

How We Travel: Insights from Household Travel Surveys and Digital Twin Modeling

Cinzia Cirillo
University of Maryland

IASS Seminar – April 21st, 2026



What is a Household Travel Survey?

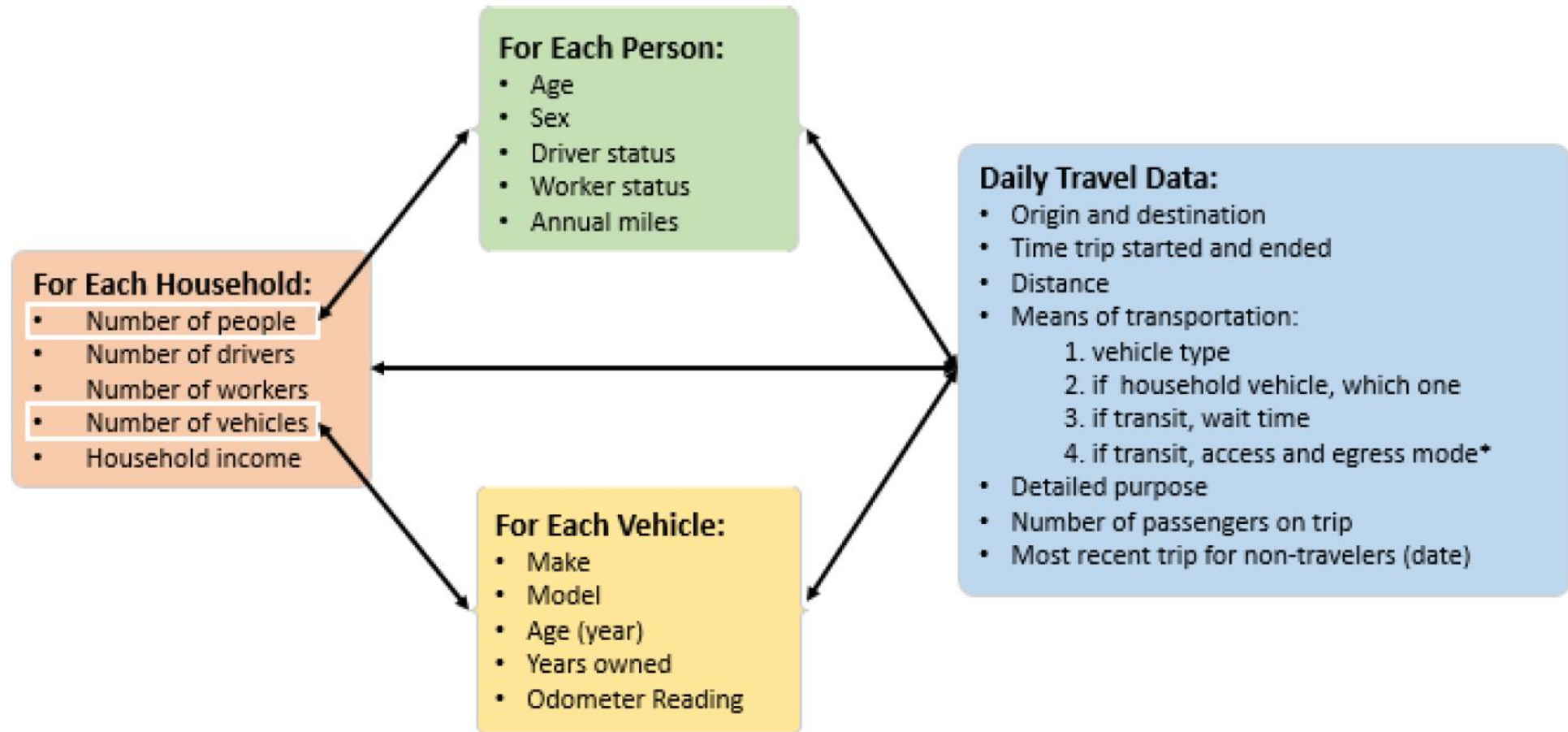
- Household travel surveys are critical tools used by transportation planners globally to understand travel behaviors, such as trip purpose, mode, and frequency, to inform infrastructure development and transportation policies.
- Major surveys exist in the US (NHTS), UK, Germany, France, the Netherlands, New Zealand, and Japan, often employing diaries or smartphone apps to track daily mobility,
- In the USA the National Household Travel Surveys (NHTS) is conducted by the Federal Highway Administration every 7-10 years.
- The NHTS is the official source of information about travel behavior in the USA.
- It is the only source of national data that allows one to analyze trends in personal and household travel.
- It includes daily travel by all modes, including characteristics of the people traveling, their household, and their vehicles.

[source: <https://nhts.ornl.gov/>]

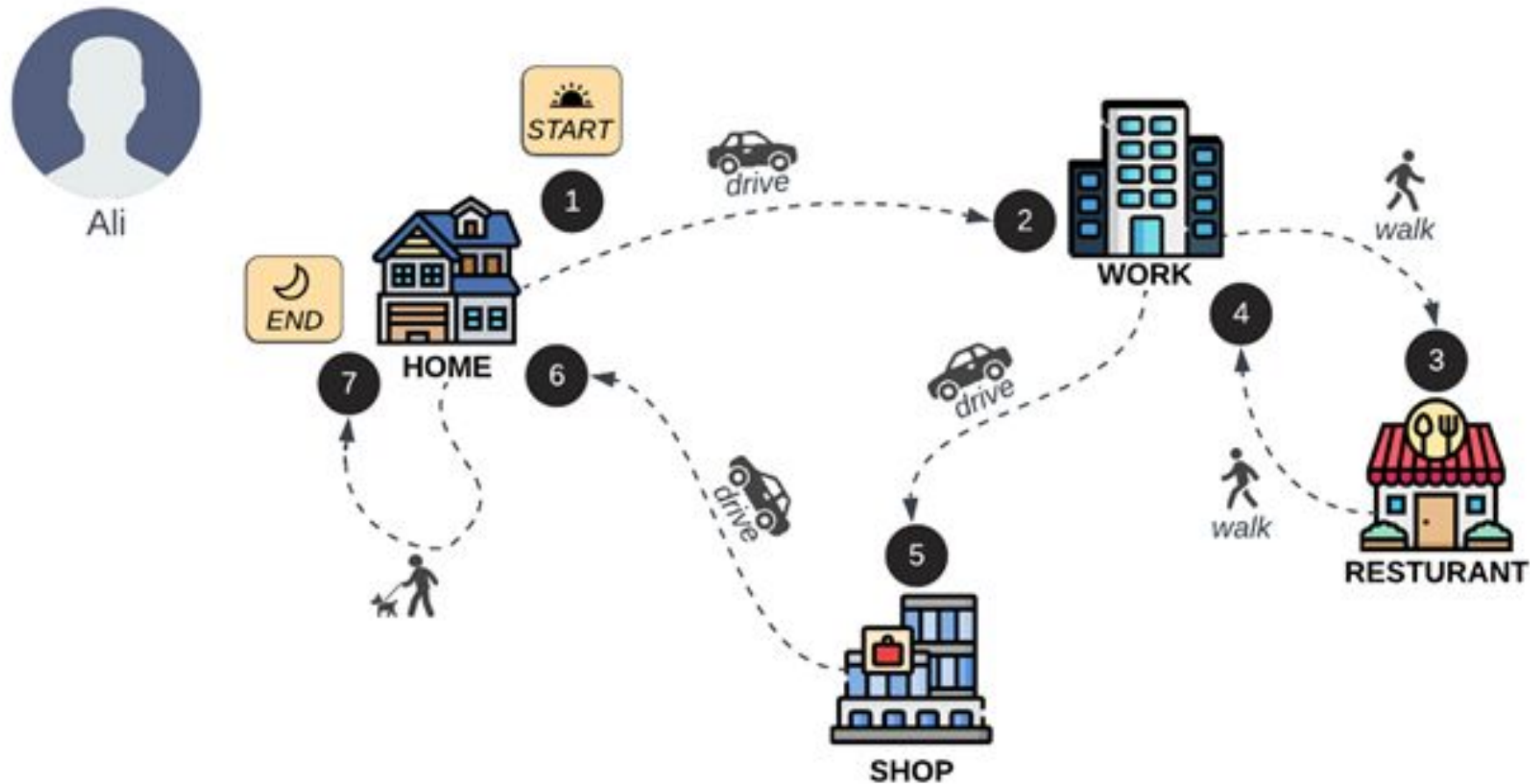
The Travel Survey over time

	Personal Travel Survey		Household Travel survey		
1969–1983	1990 NPTS	1995–2001 NPTS	2009 NHTS	2017 NHTS	2022 onwards
Face-to-face interviews (Census staff, recall-based)	RDD Telephone (CATI) Landline-only sample	Two-phase design: • Phone recruitment (RDD) • Mail travel diary	Coverage concerns: • Landline bias • Cell-only households excluded	Address-Based Sampling (ABS) Multi-mode: • Mail + Web + optional CATI	Address-Based Sampling (ABS) Multi-mode: • Mail + Web + Phone + optional CATI • Panel (every 2 years)

