Passive Data for Transportation Modelling:
imputing mode and purpose information from cellphone traces

UT-ITE Seminar
February 8, 2019

Ahmadreza Faghih Imani, Ph.D.
Postdoctoral Fellow
Dept. of Civil & Mineral Engineering
University of Toronto

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

UTTRI

Passive Data for Transportation Modelling: imputing mode and purpose information from cellphone traces

- **Acknowledgements:**
- **James Vaughan, Bilal Yusuf, Prof. Eric Miller**

- This study was part of iCity-South project funded by CAF.
- CAF: Nicolas Estupinan and Andres Alcala
- Diego Hernandez, Universidad Catolica del Uruguay
- Antonio Mauttone, Universidad de la Republica de Uruguay
- Antel: Juan Pablo Pignataro, Julio Ciambelli and Sebastian Lamas
- Intendencia de Montevideo: Richard Delgado, Nestor Sosa and Justo Onandi

UTTRI

CAF BANCO DE DESARROLLO DE AMÉRICA LATINA

Más oportunidades, un mejor futuro.

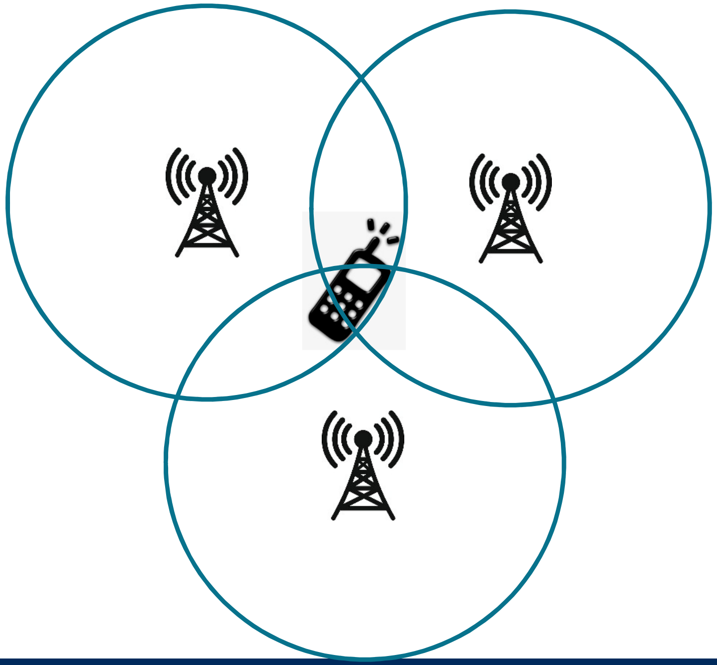
Introduction

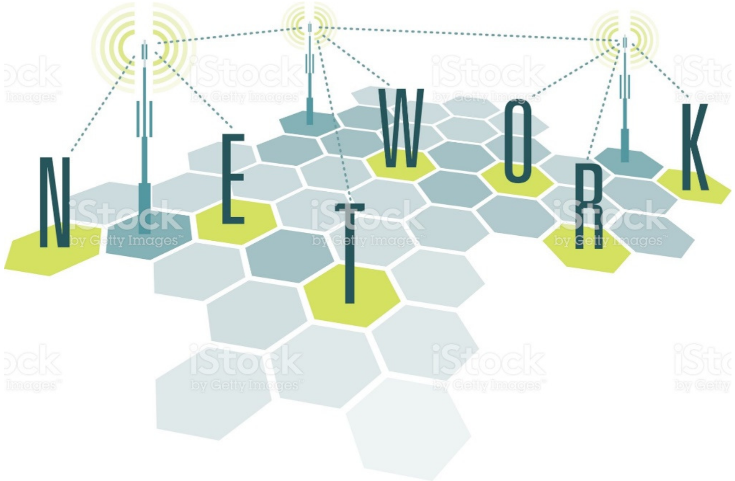
- What is passive data?
- Respondents don’t need to actively participate and report the information.
  - Typically in nature present “revealed preference”

- Examples:
- Public transit smartcard transactions or cellular network data

Cellular Data

- It is possible to identify people movements using cellular network data from the base transceiver stations (BTSs).



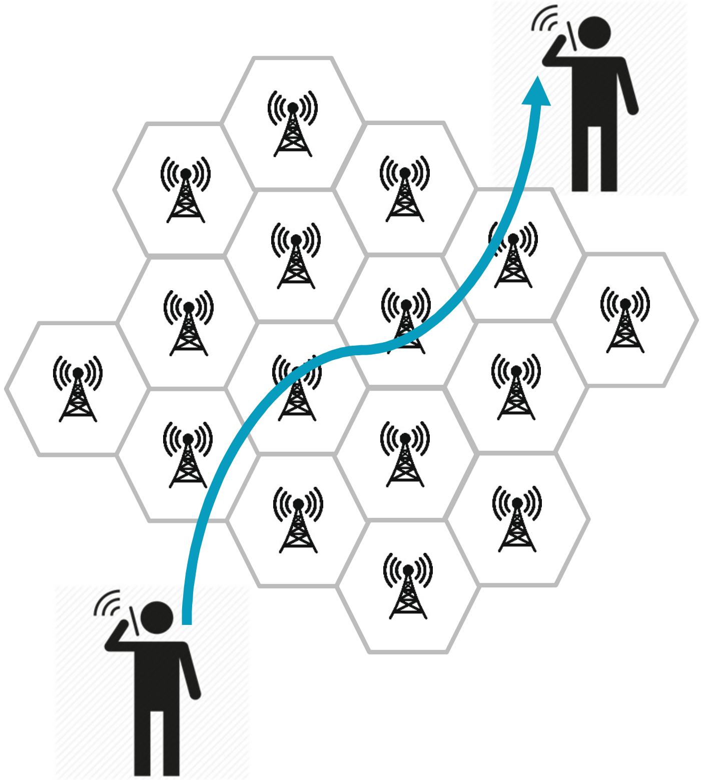


Cellular Data

- Several cell to infrastructure connection events:
  - *handovers* (HO)
  - *call detail records* (CDR)
  - *location updates* (LU)
    - HO and CDR provide data of communication events such as calls or SMSs and LU notify the cellular network when a cellphone moves from one BTS to another.

→Results in time-space traces of movements of cellphones

→can be used to impute travel movements by users.



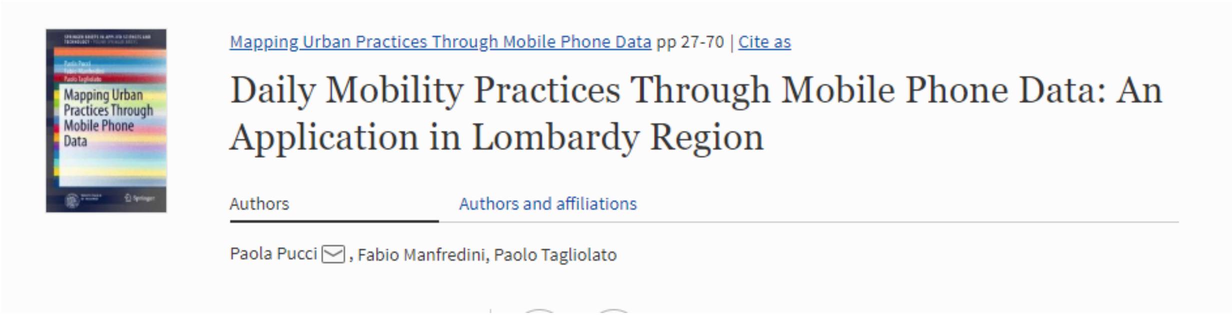
Cellular Data

Advantages
Use of existing data
Large sample size
Large coverage
Long time period
No interaction with users

Disadvantages
No individual information
Depends on cell phone market penetration
Multiple counting of a person with multiple devices
Lack of spatial precision

Earlier Studies on Cellular Data

- Mobility Patterns

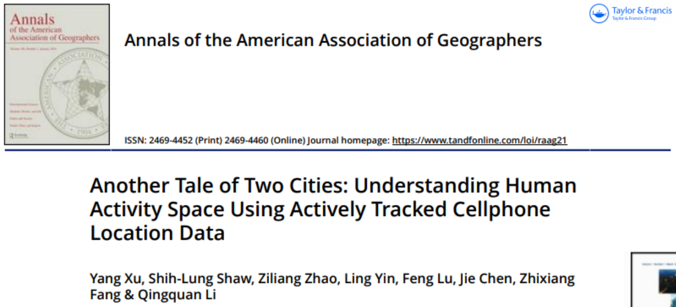


Mapping Urban Practices Through Mobile Phone Data pp 27-70 | Cite as

Daily Mobility Practices Through Mobile Phone Data: An Application in Lombardy Region

Authors   Authors and affiliations

Paola Pucci, Fabio Manfredini, Paolo Tagliolato

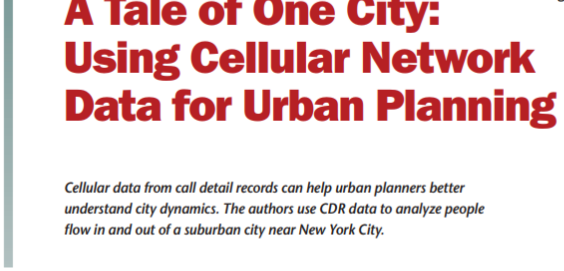


Annals of the American Association of Geographers

ISSN: 2469-4452 (Print) 2469-4460 (Online) Journal homepage: https://www.tandfonline.com/loi/raag21

Another Tale of Two Cities: Understanding Human Activity Space Using Actively Tracked Cellphone Location Data

Yang Xu, Shih-Lung Shaw, Ziliang Zhao, Ling Yin, Feng Lu, Jie Chen, Zhixiang Fang & Qingquan Li



A Tale of One City: Using Cellular Network Data for Urban Planning

Cellular data from call detail records can help urban planners better understand city dynamics. The authors use CDR data to analyze people flow in and out of a suburban city near New York City.

Journal of Location Based Services

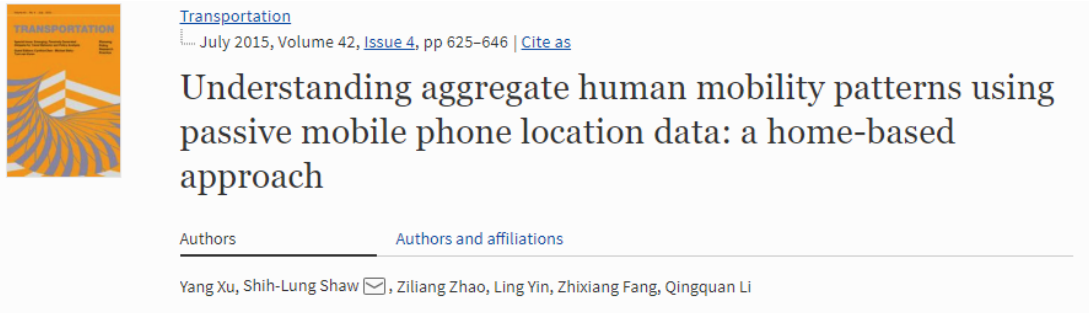
ISSN: 1748-9725 (Print) 1748-9733 (Online) Journal homepage: https://www.tandfonline.com/loi/tlbs20

Investigation of travel patterns using passive cellular phone data

S. Bekhor & I. Blum Shem-Tov

Earlier Studies on Cellular Data

- Meaningful locations:
  - Home
  - Work/School



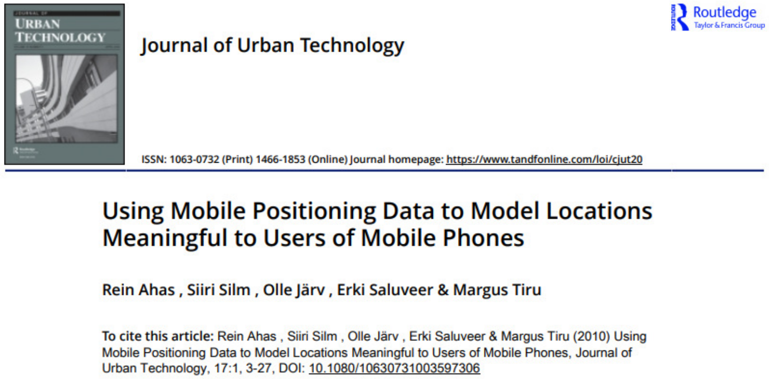
Transportation

July 2015, Volume 42, Issue 4, pp 625–646 | Cite as

Understanding aggregate human mobility patterns using passive mobile phone location data: a home-based approach

Authors Authors and affiliations

Yang Xu, Shih-Lung Shaw, Ziliang Zhao, Ling Yin, Zhixiang Fang, Qingquan Li



Journal of Urban Technology

Routledge
Taylor & Francis Group

ISSN: 1063-0732 (Print) 1466-1853 (Online) Journal homepage: https://www.tandfonline.com/loi/cjut20

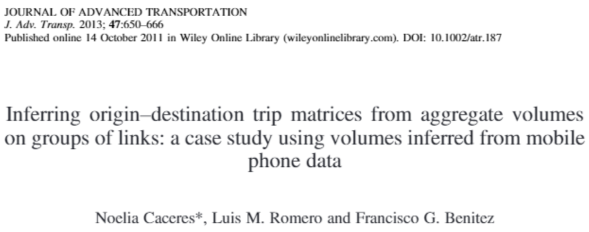
Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones

Rein Ahas , Siiri Silm , Olle Järv , Erki Saluveer & Margus Tiru

To cite this article: Rein Ahas , Siiri Silm , Olle Järv , Erki Saluveer & Margus Tiru (2010) Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones, Journal of Urban Technology, 17:1, 3-27, DOI: 10.1080/10630731003597306

Earlier Studies on Cellular Data

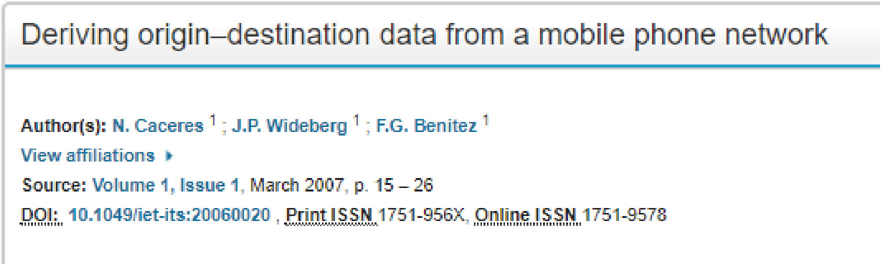
- OD estimations



JOURNAL OF ADVANCED TRANSPORTATION
J. Adv. Transp. 2013; **47**:650–666
Published online 14 October 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/atr.187

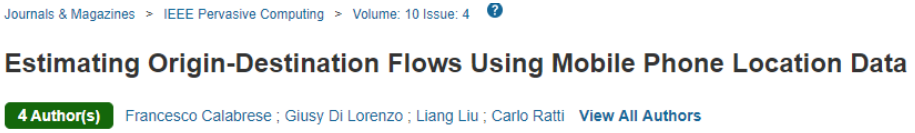
Inferring origin–destination trip matrices from aggregate volumes on groups of links: a case study using volumes inferred from mobile phone data

Noelia Caceres*, Luis M. Romero and Francisco G. Benitez



Deriving origin–destination data from a mobile phone network

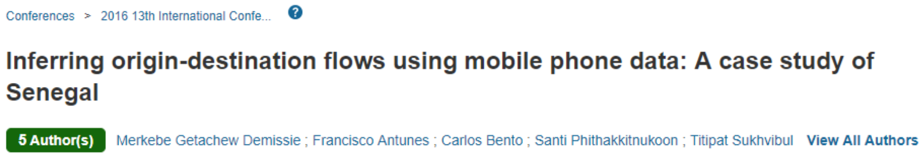
Author(s): N. Caceres [1] ; J.P. Wideberg [1] ; F.G. Benitez [1]
View affiliations ▸
Source: Volume 1, Issue 1, March 2007, p. 15 – 26
DOI: 10.1049/iet-its:20060020 , Print ISSN 1751-956X, Online ISSN 1751-9578



Journals & Magazines > IEEE Pervasive Computing > Volume: 10 Issue: 4

Estimating Origin-Destination Flows Using Mobile Phone Location Data

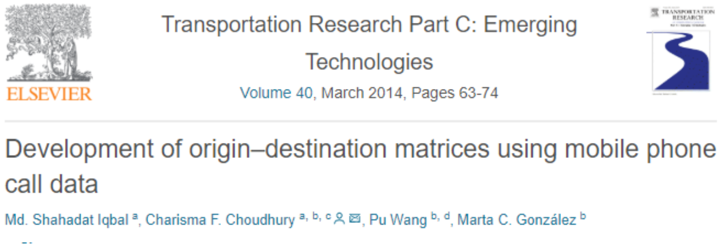
4 Author(s) Francesco Calabrese ; Giusy Di Lorenzo ; Liang Liu ; Carlo Ratti View All Authors



Conferences > 2016 13th International Confe...

