SYNERGIES BETWEEN OPEN SOURCE AND PROPRIETARY SOFTWARE

Modeling Mobility Conference • September 16, 2025

Andrew Rohne • Vince Bernardin

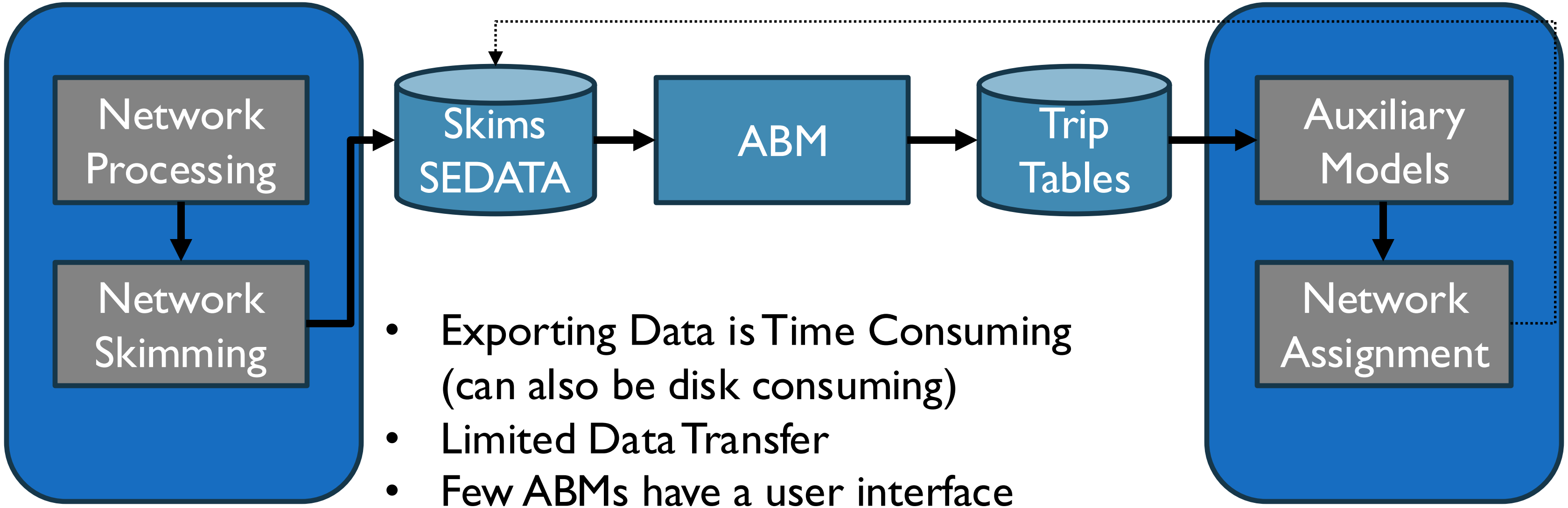
Caliper®

Transportation & Mapping Solutions

Maptitude • TransCAD • TransModeler

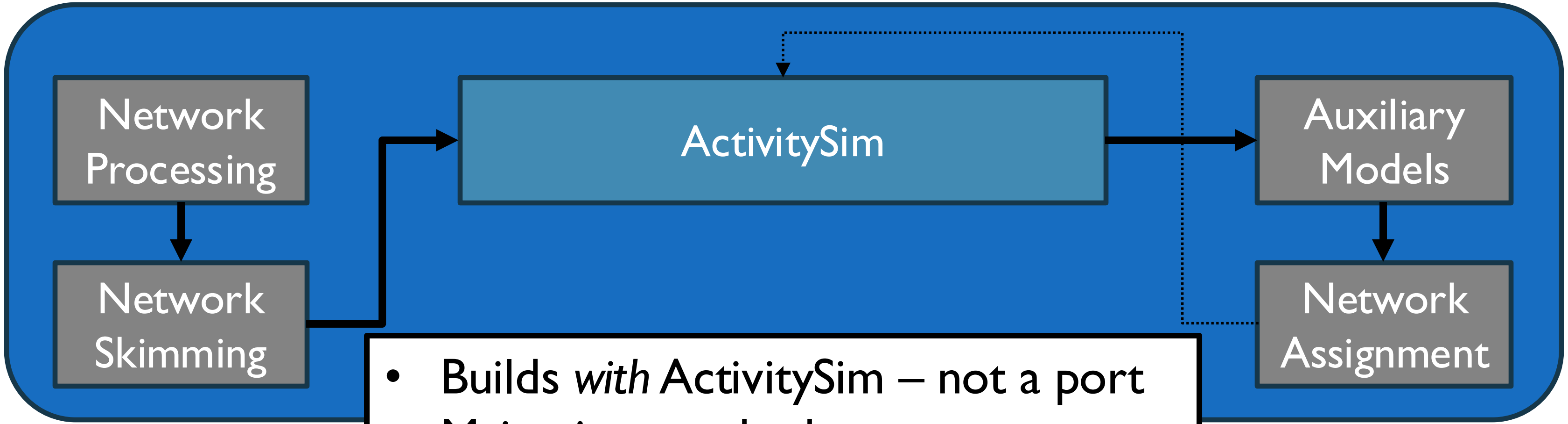
Synergy: The interaction or cooperation of two or more organizations, substances, or other agents to produce a combined effect greater than the sum of their separate effects.

