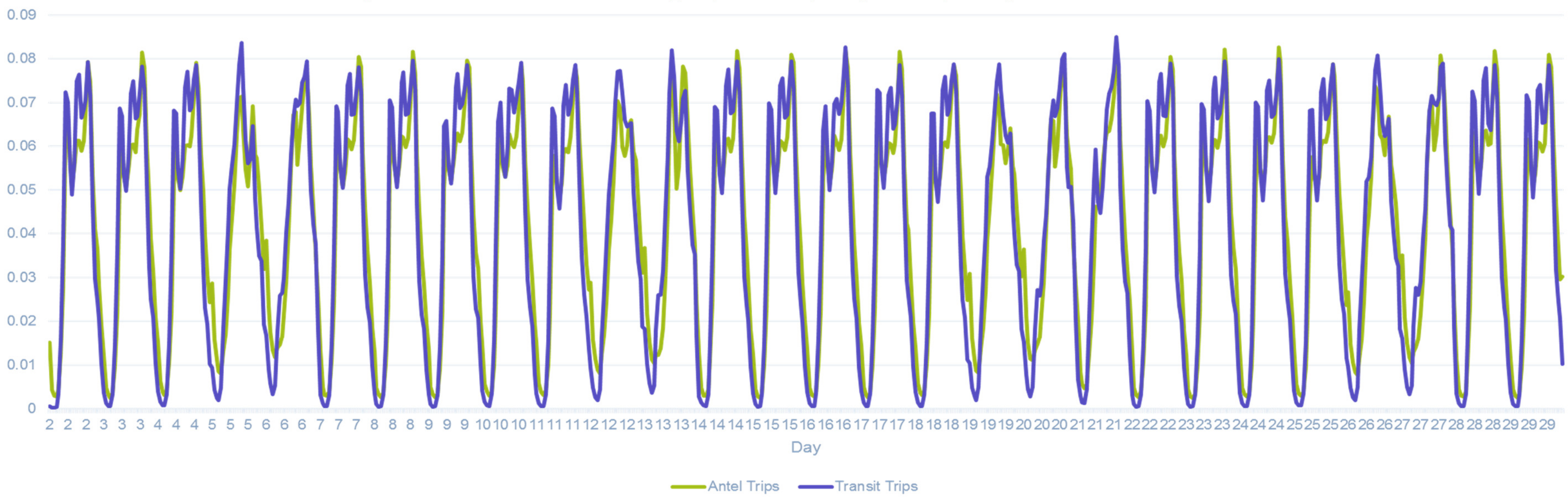

Figure 3.12: Comparison of Transit Boardings & Cellphone Trips by Time of Day & Day

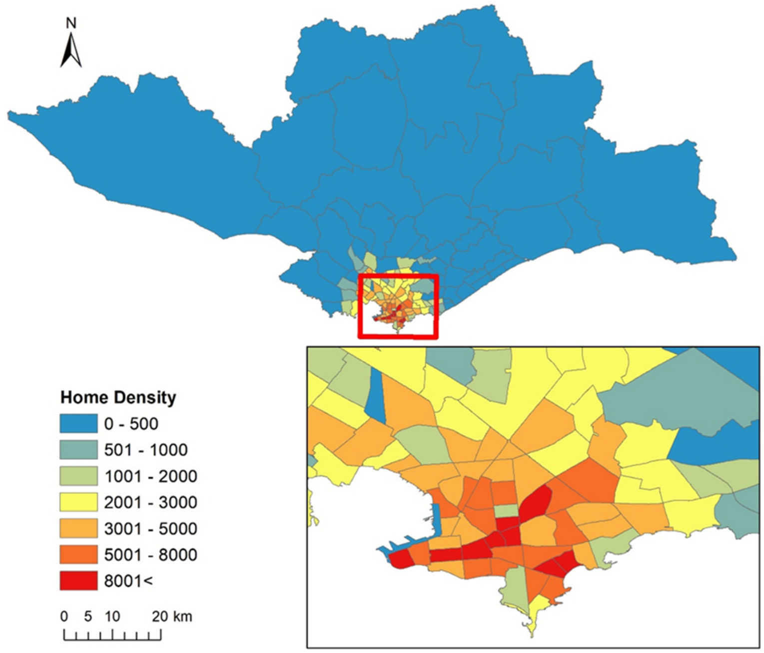

Figure 3.13: Smartcard User Home Locations

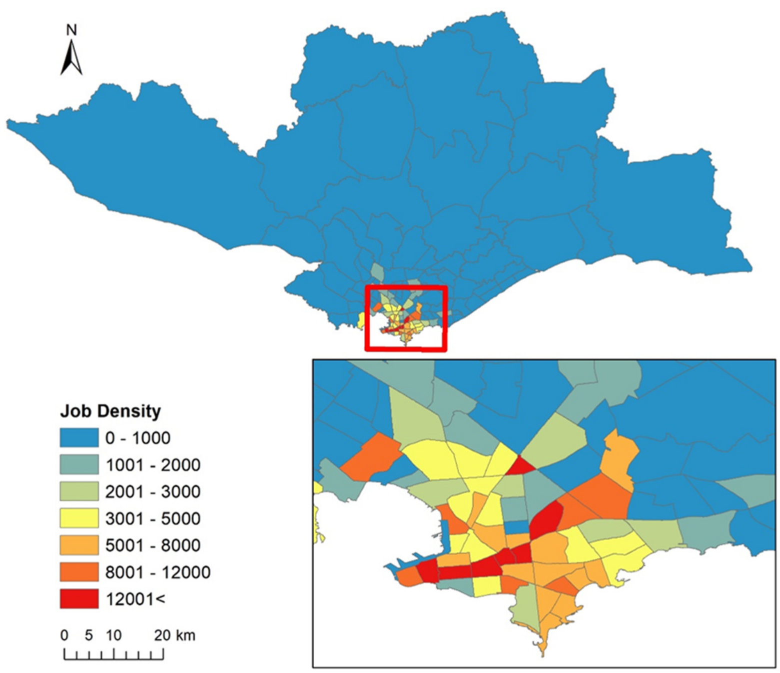

Figure 3.14: Smartcard User Work Locations

Home and work[5] locations were identified for smartcard owners making multiple trips during the sample period. Home locations were assumed to be identified if the same (stop) location was the first tap-on point of the day for at least 3 days during the month. Similarly a (stop) location was

[5] Or, possibly, a school location. More precisely, these locations should be considered as a non-home location that is visited on a systematically recurring basis during the observation period.

assumed to be the trip-maker's work location if it was observed to be the same last tap-on of the day for at least 3 days during the month. Using these rules, 75.9% of the smartcard users were assigned a home location, while 64.9% were assigned a work location. Figures 3.13 and 3.14 map the identified home and work locations for the smartcard users, which generally correspond well with census data.

Given these home and work definitions, trips by time of day to/from home and work for weekday trip-makers were identified in the smartcard data, as summarized in Figures 3.15 and 3.16.

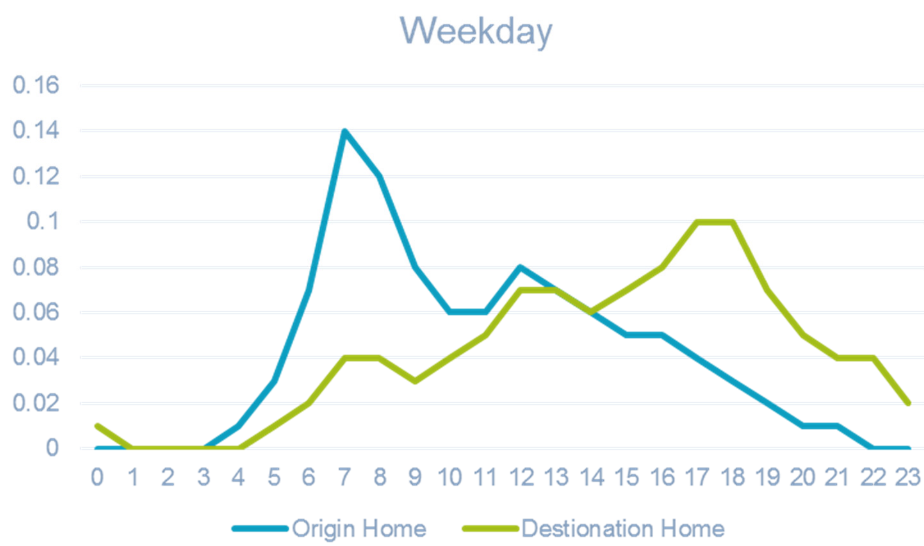

Figure 3.15: Smartcard Weekday Trips to/from Home by Time of Day

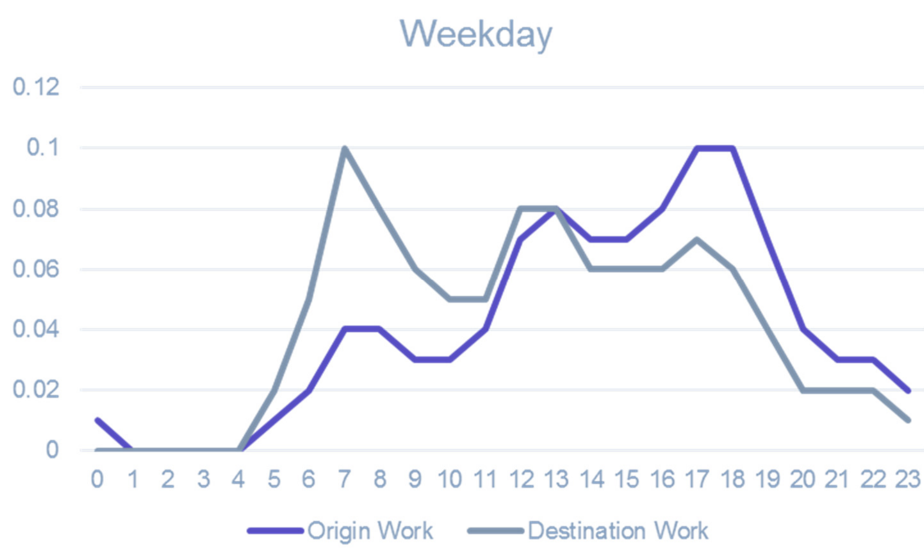

Figure 3.16: Smartcard Weekday Trips to/from Work by Time of Day

3.4 2016 MHMS DATA

The MHMS data has already been discussed in Section 2.2. The data were collected during the period of August-October 2016 in the Metropolitan Area of Montevideo. The size of this survey represents a 0.34% sample of the households in the region, with 2,230 households interviewed. These households consist of 5,946 individuals and reported 12,546 trips which results in a trip rate of 2.11. The average age of individuals in the survey is 38.8 years; 53.1% of them are female. The MHMS data is spatially aggregated to census segments. Census segments are groups of blocks (usually between 6 to 12 blocks) that are the spatial units used by the 2011 Uruguayan census. There are about 1,720 census segments in the Montevideo region with an average area of 2.91 km^2. Figure 3.17 shows the census segments in the Montevideo region. The mode share of the reported trips in MHMS dataset is presented in Table 3.1. Walking has the largest mode share, that over a third of the trips recorded are short. Combining drives and passengers, cars account for 29.2% of trips, while the bus system carries about 25.2% of the trips.

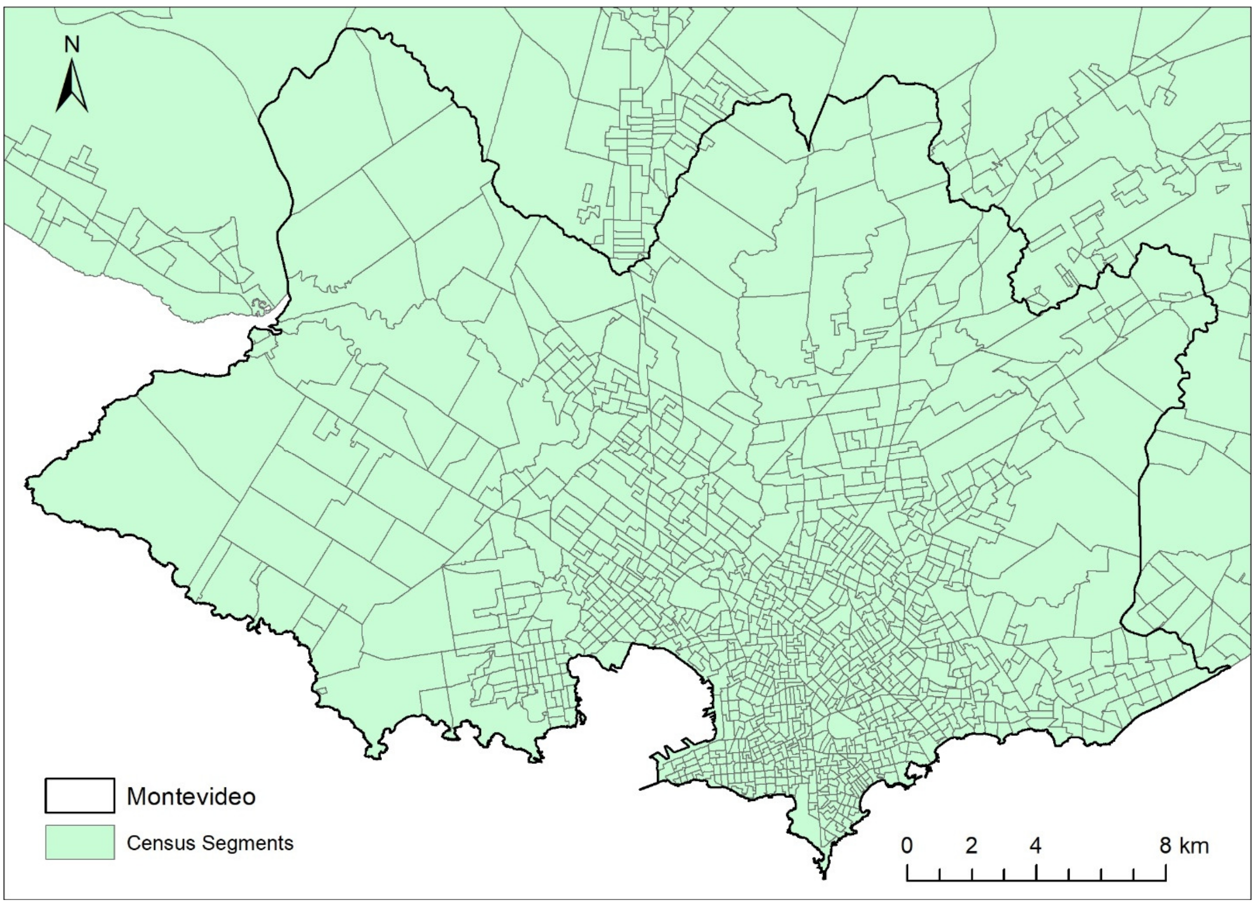

Figure 3.17: Census Segments in the Montevideo Region

Table 3.1: Mode Share of Trips in MHMS Dataset

Travel Mode	Mode Share in %	Number of Trips
Walk	34.0	4265
Bike	3.5	439
Auto Passenger	10.0	1251
Auto Driver	19.2	2410
Motorcycle	6.1	769
Bus	25.2	3166
Other	2.0	246

3.5 NETWORK DATA

Figure 3.18 presents the representation of the Montevideo road network in the Montevideo that is used in this study. It is obtained from the OpenStreetMap database. Further, the data regarding the public bus system including routes, stops and frequencies are gathered from the open data portal of Montevideo government. The bus network is presented in Figure 3.19. These two transportation networks are used in our traffic assignment model using EMME software to estimate link travel times, volumes and congestion in the network, along with O-D travel times and costs.