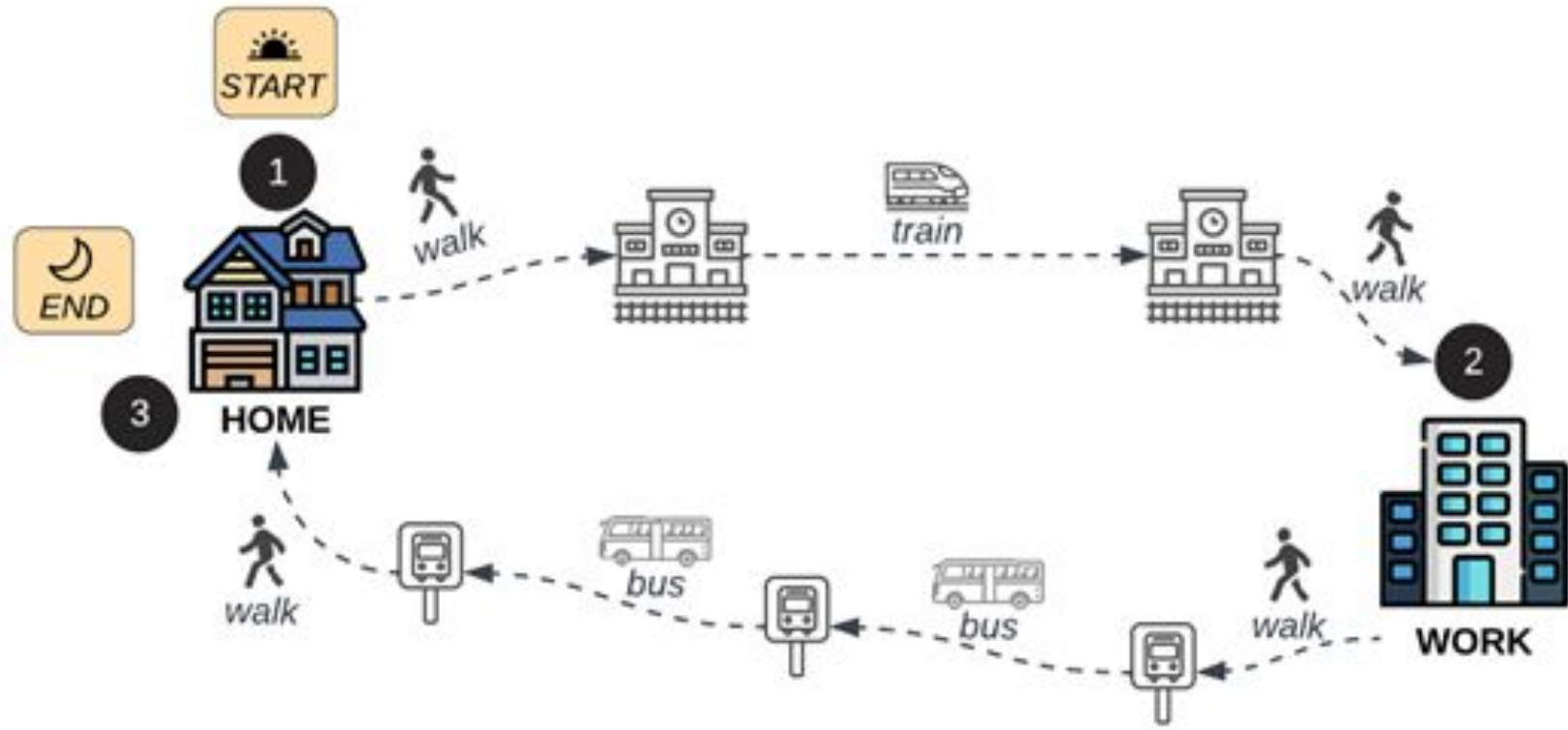
The data structure



Complex Activity Patterns - Tours



Use of Public Transportation



The Household Interview Survey (HIS) in Dubai

83% Expats

54% Male

48% are employed

- 92% White collar

Income

10% Low (0–9,999 AED)
19% Middle (10k–24.9k AED)
11% High (25k–54.9k AED)
5% Very High (55k+ AED)

Travel Behavior

Trips

Average individual makes **2.38 trips/day** and travels **~38 km**, with high variability in distance.

Car Dependency

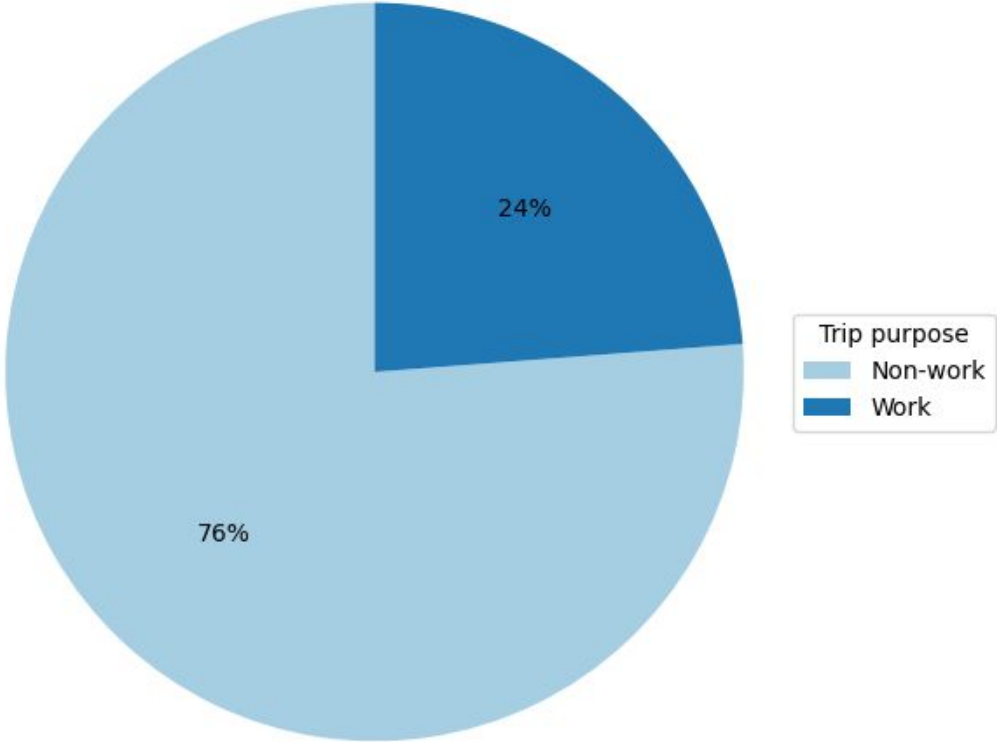
~81% own at least one vehicle, while only **~7%** use public transport.

Delivery Activity

Frequent and skewed activity averaging **5.7 deliveries/week**. Food and grocery are most common.

Work vs non-work Activities

Figure 4: Trip Purpose



Mode Choice

Figure 5: Transportation Mode Choice

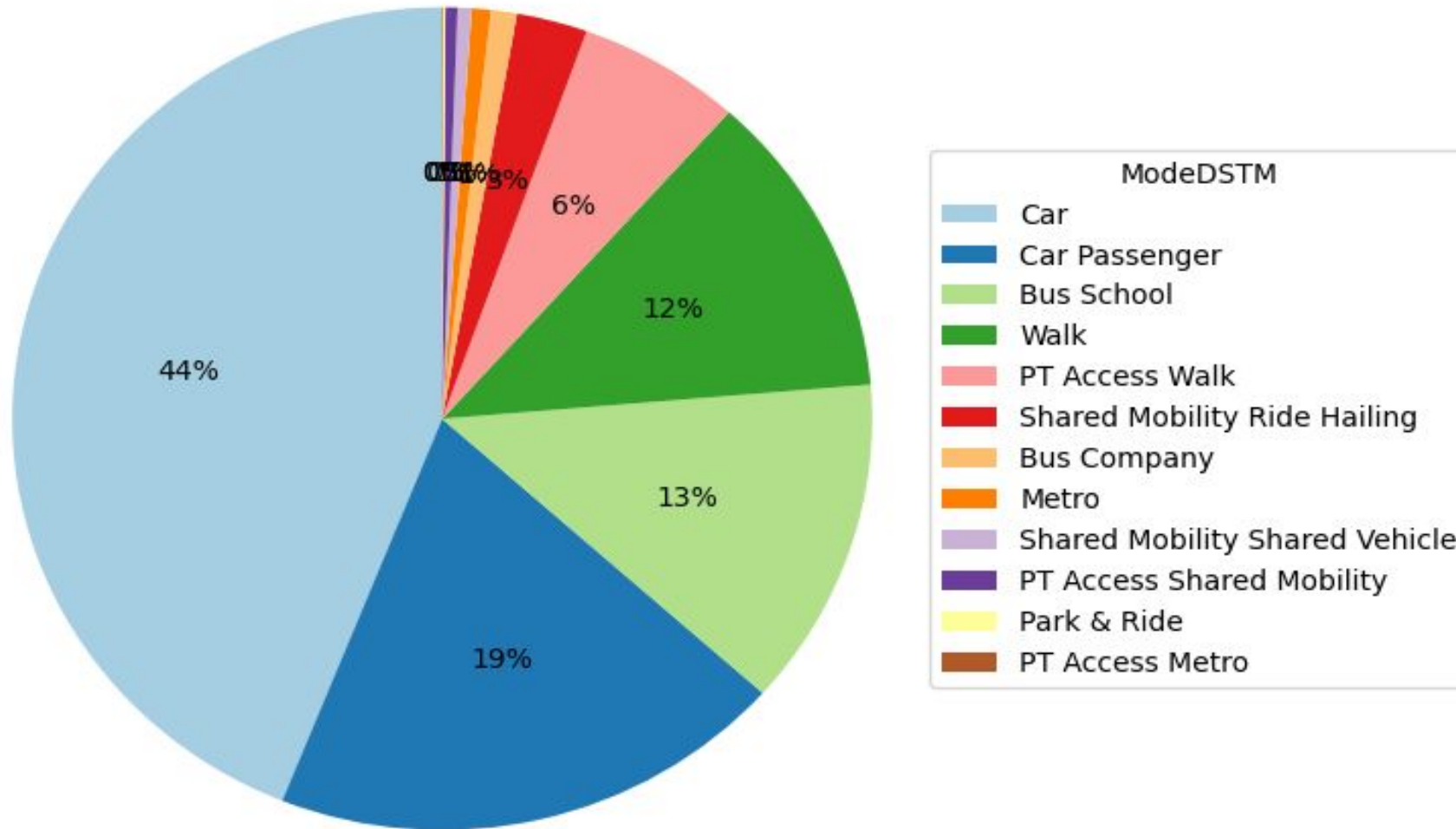


Table 3: Regression Results

	(1) OLS Num. of Trips	(2) Poisson Num. of Trips	(3) OLS Total Trip Dist. (km)	(4) OLS Vehicle Ownership	(5) OLS WFH	(6) OLS Made a Trip
Total Trip Distance (km)	0.003*** (0.0002)	0.001*** (0.0001)		0.0006*** (0.0001)	0.00005 (0.00005)	
Number of Trips				-0.009 (0.002)	0.014*** (0.003)	
Vehicle Ownership	-0.015 (0.026)	-0.005 (0.010)	17.16*** (1.280)		0.043*** (0.007)	
Deliveries Received in Past Week	0.007*** (0.002)	0.003*** (0.001)	0.208*** (0.076)	0.0062*** (0.001)	0.002*** (0.0004)	0.003*** (0.001)
Works from Home	0.207*** (0.039)	0.080*** (0.019)	3.410* (1.902)	0.0704*** (0.019)		-0.121*** (0.014)
Employed	0.134*** (0.021)	0.057*** (0.009)	19.26*** (1.019)	0.0228 (0.016)	0.123*** (0.005)	0.372*** (0.009)
Expat	0.075*** (0.027)	0.032*** (0.011)	-12.74*** (1.340)	-0.096 (0.019)	-0.090*** (0.007)	0.009 (0.010)
Student						0.414*** (0.017)
Male						0.067*** (0.007)
High Income	-0.064** (0.029)	-0.027** (0.012)	0.668 (1.41)	0.0832*** (0.018)	0.034*** (0.007)	-0.024** (0.010)
Very High Income	-0.038 (0.040)	-0.015 (0.017)	0.621 (1.97)	0.105*** (0.026)	0.053*** (0.010)	-0.002 (0.014)
Used Public Transport	-0.308*** (0.036)	-0.135*** (0.010)	-0.898 (1.75)	-0.532*** (0.020)	0.010 (0.009)	
Blue-Collar Occupation	-0.027 (0.034)	-0.011 (0.013)	-4.10** (1.647)	-0.202*** (0.013)	-0.036*** (0.009)	0.039*** (0.013)
Constant	2.296*** (0.053)	0.834*** (0.022)	24.116*** (2.585)	0.815*** (0.021)	-0.059*** (0.015)	0.334*** (0.023)
Observations	9,849	9,849	9,849	3,666	9,849	12,551
R^2 / Pseudo R^2	0.052	0.017	0.075	0.308	0.080	0.236

Notes: Standard errors in parentheses. Model (2) is Poisson with robust (HC3) standard errors. OLS models use conventional standard errors. Low income and Age 65+ are omitted reference categories. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.