Inferring origin-destination flows using mobile phone data: A case study of Senegal

5 Author(s) Merkebe Getachew Demissie ; Francisco Antunes ; Carlos Bento ; Santi Phithakkitnukoon ; Titipat Sukhvibul View All Authors

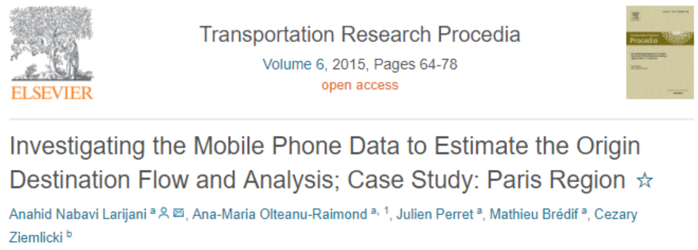


ELSEVIER

Transportation Research Part C: Emerging Technologies
Volume 40, March 2014, Pages 63-74

Development of origin–destination matrices using mobile phone call data

Md. Shahadat Iqbal [a], Charisma F. Choudhury [a, b, c], Pu Wang [b, d], Marta C. González [b]

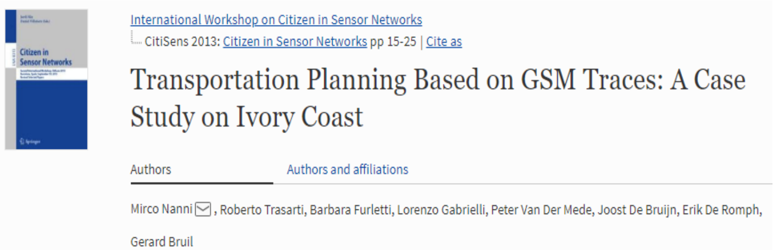


ELSEVIER

Transportation Research Procedia
Volume 6, 2015, Pages 64-78
open access

Investigating the Mobile Phone Data to Estimate the Origin Destination Flow and Analysis; Case Study: Paris Region ☆

Anahid Nabavi Larijani [a], Ana-Maria Olteanu-Raimond [a, 1], Julien Perret [a], Mathieu Brédif [a], Cezary Ziemlicki [b]

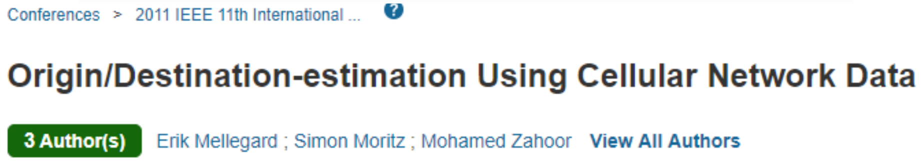


International Workshop on Citizen in Sensor Networks
CitiSens 2013: Citizen in Sensor Networks pp 15-25 | Cite as

Transportation Planning Based on GSM Traces: A Case Study on Ivory Coast

Authors | Authors and affiliations

Mirco Nanni, Roberto Trasarti, Barbara Furletti, Lorenzo Gabrielli, Peter Van Der Mede, Joost De Bruijn, Erik De Romph, Gerard Bruil



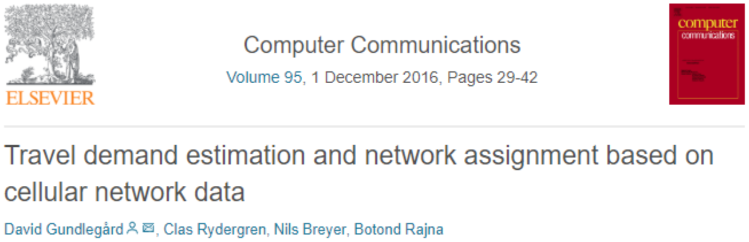
Conferences > 2011 IEEE 11th International ...

Origin/Destination-estimation Using Cellular Network Data

3 Author(s) Erik Mellegard ; Simon Moritz ; Mohamed Zahoor View All Authors

Earlier Studies on Cellular Data

- OD estimations
  - Estimating traffic flows
  - Optimizing public transport network
  - Activity-based travel demand models



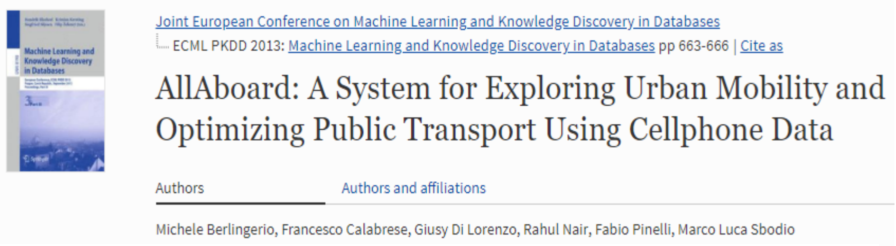
ELSEVIER

Computer Communications

Volume 95, 1 December 2016, Pages 29-42

Travel demand estimation and network assignment based on cellular network data

David Gundlegård, Clas Rydergren, Nils Breyer, Botond Rajna



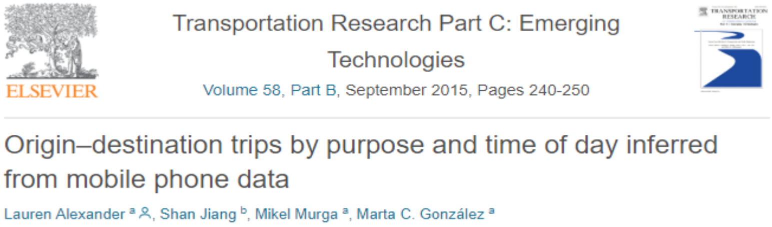
Joint European Conference on Machine Learning and Knowledge Discovery in Databases

ECML PKDD 2013: Machine Learning and Knowledge Discovery in Databases pp 663-666 | Cite as

AllAboard: A System for Exploring Urban Mobility and Optimizing Public Transport Using Cellphone Data

Authors | Authors and affiliations

Michele Berlingerio, Francesco Calabrese, Giusy Di Lorenzo, Rahul Nair, Fabio Pinelli, Marco Luca Sbodio



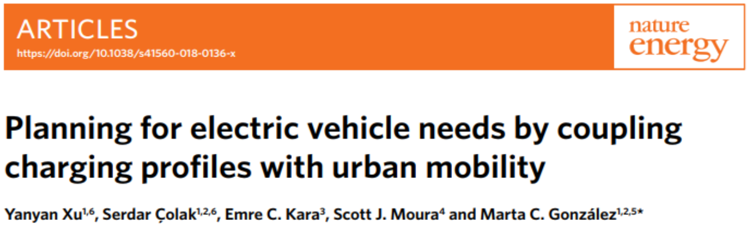
ELSEVIER

Transportation Research Part C: Emerging Technologies

Volume 58, Part B, September 2015, Pages 240-250

Origin–destination trips by purpose and time of day inferred from mobile phone data

Lauren Alexander [a], Shan Jiang [b], Mikel Murga [a], Marta C. González [a]



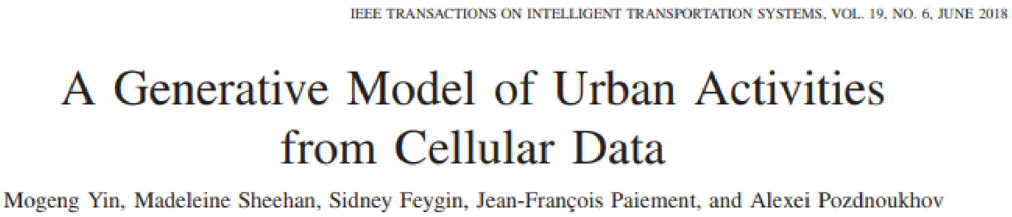
ARTICLES

https://doi.org/10.1038/s41560-018-0136-x

nature energy

Planning for electric vehicle needs by coupling charging profiles with urban mobility

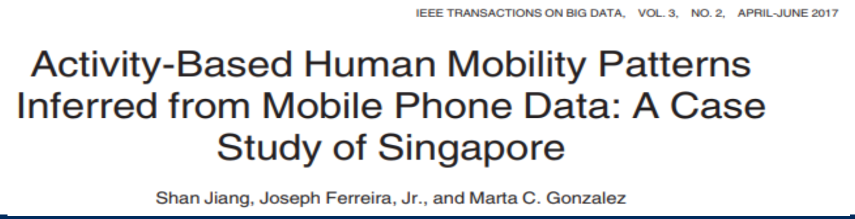
Yanyan Xu[1,6], Serdar Çolak[1,2,6], Emre C. Kara[3], Scott J. Moura[4] and Marta C. González[1,2,5*]



IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, VOL. 19, NO. 6, JUNE 2018

A Generative Model of Urban Activities from Cellular Data

Mogeng Yin, Madeleine Sheehan, Sidney Feygin, Jean-François Paiement, and Alexei Pozdnoukhov



IEEE TRANSACTIONS ON BIG DATA, VOL. 3, NO. 2, APRIL-JUNE 2017

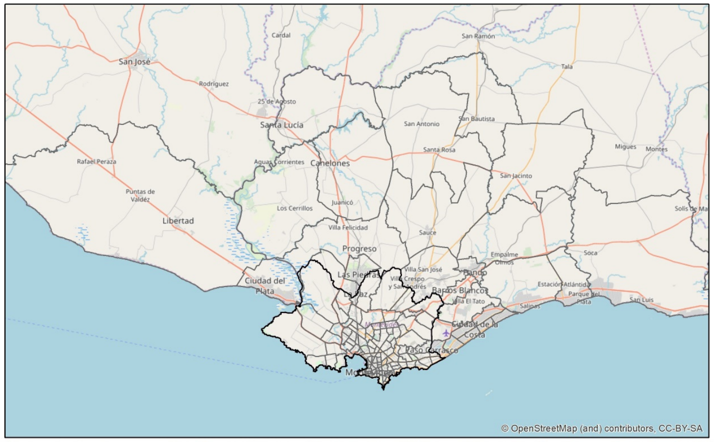
Activity-Based Human Mobility Patterns Inferred from Mobile Phone Data: A Case Study of Singapore

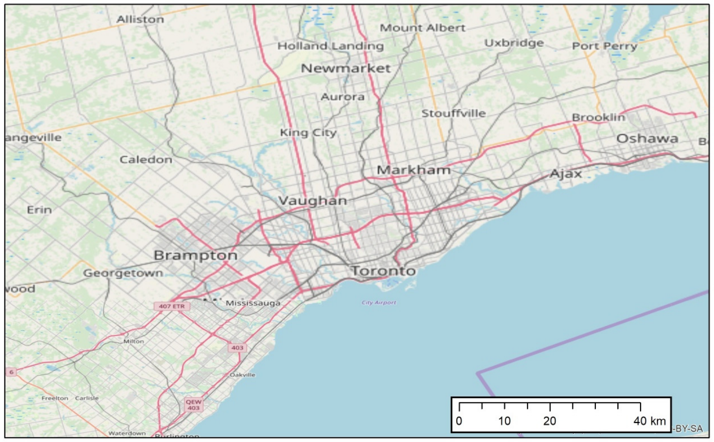
Shan Jiang, Joseph Ferreira, Jr., and Marta C. Gonzalez

Objectives

1. Identifying trips and activities
2. Detecting home and work location
3. Imputing mode of travel This is new; primary objective!

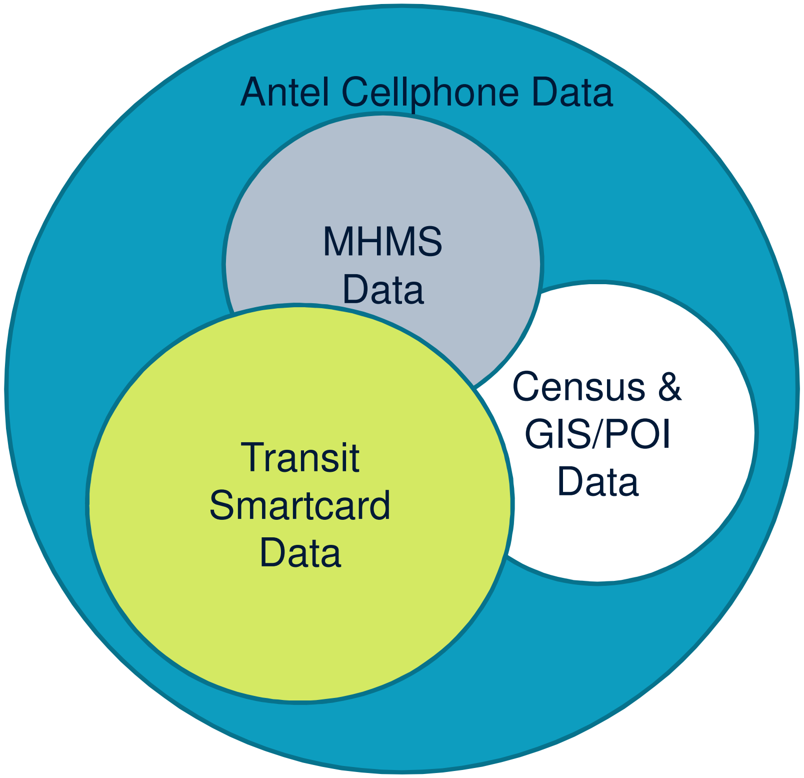
Colombia
Ecuador
Peru
Brazil
Bolivia
Paraguay
Chile
Argentina
Buenos Aires
Montevideo
São Paulo
Salvador
Brasilia
Santiago
Córdoba
Asuncion
Lima District
Quito
Suriname





- Montevideo, Uruguay
  Population:
  City: 1.38 million
  Metro: near 2 million
  Country: 3.4 million

Data

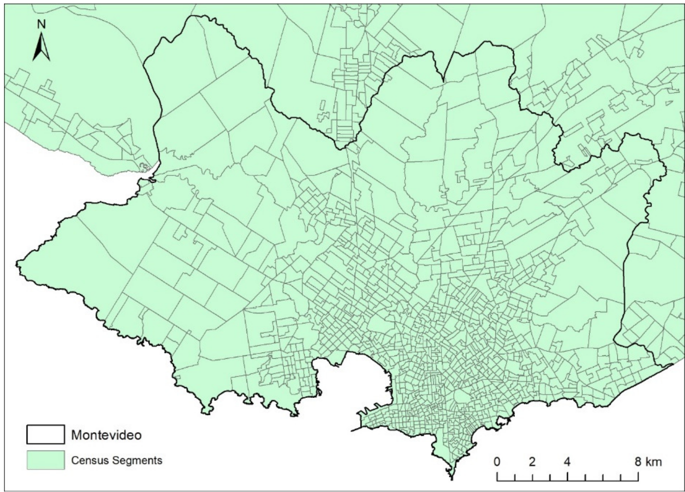


- 2016 Montevideo Household Mobility Survey (MHMS) records.
- Road and transit network data.
- Census data.
- A very large sample of Antel cellphone traces, consisting of 40% of all such traces for four weeks.
- All public transit fare transaction records for the same time period, provided by the Intendencia de Montevideo.

MHMS Data

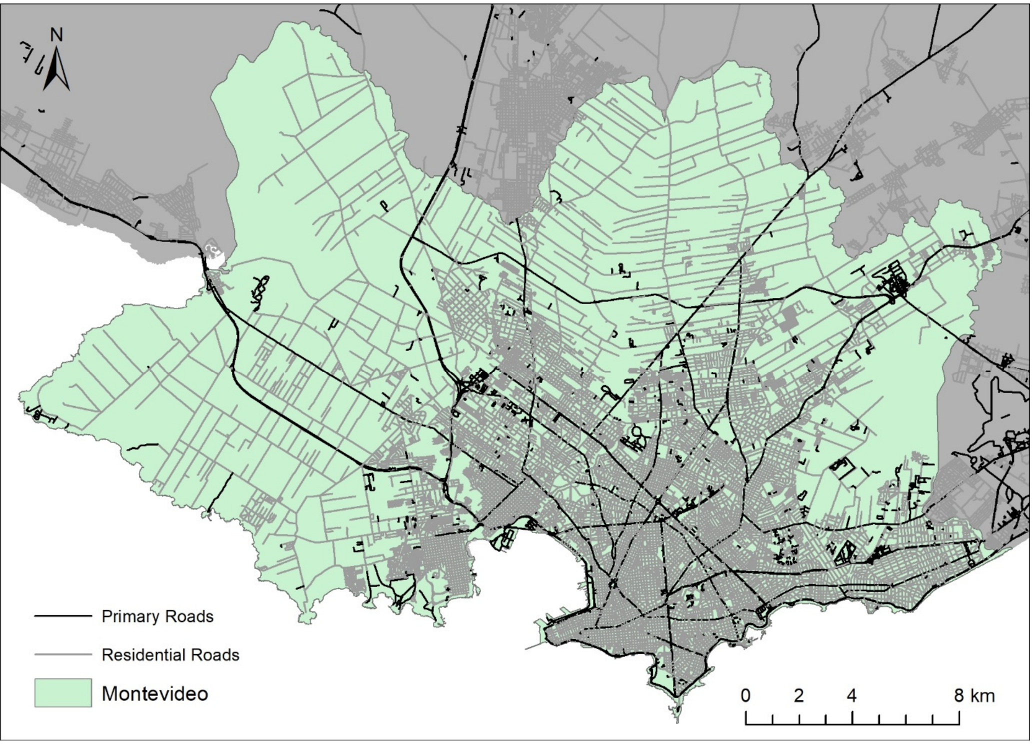
- August-October 2016 in the Metropolitan Area of Montevideo.
- 0.34% sample of the households in the region:
  - 2,230 households.
  - 5,946 individuals.
  - 12,546 reported trips.
  - Average daily rate of 2.11 trips/person.
- Spatially aggregated to census segments.