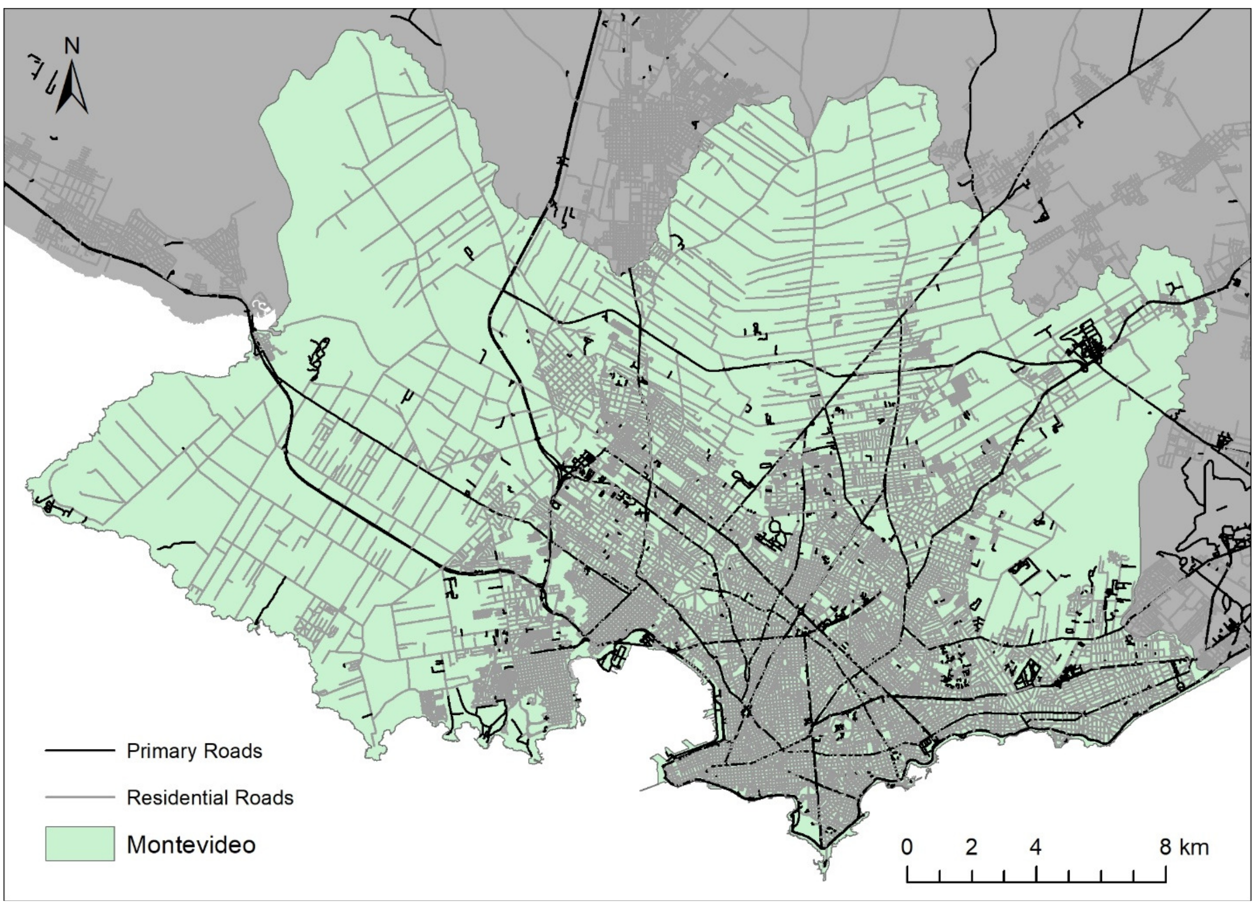

Figure 3.18: Montevideo Road Network

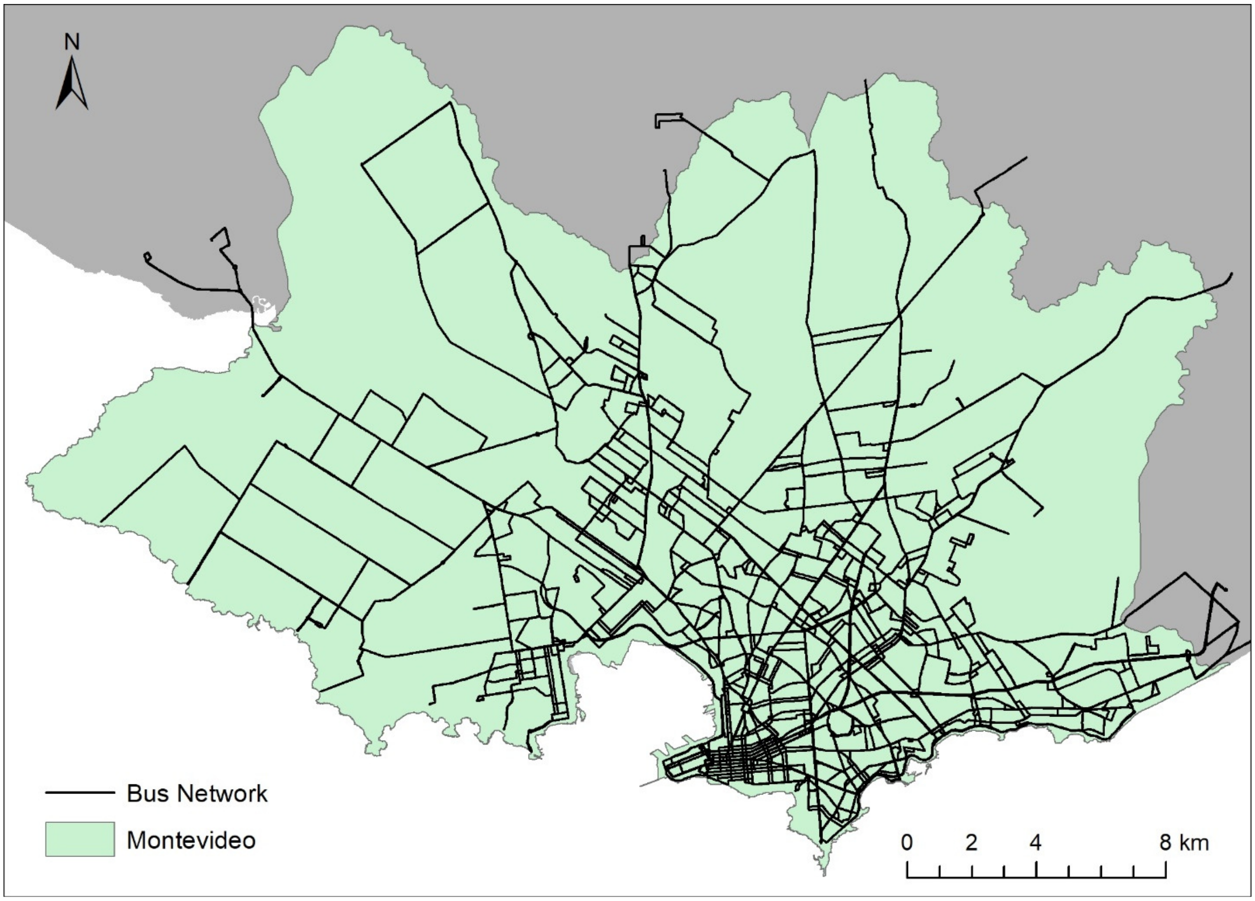

Figure 3:19: Montevideo Bus Network

3.6 OTHER DATA

The latest census in Montevideo region was conducted in 2011. Data of interest to this study include the number of persons and households in each zone. Socio-demographic information such as population age and gender distributions in the region is also available. In our analysis, the census data is used to validate our estimation of home locations of cellphone users in our sampled data, based on the Antel cellphone traces.

CHAPTER 4
ANALYSIS OF CELLPHONE TRACES: METHODS

4.1 INTRODUCTION

As demonstrated in both the preliminary analysis of a small subset of Antel data (Section 2.4) and the analysis of the large May, 2018 dataset (Section 3.2), cellphone trace data can be used to generate robust origin-destination (O-D) trip matrices by traffic zone by time of day that are sufficiently precise for travel demand modelling and other transportation planning analyses. They lack, however, two key trip attributes that are required if these data are to be used for modelling purposes: trip mode and trip purpose.

The objectives of the analysis undertaken in this study are to impute for the Antel cellphone trace data:

1. Home and work locations for trip-makers.
2. Travel mode (auto, transit, etc.) using the 2016 MHMS data, which contains mode choice attributes for each observed trip.

Section 4.2 describes the methods used to impute home and work locations. Section 4.3 describes the methods used to impute trip mode. All results from the application of these methods to the cellphone traces are presented in Chapter 5.

4.2 IMPUTING HOME & WORK LOCATIONS

4.2.1 Home Locations

A variety of criteria were tested to identify the home location of each cellphone user in the dataset. All depended on having multiple-day records of cellphone usage, so that usage patterns could be identified. The criteria tested are:

1. The home zone is the zone with greatest total duration over the four-week observation period (i.e., the cellphone is located in the zone for a greater total amount of time than any other zone). Durations were computed for:
    a. All times during all days (total time).
    b. Weekends only.
    c. Nights only (using both 2100-0700 and 1900-0900 time periods).
2. Count the number of times that a zone is the origin of the first trip of the day plus the number of times it is the destination of the last trip of the day. The zone with the highest count of first/last trips is defined as the home zone. For this analysis, the "day" was defined as starting at 0400 on one calendar day and going to 0359 the next calendar day.
3. Count the number of times that a zone is the trip destination. The zone with highest count is labelled as the home zone. As with durations, this count was computed for all days/times, weekends only and nights only (for 2100-0700 and 1900-0900 time periods).

In all cases, a minimum number of observations were required for the criterion to be applied.

4.2.2 Work Locations

Similar to the logic used in identifying home locations, it was assumed that the second most-visited zone is very likely to be their usual place of work/school. It is not possible within the dataset to differentiate between work and school activities. Indeed, it is possible for a non-worker/student who visits a given location on a repeated, very frequent basis to be captured in this analysis. For simplicity of discussion, however, all locations identified in this analysis are labelled as “work”.

The rule for work location identification is simply the zone which is stayed at for the longest duration during the day (defined as 0800-1800), with a minimum of a one-hour stay. The rule was tested for all seven days of the week and for weekdays (Monday to Friday) only.

Cellphone users for whom no work location could be identified through this method are assumed to not be workers/students.

4.2.3 Origin-Destination (O-D) Trip Matrices

Given the imputation of home and work locations, all cellphone traces can be categorized by their origin and destination purposes of home, work and “other”. Given available data, it is not possible within this study to identify “other” activity types (shopping, recreation, etc.) in greater detail. Origin-destination (O-D) trip matrices can then be constructed (by time of day, as desired) for the following trip categories:

- Home-to-work.
- Home-to-other.
- Work-to-home.
- Work-to-other.
- Other-to-home.
- Other-to-work.
- Other-to-other.

4.3 Imputing Trip Mode

Figure 4.1 displays the overall approach to imputing trip mode for the cellphone traces. The first, key (and novel) step in this procedure is to convert MHMS O-D trips into pseudo cellphone traces; i.e., for each trip, convert it into a “trace” at the same level of spatial and temporal aggregation as the actual Antel traces. These MHMS pseudo-traces were then used as labelled input data to train the neural network model.

In order to construct these traces, the MHMS O-D trips must first be assigned to paths (routes) through the road and transit networks (depending on each trip’s chosen mode). To do this, road and transit networks for Montevideo where constructed within the Emme network modelling software system, as briefly described in Section 3.5. Maximum utility paths through the road and transit networks were found for auto (drive, passenger, taxi, motorcycle) and public transit trips, respectively. Active mode traces were constructed by taking the shortest distance paths through the road network at an assumed speed of 4 kph. Since it was not possible to calibrate Montevideo-specific assignment model parameters within this study, parameters from Toronto’s GTAModel V4.0 were used. It is not expected that the use of the Toronto parameters significantly affected the results.