Regression Results

- Deliveries → Number of Trips
 - Positive and highly significant ($p < 0.01$):
 - OLS: 0.007
 - Poisson: 0.003
- Deliveries → Travel Distance
 - Positive and significant (0.208 km per delivery)

Vehicle Ownership and Key Controls

- Deliveries → Vehicle Ownership
 - Positive and highly significant (0.0062)
 - Households with more deliveries are more likely to own cars
- Important control variables:
 - Employment: increases trips (0.134), distance (19.26 km)
 - WFH: increases trips (0.207), but reduces probability of making a trip
 - Public Transport Use: strongly reduces trips (−0.308)
 - Income: higher income → higher vehicle ownership
- More mobile, higher-income households have both more deliveries and more travel

Interpretation

- Overall: More deliveries → more travel
- An increase in deliveries results in:
 - Increased number of trips
 - Increased travel distance
 - Increased vehicle ownership
- Takeaway: e-commerce complements, not substitutes, travel

Table 4: Effects of Delivery Type on Travel Outcomes

	(1) OLS Num. of Trips	(2) Poisson Num. of Trips	(3) OLS Total Trip Dist. (km)	(4) OLS Vehicle Ownership	(5) OLS WFH	(6) OLS Made a Trip
Parcel	0.002 (0.005)	0.001 (0.002)	0.190 (0.232)	0.012*** (0.002)	0.007*** (0.001)	0.004** (0.002)
Food	0.010** (0.005)	0.004** (0.002)	0.350 (0.224)	0.004** (0.002)	0.002 (0.001)	0.003 (0.002)
Grocery	0.008* (0.004)	0.003* (0.002)	0.136 (0.202)	0.004** (0.002)	-0.001 (0.001)	0.003** (0.001)
Other	-0.010 (0.015)	-0.004 (0.006)	-1.013 (0.750)	0.034*** (0.006)	0.019*** (0.004)	-0.016*** (0.005)

Notes: Each column corresponds to the same specification as in Table 3, but replaces total home deliveries with delivery type indicators. Standard errors in parentheses. Model (2) is Poisson with robust (HC3) standard errors. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Delivery Type Differences

- Food and grocery:
 - Stronger link to more trips
- Parcel deliveries
 - Stronger link to vehicle ownership

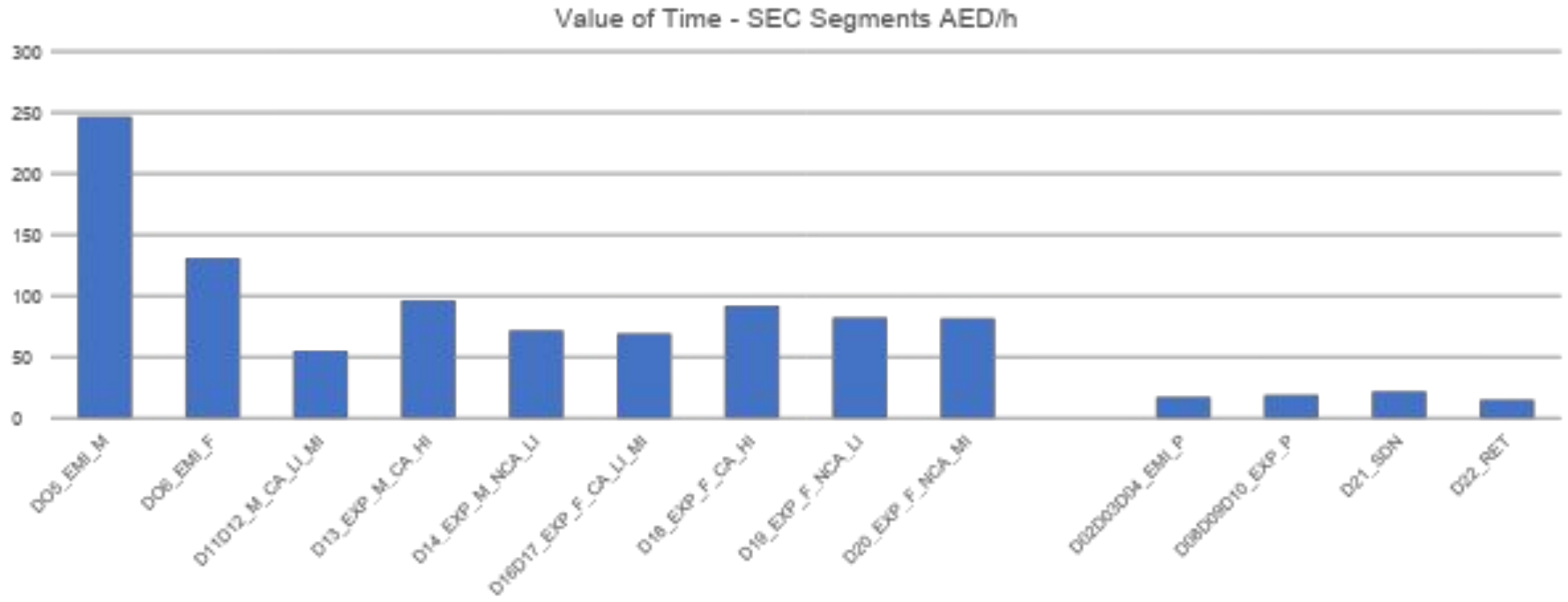
Elasticities

- 10% ↑ deliveries → 0.28% ↑ trips
- 10% ↑ deliveries → 0.51% ↑ travel distance

What is the Value of Travel Time Savings

- This is a measure of the willingness to pay to save a unit of travel time (i.e. \$/h)
- It is usually the outcome of discrete choice models.
- It is used for projects appraisal, to measure the benefits deriving from a reduction in travel time resulting from investment in transportation.

VTTs by Socio-Demographic Characteristics



VTTS by mode of transportation



Population Synthesis and Microsimulation



UNIVERSITY OF
MARYLAND

A. JAMES CLARK
SCHOOL OF ENGINEERING

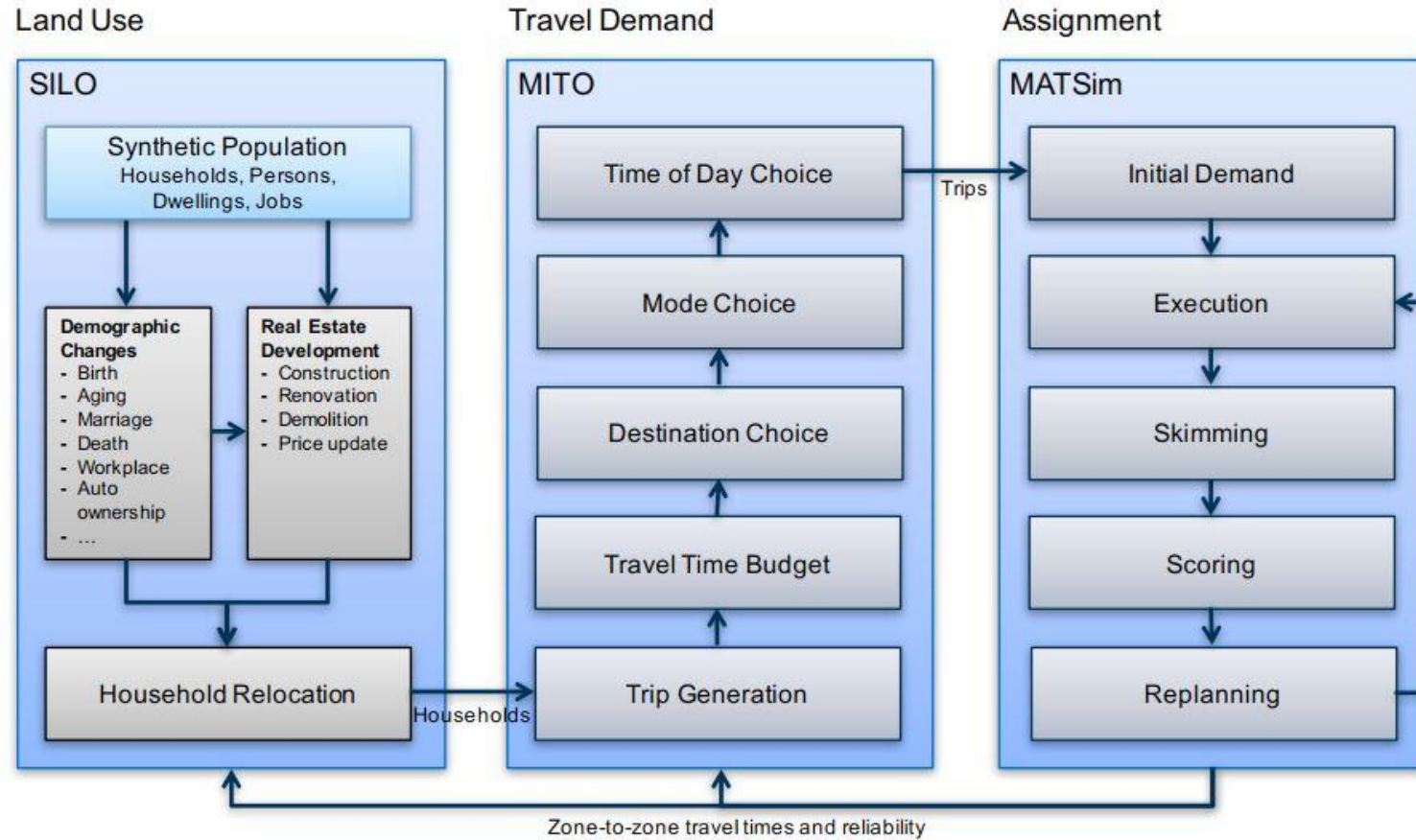


CMMM

Center for Multi-Modal
Mobility in Urban, Rural
and Tribal Areas

Modeling Framework

Open-sourced land-use, travel demand model and transport simulator



Source: <https://www.msm.bgu.tum.de/en/research/modeling/integration/>



CMMM

Center for Multi-Modal
Mobility in Urban, Rural
and Tribal Areas

Simple Integrated Land Use Orchestrator (SILO)

Powerful land use model; can be integrated with travel demand models

- Synthetic Population
 - ACS PUMS data used, 5 percent sample used for increased spatial resolution
 - Updates population, households, dwellings and jobs annually.
 - Socio-demographics: age, gender, drivers license, income, employment type, hhsiz, hhveh, dwelling type.

Demography

- Aging, birth, leave parental household, marriage, divorce, income, employment change, death
- Modeled using Markov models that apply transition probabilities to select whether events happen or not

Real Estate

- Build new dwelling, renovate, deterioration, demolish dwellings

Household relocation

- Logit model
- Simulation - decision, first selects a region, and then a dwelling within that region is chosen.