MOTIVATION



- Exporting Data is Time Consuming (can also be disk consuming)
- Limited Data Transfer
- Few ABMs have a user interface
- ABMs tend to enforce standard data structures

WHAT IF…



- Builds *with* ActivitySim – not a port
- Maintains standards

- ***Efficiency***
- ***Engineering***
- ***Experience***

INTEGRATION – NOT REPLICATION

- ActivitySim and TransCAD are both flexible and extensible modeling systems
  - Call TransCAD functionality from Python
  - Call Python libraries & functionality from TransCAD
- Freedom to use individual components from ActivitySim together with components from TransCAD
- Minimalist approach to replacing ActivitySim code with TransCAD calls – only where it offers a computational or methodological advantage

COMPUTATIONAL EFFICIENCY

Caliper

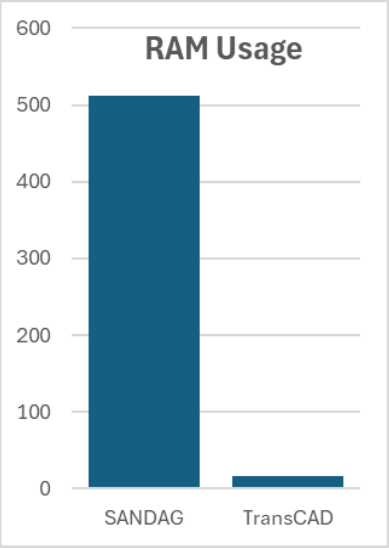
EFFICIENCY

- No exporting (huge time saver!)
- Access all data in network and ABM
  - Call matrices as needed, no need to hold in memory
- Works to reduce resource needs
- Leverage both the benefits of ActivitySim & TransCAD
  - Open-source contributions from consortium
  - Proprietary contributions from Caliper

SOFTWARE ENGINEERING

SOFTWARE ENGINEERING

- Initial focus on RAM requirements
  - SANDAG with 5,000 TAZ plus MAZ/TAPs needs minimum **512 GB of RAM**
  - RAM required to deal with over a thousand matrices (primarily transit skim matrices) loaded into memory at the beginning of the model, their primary use being tour mode choice
  - Replaced existing native python ActivitySim tour mode choice model with python script using TransCAD NL procedure and matrices implementing the same tour mode choice specification
  - Reproduce native ActivitySim results with similar runtime but **only 16 GB RAM!**