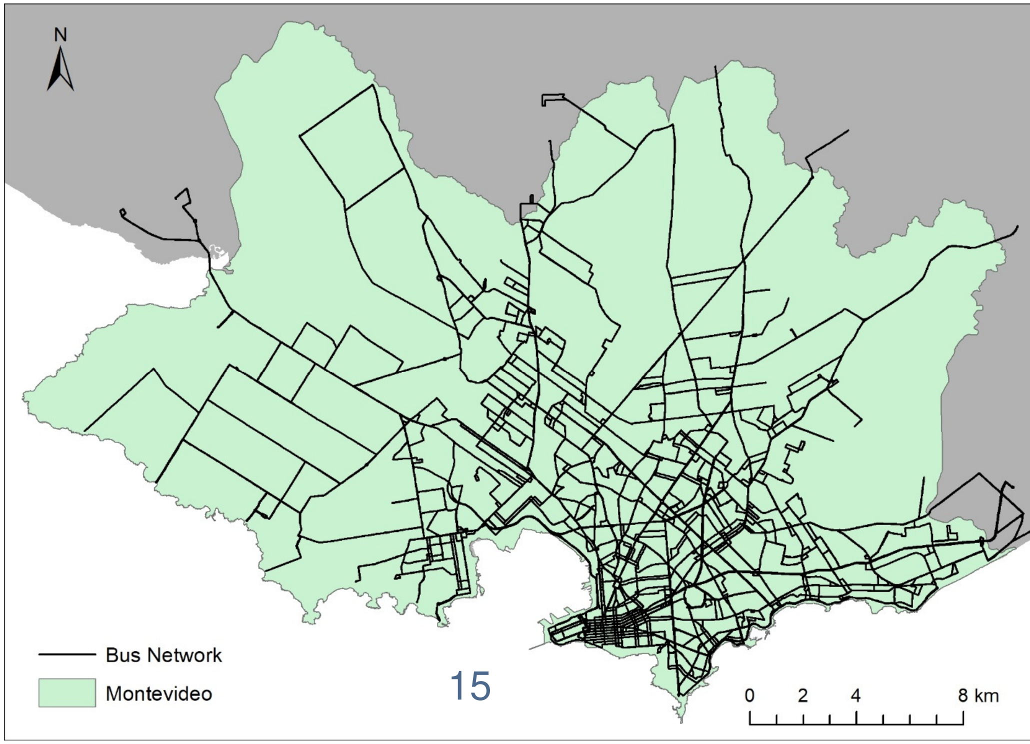
Travel Mode	Mode Share in %	Number of Trips
Walk	34.0	4265
Bike	3.5	439
Auto Passenger	10.0	1251
Auto Driver	19.2	2410
Motorcycle	6.1	769
Bus	25.2	3166
Other	2.0	246



Network Data

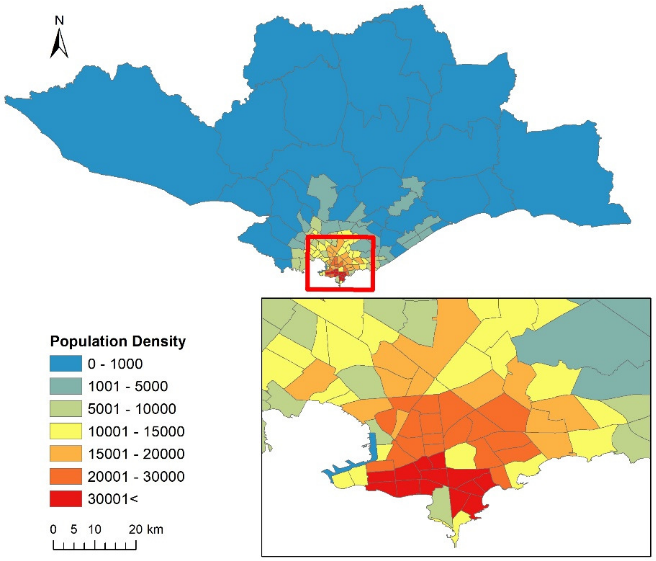
- Computerized representations of the Montevideo road & transit networks were constructed in Emme for modelling purposes. Sources:
  - Roads: OpenStreetMap database.
  - Bus Network: Open data portal of Montevideo government.

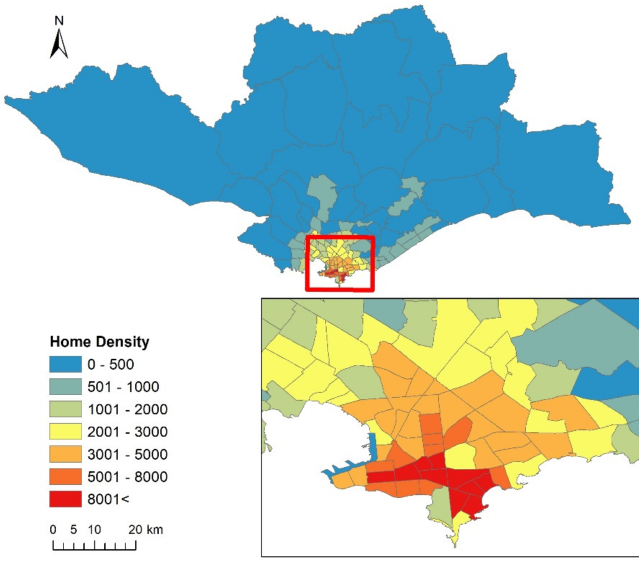


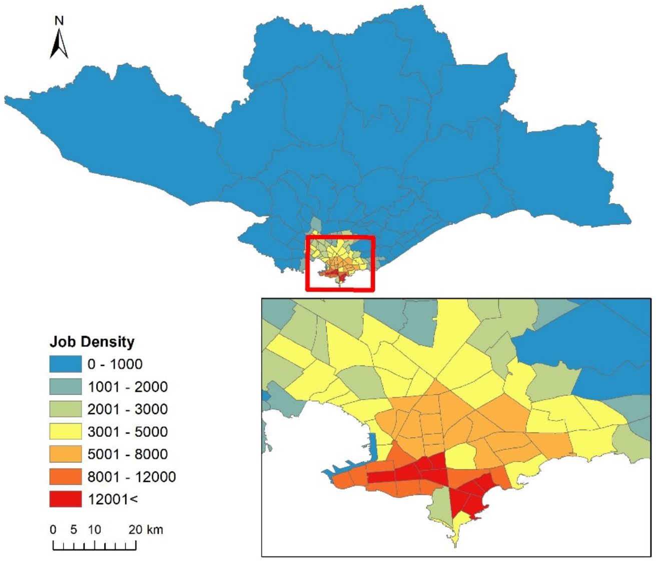


Census Data

- Population, Home and Jobs at Census Segment level.







Antel Data (1)

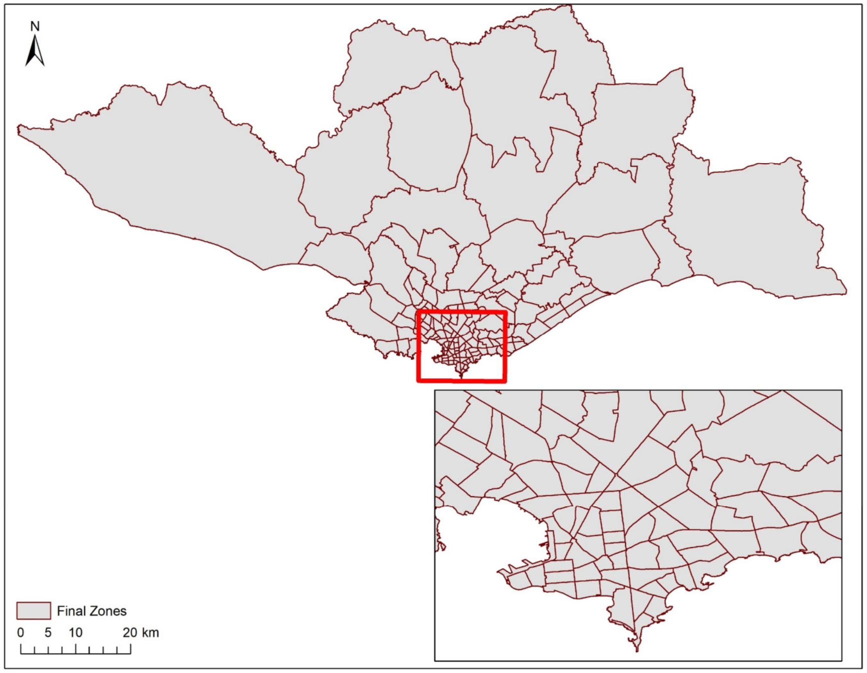
- Antel is the primary cellular telecommunications company in Uruguay.
- Traces of a random sample of 40% of mobile phones within Montevideo and the surrounding metropolitan area.
- May 2nd to May 29th, 2018.

- Raw data processed by Antel:
  - To eliminate as much noise in the data as possible & to preserve user anonymity.
  - Trace data are temporally reported in minutes.
  - Spatially aggregated to 135 zones in the Montevideo region

- 117,862,000 cellphone traces for about 948,600 unique cellphones.

Challenges with Data

- Spatial precision

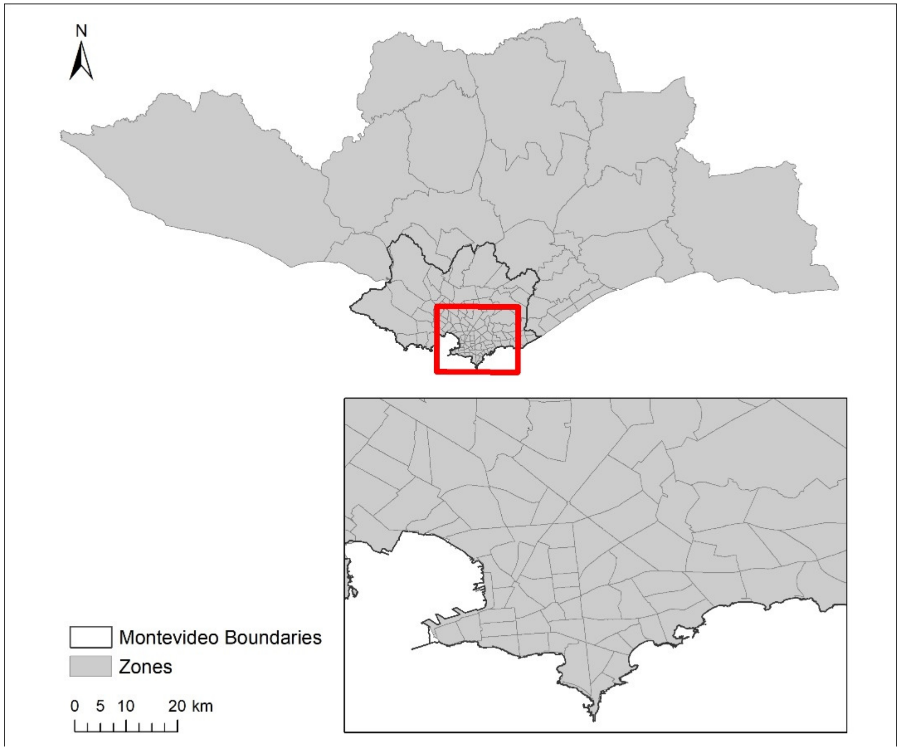


135 Zones

The Montevideo has 186 TAZs

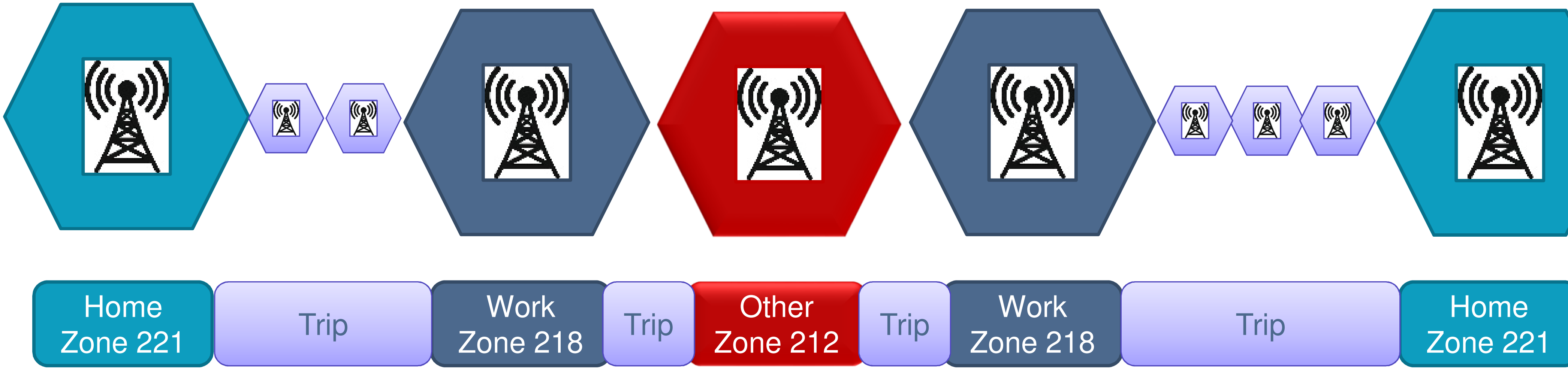
Antel Data (2)

- Zone size:
  - Average: 37.1 km^2
  - Min: 0.38 km^2 , Max: 1,020 km^2.
- **Activity**: any time a cellphone user stays more than 30 minutes in a zone.
- **Trips**: when a user moves between two activities with the first activity location identified as the trip origin and the second activity location as the trip destination.



Method - Example

Time	…	7:45	8:00	8:15	8:30	…	12:00	12:15	…	13:15	13:30	…	18:15	18:30	18:45	19:00	…
Zone	…	221	230	229	218	…	218	212	…	212	218	…	218	236	230	221	…
	Home		Travel		Activity 1			Activity 2			Activity 3			Travel		Home	



Home Zone 221

Trip

Work Zone 218

Trip

Other Zone 212

Trip

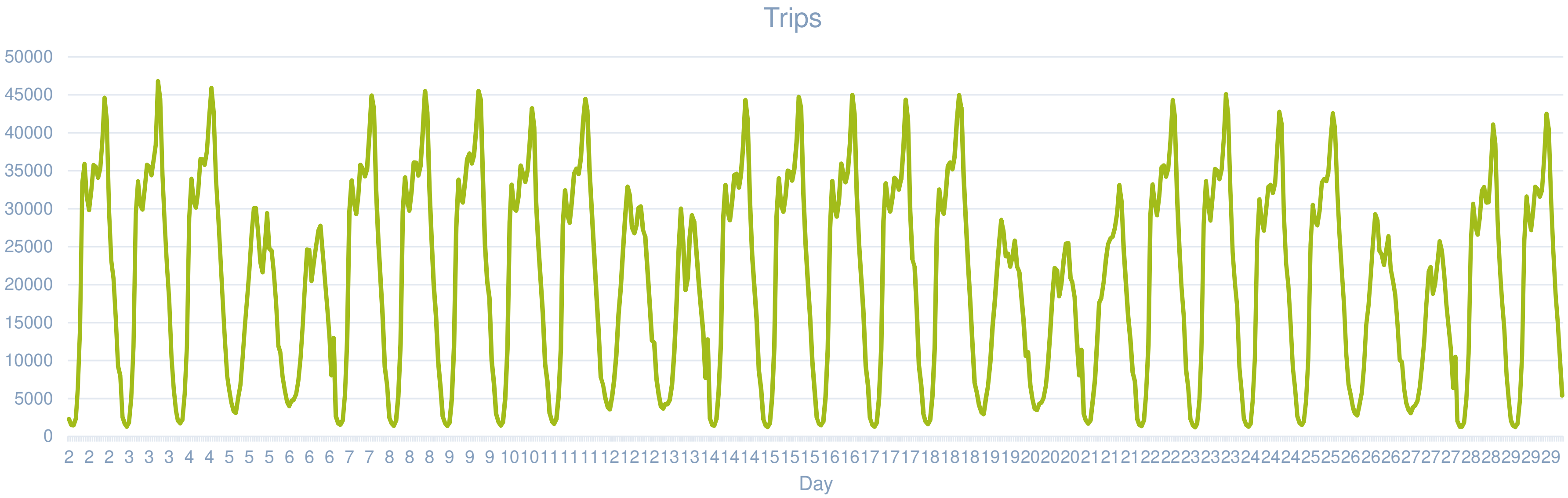
Work Zone 218

Trip

Home Zone 221

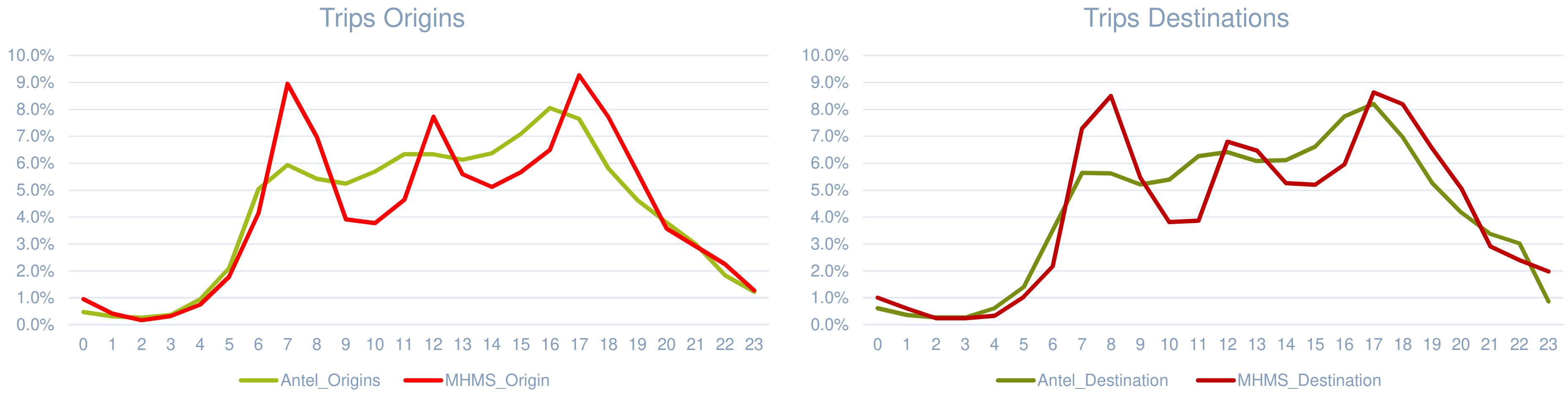
Antel Trips (1)

- The data were cleaned to remove some inconsistent trips such as jumps in locations and trips with zero duration with the same origin and destination.