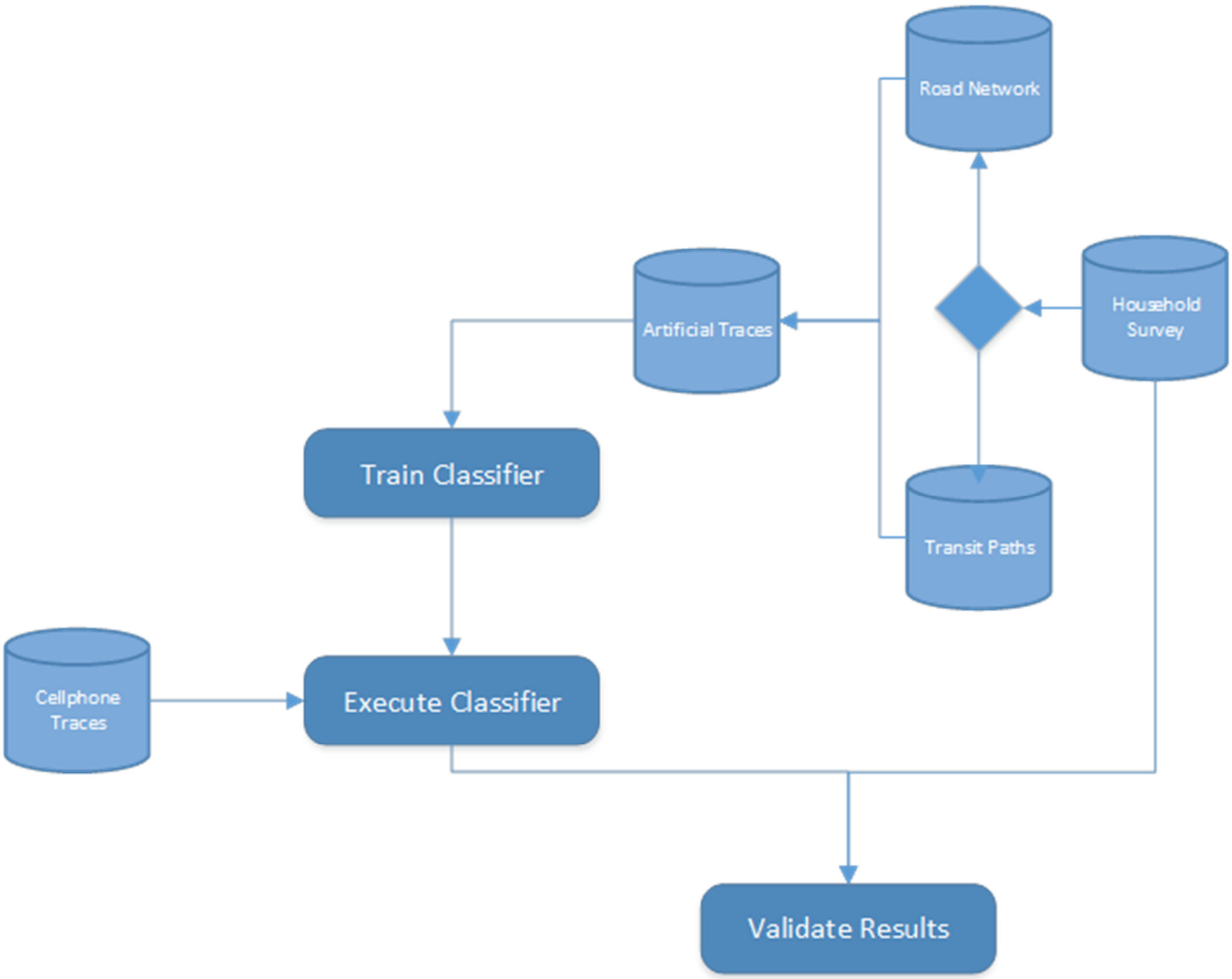

Figure 4.1: Analysis Approach

Emme is used for this purpose since the study team is very familiar with this network assignment software. In the longer run, any similar network modelling package could be used for this purpose. Emme is a commercial software package, developed and marketed by Inro, a Canadian software and consulting firm (https://www.inrosoftware.com/). It is used by transportation planning agencies world-wide, including in the Greater Toronto-Hamilton Area. It is also the software used in the development of the Asunción prototype travel demand model system (Miller, et al., 2017a,b).

In this model, the day is divided into five-minute segments. A trip is defined by (see Figure 4.2):

1. Whether it is occurring during a given five-minute segment (=1) or not (=0); i.e., whether the person is moving during this time segment.
2. The distance travelled during the five-minute segment.

The job of the neural net model is to determine that probability of each trip being made by the auto, transit or active modes, given the time of day, trip length (in time and distance) and distances travelled per time segment (approximate speed) for the trip.

	Time of Day					
	0:00	0:05	0:10 …		23:50	23:55
Distance	0	0	0	0.002	0.002	0
Active	0	0	0	1	1	0

Figure 4.2: Trip Trace Representation

The neural net classifier model developed in this study consisted of three hidden layers, with 400 neurons per layer. As illustrated in Figure 4.3, each neuron receives input signals from all the neurons from the layer above. In the case of the top hidden layer, these are the values of the feature inputs (the attributes of each trace that are used to impute its mode). These signals are combined in a weighted score (equation 4.1), which is then transformed by an *activation function* (equation 4.2) that generates the neuron's output signal, which is then transmitted to the next lower level's neurons.

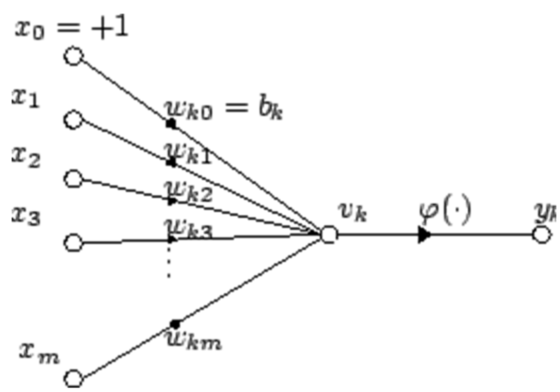

Figure 4.3: Inputs & Outputs from a Single Neuron
(Source: https://en.wikipedia.org/wiki/File:Artificial_neuron.png)

$$Z_k = b_k + \Sigma_i w_{ki}x_i \qquad (4.1)$$

where for neuron k:
x_i = Output signal from upper-level neuron i
w_{ki} = Weight attached to the signal from upper-level neuron i to lower-level neuron k
b_k = Bias term for neuron k
Z_k = Weighted input signal to neuron k

$$y_k = \varphi(Z_k) \qquad (4.2)$$

where:
y_k = Output signal from neuron k
$\varphi(\ )$ = Activation function for neuron k

Weights were randomly initialized from a normal distribution. Backpropagation combined with stochastic gradient descent (SGD) was used to update the weights in each iteration of the training session. Weights are chosen to maximize a cross-entropy (effectively a log-likelihood) function (equation 4.3):

$$H(p,q) = - \Sigma_t \Sigma_m p_t(m)\log(q_t(m) \qquad (4.3)$$

Where:

t	=	Trip t
m	=	Mode m (auto, transit, active)
$p_t(m)$	=	1 of mode m is used for trip t; = 0 otherwise
$q_t(m)$	=	Predicted probability of mode m being used for trip t

A linear rectifier activation function (equation 4.4) is used for the hidden layer neurons.

$$\varphi(Z) = \begin{cases} 0 \text{ if } Z < 0 \\ Z \text{ if } Z \geq 0 \end{cases} \qquad (4.4)$$

where Z is the weighted sum of the inputs to the neuron and φ(Z) is the neuron's output signal.

A softmax activation function (equation 4.5) is used for the output layer, in order to generate probabilities to assign to the three modes. Note that the softmax activation function is effectively a logit model, where $V_{tm,}$ the "systematic utility" function in the logit model, is the weighted input from the final hidden layer for mode m.

$$q_t(m) = \frac{e^{V_{tm}}}{\Sigma_{m'} e^{V_{tm}}} \qquad (4.5)$$

Chapter 5
Analysis of Cellphone Traces: Results

5.1 Introduction

This chapter presents the results of the analysis of Antel cellphone traces using the methods described in Chapter 4. Section 5.2 presents the results of imputing home and work locations, while Section 5.3 presents the trip mode imputation results.

5.2 Home & Work Locations

Table 5.1 presents the results obtained using the various criteria for imputing home locations described in Section 4.2. The first column of the table describes the rule. The second shows the percentage of cellphone users for which a home zone could be imputed using the given rule. The third column shows the correlation of the spatial distribution of population residential locations imputed by the cellphone data (using the given rule) with 2011 Census data. As can be seen from the table the total duration spent in the zone during the weekends gives the highest correlation with census data.

Table 5.1: Home Location Imputation Results

Type of Algorithm	**% of users home identified**	**Correlation with census home**
most stayed duration	100.0%	0.525
most stayed duration during night time (19 to 9)	90.0%	0.545
most stayed duration during night time (21 to 7)	74.9%	0.555
most stayed duration during weekends	86.4%	0.572
most first trips originated and lasted trips destined	43.3%	0.390
most trip destined	76.7%	0.436
most trip destined during night time (19 to 7)	56.2%	0.488
most trip destined during weekend	52.8%	0.497

Figure 5.1 shows the distribution of residential population as imputed for the cellphone users in the sample.

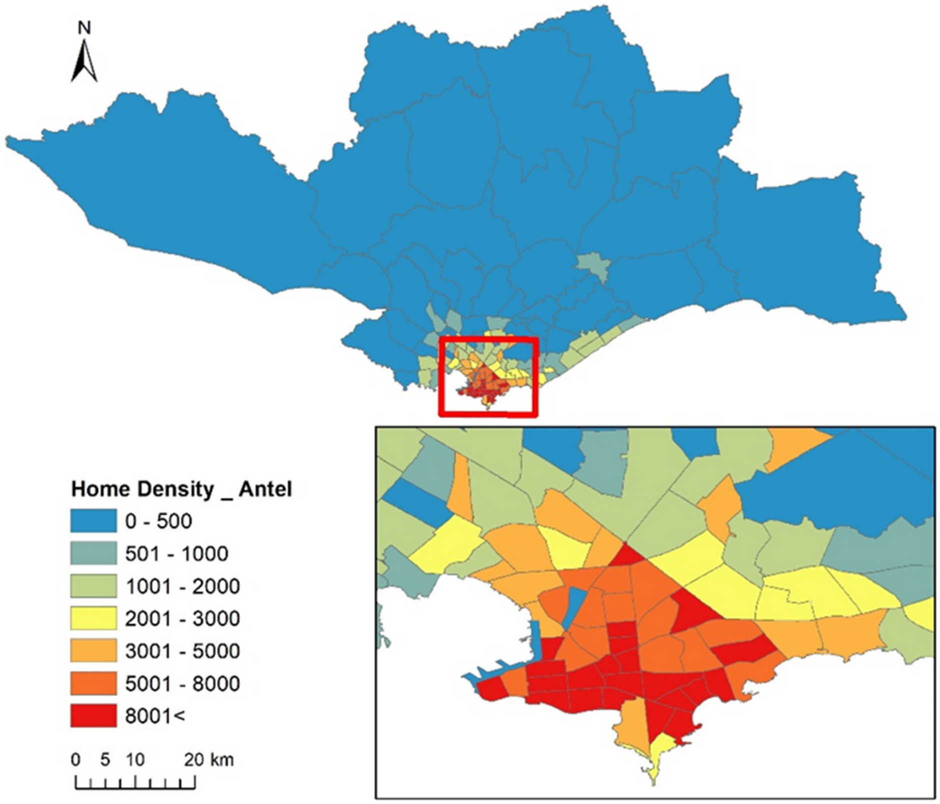

Figure 5.1: Imputed Distribution of Cellphone Users Residential Locations

Table 5.2 similarly presents the results obtained for imputing work locations using the two criteria tested, both of which give very similar results, but with the weekday-only criterion generating a slightly higher correlation with Census data. Figure 5.3 displays the imputed distribution of jobs based on the imputed cellphone work locations.

Table 5.2: Work Location Imputation Results

Type of Algorithm	% of users home identified	Correlation with census jobs
most stayed duration during day time (8 to 18), minimum of 1 hour stay	68%	0.471
most stayed duration during day time during weekdays (8 to 18), minimum of 1 hour stay	66%	0.474

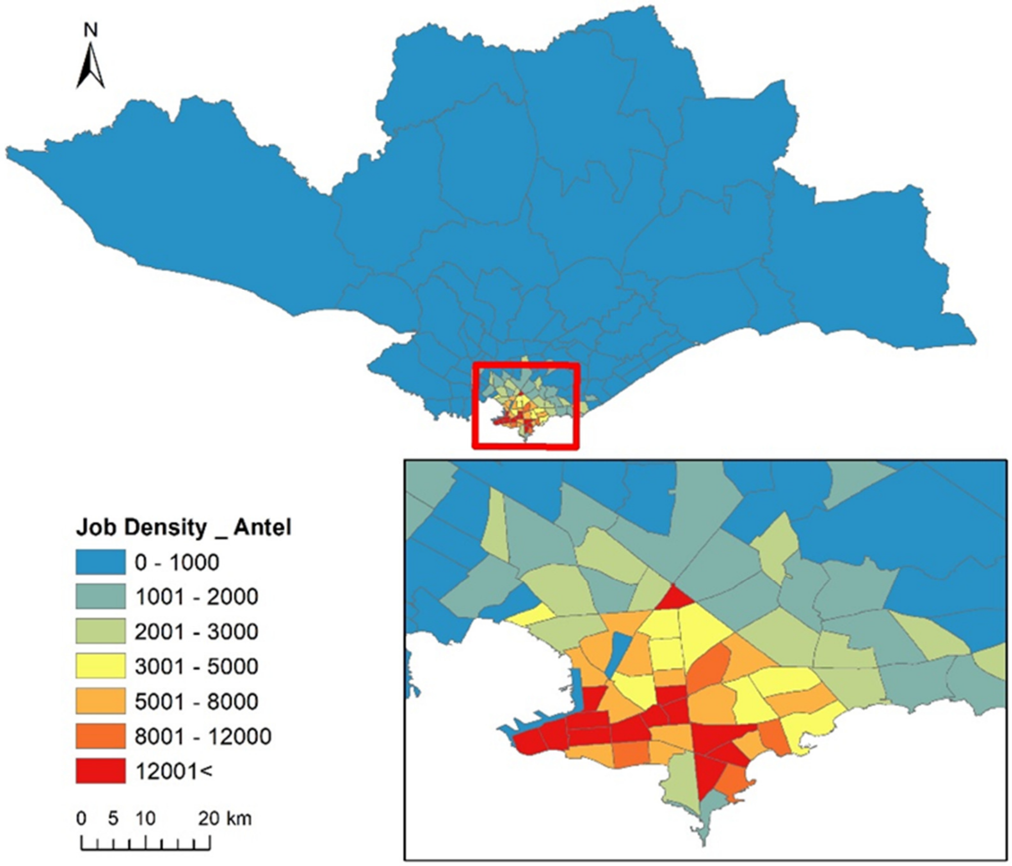

Figure 5.2: Imputed Distribution of Cellphone Users Job Locations

Once home and work locations have been identified, trips to/from home and work can be extracted from the overall trip set. Figures 5.3 and 5.4 display cellphone trips to and from home by start time and arrival time, respectively, compared to MHMS trips. As with total cellphone trips, it is seen that the time of day correspondence is good, except for the unexplained deviation in morning peak period trips.