Travel Demand

- Travel demand models can be integrated with SILO



CMMM

Center for Multi-Modal
Mobility in Urban, Rural
and Tribal Areas

Microscopic Transportation Orchestrator (MITO)

Hybrid trip-based model that combines the strengths of trip-based and activity-based models

- **Trip Generation**

- Monte Carlo Sampling
- Microscopic trip generation model, trips by purpose are generated individually for each household
- Used observed trip frequencies of household travel survey

- **Travel Time Budgets**

- Estimates the total time by purpose that households allocate to travel within a day

Destination Choice Model

- Trips are assigned to persons within the household using a rule-based approach
- Logit model used

Mode Choice model

- MNL models by purpose for drive alone, car passenger, bus, train, bike, walk and shared ride.

Time of the Day Choice

- Empirical distribution of arrival times and duration for purposes, from RTS
- Times are selected by means of a probabilistic choice; resolution- 1 min



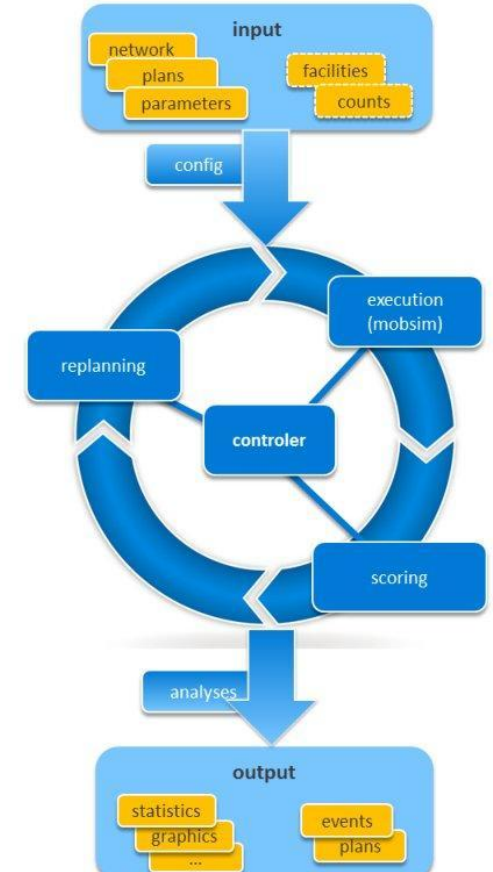
CMMM

Center for Multi-Modal
Mobility in Urban, Rural
and Tribal Areas

Multi-Agent Transport Simulation (MATSim)

An open-source, agent-based transport simulation framework.

- Designed for large-scale, high-resolution transportation modeling
- Simulates individual travelers and their daily activity plans
- Multi-modal transportation support
- The probability of choosing a plan is based on **logit choice mode**
- Queue-Based Model for Traffic Flow Model
- Dijkstra's Shortest Path Algorithm for Route Choice Model
- Re-planning Strategies – Time Mutation, Route Choice, Mode Choice, Sub-tour Mode Choice

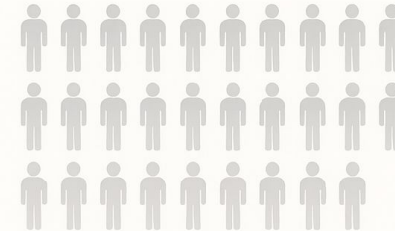


Sparse Datasets, Dense Populations

- Agent-based Modeling enables simulation of complex and complicated systems (Traffic, Land Use, Growth, etc.)
- Surveys do not provide enough data (at face-value) for population-scale number of agents
- Data for many-attributed population from surveys is sparse (lots of sampling zeroes)
- We need a realistic input population for agent-based modeling simulations to be useful

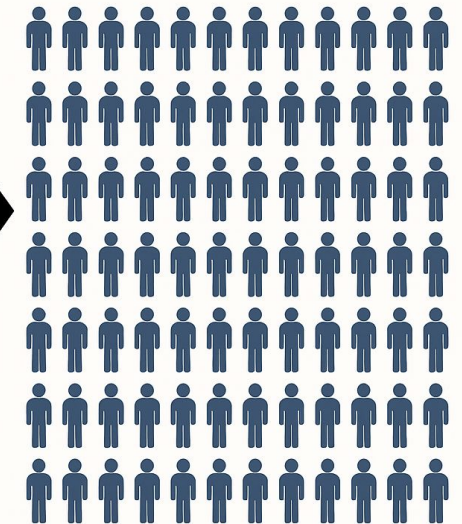
Sparse Survey Data

Age	Income	Car
35	\$50.000	No
44	\$80.000	Yes
29	\$35.000	Yes
52	\$50.000	No
40	\$75.000	Yes



Sparse Survey Data

Synthetic Population



Simulation



Population Attribute Data Sources

American Community Survey (ACS):

- Approx 3.5mil addresses sampled annually
- Attributes collected include ancestry, citizenship, educational attainment, income, language proficiency, migration, disability, employment, and housing characteristics

National Household Travel Survey (NHTS):

- Household-level attributes include number of household members, income, number of vehicles, housing characteristics, land use
- Person-level attributes include age, gender, work status, has license, etc.

Census:

- 10-question questionnaire sent to every US household every 10 years
- Information collected includes number of household members, presence of children, gender, age, race, etc.

IRS:

- Income Data not available in Census
- Number of people in each ZIP Code in a particular income band is available

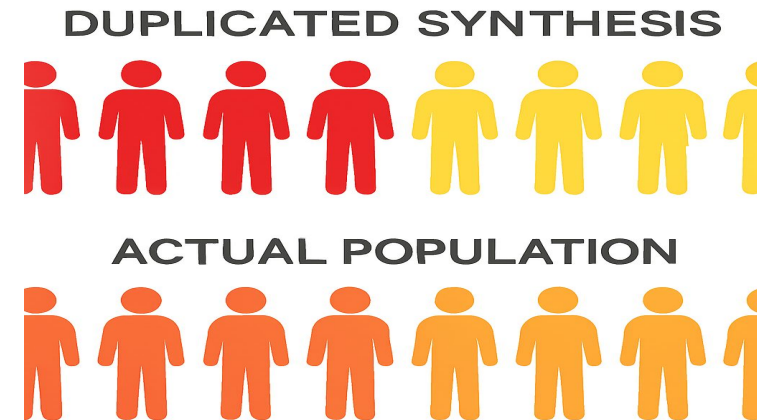
Duplication

Naïve solution is to just duplicate the population according to the survey weights

This yields a full-scale synthetic population with exact matches for joint, marginal, and conditional distributions

But this output doesn't "interpolate" (e.g. if a lot of people surveyed are 21, 22, 24, and 25 years old (but none are 23), we'd reasonably expect a lot of people in the actual population to be 23, but a duplicated survey dataset will not contain any 23-year-olds

Moreover, if we wanted to eventually run simulations from the data, duplication of trip information would yield overly "peaky" simulation conditions (e.g. everyone taking the same routes at the same times)



Bayesian Network: Motivation and Procedure

- Bayesian Networks model joint distributions as dependency graphs.
- Factorize high-dimensional spaces efficiently.
- Avoid exponential table growth from IPF.
- Capture non-linear and multiway dependencies among attributes.
- Generate unseen but plausible attribute combinations.
- Handle sparse samples by leveraging conditional independence.

Bayesian Network: Procedure

Two learning stages:

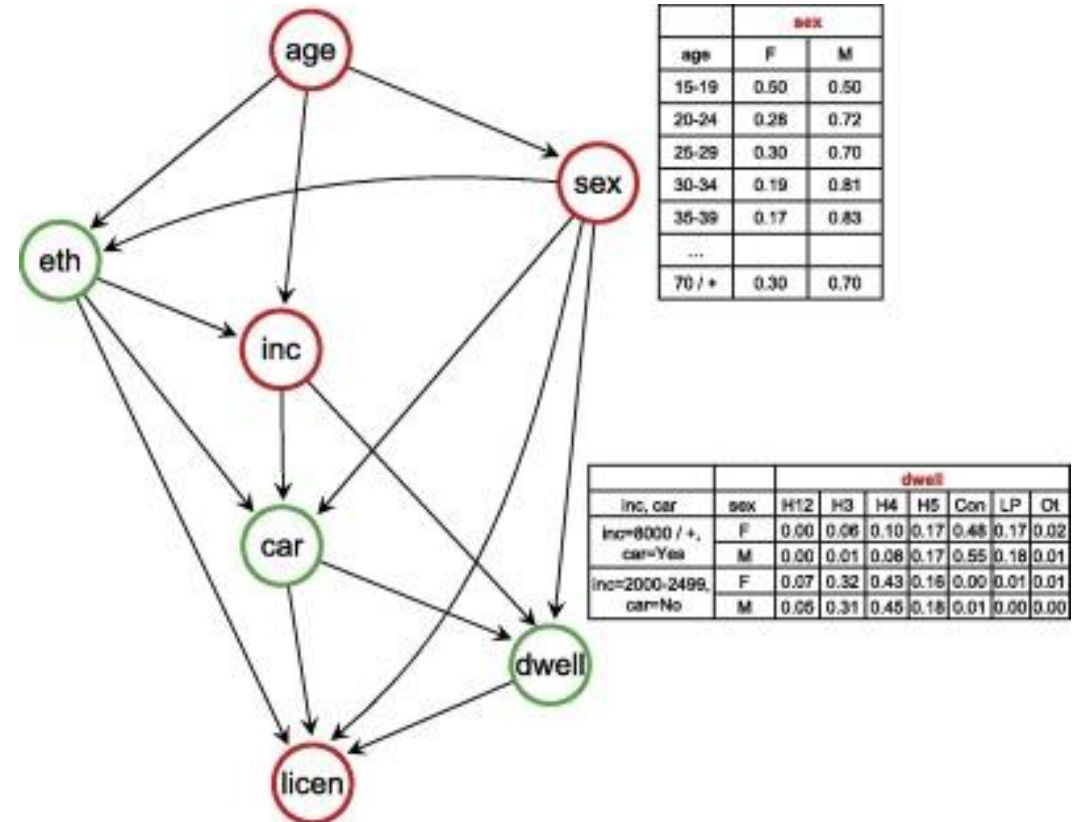
Structure learning – figure out which variables depend on which.

- Score-based algorithms (e.g., Bayesian Information Criterion).
- Constraint-based algorithms (e.g., conditional independence tests).
- Or manual expert design (e.g., socio-demographic logic).

Parameter learning – estimate each node's CPT from data.

- For discrete variables → frequency counts.
- For continuous variables → regression or Gaussian models.

The result: a full probabilistic model describing how all attributes co-vary.



Integration of Copula concept

- Copula = function linking marginals + dependencies (Sklar 1959).
- Separates marginal shapes from dependency structure.
- Solves issue of ML models reproducing sample marginals only.
- Allows combining ML-learned dependencies with census marginals.
- Enables reuse of dependency patterns across regions.
- Key idea: learn dependency once, apply different marginals.