Antel Trips (2)

- The general pattern of trip-making observed in the Antel data is similar to the MHMS pattern, except the MHMS AM trips are more peaked. This may be due to:
  - Antel traces missing some morning trips (cellphones not turned on when leaving home in the morning?).
  - MHMS missing mid-day trips (which would skew the MHMS distribution).
- Cell traces from weekdays only used to be comparable to household travel survey.



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

Home Location (1)

- From cell traces, it is possible to infer and make high accuracy estimates of people's primary locations such as their place of residential location.
- An Antel zone is labeled as the cellphone user's home zone using different approaches and then compared with the census.
  - Home is the zone with most stayed duration.
  - Home is the zone with most stayed duration during specific periods.
  - Home is the zone with most first trips originated and last trips destined.
  - Home is the zone with most trips destined there.

Home Location (2)

- Most stayed duration in weekends provides the highest fit to the distribution of the homes from census.
  - For some (13.6%), this method could not clearly identify a location for home (for example because of no traces recorded during weekends)
  - For these group, the most stayed location is selected as home zone.

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

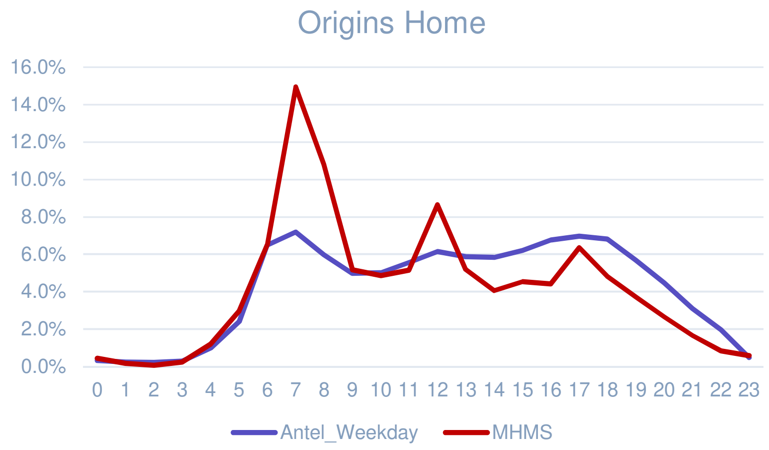
Work Location

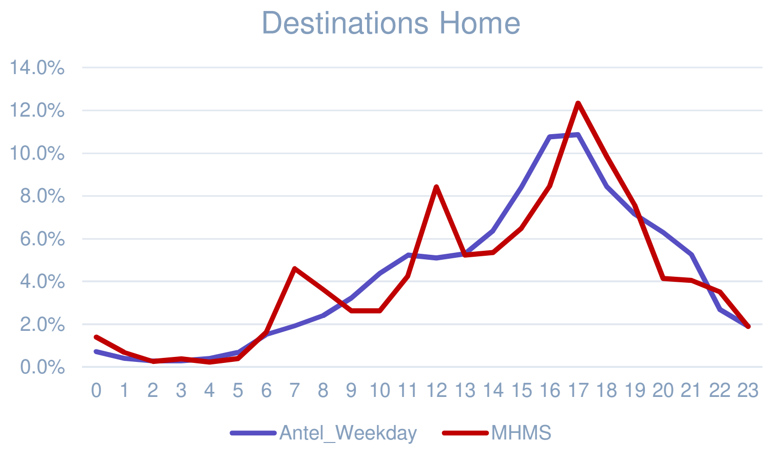
- The second most visited location for cell phone users is very likely to be their usual place of work/school.
- Similar approach to home location detection. At the end, 66% of cellphones were given a work/school location.
  - Most stayed location during weekdays daytime (8 to 18) with minimum of 1 hour stay.
    - A threshold of minimum of 1 hour stay in a day in average is also used to remove the locations with frequent visits but with short duration.
    - For those users that work location were not identifiable, it is assumed that those users are not workers/students.

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

Home based trips

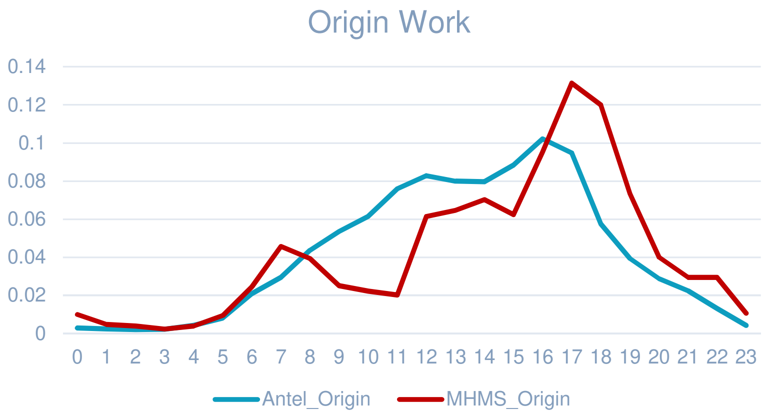
- Antel home destinations correspond well to the MHMS distribution.
- Again, AM trip origins from home seem to be under-represented in the Antel data. The rest of the day looks good.

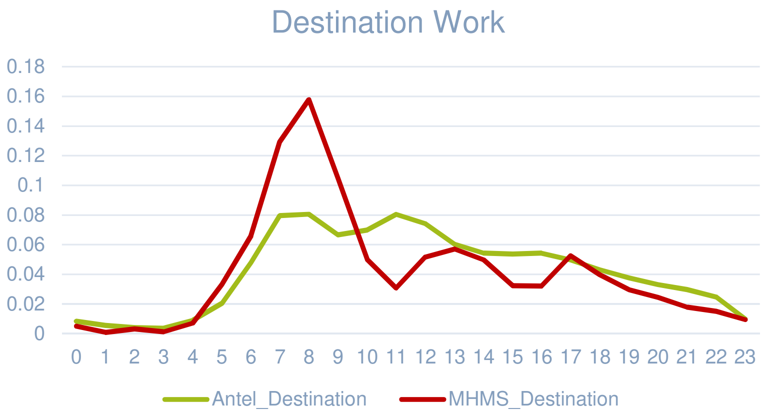




Work Based Trips

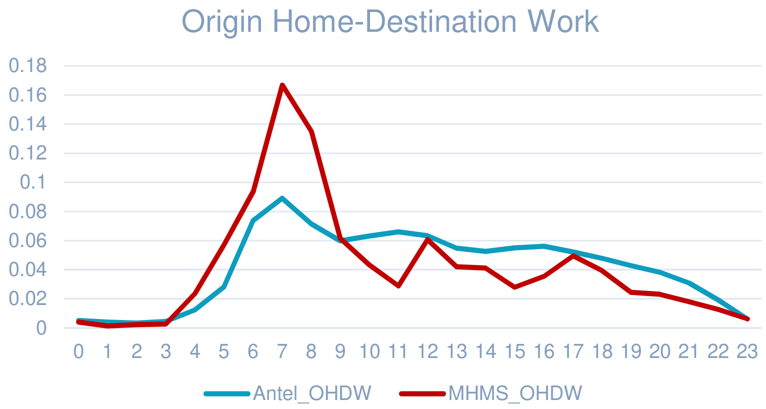
- MHMS data may be missing morning/mid-day work trip origins?
- Again, Antel AM work destinations appear to be low.

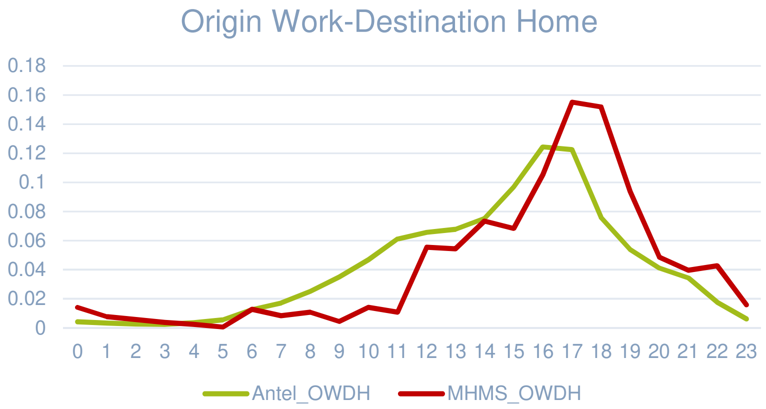




Home-Work Trips

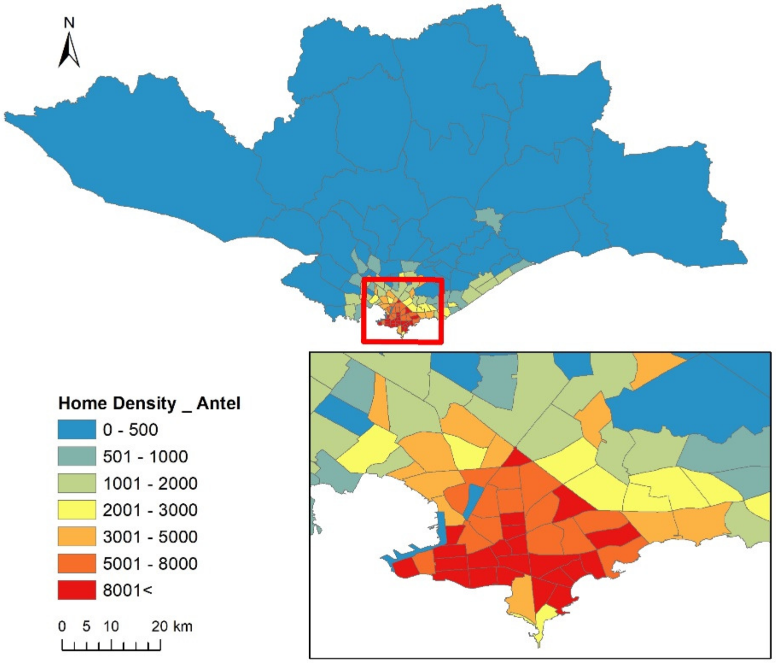
- Note that these are not tours; i.e., these are one-way trips
- Comparison to MHMS trips is generally good, except, again for AM home-work trips.


Origin Home-Destination Work


Origin Work-Destination Home

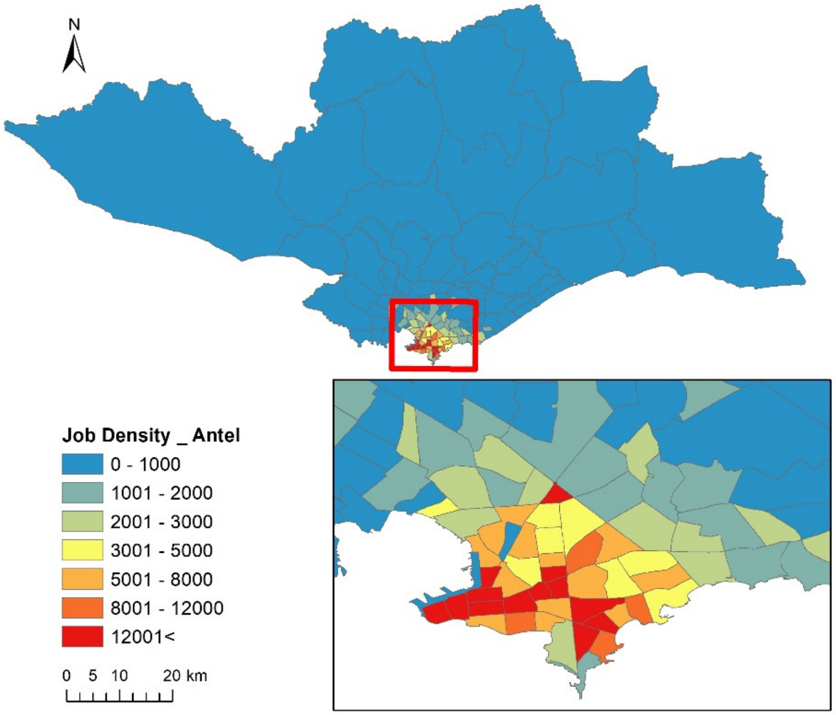
Home Density

- Household residential are computed based on the home zones identified for the 40% sample of Antel cellphone users.
- It is individual based and not household based; i.e. homes for individuals in the same household are double-counted.
- Still the spatial distribution resembles the census distribution.



Job Density

- Similarly, employment zone job densities are computed from the 40% Antel user sample.
- It is based on the work/school locations identified, which are basically the second primary location where a user spends time.
- Still the spatial distribution resembles the census distribution.

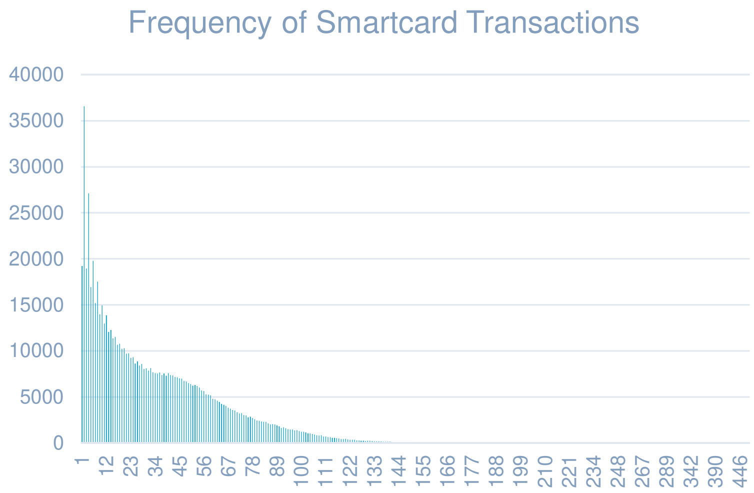


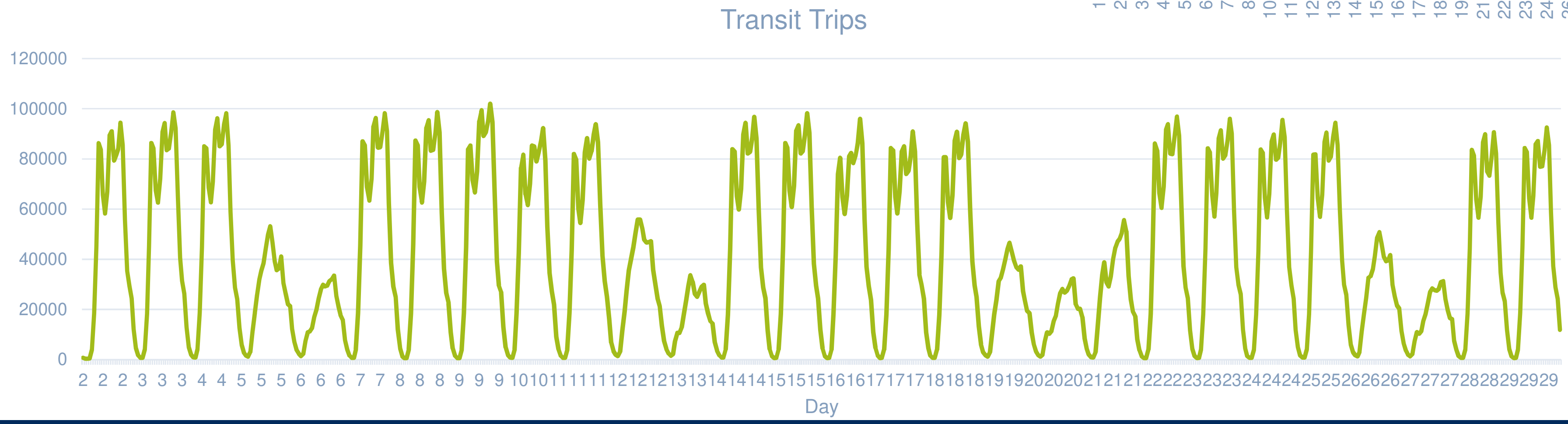
Transit Data

- The Intendencia de Montevideo provided data for Montevideo’s integrated public transit system (STM) for the same period in May 2018.
- Every transit boarding is available: smartcard and cash payments.
- The transit system is tap-on only, hence destinations must be estimated.
- Intendencia de Montevideo has developed their own transaction data processing methods to estimate alighting stops.
  - 62% alighting stops have been successfully estimated.