Figures 5.5 and 5.6 display similar plots for trips originating and destined to work. Again, the distribution for work trip origins corresponds well to MHMS data, while trip arrivals at work appear to be low in the morning peak period.

Finally, Figures 5.7 and 5.8 bring home and work places together to display the time of day distribution (by trip start time) for home-to-work and work-to-home trips, respectively, with similar findings to those discussed above.

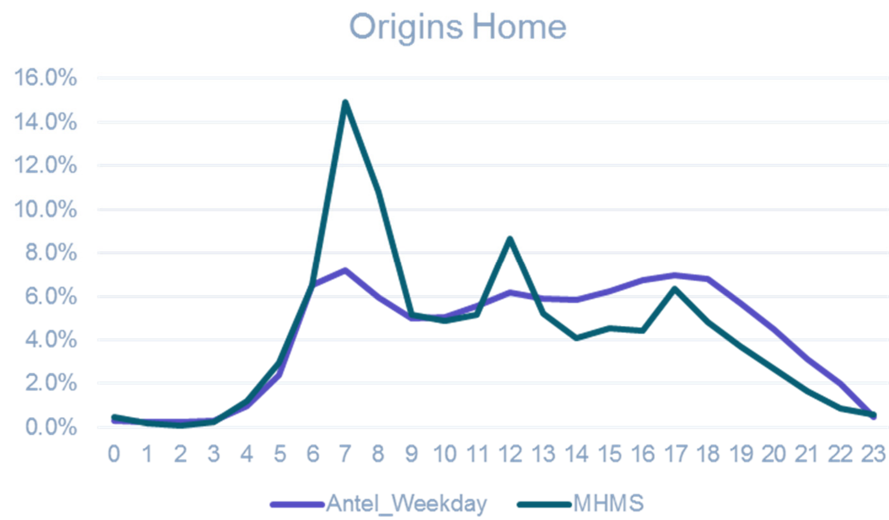

Figure 5.3: Cellphone vs. MHMS Trip Start Times, Home as Origin

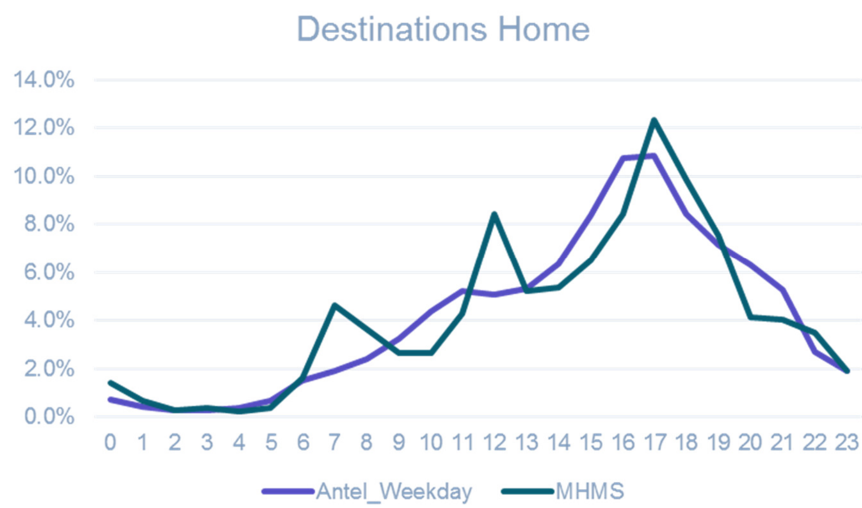

Figure 5.4: Cellphone vs. MHMS Trip Arrival Times, Home as Destination

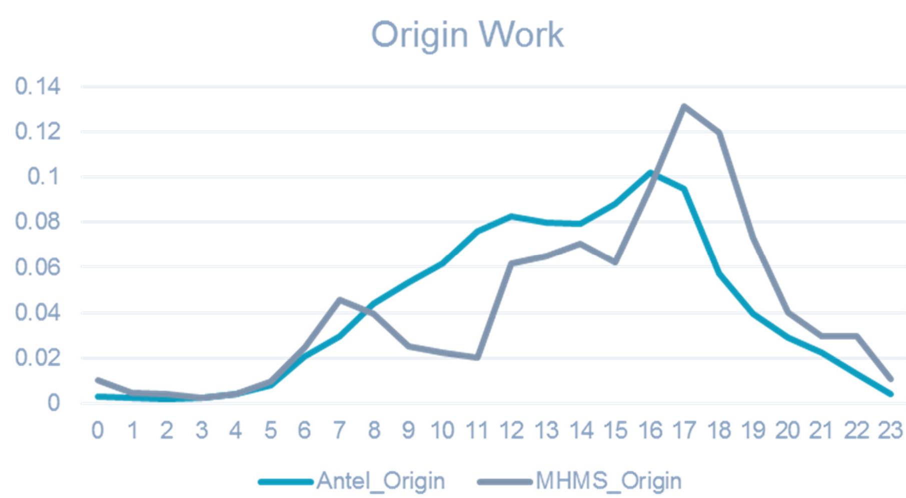

Figure 5.5 Cellphone vs. MHMS Trip Start Times, Work as Origin

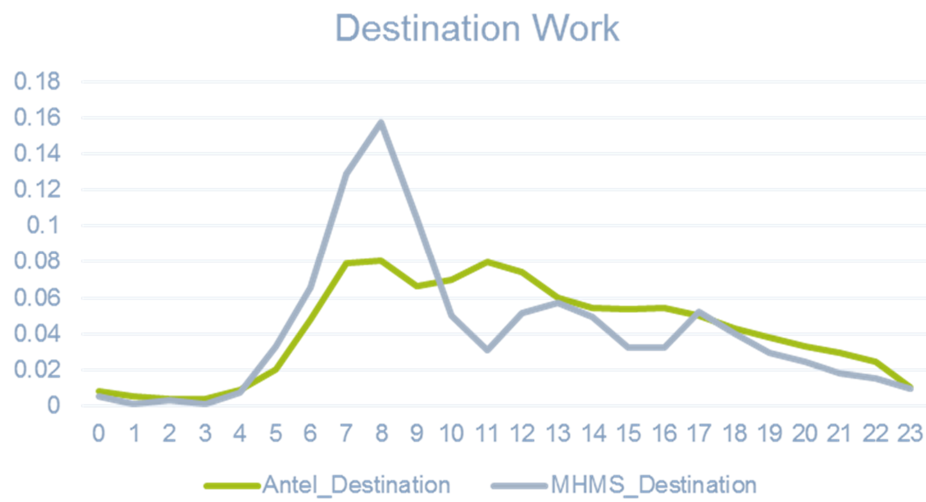

Figure 5.6: Cellphone vs. MHMS Trip Arrival Times, Work as Destination

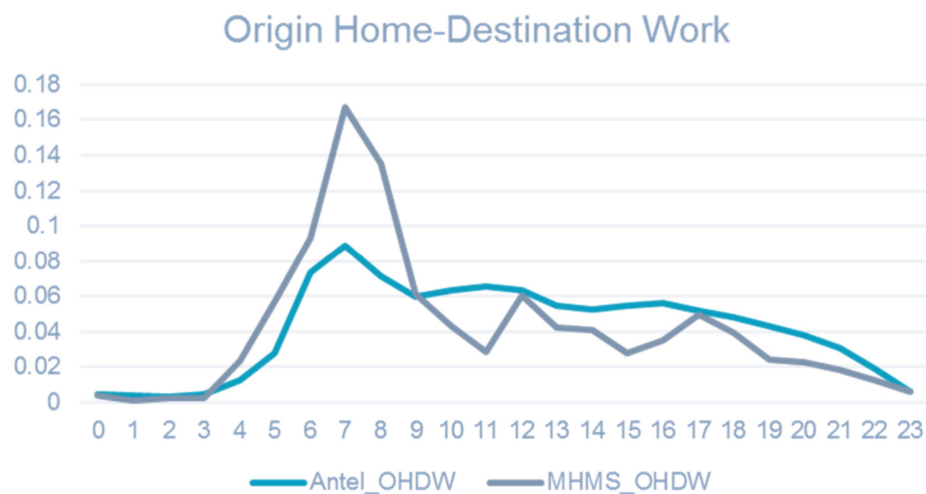

Figure 5.7: Cellphone vs. MHMS Home-to-Work Trip Start Times

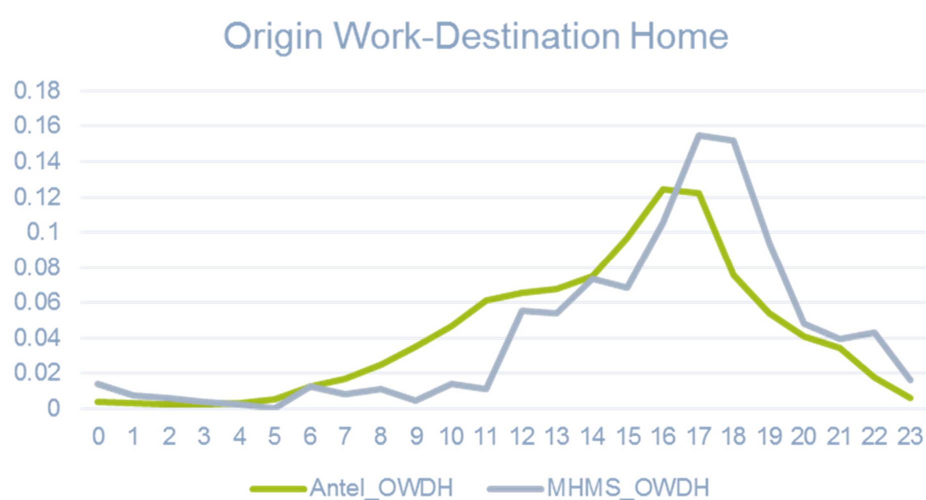

Figure 5.8: Cellphone vs. MHMS Work-to-Home Trip Start Times

5.3 Cellphone Trip Mode

5.3.1 Introduction

Three modes were imputed for cellphone trips in this analysis:

- ***Auto:*** Any trip made in a passenger car or equivalent (e.g., a light truck). This includes auto-drivers and auto passengers in private cars and taxi/Uber passengers. Note that some double-counting may occur in the dataset, given that some taxi/Uber drivers may also be included in the dataset. Their trips while carrying passengers should not be included in the trip count. It may be possible to identify these drivers in the dataset and delete them from the mode imputation analysis, but this has not been attempted in this study.
- ***Public transit:*** Any trip using the STM bus system.
- ***Active travel:*** Any trip all-way from origin to destination by walking or bicycle modes. Given the spatial precision of the cellphone data it is not possible to distinguish between walk and bicycle trips with any confidence.

Sub-section 5.3.2 presents the results of neural network model training using MHMS pseudo-traces. Sub-section 5.3.3 then presents the results of the applying the trained neural network to the full cellphone sample.

5.3.2 Neural Net Training Results

1,703 MHMS pseudo-traces were constructed. 50% (5,820 traces) were used to train the neural net model, while the other 50% (5,883 traces) were used to validate the trained model.

An excellent fit of 98% correct predictions by the trained model on its training set was achieved. When the trained model was applied to the MHMS validation set a similarly excellent fit of 86.9% was achieved, as shown in Table 5.3.

Table 5.3: Neural Net Validation Using the MHMS Sample

	Observed Modes		
Predicted	Auto	Transit	Active
Auto	33.65%	1.08%	2.54%
Transit	1.76%	21.36%	1.92%
Active	2.91%	2.88%	31.91%

5.3.3 Cellphone Imputed Trip Mode Results

Appendix I contains plots of the spatial distribution within the Montevideo region of market mode shares for auto, transit and active travel modes as predicted by the neural network model for the cellphone traces. For each mode maps are shown for total 24-hour trips, trips by time period and all-day cellphone trips compared to the MHMS distributions. In all cases, the spatial distributions are plausible, with more auto trips occurring in rural and suburban locations and more transit and active trips occurring in more central locations. The appendix also displays plots of the spatial distributions of imputed cellphone transit trips by time of day versus the corresponding fare transaction distributions. While differences exist, the overall patterns are similar.

Table 5.4 compares the overall predicted cellphone mode shares with MHMS mode shares. Overall, auto trips are being very well-predicted by the neural net model, but transit trips appear to be over-predicted, with a corresponding under-prediction of active trips.

Table 5.4: Predicted Cellphone vs. Observed MHMS Aggregate Mode Shares

Cell Traces			MHMS (no intrazonal)		
Auto	Transit	Active	Auto	Transit	Active
4083414	5159537	2515295	4250	2992	4461
34.73%	43.88%	21.39%	36.32%	25.57%	38.12%

Figure 5.9 shows predicted cellphone and observed MHMS trips by mode and distance. One can note in the figure that short-distance (0-3km) are over-predicted for transit and under-predicted for active travel relative to MHMS data. Thus, it appears that the neural net model is somewhat "confused" in its attempts to differentiate between transit and active modes for short distances. This is a not particularly surprising result, given the size of the traffic zones used in the analysis and the very simple set of explanatory factors being used in the current model.

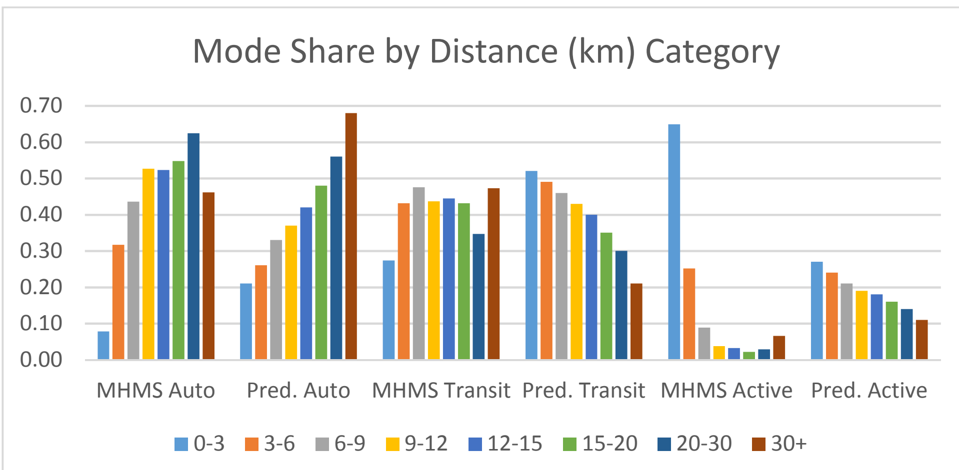

Figure 5.9: MHMS & Predicted Cellphone Trip Modes by Trip Distance

Predicted auto trips by distance show a relatively similar pattern relative to MHMS data, except that many more very long trips occur in the cellphone data – a very reasonable result.