Copula procedure

- Step 1: transform data \rightarrow uniform $[0,1]$ via CDF.
- Step 2: train Bayesian model on these ranks.
- Step 3: sample new uniform vectors from model.
- Step 4: map back via inverse CDFs of target marginals.
- Guarantees correct one-dimensional marginals + realistic dependence.
- Powerful framework combining ML + statistical control.



CMMM

Center for Multi-Modal
Mobility in Urban, Rural
and Tribal Areas

Predicting Transit Ridership of Purple Line





CMMM

Center for Multi-Modal
Mobility in Urban, Rural
and Tribal Areas

Motivation

- Commonly techniques include travel demand models, professional judgment, service elasticity analysis, and regression models
- Demand model approaches are prevalent for proposed transit lines
- A significant challenge in estimating the sub-models- trip generation, attraction, and mode choice patterns in the area serviced by the proposed transit line.
- Traditional four-step models and activity/agent-based models can lead to varying estimation results because of their aggregate and disaggregate nature.



CMMM

Center for Multi-Modal
Mobility in Urban, Rural
and Tribal Areas

- MDOT is constructing the Purple Line within Prince George's and Montgomery County
- 16.2 miles line will have 21 stations and connect 4 metro lines of Washington DC
- Improve accessibility, connects to commuter rail and local transit
- **What will be the ridership for Purple Line?**

Purple Line



Photo credit:
https://www.purplelinemd.com/images/pl_opengraph.jpg

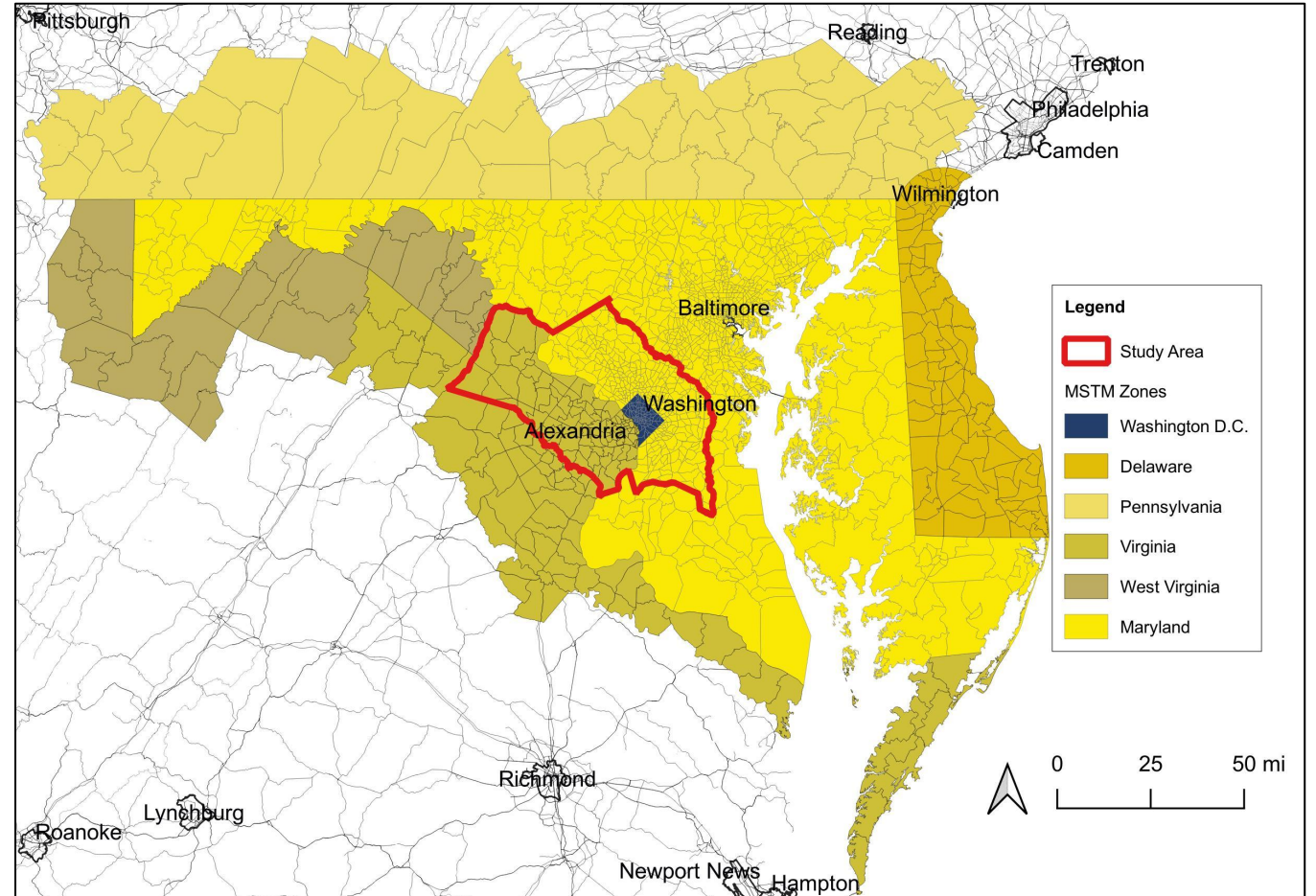


CMMM

Center for Multi-Modal
Mobility in Urban, Rural
and Tribal Areas

- SILO and MITO (for Travel demand model)
 - Maryland Statewide Transportation Model Zones
 - Maryland, Washington D.C. and Part of West Virginia, Virginia, Pennsylvania, Delaware
- MATSim (For scenario testing)
 - DMV Region
 - Washington D.C., Maryland and Virginia

Study Area





CMMM

Center for Multi-Modal
Mobility in Urban, Rural
and Tribal Areas

Modeling Framework

SILO

Synthetic
population:
2016

Simulation years:
2016 to 2035

2019 base-year
scenario
• Model validation

2027 Purple Line
scenario
• Syn pop generation

MITO

2019 base-year
scenario
• Model validation

2027 Purple Line
scenario
• Trips data

MATSim

2019 base-year
scenario
• Model validation

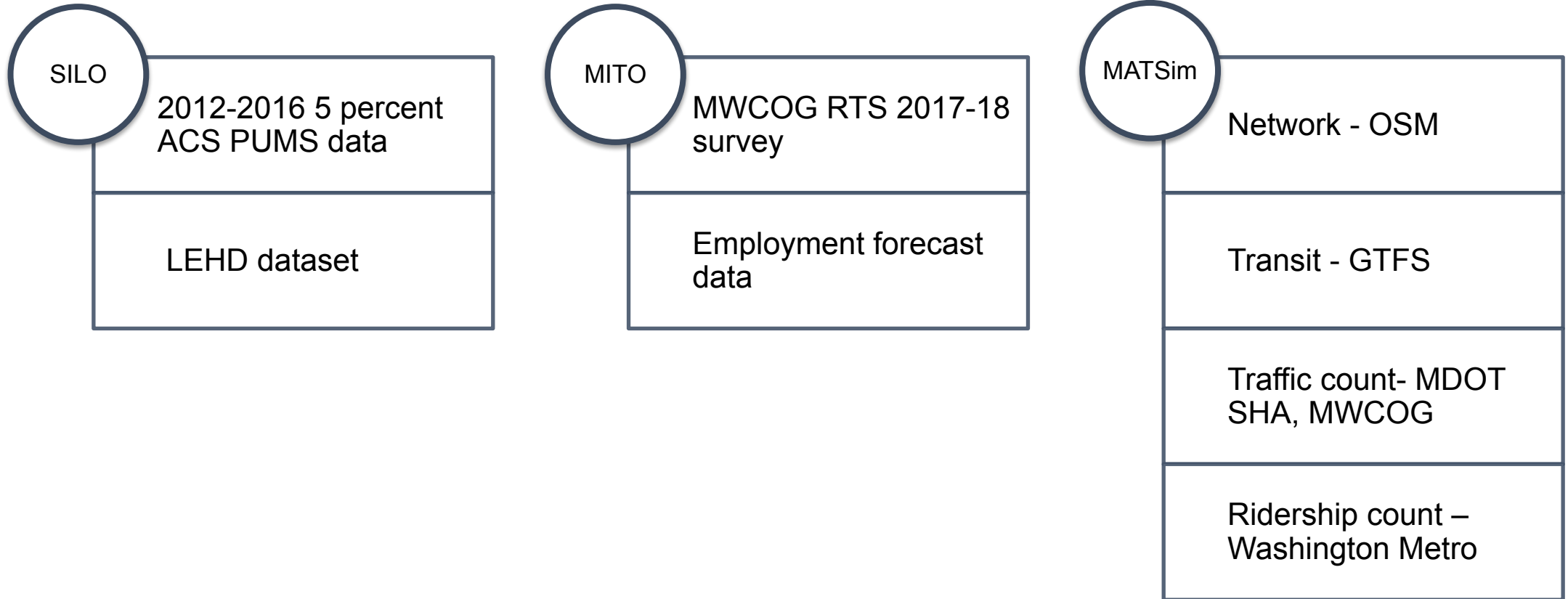
2027 Purple Line
scenario
• **Ridership Prediction**



CMMM

Center for Multi-Modal
Mobility in Urban, Rural
and Tribal Areas

Major Datasets





CMMM

Center for Multi-Modal
Mobility in Urban, Rural
and Tribal Areas

	Persons - 11,971,250
	Households 4,844,716
	Dwelling - 5,533,306

Base Scenario - SILO

Comparison of synthetic population data

	Synthetic Population	PUMS 2016	Difference	Area
Person	5,511,337	5,512,585	-0.02%	Maryland
household	2,176,757	2,177,487	-0.03%	Maryland

Comparison of SILO simulation and PUMS data for 2019

	SILO 2019	PUMS 2019	Difference	Area
Person	5,627,993	5,577,523	0.90%	Maryland
Household	2,265,621	2,205,203	2.73%	Maryland

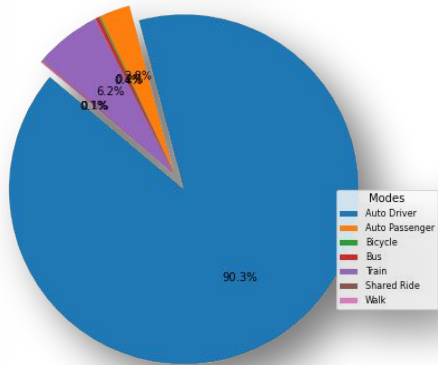


CMMM

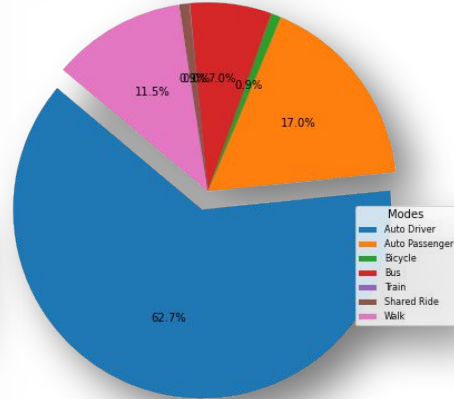
Center for Multi-Modal
Mobility in Urban, Rural
and Tribal Areas