Transit Data

- 29,868,716 recorded transactions.
- 82.5% by smartcards.
- 734,569 unique smartcards.
- On average, a smartcard was used for 33.6 transactions (SD=29.3) in May.



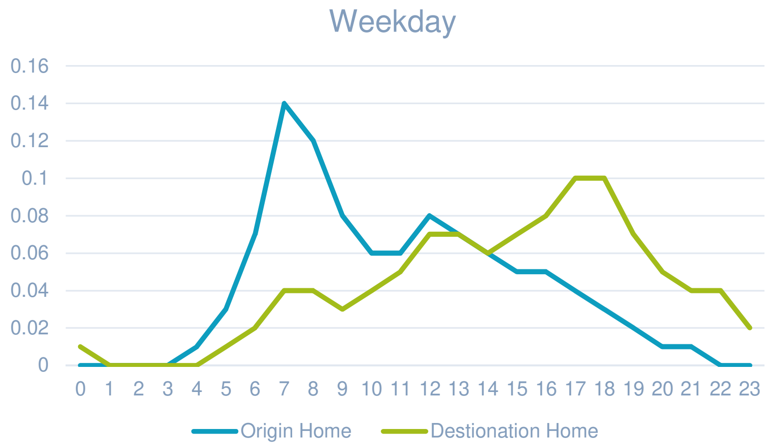


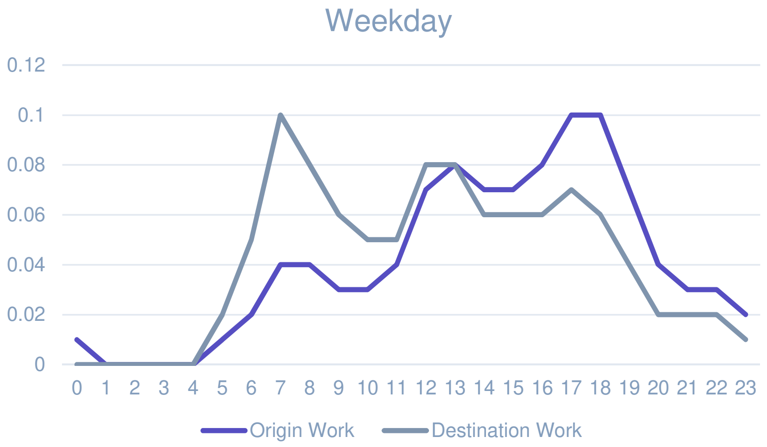
Transit Data – Identifying locations

- For the smartcard transactions, it is possible to investigate the frequency of bus stops by time of the day.
- Assuming the first transaction of the day is usually made at the stops near home location, we can label that stop (and the zone) as the home location of the rider.
  - Since some of the smartcard users do not have a repetitive travel pattern (e.g. 7.6% of the users only have 1 or 2 transaction in the month), we keep a threshold of minimum three trips originated from the stop to be considered for home location.
- Similarly, the last transaction of the day is labeled as work location (with the same minimum 3 transaction threshold)
- Overall, 76% of smartcards have assigned a home zone and 65% have a work zone assigned.

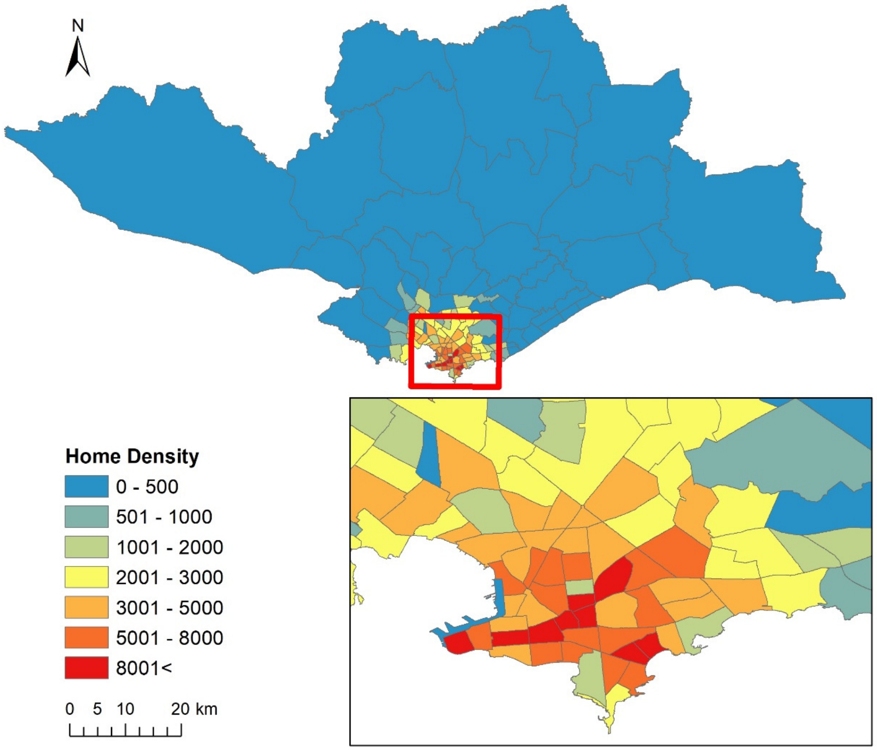
Transit Data - Trips

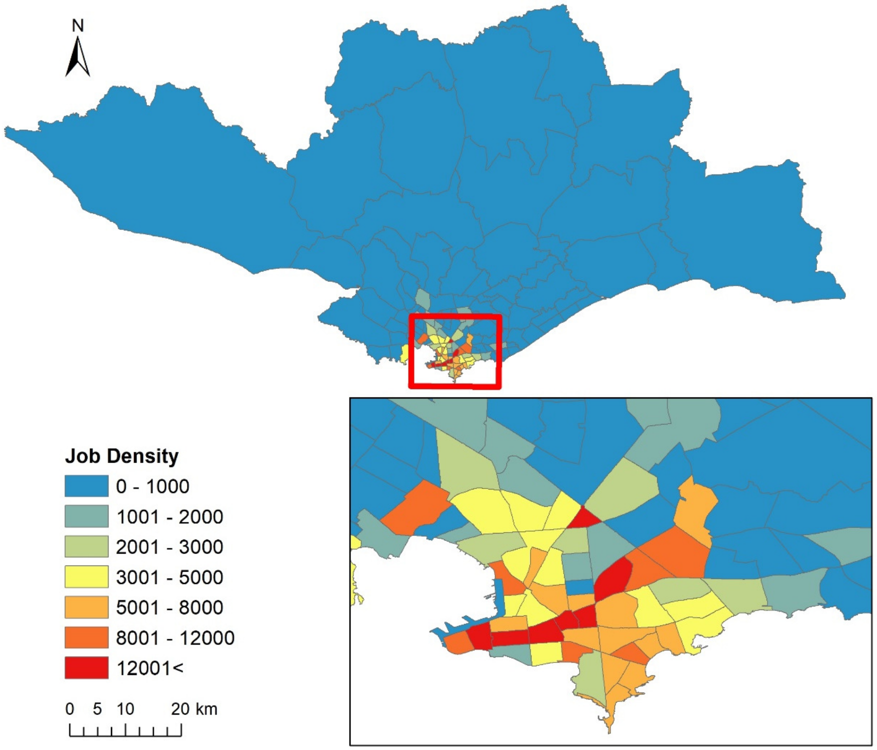
- Hourly distribution of trips with origin/destination at home/work during weekdays.





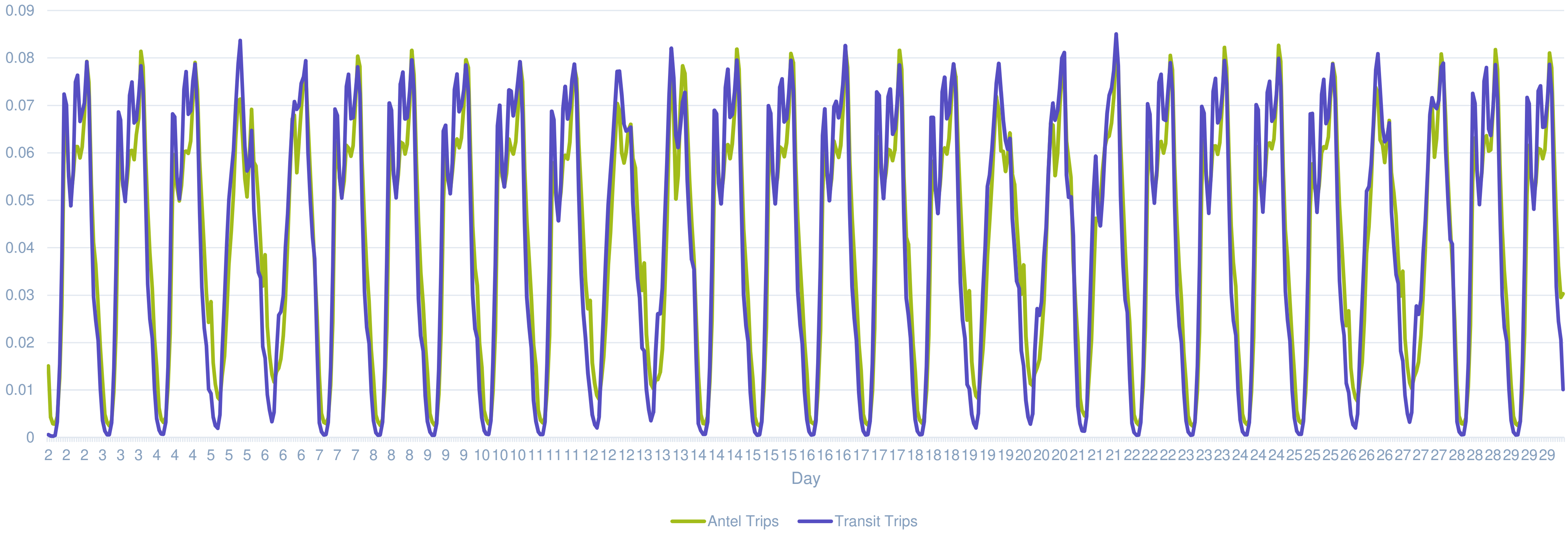
Transit Data - Home & Work Locations



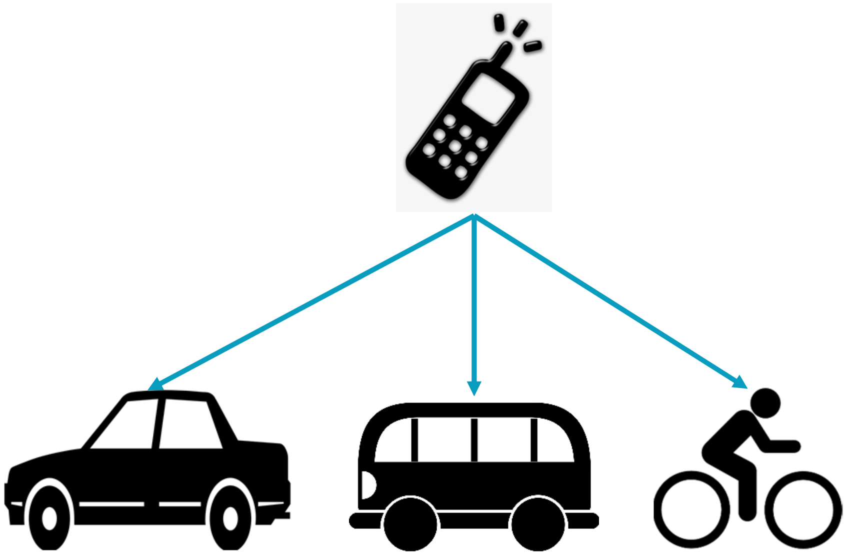


Transit Data (2)

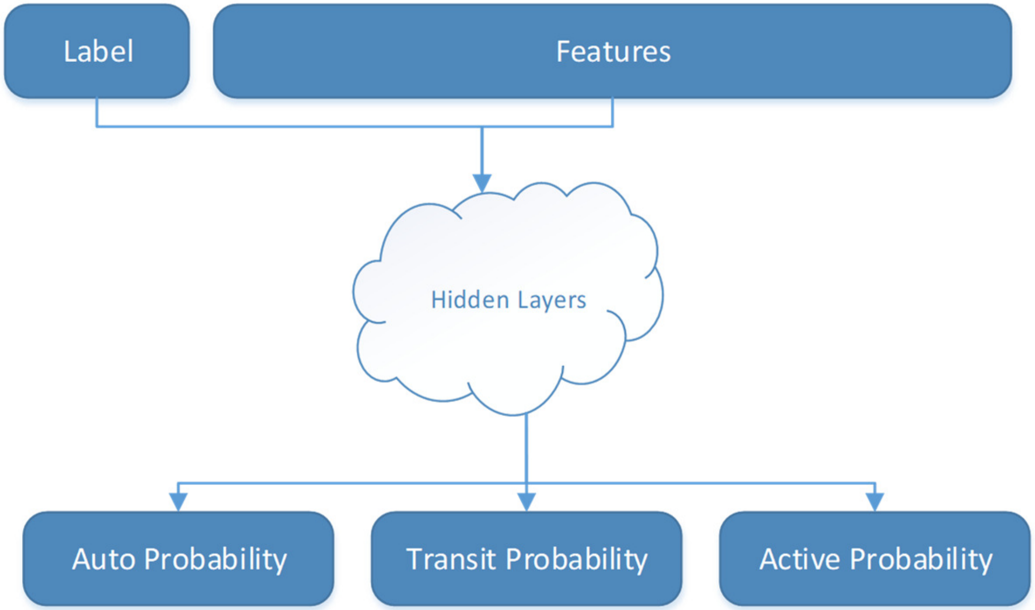
- Comparing the daily frequency of trips based on the Antel cellphone data and smartcard transaction data.



Imputing **travel mode** for Antel cellphone traces

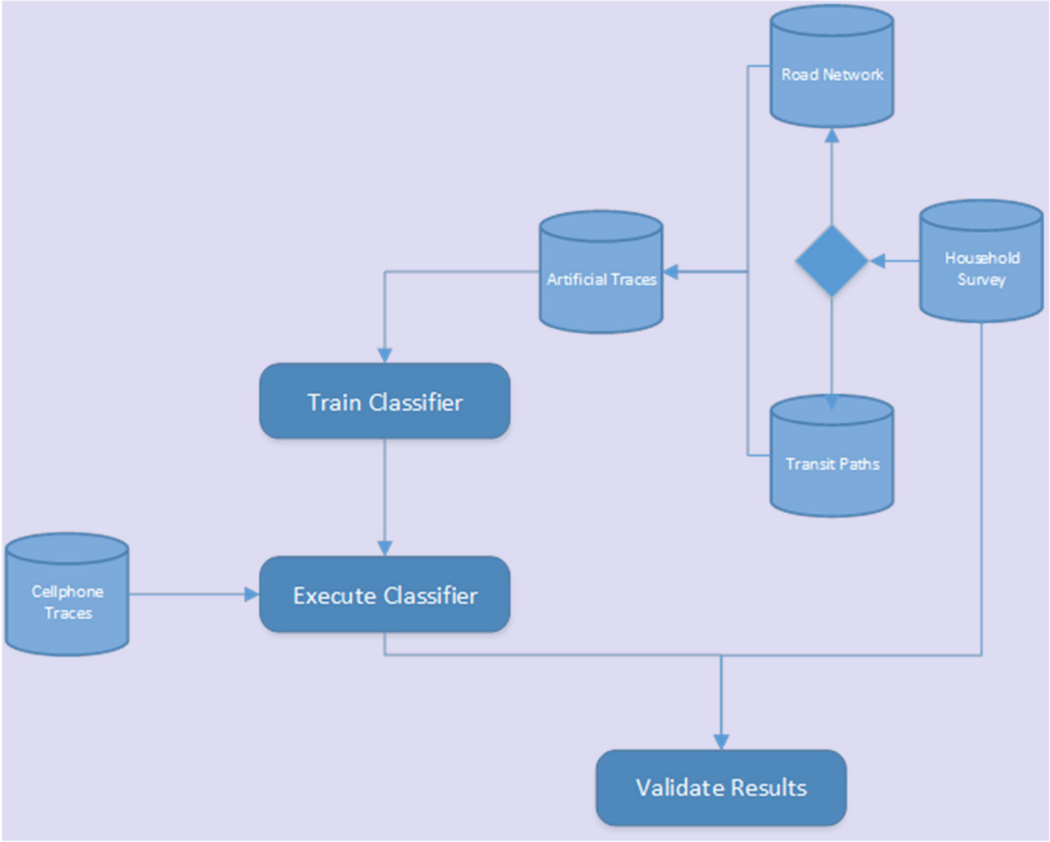


- Neural Network model.