Figure 5.10 compares cellphone and MHMS mode shares by time of day (trip start time). The patterns are generally similar, with the exception that the auto mode share is much higher in the cellphone data than MHMS for the very early hours of the morning. Given that the MHMS may well under-report trips during this time period, this is likely to be a reasonable result. In both datasets, transit mode shares are highest during the morning peak period – an expected result.

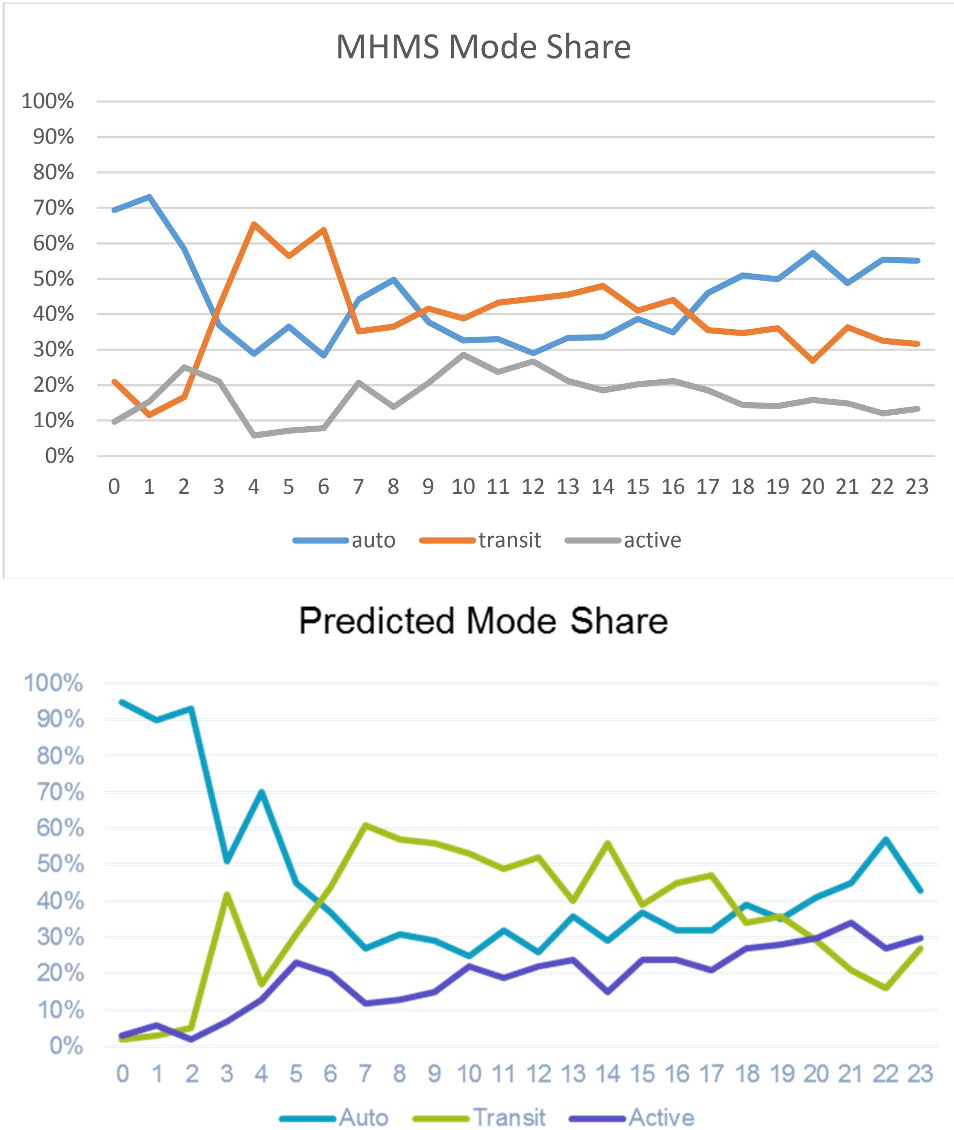

Figure 5.10: Observed MHMS & Predicted Cellphone Mode Shares by Trip Start Time

Finally, Figure 5.11 plots the estimated number of transit trips in the cellphone data by trip start time along with transit boardings. In order to weight up the cellphone sample, the simple weight defined by equation 5.1 was uniformly applied to all cellphone records:

Weight = 1.0 / [(Sample fraction of total Antel customers * (Antel cellphone market share)]
= 1.0 / (0.40)*(0.53) = 4.72 (5.1)

The temporal distribution of the estimated cellphone transit trips compared to the smartcard boarding data is excellent. Consistent with the comparisons to MHMS data, it appears that total transit trips are being over-predicted somewhat. In addition to issues previously discussed concerning possible confusion in the neural net's classification of transit vs. active trips, it may be that a more sophisticated weighting of cellphone trips is required in order to properly factor up the sample observations to total population trips. This is an issue for further investigation.

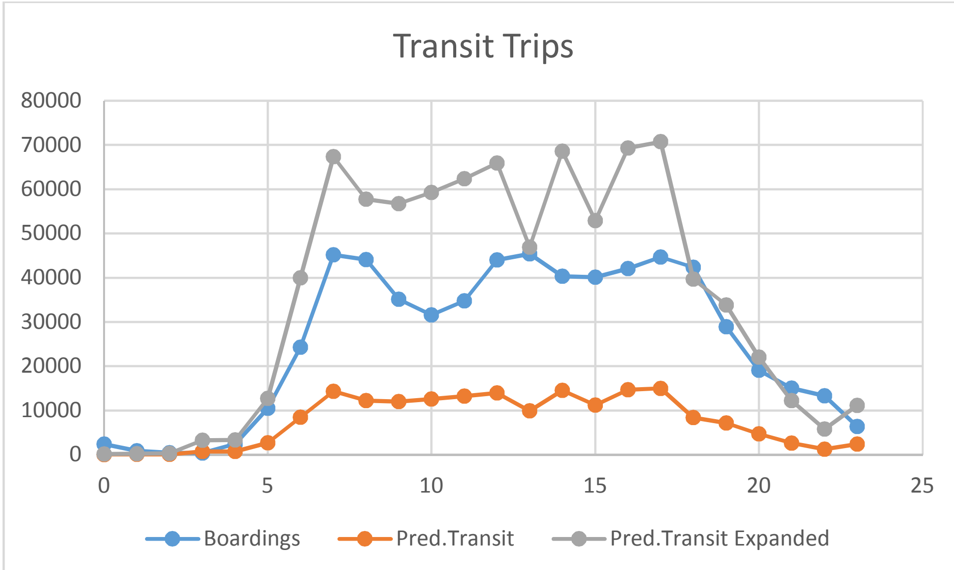

Figure 5.11: Transit Boardings & Weighted Cellphone Transit Trips by Trip Start Time

CHAPTER 6
SUMMARY & POSSIBLE NEXT STEPS

6.1 SUMMARY OF RESULTS

This study has developed a novel and powerful procedure for fusing cellphone trace data with traditional home interview survey records to create an enhanced representation of origin-destination trip-making by time of day, trip purpose and mode for the Montevideo region. This cross-sectional dataset is suitable for developing a travel demand forecasting model system for the Montevideo region (see Section 6.3 below). It can also support a wide variety of transportation policy related analyses.

The procedure developed was demonstrated using May, 2018 cellphone and transit data, along with 2016 MHMS data. It can be used to generate similar datasets for additional periods of time providing that similar cellphone and transit data are available for these time periods. Thus, a time-series of detailed cross-sectional "snapshots" of travel behaviour can be constructed over time (e.g., perhaps on an annual basis). Note that this depends on the assumption that the trip purpose and mode relationships established based on the 2016 MHMS data hold into future time periods. As one moves further into the future, this assumption, of course, will become somewhat more difficult to maintain. Thus, a need for at least an occasional small sample update of survey data over time that provides information concerning trip purposes and modes may well continue. As discussed in the next section, however, new methods for collecting such data may become increasingly viable in the near future that may be able to cost-effectively address this need.

6.2 POSSIBLE NEXT STEPS: DATA COLLECTION & FUSION METHODS

In addition to use of the procedure developed in this study by Montevideo agencies to update their travel behaviour database over time, a number of possible next steps exist that CAF may wish to consider with respect to further development of data collection methods in Latin American in support of their UMO program.

The first of these is to test the transferability of the Montevideo procedure to another (probably larger) Latin American urban region. Given the increasingly ubiquitous availability of both cellphone and smartcard data in many cities, it should be possible to apply the Montevideo procedure in any city for which such data can be obtained from both a cellular provider and the municipal transit authority. The availability of a recent home-interview survey (such as was the case in Montevideo) would be extremely advantageous, but other sources of supplementary data to update and test the procedure in the new urban region could be investigated as well. If the procedure could be tested and validated in at least one other urban region (or, preferably, a small handful of cities of different sizes, etc.) then the potential for developing a "universal" procedure is presumably very high.

Second, as noted in Chapter 4, data and time limitations within this study meant that it was unable to undertake the following three additional data imputation tasks:

1. Impute trip-maker socio-economic attributes (age, income), etc. for persons in either the cellphone or the transit transaction datasets.
2. Differentiate between work and school trips in the two datasets.
3. Estimate types of NWS trips (shopping, recreation, etc.)

Ideally, all three tasks are desirable in order to convert the fused trip dataset into a maximally useful database for activity-based travel model development. With more time and access to more detailed GPS-based "Point of Interest" (POI) data, all three should be doable. In particular, imputing person and household attributes given identified home zones in the Antel dataset,[6] combined with MHMS and census data is a feasible task to undertake. More detailed estimation of trip purposes (activity types) requires probabilistically assigning purposes based the land uses / building types in observed destination zones. This is not yet a well-developed process and is likely to be subject to a fair bit of modelling error. Nevertheless, this is a topic worth further investigation. One promising approach may be to redefine trip/activity purposes explicitly in terms of land use, rather than activity type per se. For example, in this approach a trip purpose might be "go to a shopping mall", rather than "shopping", "social" ("meet friends for lunch at the mall"), etc.

Third, this study has focussed on the use of cellphone and smartcard data as alternative data sources relative to traditional home-interview surveys. Two other promising data collection technologies, however, also exist: smartphone travel data collection apps, and web-based survey methods. These were not investigated in this study for several reasons, including:

- The much lower market penetration rates of smartphones and home computers within Montevideo relative to cellphones.
- Lack of project resources to investigate these technologies in a meaningful way.
- Although developing at a rapid rate, these technologies are not yet fully mature in terms of their capabilities.

Nevertheless, both technologies are, indeed, advancing rapidly, and both hold significant promise as complements to, or even substitutes for, traditional home survey methods. As illustrated within this study, home interview type data – with its complete information concerning trip and trip-maker attributes – are still extremely valuable for augmenting cellphone and smartcard data. If some combination of smartphone apps and web-based surveys can provide similar information cost-effectively for significant sub-samples of the population, then they may well be valuable additions to the multi-instrument data collection design which CAF is interested in developing.

UTTRI has been investigating both technologies extensively over the past few years, including extensive, systematic testing of smartphone apps, leading to detailed design recommendations for future app development (Harding, et al., 2016a, 2016b; Faghih-Imani, et al. 2018) and developing new, powerful software for multi-platform web-based surveys (Chung, et al., 2017).

[6] Imputing attributes for trip-makers in the transit transaction data will be subject to greater modelling error, since their home zones are also modelled, based on their transit stop locations, thereby introducing greater uncertainty concerning their actual home zones.

Testing of both technologies in a Latin American context would appear to be well worth the effort in terms of the lessons that could be learned and the eventual operational applications of these methods.

6.3 POSSIBLE NEXT STEPS: MODELLING

It is important to note that the data fusion procedure developed in this study is not a forecasting tool. Rather, it creates an enhanced representation of current travel conditions. To forecast future travel for strategic planning purposes still requires the development of a formal travel demand forecasting model system. What the procedure developed in this study does do, however, is create a significantly enhanced base dataset to support the development of such a forecasting model system.

The potential for developing an agent-based microsimulation travel demand model system for Montevideo similar to the SATA model developed for Asunción (Miller, et al. 2017a,b) has been briefly discussed in *Report 2* of this project (Miller, et al., 2017c). The fused travel behaviour dataset developed in this project can be used directly to develop the individual behavioural choice models (trip generation, distribution, mode choice) within the overall model system. In addition, the Emme road and transit networks developed in this project to generate the MHMS trip pseudo traces provide an excellent starting point for developing the "operation grade" networks needed for an operational travel demand model system. Thus, this project, combined with the lessons learned during the Asunción prototype model system development, has placed Montevideo in an excellent position to rapidly and cost-effectively develop a state-of-the-art travel demand forecasting system for the region. Such a model system could play an extremely useful role in a wide range of strategic transportation policy analyses, including transit network design; transit fare policies; assessment of major transportation infrastructure investments (road and transit); etc.

REFERENCES

Chung, B., S. Srikukenthiran, K.M.N. Habib & E.J. Miller (2017) *Development of a Web Survey Builder (TRAISI): Designing Household Travel Surveys for Data Accuracy and Reduced Response Burden*, Transportation Tomorrow Survey 2.0 research report, Toronto: University of Toronto Transportation Research Institute, November.

Encuesta de movilidad en el Área Metropolitana de Montevideo 2016, Principales resultados e indicadores, Montevideo: July, 2017.

Faghih-Imani, A., C. Harding, M. Taws, S. Srikukenthiran, K. Nurul Habib and E.J. Miller (2018) *Large-Scale Test of Smartphone Apps for Travel Data Collection*, Transportation Tomorrow Survey 2.0 research report, Toronto: University of Toronto Transportation Research Institute, August.