Total Trips - 44,713,290

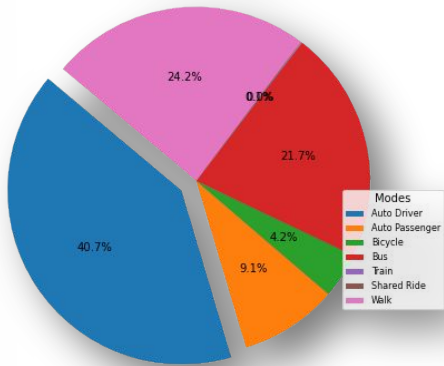
Mode Share_HBW



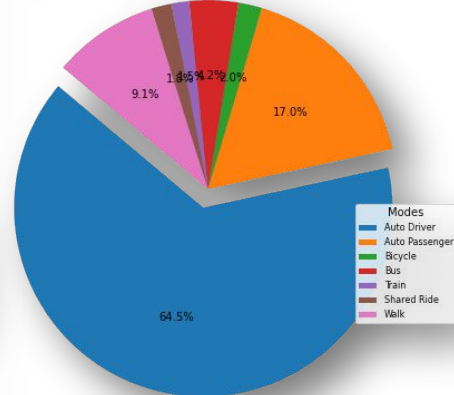
Mode Share_HBO



Mode Share_NHBO

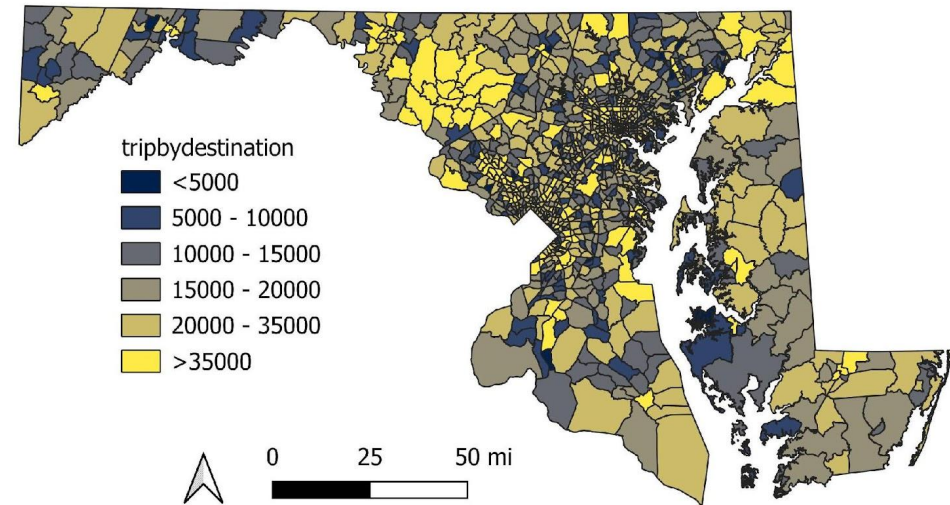
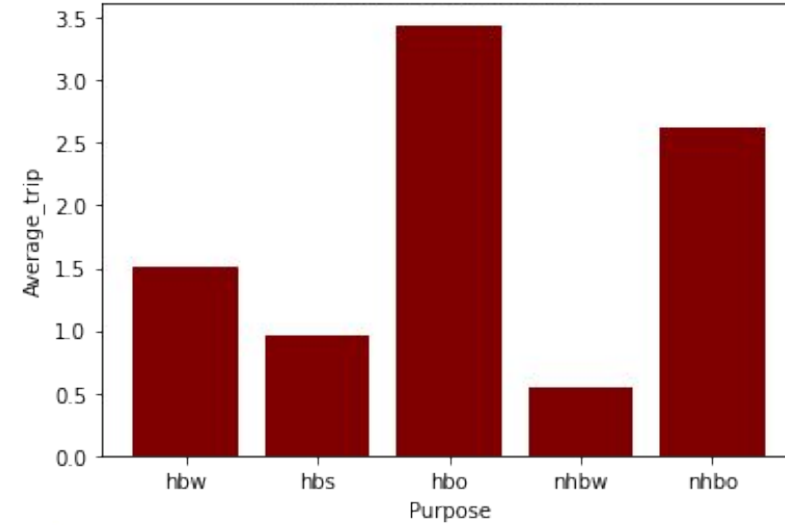


Mode Share_HBS



Base Scenario - MITO

Average household trips



Trip Attraction by zones

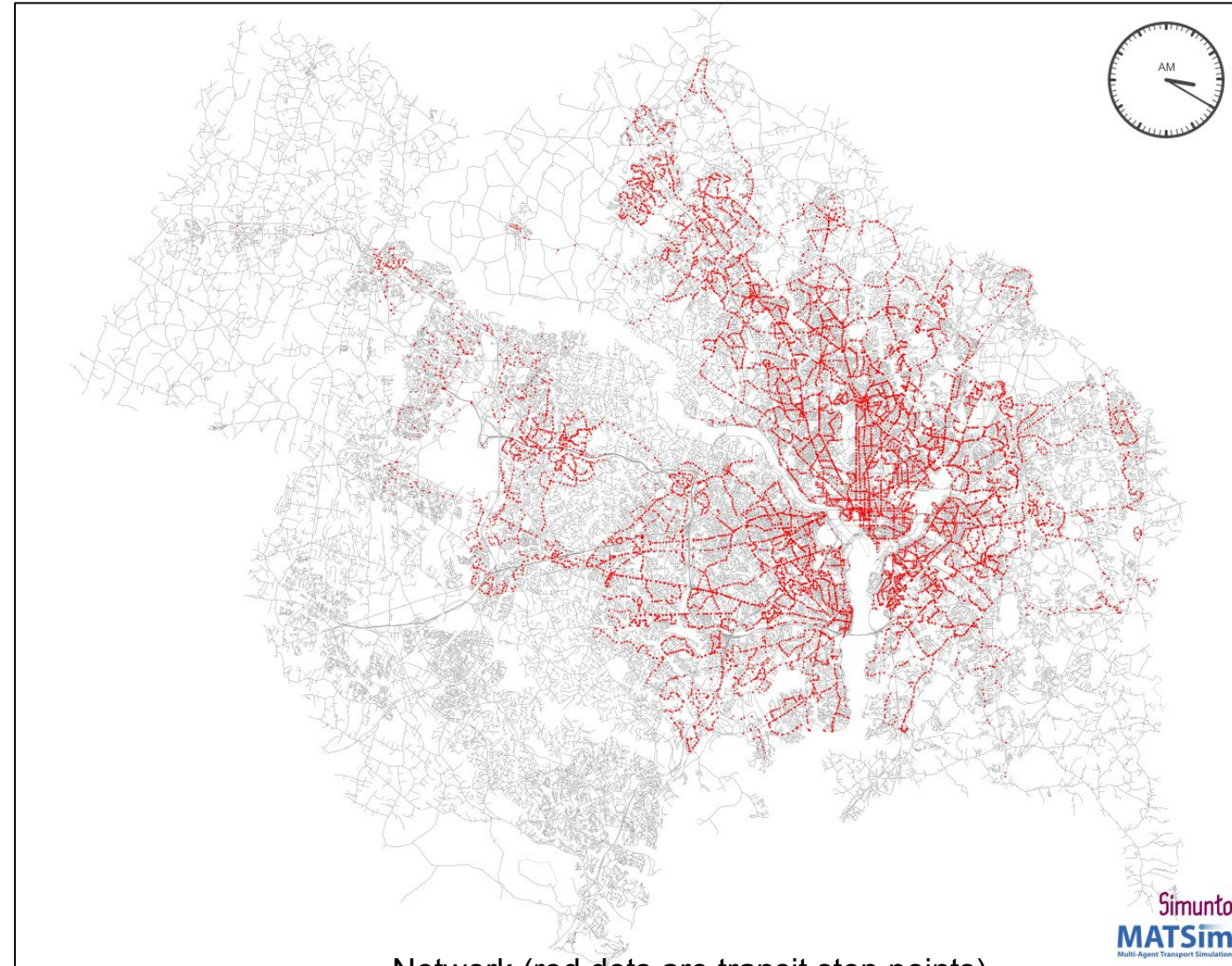


CMMM

Center for Multi-Modal
Mobility in Urban, Rural
and Tribal Areas

- Network - DMV (Washington D.C., Maryland and Virginia) network from OSM
 - 192,047 nodes and 424,799 links
- Simulation Run 10 percent of population
- Education trips excluded
- Only intra-zonal trips included
- Flow Capacity and Storage Capacity scaled down

Base Scenario - MATSim



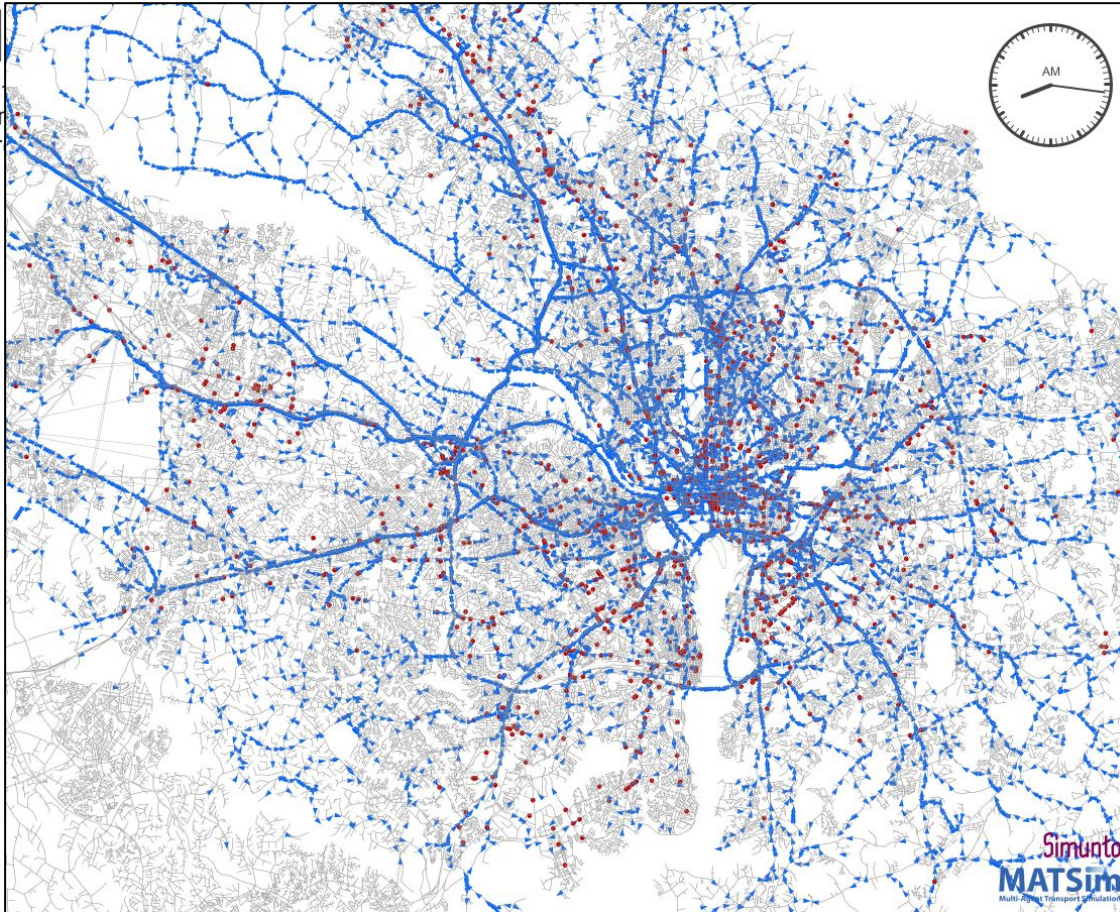
Network (red dots are transit stop points)