- Runtime
  - Plan to test distributed computing

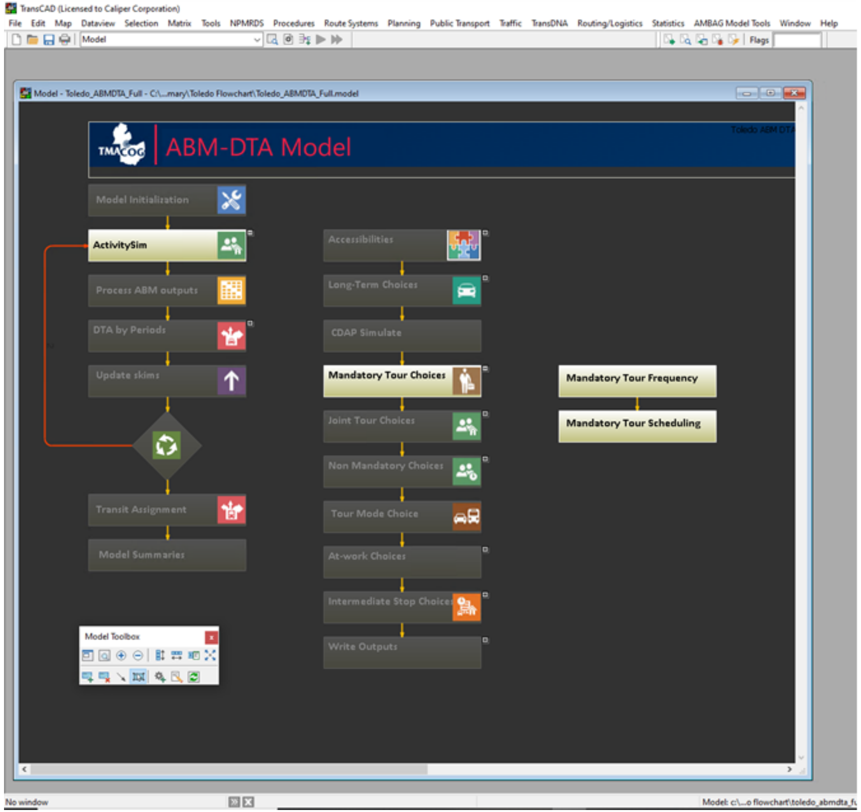
TRANSCAD'S POPULATION SYNTHESIS

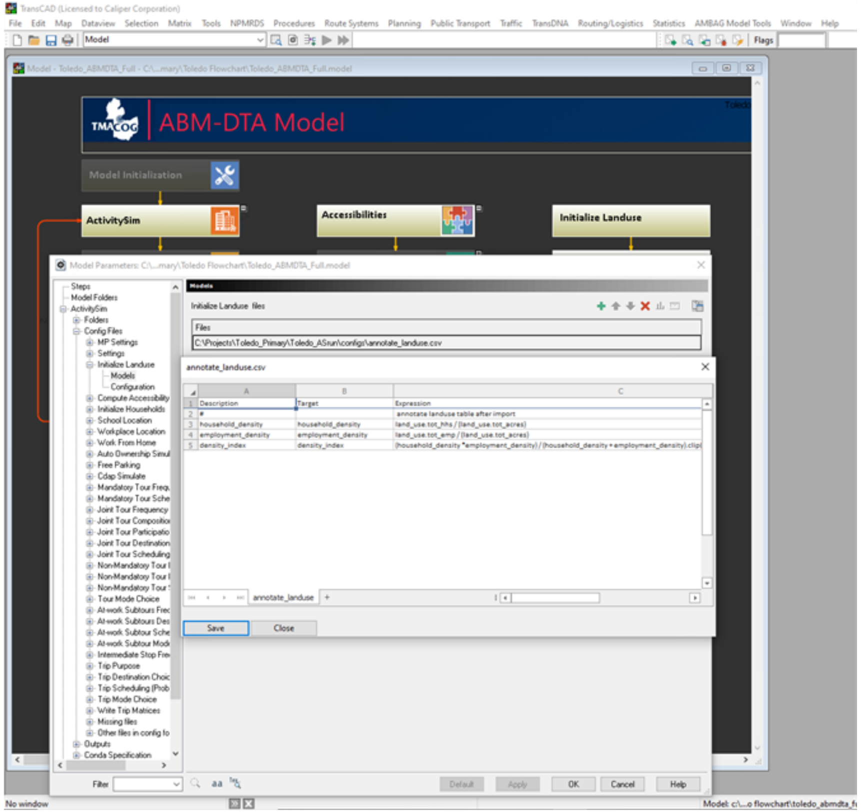
- Most new travel models in the US, whether trip-based or activity-based, use a population synthesizer
- TransCAD's developed our own FAST Iterative Proportional Updating algorithm
  - Household and Person level controls
  - Controls at multiple levels of geography
  - Subarea re-synthesis
  - More stable than PopulationSim
  - **Can synthesize 1 million people a minute!!!**