Faghih-Imani, A. and E.J. Miller (2018) *Preliminary Analysis of Montevideo Cellular Data Records, Report 4: iCity-South: Urban Informatics for Sustainable Metropolitan Growth in Latin America*, report to CAF, the Development Bank of Latin America, Toronto: University of Toronto Transportation Research Institute, January.

Harding, C., S. Srikukenthiran, Z. Zhang, K.M.N. Habib and E.J. Miller (2016a) *Smartphone App and Trace Processing Assessment, Volume 1*, Transportation Tomorrow Survey 2.0 research report, Toronto: University of Toronto Transportation Research Institute, November.

Harding, C., S. Srikukenthiran, Z. Zhang, K.M.N. Habib and E.J. Miller (2016b) *Smartphone App and Trace Processing Assessment, Volume 2*, Transportation Tomorrow Survey 2.0 research report, Toronto: University of Toronto Transportation Research Institute, November.

Miller, E.J., F.F. Calderón Peralvo, J. Vaughan and B. Yusuf (2017a) *SATA: Simulador de Actividad de Transporte de Asunción, Development of the SATA Prototype Volume I: Final Report*, report to CAF, Toronto: University of Toronto Transportation Research Institute.

Miller, E.J., F.F. Calderón Peralvo, J. Vaughan, B. Yusuf and A. Faghih-Imani (2017b) *SATA: Simulador de Actividad de Transporte de Asunción, Development of the SATA Prototype Volume II: Technical Appendices*, report to CAF, Toronto: University of Toronto Transportation Research Institute.

Miller, E.J. and K.N. Habib (2017) *The State of the Art in Urban Transportation Data Collection, Report 1: iCity-South: Urban Informatics for Sustainable Metropolitan Growth in Latin America*, report to CAF, the Development Bank of Latin America, Toronto: University of Toronto Transportation Research Institute, September.

Miller, E.J., K.N. Habib, M. Lee-Gosselin, C. Morency, M.J. Roorda and A.S. Shalaby (2012) *Changing Practices in Data Collection on the Movement of People, Final Report*, report to the Transportation Association of Canada, Île d'Orléans, Québec: Lee-Gosselin Associates Ltd.

Miller, E.J., C. Parada Hernandez and K.M.N. Habib (2017c) *Review of the Montevideo Home Mobility Survey, Report 2: iCity-South: Urban Informatics for Sustainable Metropolitan Growth in Latin America*, report to CAF, the Development Bank of Latin America, Toronto: University of Toronto Transportation Research Institute.

Parada Hernandez, C. (2018) *Analysis of Automated Fare Collection Data from Montevideo , Uruguay for Planning Purposes*, MASc thesis: Toronto: Department of Civil Engineering, University of Toronto.

Parada Hernandez, C. and E.J. Miller (2017) *Analysis of Montevideo Smartcard Data, Preliminary Report, Report 3: iCity-South: Urban Informatics for Sustainable Metropolitan Growth in Latin America*, report to CAF, the Development Bank of Latin America, Toronto: University of Toronto Transportation Research Institute, September.

Parada Hernandez, C. and E.J. Miller (2018) *Analysis of Montevideo Automated Fare Collection Data, Final Report, Report 5: iCity-South: Urban Informatics for Sustainable Metropolitan Growth in Latin America*, report to CAF, the Development Bank of Latin America, Toronto: University of Toronto Transportation Research Institute, June.

APPENDIX I: CELLPHONE TRIP MODE SHARES

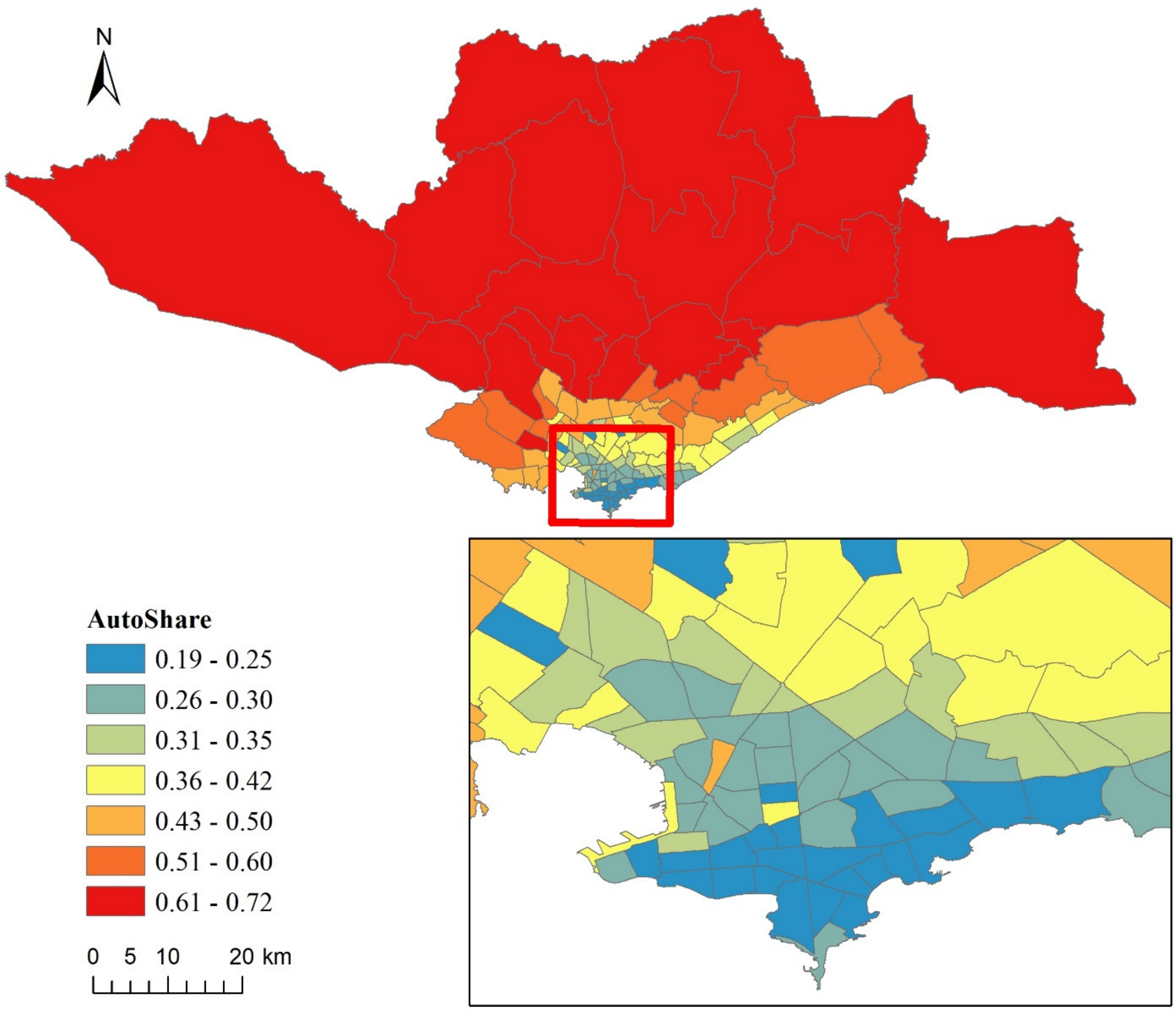

Figure I.1: Weekday Auto Mode Share, All-Day by Trip Origin

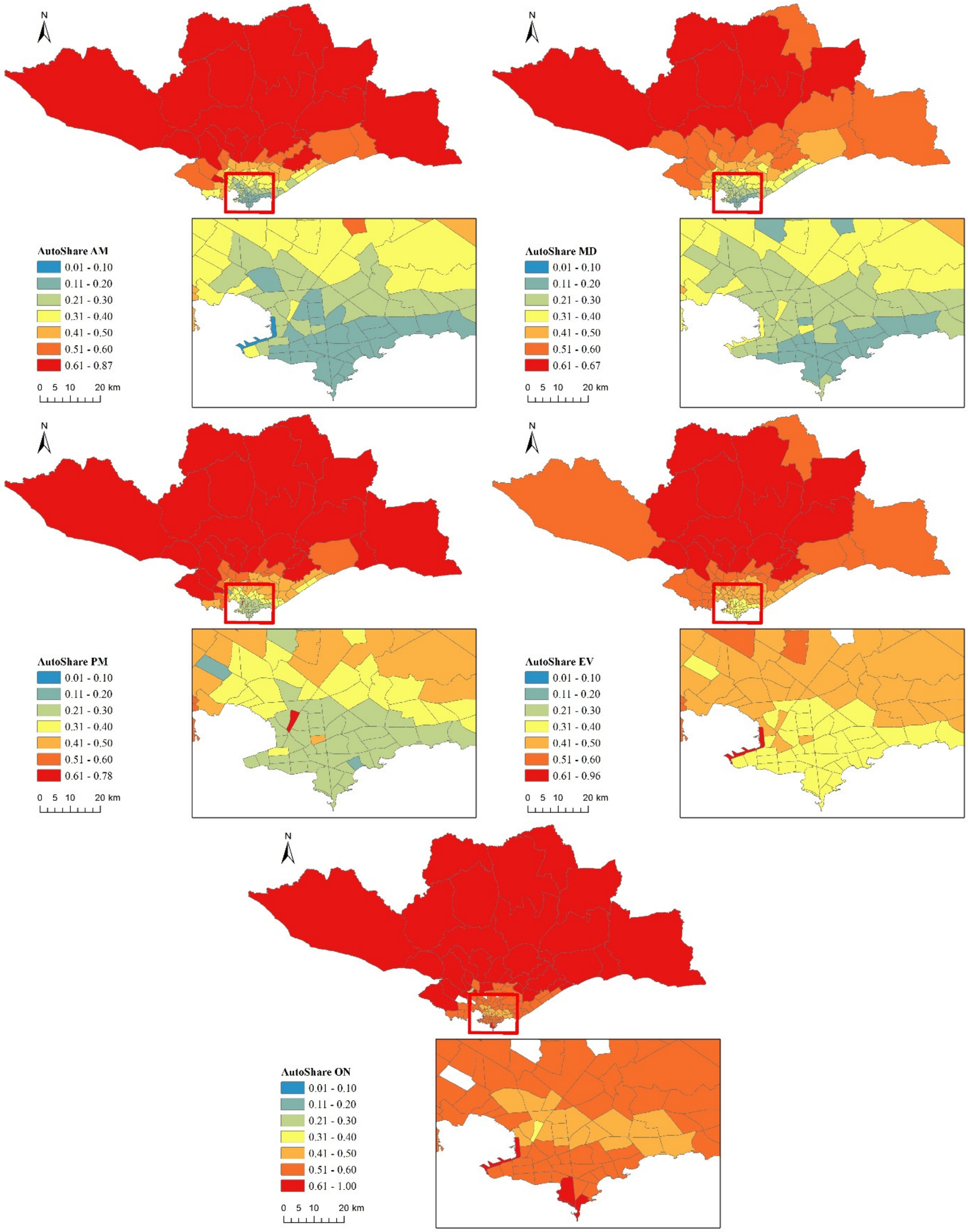

Figure I.2: Weekday Auto Mode Share, by Time of Day & Trip Origin

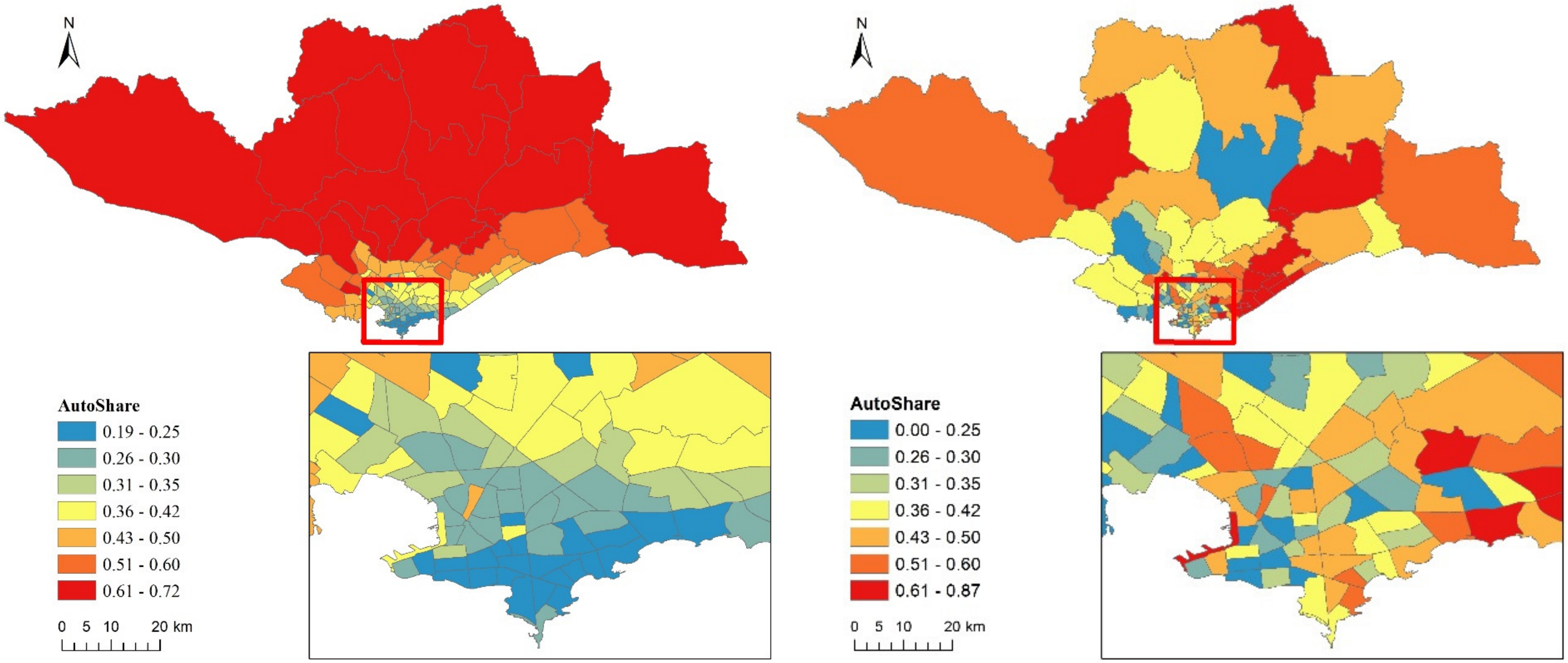

Figure I.3: Antel-MHMS comparison: Weekday Auto Mode Share, by Time of Day & Trip Origin