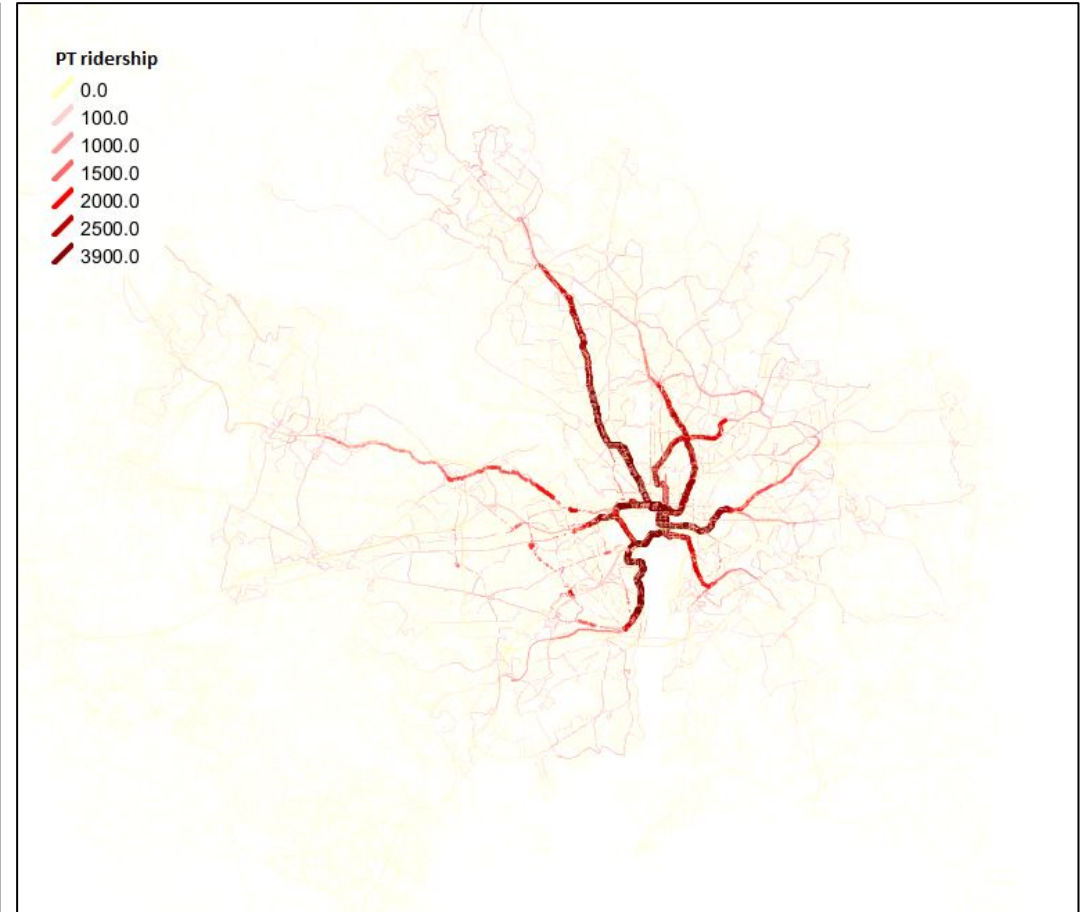
CMM

Center for Multi
Mobility in Urban
and Tribal Ar

Base Scenario - MATSim



Vehicle on network



Public Transit
Ridership



Vehicles from base
model



CMMM

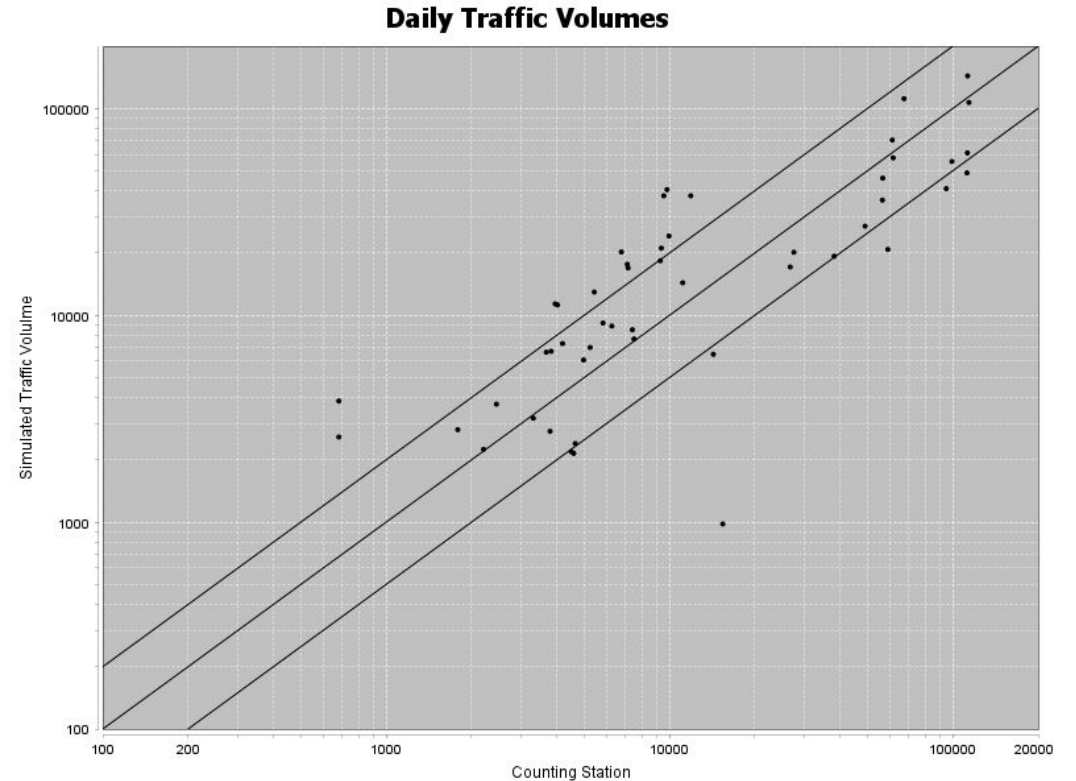
Center for Multi-Modal
Mobility in Urban, Rural
and Tribal Areas

- 1) The simulated ridership against the Washington D.C. metro system's average ridership data and
- 2) Hourly Traffic count data on network links.

Washington Metro System ridership Per Day, 2019

Actual Ridership	497,439
Simulated Ridership	460,660
Difference	-7.39%

Base Scenario - MATSim



traffic volumes on validation links

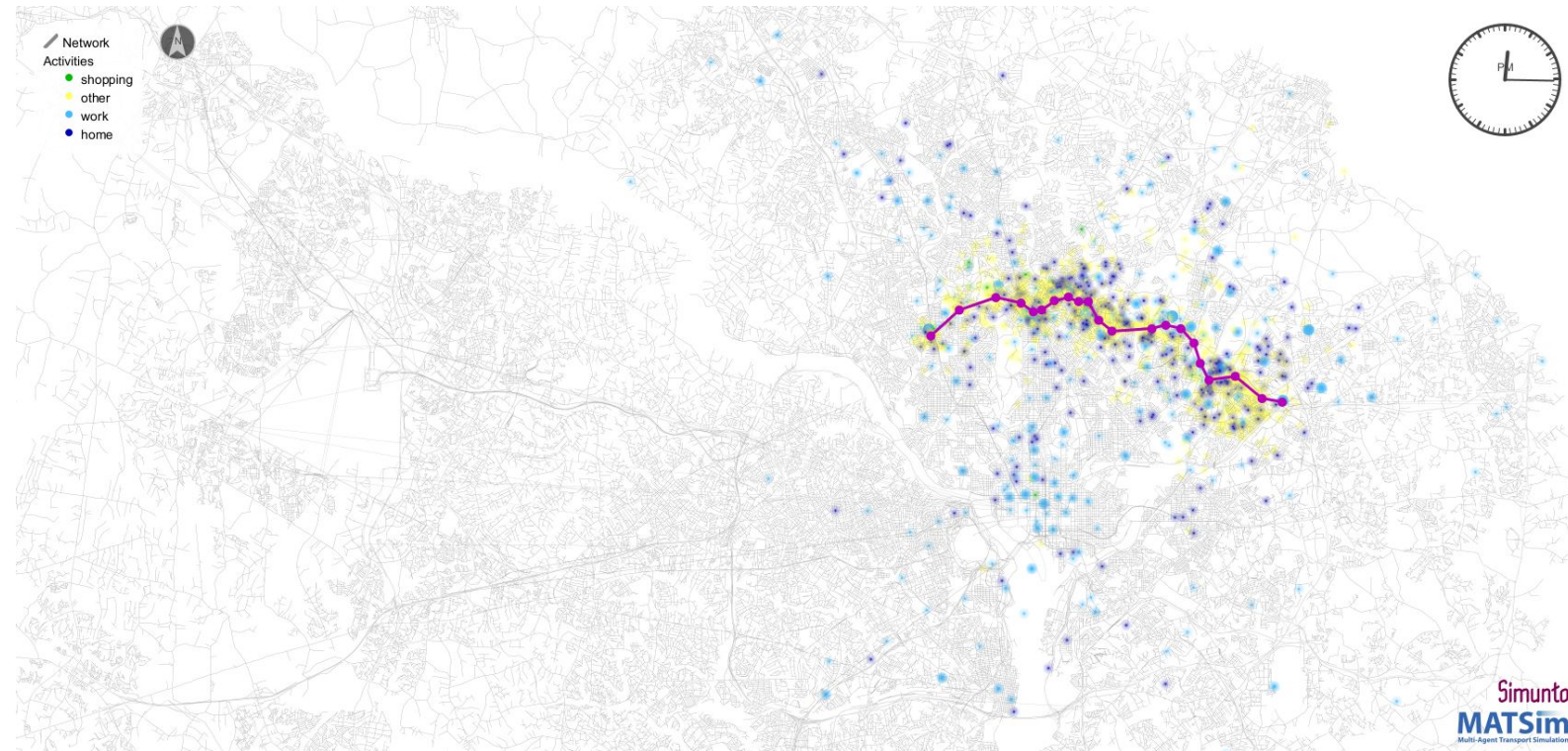


CMMM

Center for Multi-Modal
Mobility in Urban, Rural
and Tribal Areas

- MITO 2027 data
 - New PT travel time
- Added purple line network in OSM
- Added basic GTFS for purple line
 - 21 Stations
 - 15 minutes headway
- Run for 200 iterations