hghhgh ghghg

USER EXPERIENCE

TRANSCAD ACTIVITYSIM GUI





SANDAG MODE CHOICE PARAMETERS

Model Parameters: C:\...\ActivitySimModeChoice\ActivitySimTourMC.model

- Runtime
- Parameters
- Tour Mode Choice
  - Work
    - Modes
    - Auto
    - Non-Motorized
    - Ride Hail
    - Micromobility
    - Transit Walk Access
    - Transit PNR Access
    - Transit KNR Access
    - Transit TNC Access
- Display Properties
- Help Document

Filter

Modes

Tour Work Modes

Parent	Alternatives	ParentNestCoeff
Root	AUTO, NONMOTORIZED, MICROMOBILITY, TRANSIT, RIDEHAIL	
AUTO	DA, SR2, SR3	0.5
NONMOTORIZED	WALK, BIKE	0.5
MICROMOBILITY	EBIKE, ESCOOTER	0.5
TRANSIT	WALKACCESS, PNRACCESS, KNRACCESS, TNCACCESS	0.5
RIDEHAIL	TAXI, TNC_SINGLE, TNC_SHARED	0.5
WALKACCESS	WALK_LOC, WALK_PRM, WALK_MIX	0.5
PNRACCESS	PNR_LOC, PNR_PRM, PNR_MIX	0.5
KNRACCESS	KNR_LOC, KNR_PRM, KNR_MIX	0.5
TNCACCESS	TNC_LOC, TNC_PRM, TNC_MIX	0.5

aa Tex Default Apply OK Cancel Help

SANDAG MODE CHOICE PARAMETERS

Model Parameters: C:\...jects\ActivitySimTourMC\ActivitySimTourMC.model

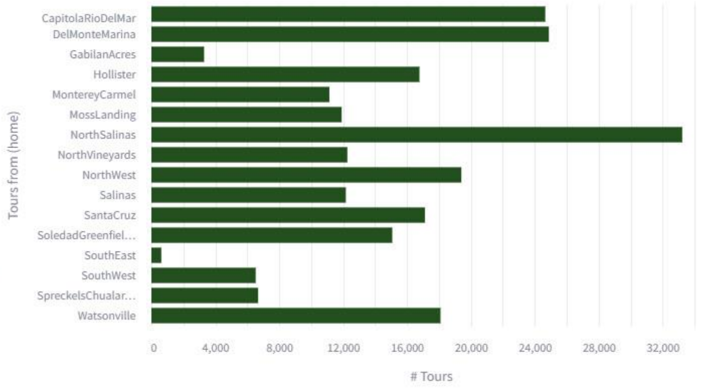
- Runtime
- Parameters
- Tour Mode Choice
  - Work
    - Modes
    - Auto
    - Non-Motorized
    - Ride Hail
    - Micromobility
    - Transit Walk Access
    - Transit PNR Access
    - Transit KNR Access
    - Transit TNC Access
- Display Properties
- Help Document