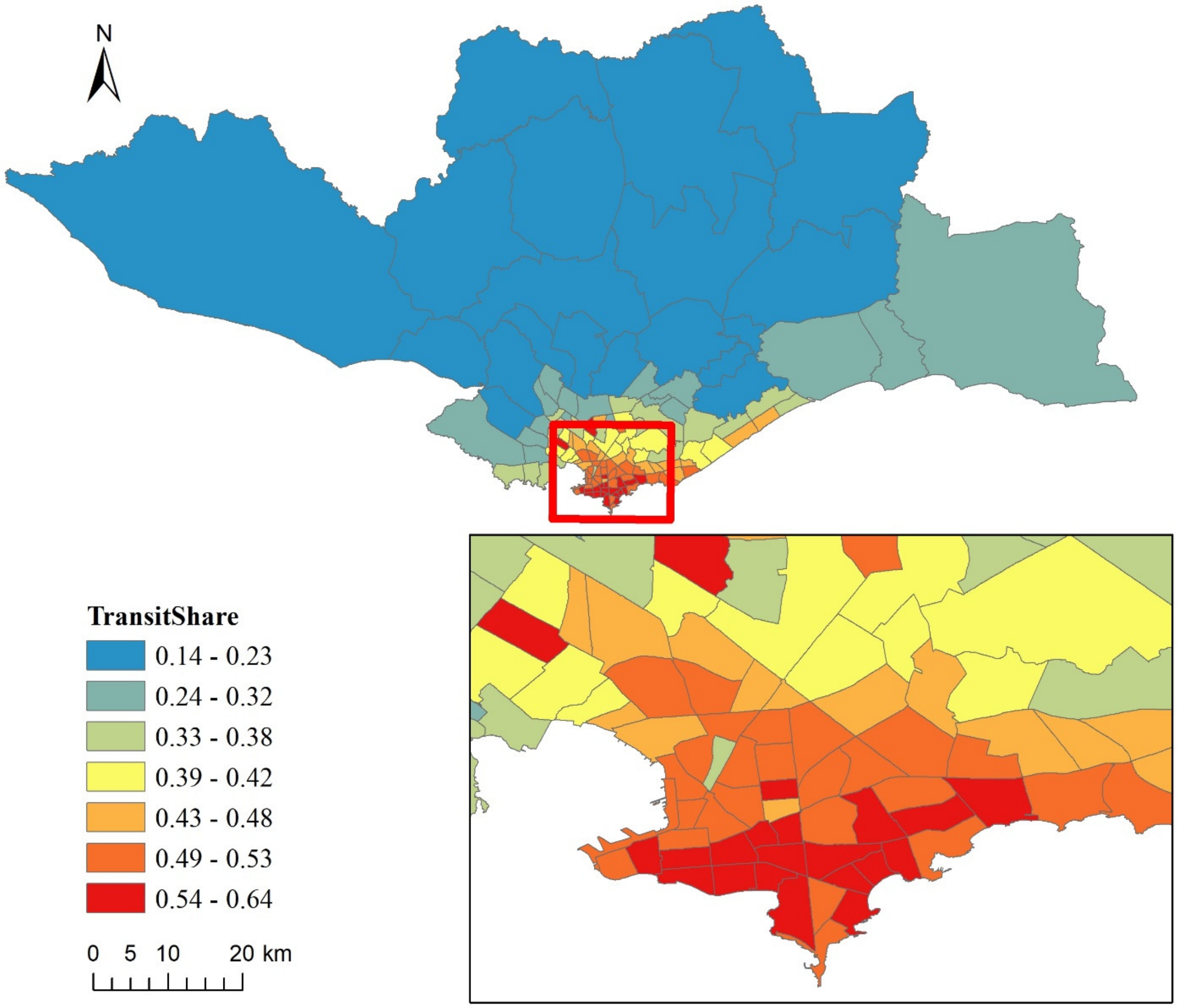

Figure I.4: Weekday Transit Mode Share, All-Day by Trip Origin

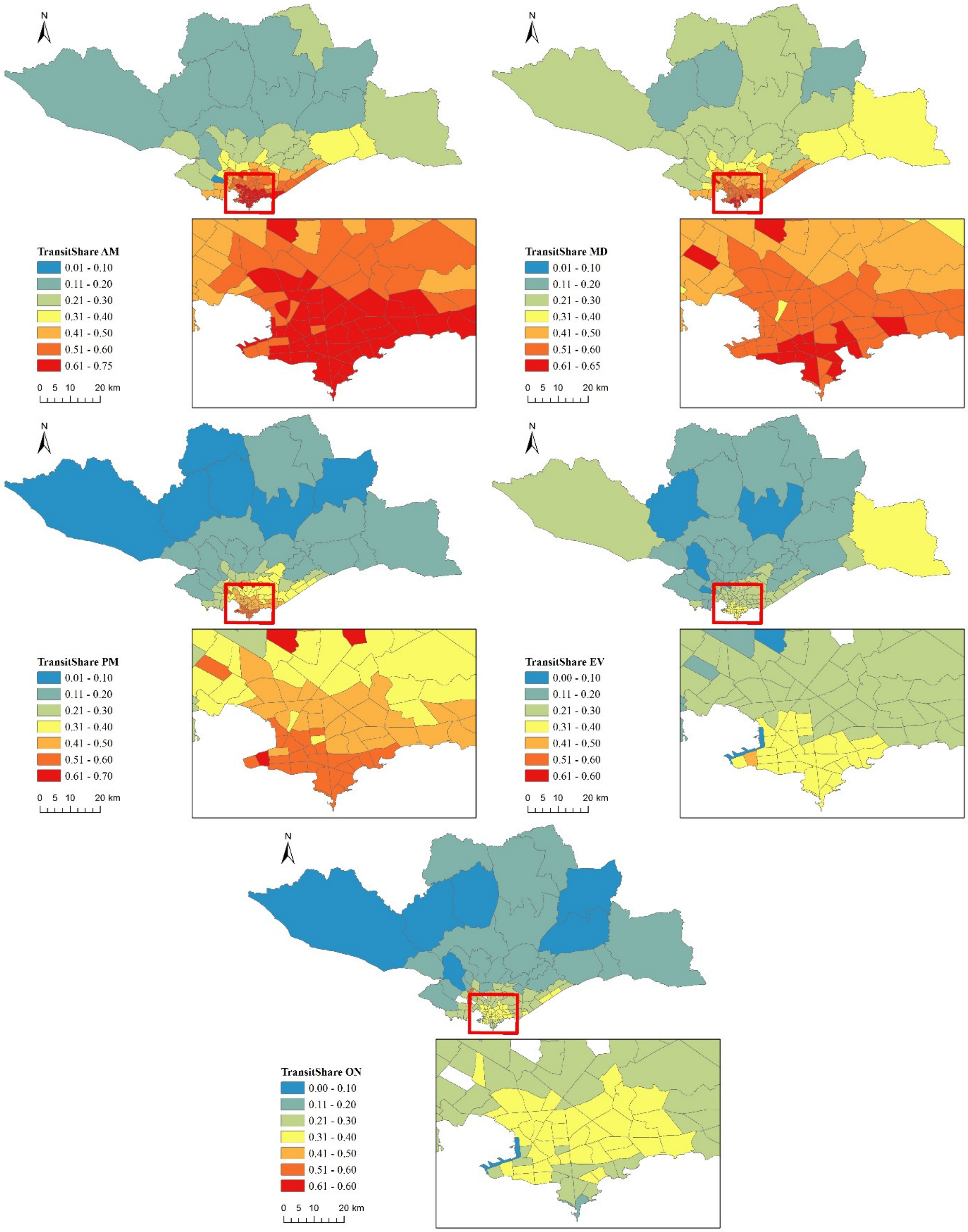

Figure I.5: Weekday Transit Mode Share, by Time of Day & Trip Origin

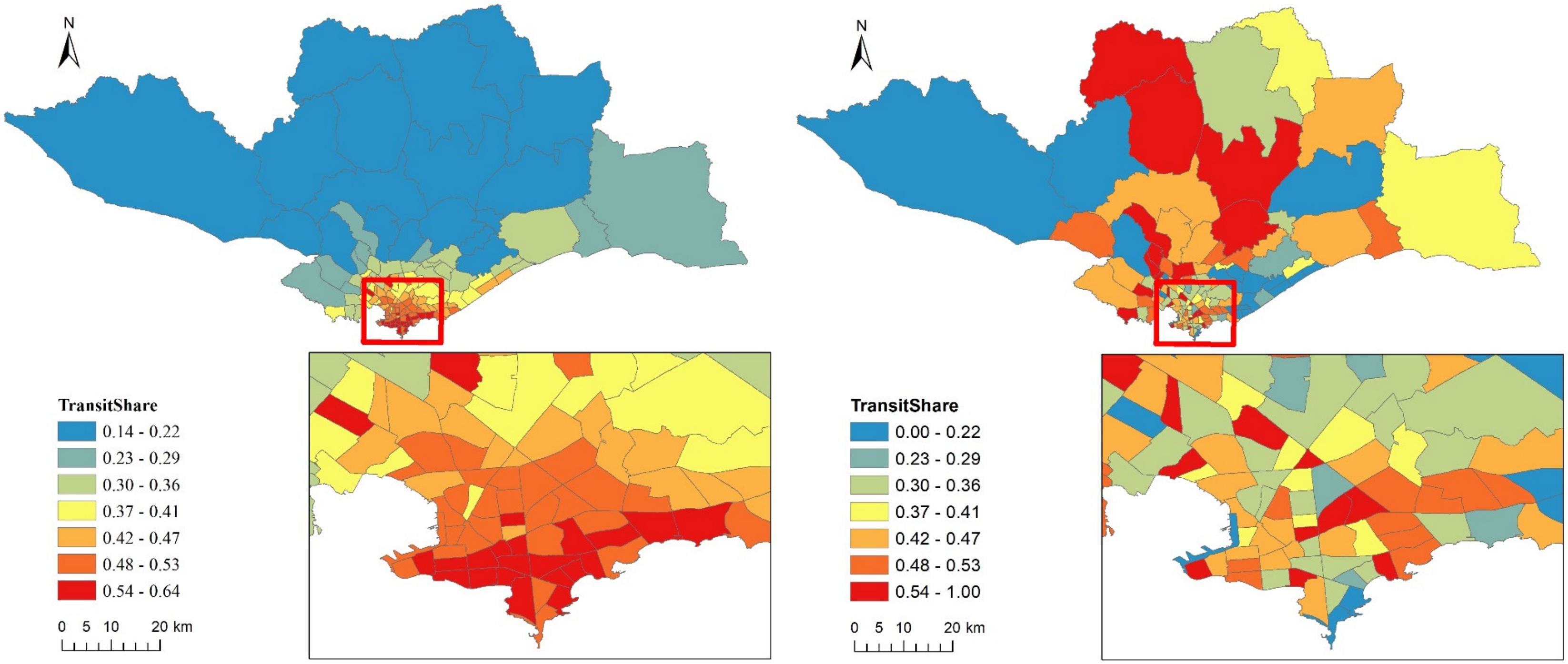

Figure I.3: Antel-MHMS comparison: Weekday Transit Mode Share, by Time of Day & Trip Origin