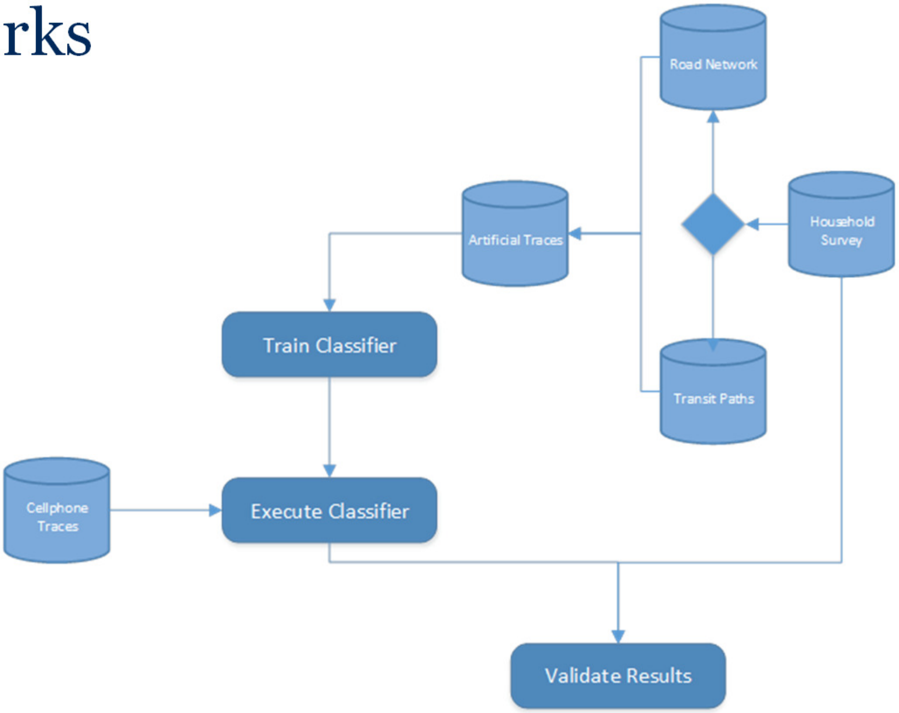
Analysis Method

1. Convert MHMS O-D trips into pseudo-cellphone traces.
2. Train a neural net on the MHMS pseudo-traces to identify travel mode.
3. Apply the trained model to the Antel cellphone traces to classify them by mode.



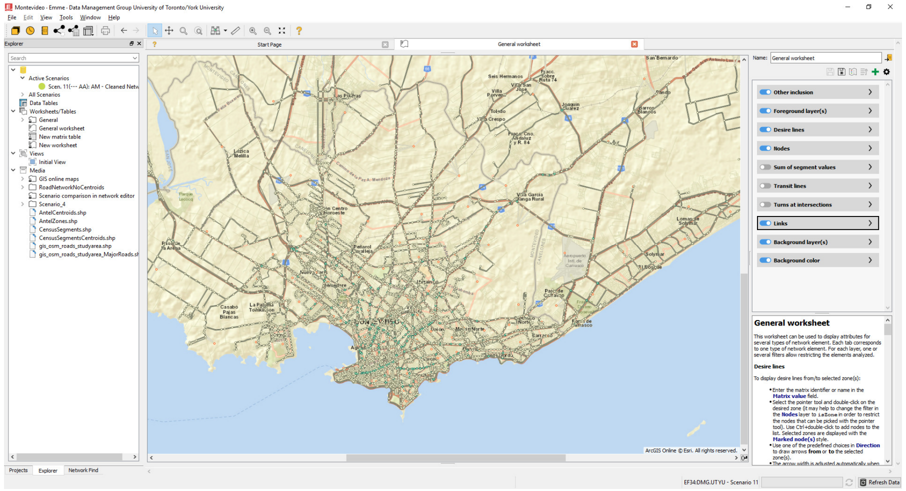
Steps in Constructing the Mode Choice Model

- Model Road & Transit Networks
- Create Artificial Traces
- Train Classifier
- Execute Classifier
- Validate Results



Create the Network Models

- Based on Open Street Map
- Transit lines loaded from GTFS & Open Portal and then manually inspected removing artifacts from the importing process
- Demand is loaded from MHMS for each time period
- Route assignment parameters are taken from GTModel V4.0.2



General worksheet

This worksheet can be used to display attributes for several types of network element. Each tab corresponds to one type of network element. For each layer, one or several filters allow restricting the elements analyzed.

Desire lines

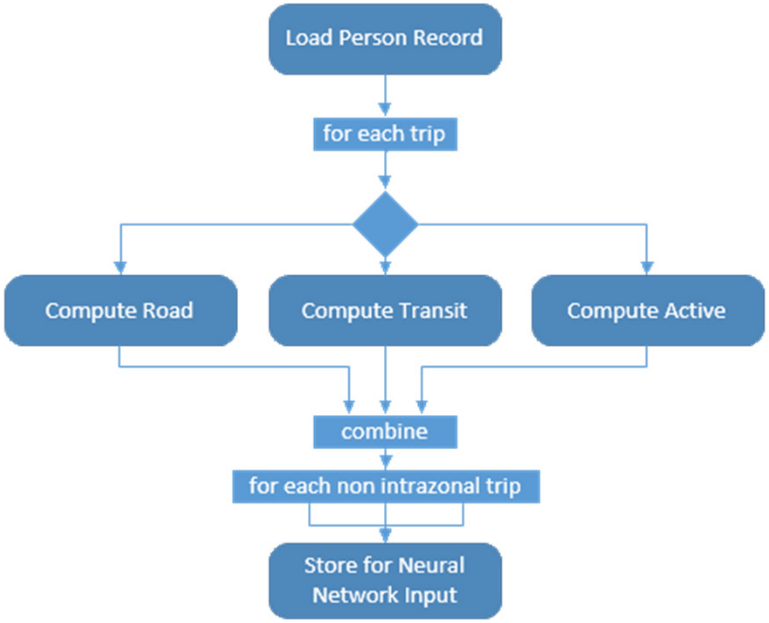
To display desire lines from/to selected zone(s):

- Enter the matrix identifier or name in the **Matrix value** field.
- Select the pointer tool and double-click on the desired zone (it may help to change the filter in the **Nodes** layer to `isZone` in order to restrict the nodes that can be picked with the pointer tool). Use Ctrl+double-click to add nodes to the list. Selected zones are displayed with the **Marked node(s)** style.
- Use one of the predefined choices in **Direction** to draw arrows **from** or **to** the selected zone(s).
- The arrow width is adjusted automatically when

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

Creating Artificial Traces

- All of the trips for a person in the MHMS are loaded
- Each trip is then processed assign the distance travelled when crossing an Antel zone boundary assigning that value to a five minute bin



Load Person Record

for each trip

Compute Road

Compute Transit

Compute Active

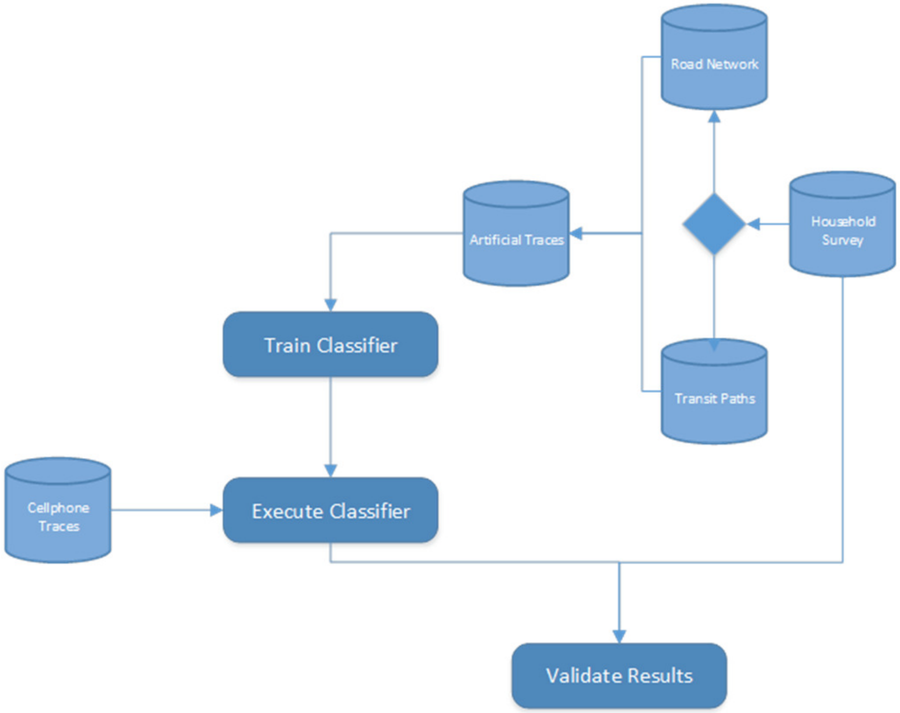
combine

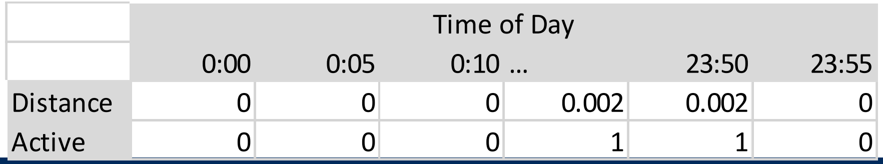
for each non intrazonal trip

Store for Neural Network Input

Neural Network Features

- Each artificial trace has two values for each five minute bin of time during the day.
    - The first value contains the distance travelled when crossing boundaries.
    - The second value contains whether or not this time bin occurs during trip to assign.
- Each record contains the **full day**'s worth of trip distance features.





	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

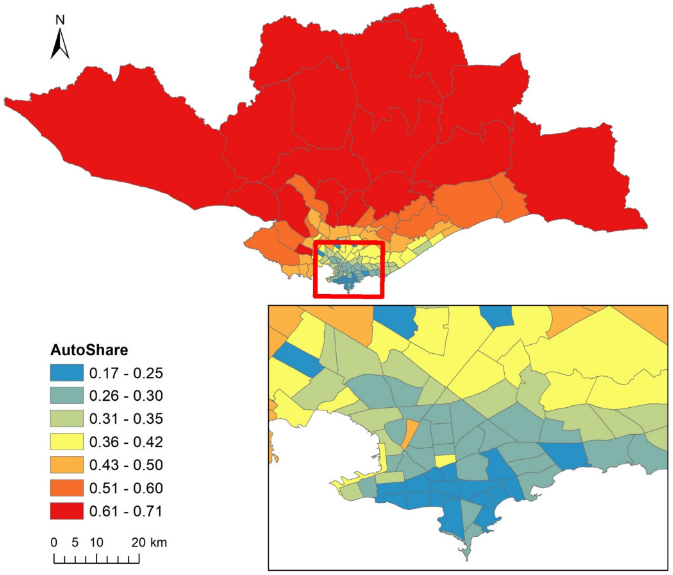
Train Classifier

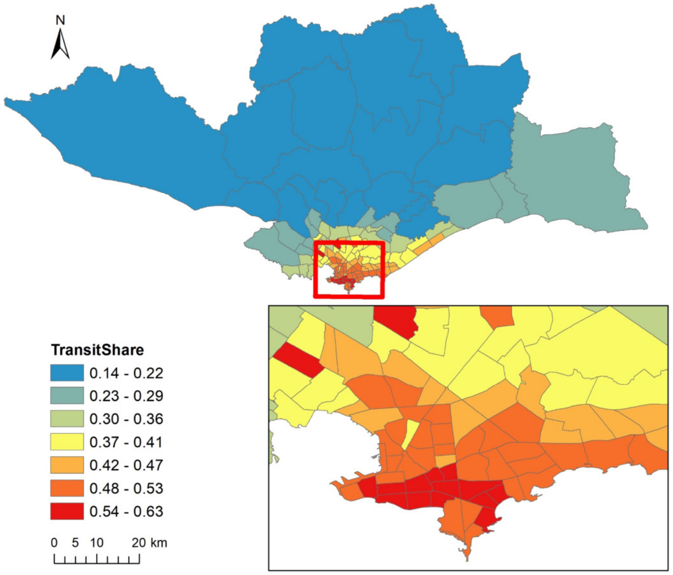
- The neural net was trained using Microsoft's Cognitive Neural Toolkit (CNTK)
- Artificial traces were split into two datasets, one for estimation and one for testing.
- On the estimation set an accuracy of 0.98 was reached.
- On the testing set an accuracy of 87% was reached.

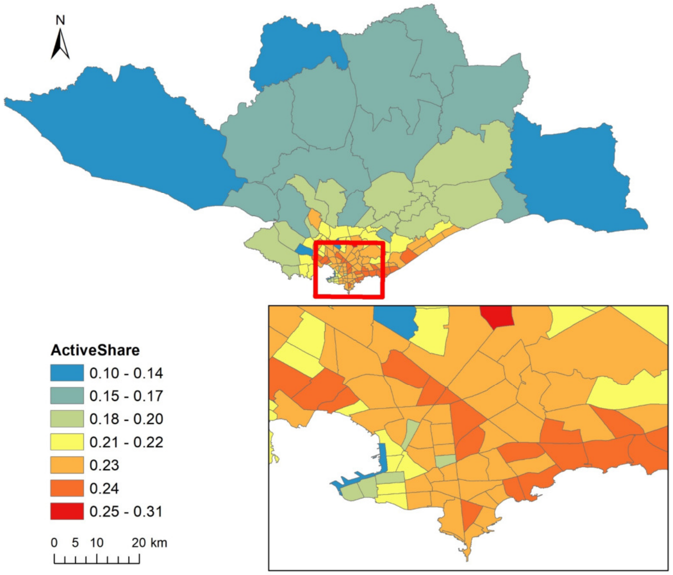
Predicted vs. Observed Mode (% of total trips)

Predicted	Observed Mode			
Mode	Auto	Transit	Active	Total
Auto	33.65	1.08	2.54	37.27
Transit	1.76	21.36	1.92	25.04
Active	2.91	2.88	31.91	37.7
Total	38.32	25.32	36.37	100

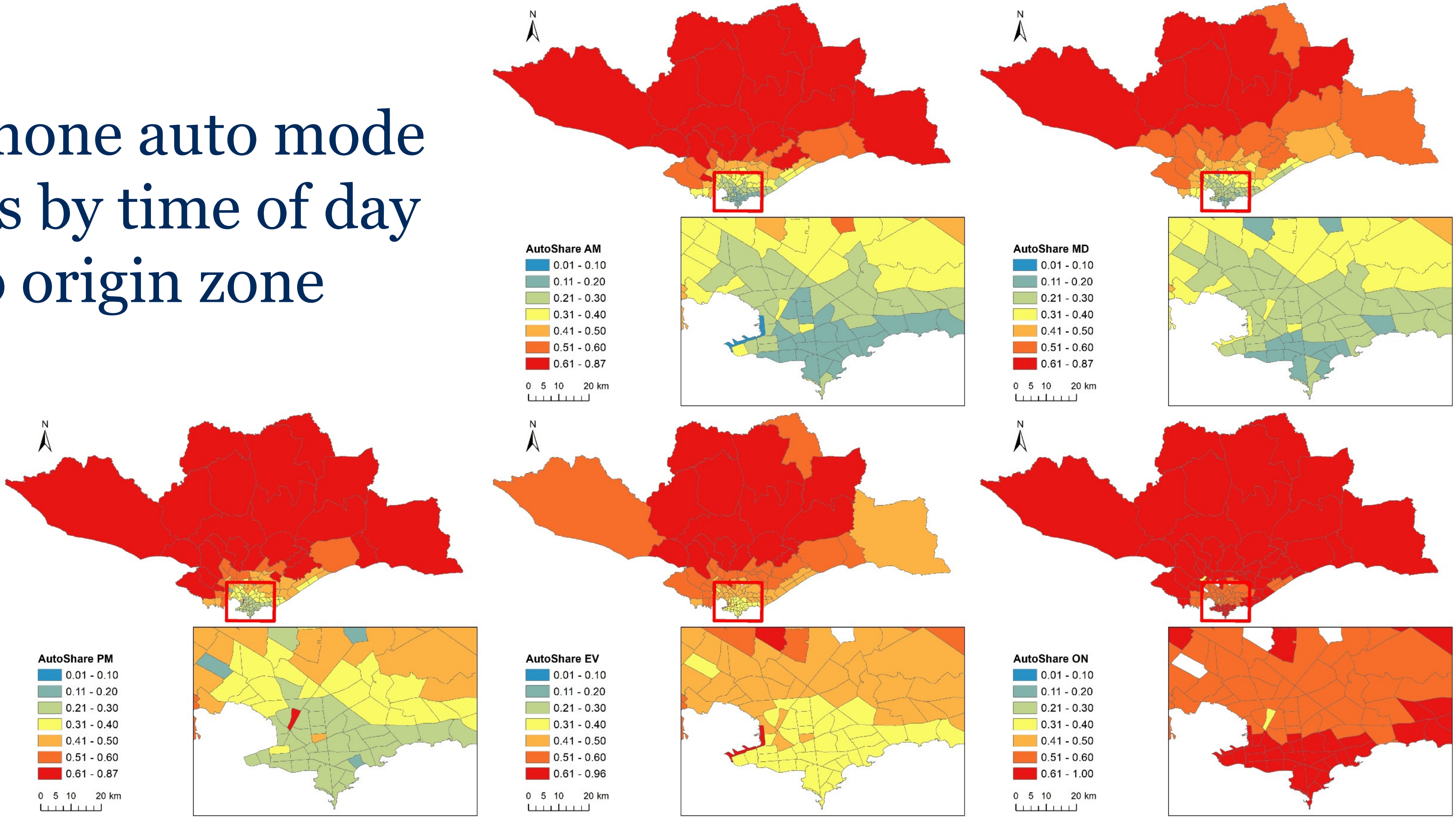
Cellphone mode results: share, all day, by trip origin