Purple Line Scenario - MATSim



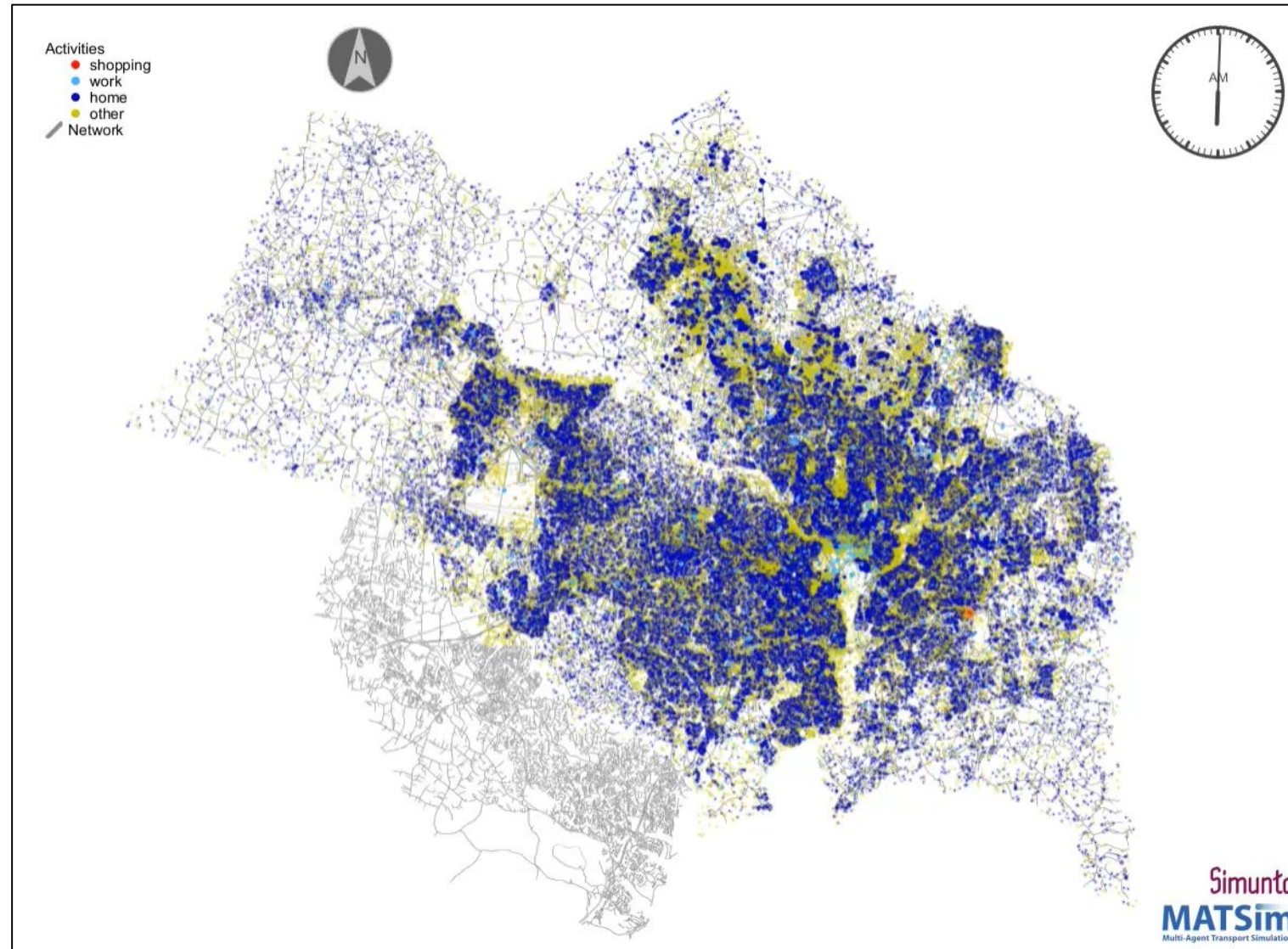
Activities of the purple line users



CMMM

Center for Multi-Modal
Mobility in Urban, Rural
and Tribal Areas

Purple Line Scenario - MATSim



Agents' activities



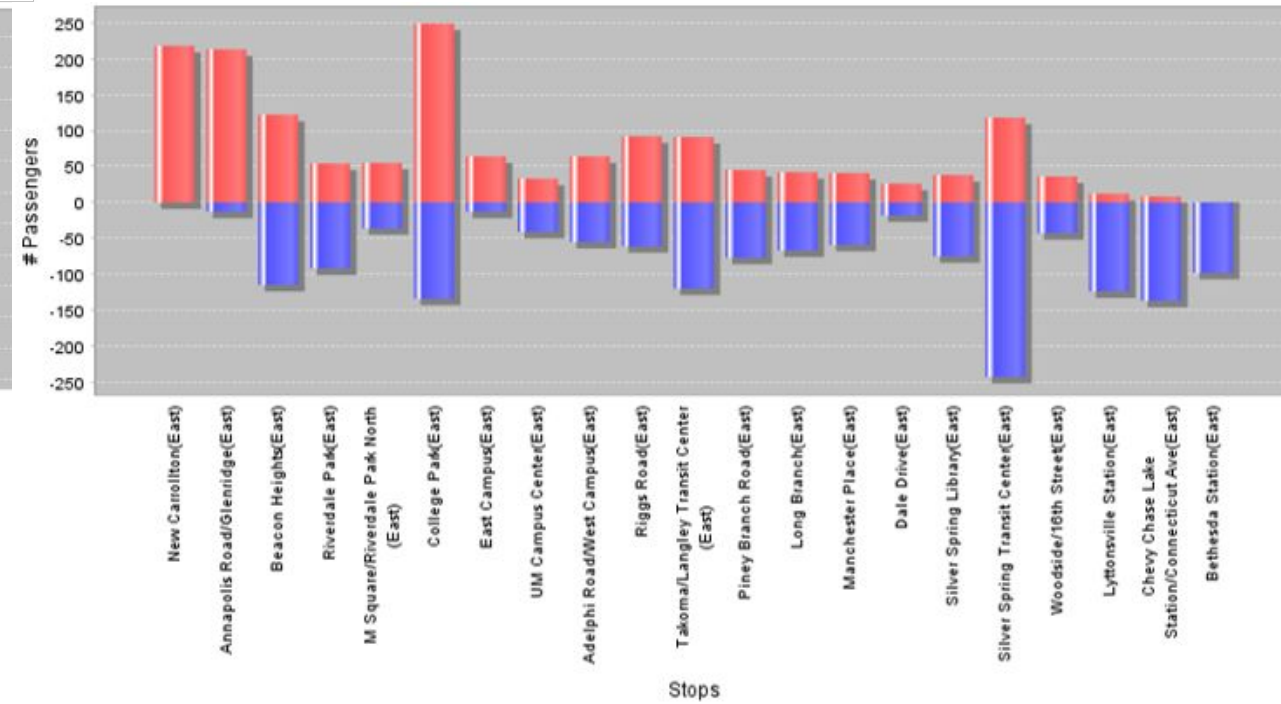
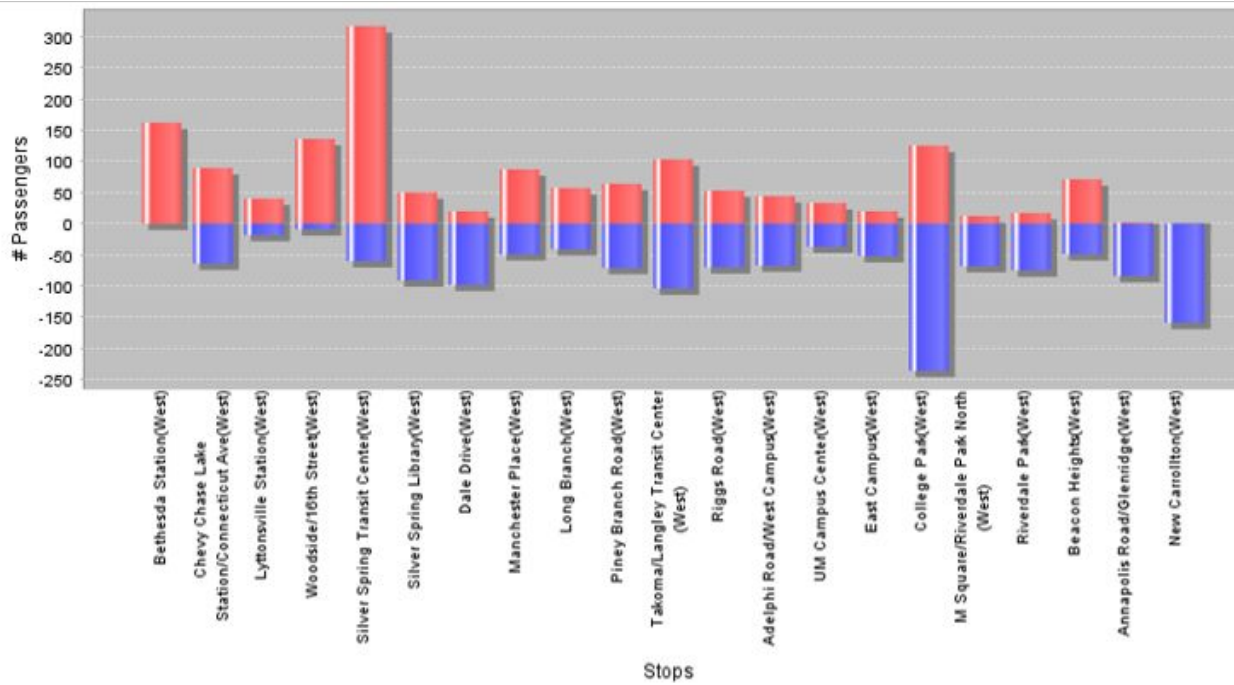


CMMM

Center for Multi-Modal
Mobility in Urban, Rural
and Tribal Areas

Purple Line Scenario - MATSim

- Daily Ridership- 31, 230
- Higher boarding and alighting at transfer points- Silver spring and College Park Station (illustration)



Purple Line Ridership (Illustration for 10 percent)



CMMM

Center for Multi-Modal
Mobility in Urban, Rural
and Tribal Areas

Conclusion

- According to FTA, purple line will serve 56,100 trips on average weekday in 2035.
- From simulation, weekday ridership of 31,230 on the purple line in 2027 for the opening year.
- This is conservative estimation
 - This excludes the education trips.
 - Difficult to adapt the UMD students and faculty-staffs trips which will have significant impact.
- This modeling suite works in disaggregate level, which give more accurate and realistic results for ridership estimation
- Policy sensitive tool to test different scenario.

This is joint work with
Maryam Khan and Mahmudul Chayan

Thank you!