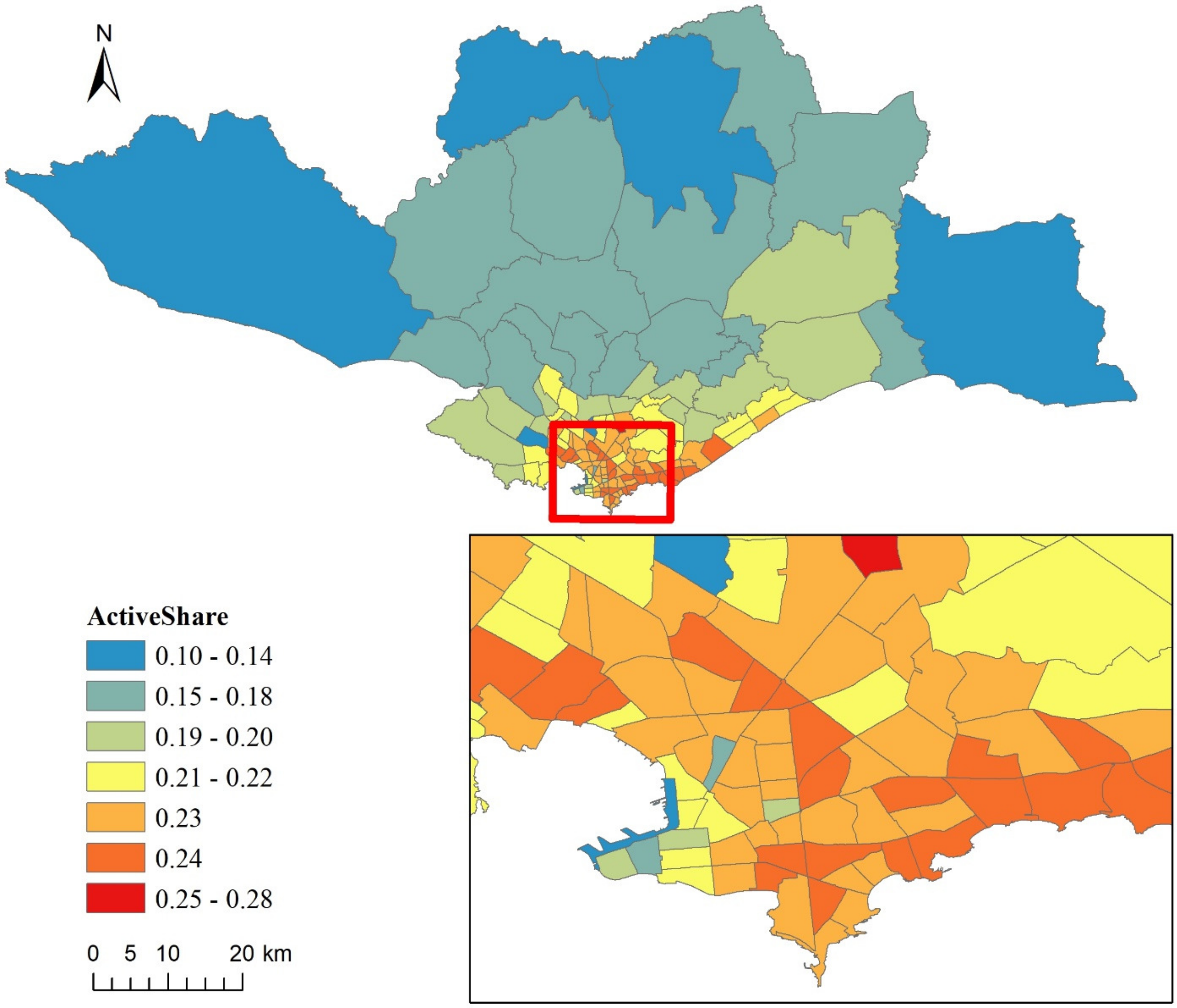

Figure I.7: Active Travel Mode Share, All-Day by Trip Origin

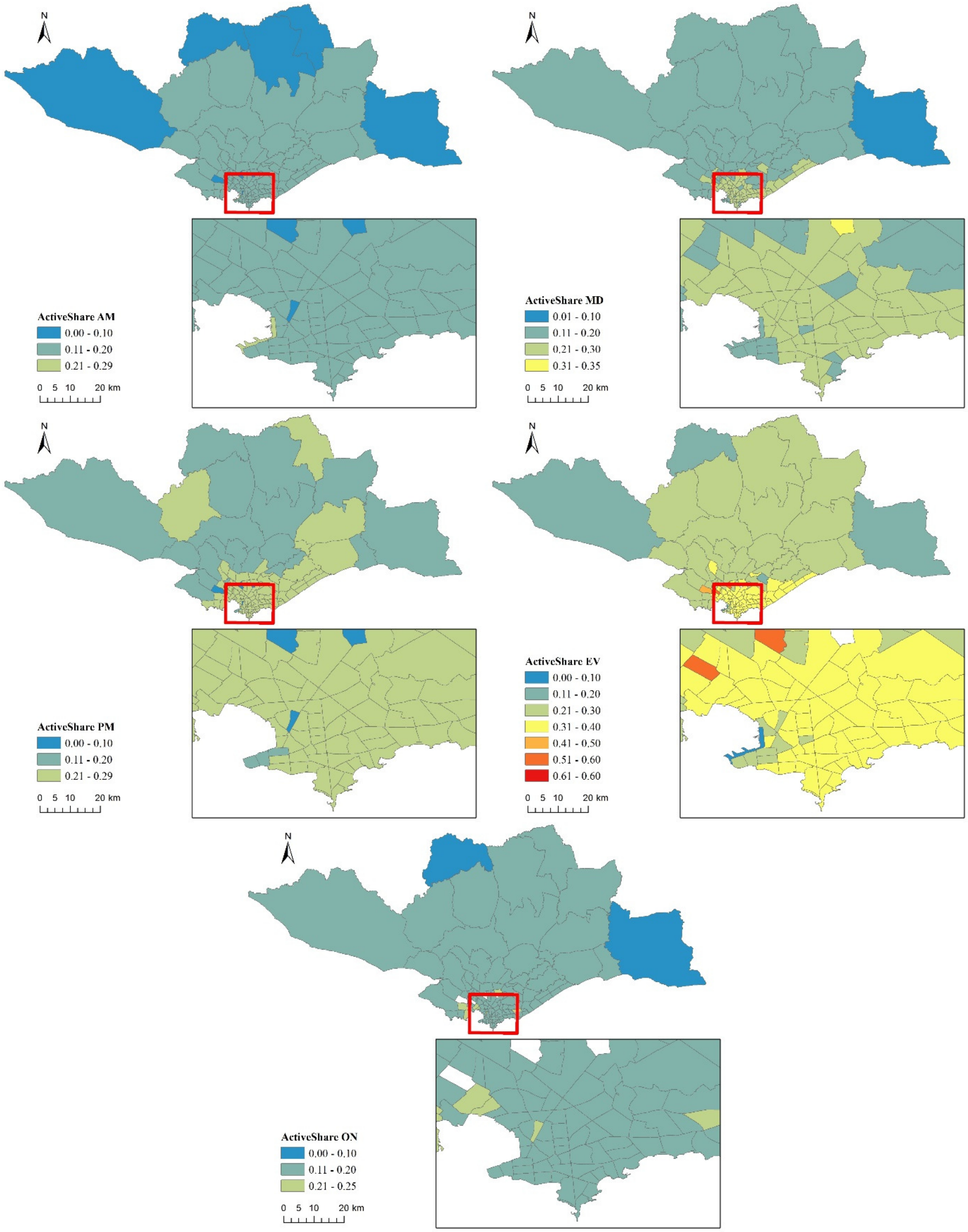

Figure I.8: Weekday Active Travel Mode Share, by Time of Day & Trip Origin

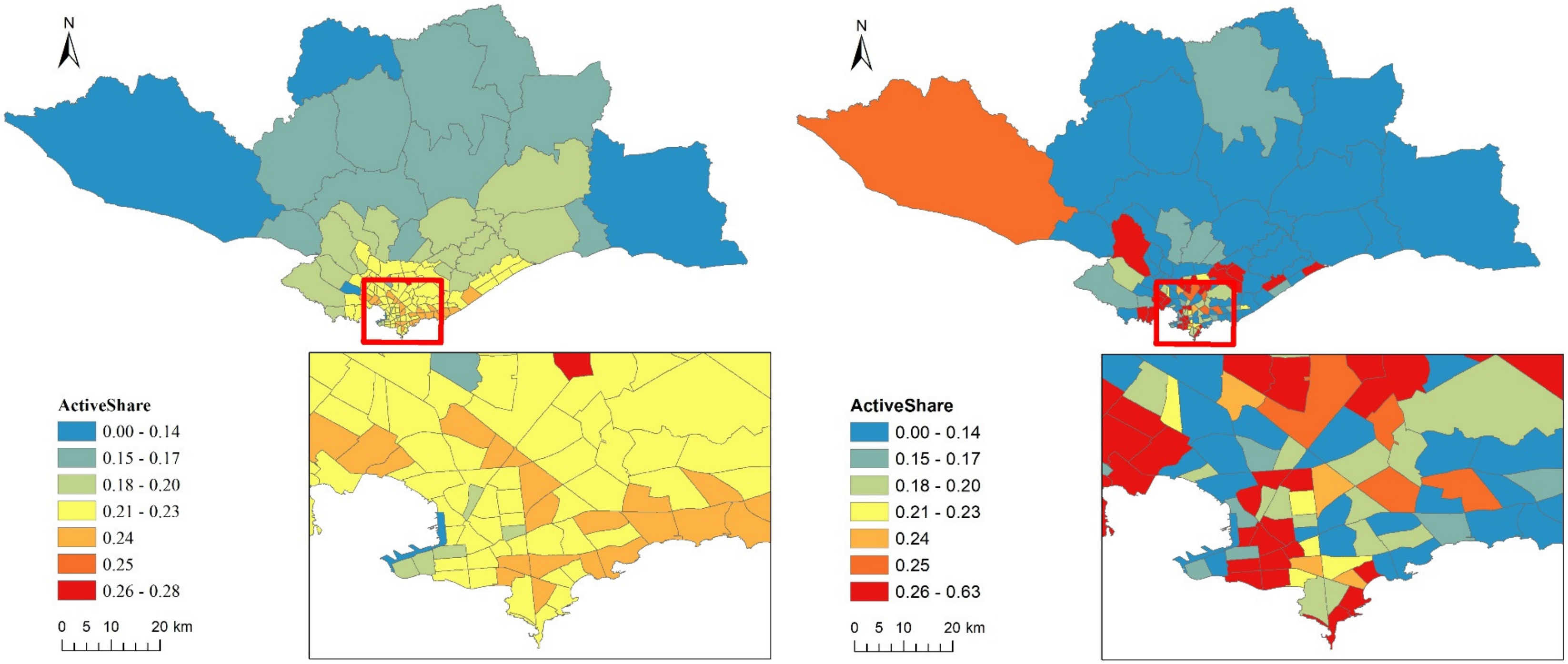

Figure I.3: Antel-MHMS comparison: Weekday Active Mode Share, by Time of Day & Trip Origin

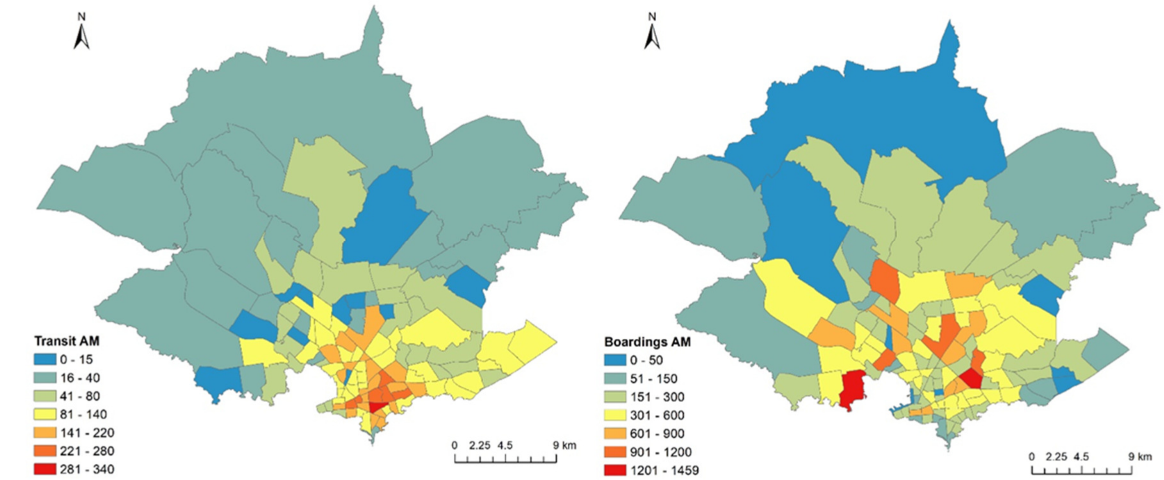

Figure I.10: Cellphone Transit Trips (left) vs. Smartcard Trips (right), AM Peak Period

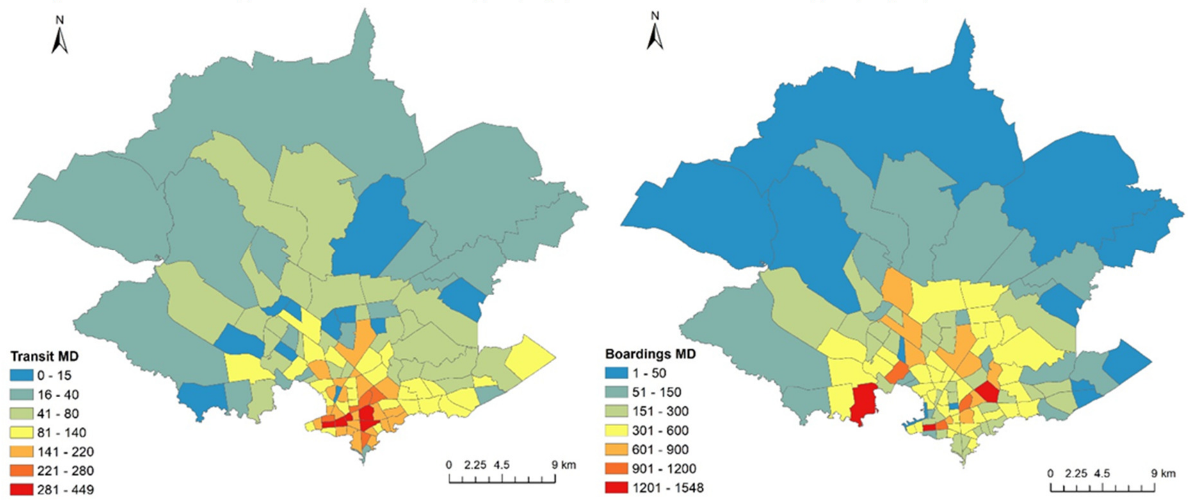

Figure I.11: Cellphone Transit Trips (left) vs. Smartcard Trips (right), Mid-Day Period

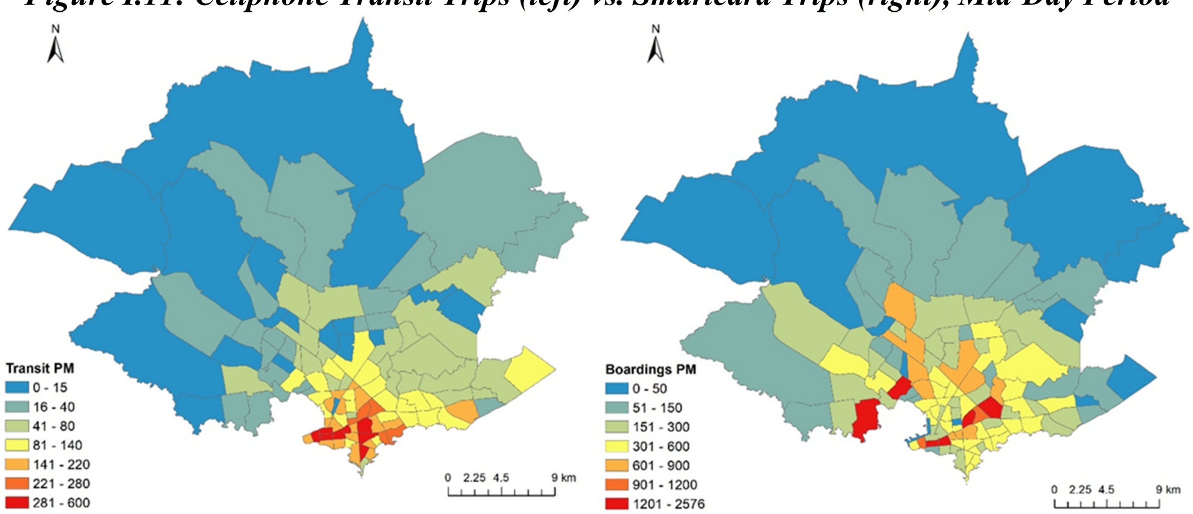

Figure I.12: Cellphone Transit Trips (left) vs. Smartcard Trips (right), PM Peak Period