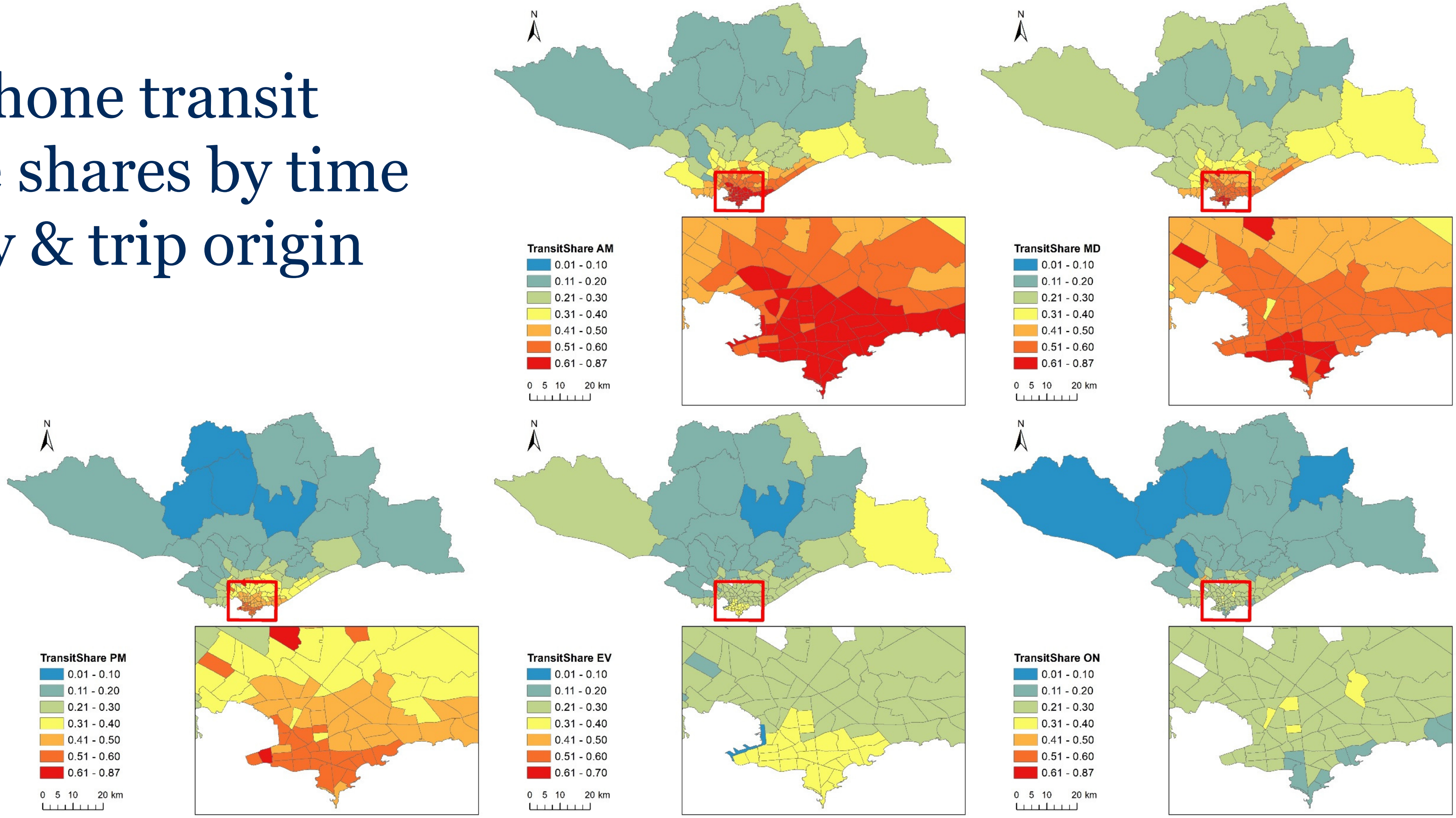




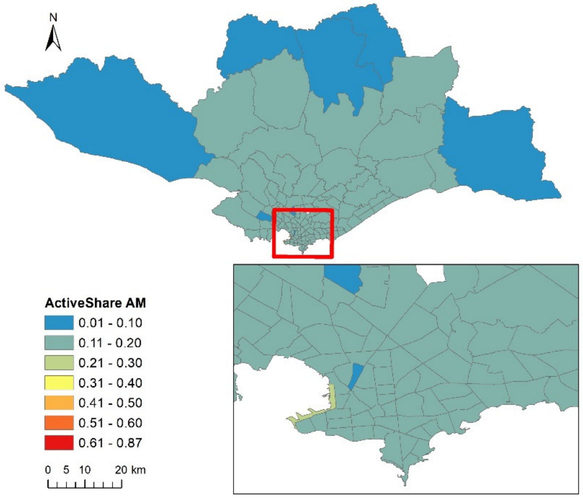
Cellphone auto mode shares by time of day & trip origin zone

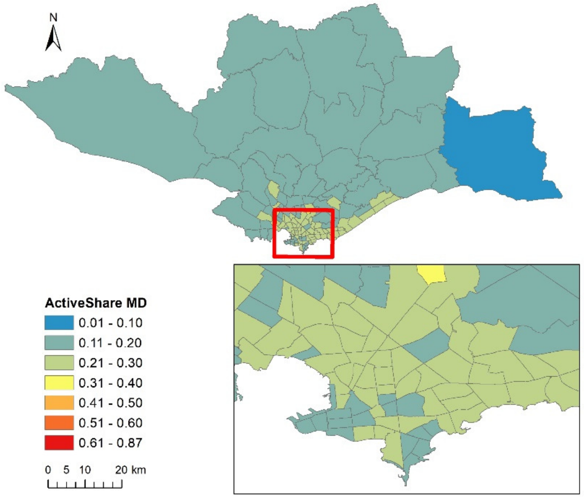


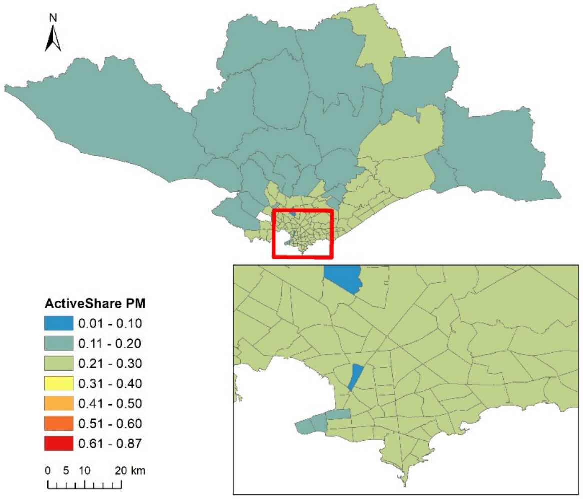
Cellphone transit mode shares by time of day & trip origin zone

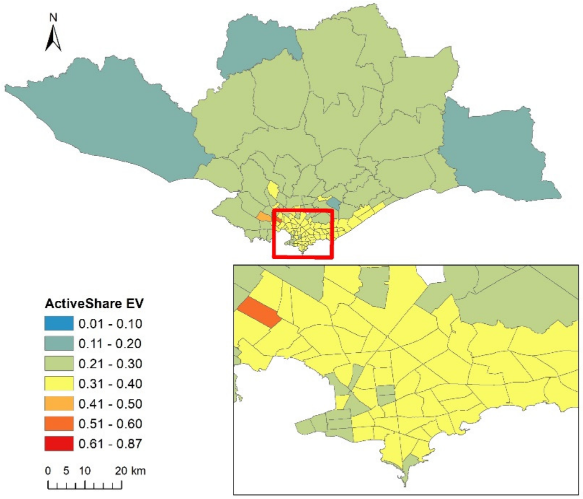


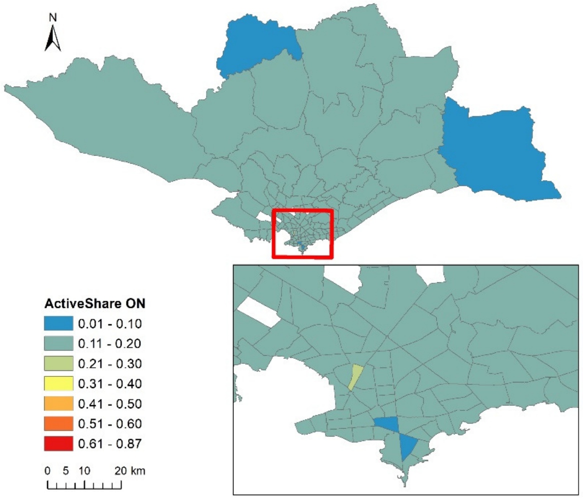
Cellphone active mode shares by time of day & trip origin zone











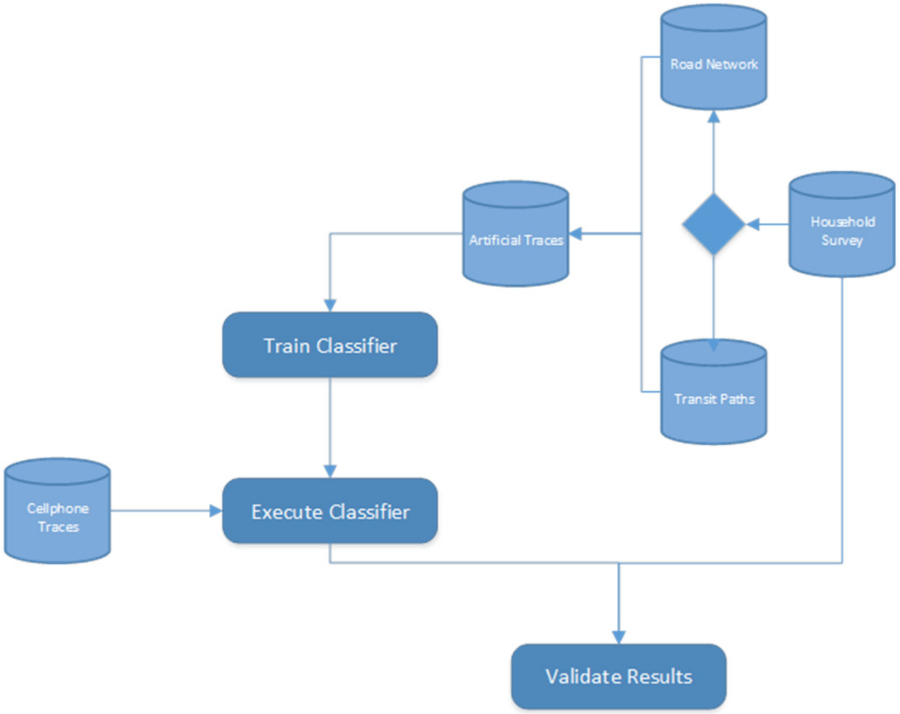
Validation Results (1)

- Aggregate numbers are then compared against the MHMS

Cell Traces		
Auto	Transit	Active
4083414	5159537	2515295
34.73%	43.88%	21.39%

MHMS (no intrazonal)		
Auto	Transit	Active
4250	2992	4461
36.32%	25.57%	38.12%

- It seems that the model has confusion between transit and active transportation however the split between auto and non-auto looks to be good.



Road Network

Artificial Traces

Household Survey

Transit Paths

Train Classifier

Cellphone Traces

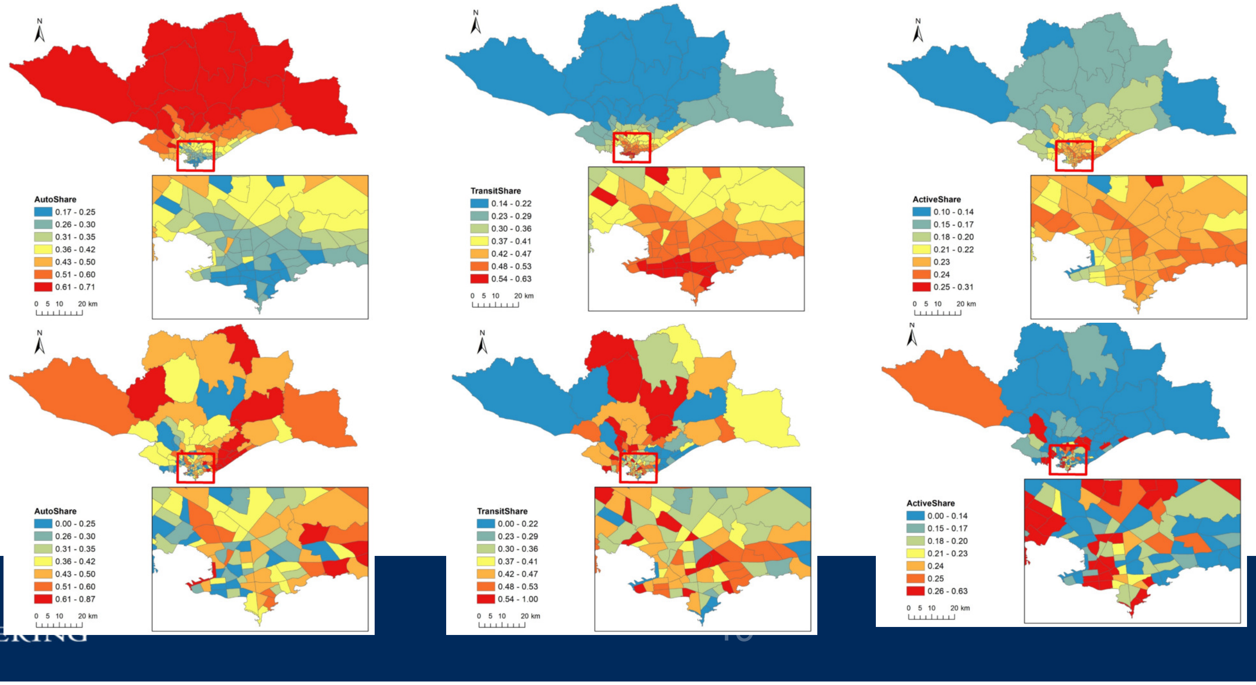
Execute Classifier

Validate Results

Validation Results (2)

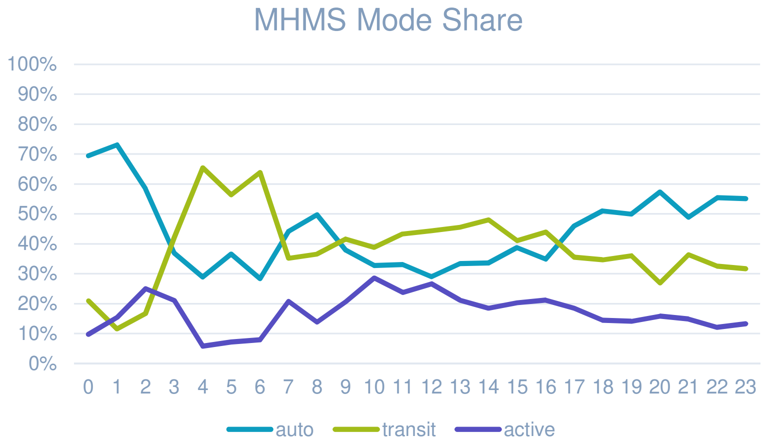
- Observed mode share by origin zone vs predicted
  - Exclude intrazonal trips in MHMS.
    - less than 0.16 for auto and transit and about 0.1 for active mode.

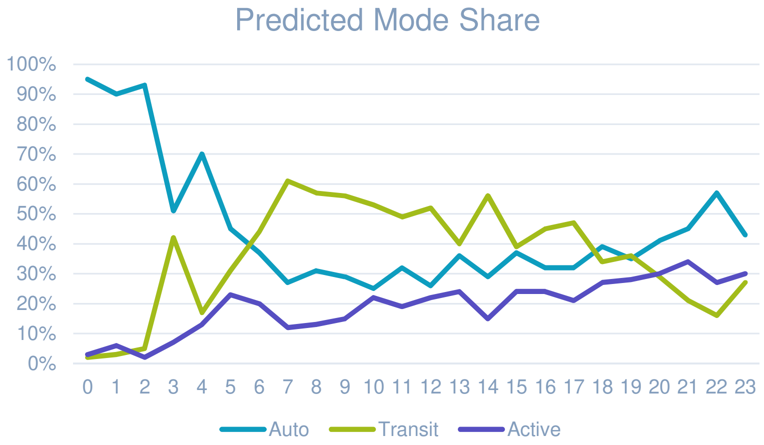
MODE SHARE	MAE
AUTO	0.1593
TRANSIT	0.1597
ACTIVE	0.1046



Validation Results (3)

- Observed mode share by time of the day vs predicted.



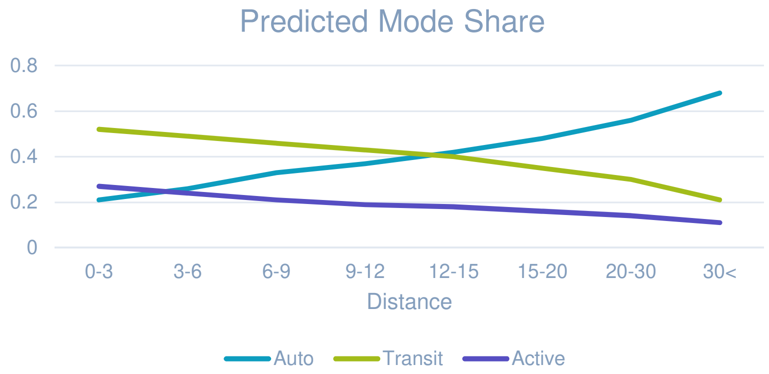


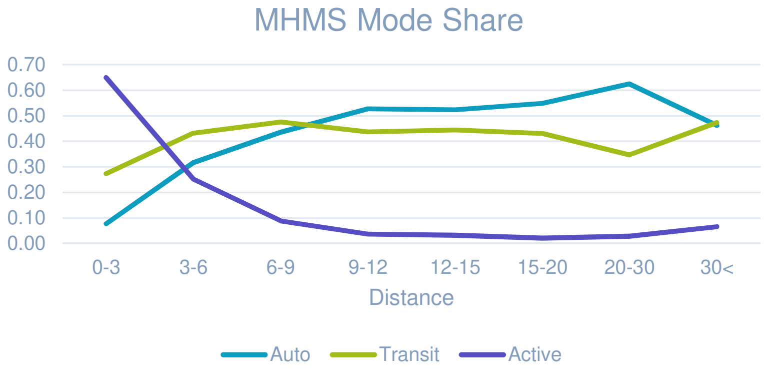
Observed	AUTO	TRANSIT	ACTIVE
AM	43%	42%	16%
MD	33%	44%	23%
PM	41%	40%	20%
EV	52%	34%	14%
ON	45%	44%	11%

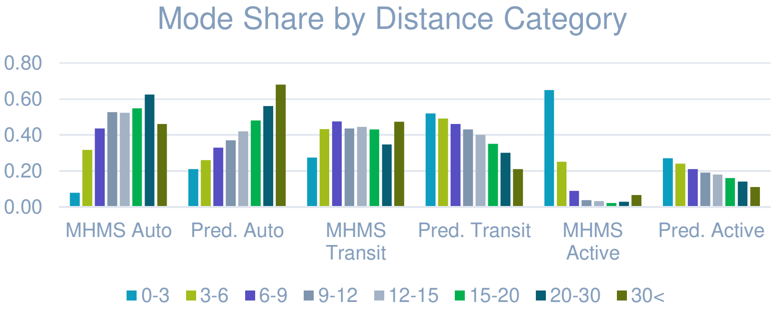
Predicted	AUTO	TRANSIT	ACTIVE
AM	31%	54%	14%
MD	29%	51%	20%
PM	35%	42%	24%
EV	42%	28%	30%
ON	63%	22%	15%

Validation Results (4)

- Observed mode share vs predicted by trip length.



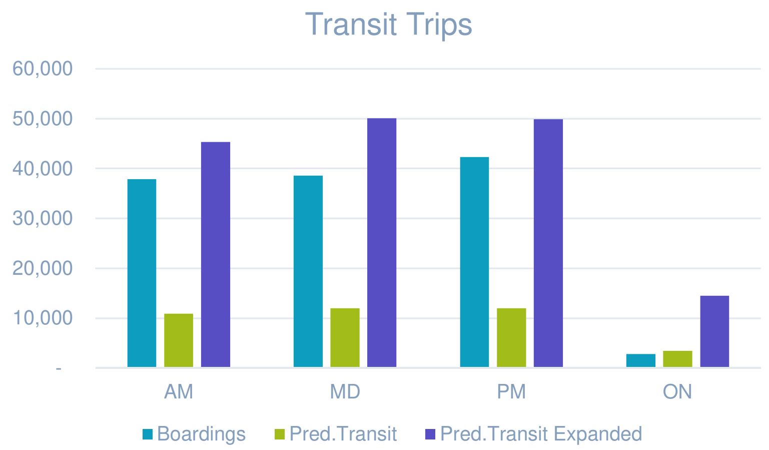


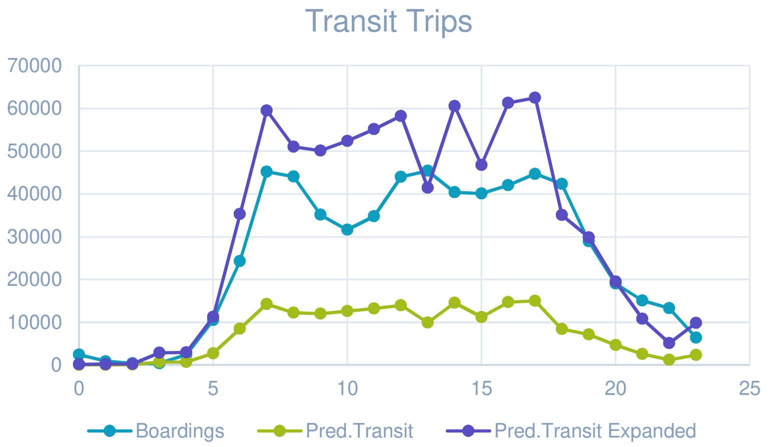


Short transit trips are over-predicted & short active trips are under-predicted.

Validation Results (5)

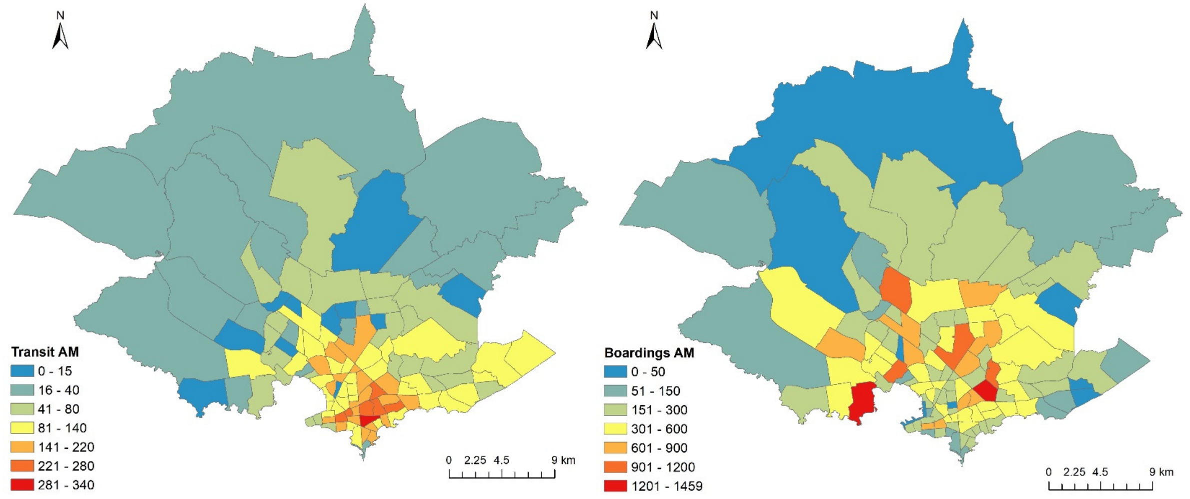
- Predicted cellphone transit trip vs smartcard data
  - Average peak hour trips.
  - For expansion of the Antel sample to the population, it is assumed:
    - 40% sample.
    - 60% market penetration.
- Temporal pattern is very good.
- Consistent with other results, it appears that we are over-estimating transit trips in the cellphone data.



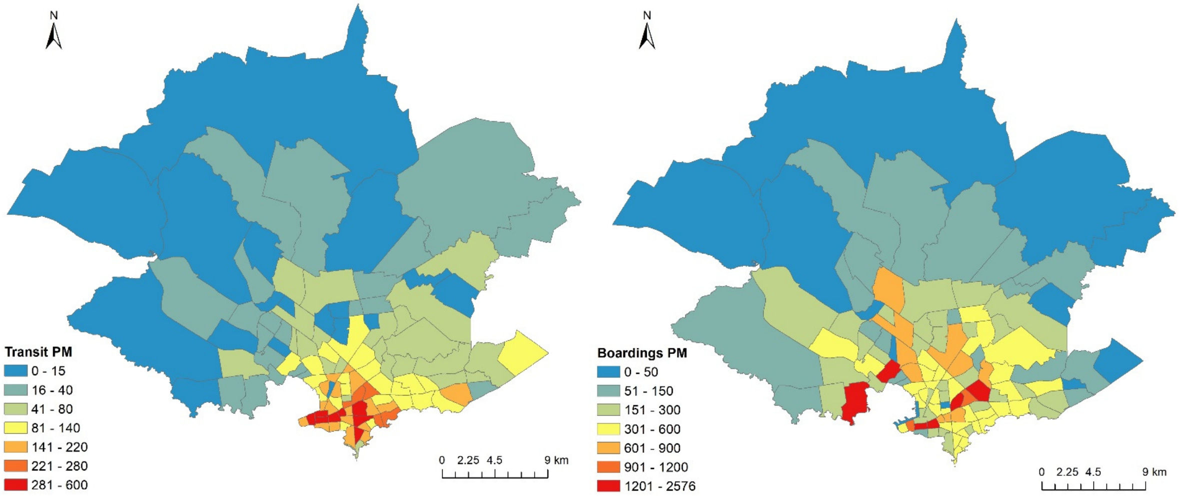


Validation Results (5)

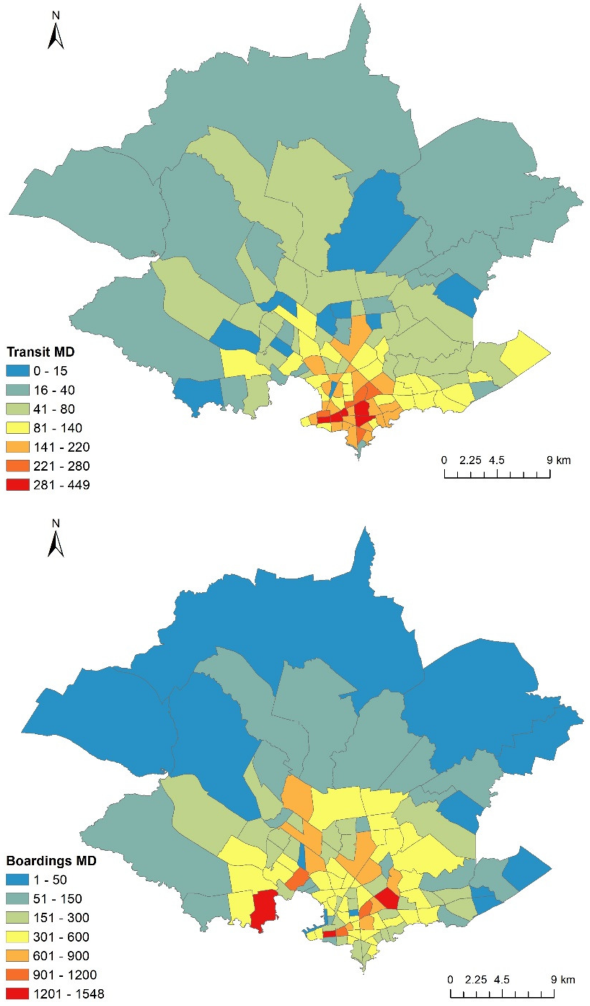
Cellphone inferred transit trips vs. smartcard trips, AM Peak



Cellphone inferred transit trips vs. smartcard trips, PM Peak



Cellphone inferred transit trips vs. smartcard trips, Mid-Day



Conclusions

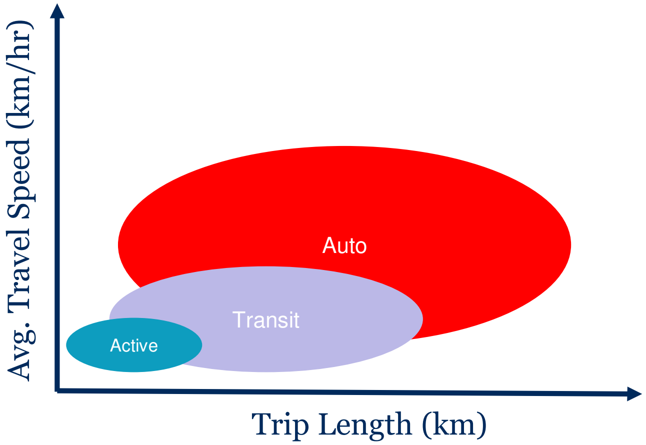
Next Steps?

Primary Study Conclusions

- O-D trip matrices can be constructed at the traffic zone level from both cellphone trace and smartcard data.
- Trip mode can be inferred for cellphone data using a neural network classifier trained on MHMS survey data.
  - This is a "first of the kind" result.
  - Classifier worked well for auto vs non-auto trips
- Home & work locations can be inferred for both cellphone trace and smartcard data, providing that records are available for a sufficient number of days to identify trip patterns.

Caveats & Further Analysis (1): Transit Mode Share

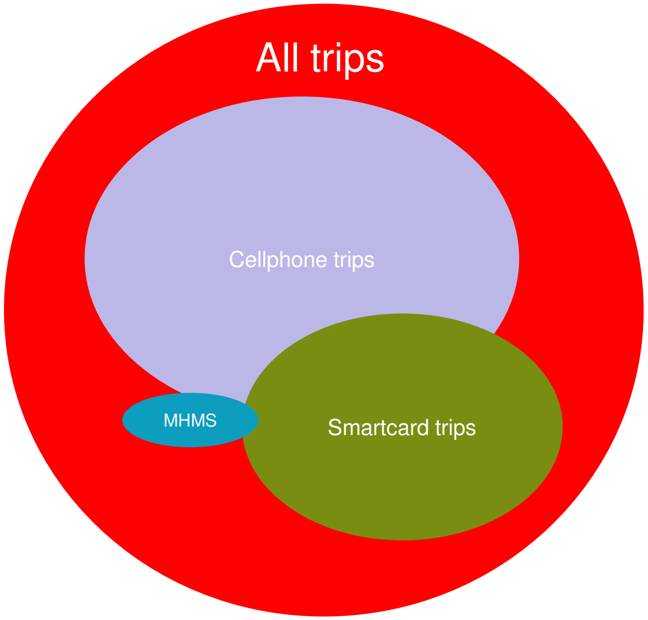
- Transit trips are currently being over-predicted for the cellphone traces; largely due to “confusion” with active trip-making.
  - This is a not unexpected result.
- NEXT STEP: We are now trying to add smartcard data to the analysis to improve transit trip classification.



Caveats & Further Analysis (2): Home & Work Locations

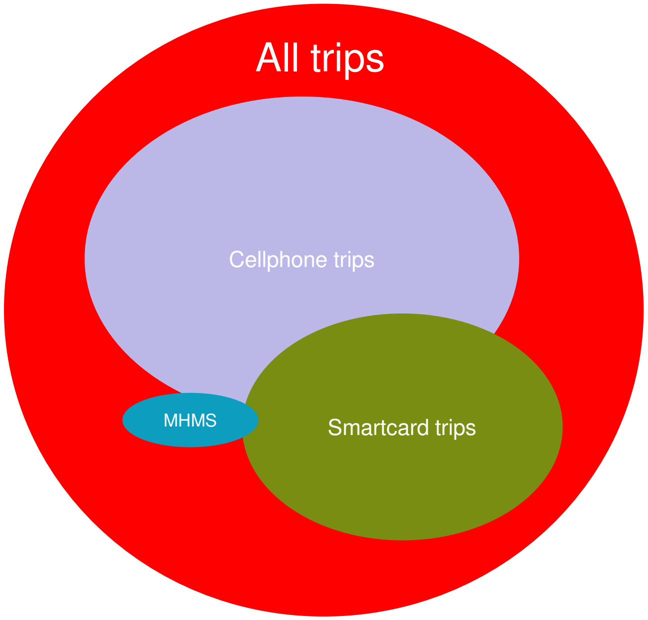
- Home & work locations currently are identified based on regularity of trip patterns within the cellphone traces & smartcard records.
- Not all cases can be identified with certainty.
- NEXT STEP: Use the records for which home (work) locations are identified with high confidence as a labelled training set to develop a neural net classifier that can be applied to the remaining record to improve their inferred home (work) locations.

Caveats & Further Analysis (3): O-D Trip Matrices



- MHMS, cellphone traces & smartcard records all provide O-D trip matrices.
- Each, however, is a sample from the actual population of trips.
- Each has its sampling errors:
  - MHMS is detailed & precise but very small sample (& people may under-report short and/or non-home-based trips).
  - Cellphone traces: Not all people have cellphones; not all people have cellphones on all the time; double counting for people two cellphones.
  - Smartcard records: Only transit trips; destinations need to be inferred.

O-D Trip Matrices, cont’d



- NEXT STEP: Attempt to combine all data to yield a best estimate of total trip making (by mode, time of day & trip purpose) for Montevideo.
  - Need to account for possible double-counting of trips across the datasets.

Thank you.

Questions?

a.faghihimani@utoronto.ca