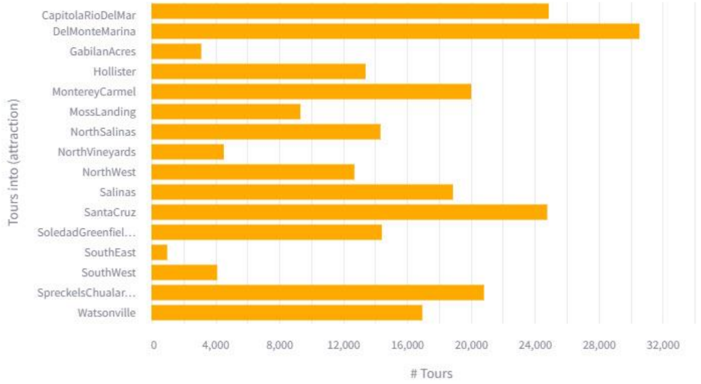
Non-Motorized

Work Mode Utility - Non-Motorized Modes

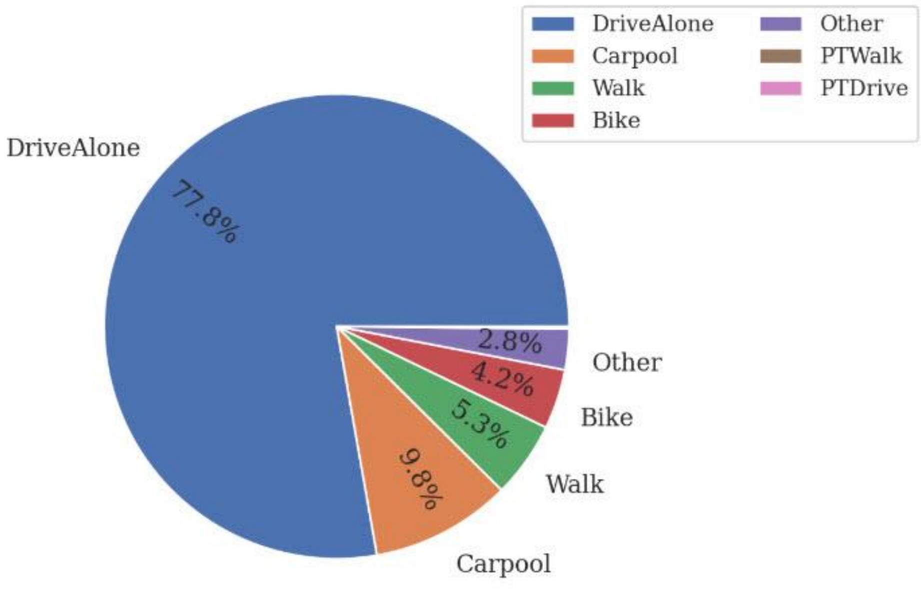
Description	Expression	Coefficient	WALK	BIKE
Walk: Unavailable	df.walkavailable = 0	-99.0000	☑	☐
Walk: Age < 16	df.age < 16	2.1884	☑	☐
Walk: Age [16, 24]	(df.age >= 16) and (df.age <= 24)	1.4306	☑	☐
Walk: Age [41, 55]	(df.age >= 41) and (df.age <= 55)	-0.4321	☑	☐
Walk: Age [56, 64]	(df.age >= 56) and (df.age <= 64)	-0.5200	☑	☐
Walk: Age >= 65	df.age >= 65	-0.8316	☑	☐
Walk: Income [60000, 99999]	(df.income > 59999) and (df.income < 100000)	-0.2084	☑	☐
Walk: Normalized Landuse Var Sum [Origin Intersection + DU]	df.LUVarsNormalized_walk	0.0093	☑	☐
Walk: Normalized Dest Emp Density	df.dMGRAEmpDenNorm_walk	0.0998	☑	☐
Walk: Time	(df.walk_time_skims_inb + df.walk_time_skims_out) * df.time_factor	-0.0424	☑	☐
Walk: ASC Zero Auto HH	df.is_indiv and df.auto_ownership = 0	2.6903	☑	☐
Walk: ASC Auto Insuff HH	df.is_indiv and (df.auto_ownership < df.num_adults) and (df.auto_owners	-0.7049	☑	☐
Walk: ASC Auto Suff HH	df.is_indiv and (df.auto_ownership >= df.num_adults) and (df.auto_owne	-3.4983	☑	☐
Bike: Unavailable	df.bikeavailable = 0	-99.0000	☐	☑
Bike: Age [41, 55]	(df.age >= 41) and (df.age <= 55)	-0.7311	☐	☑
Bike: Age [56, 64]	(df.age >= 56) and (df.age <= 64)	-0.6435	☐	☑
Bike: Age >= 65	df.age >= 65	-1.5487	☐	☑
Bike: Female	df.female = 1	-1.1936	☐	☑
Bike: Income 100k+	df.income > 100000	0.6414	☐	☑
Bike: Normalized Landuse Var Sum [Origin Intersection + DU]	df.LUVarsNormalized_bike	0.0833	☐	☑
Bike: Logsum	df.bikeLSI + df.bikeLSO	0.1343	☐	☑
Bik: Miles to coast from origin MGRA	df.DistanceToCoast	-1.4202	☐	☑
Bike: Miles greater than 2 to coast from origin MGRA	Max(df.DistanceToCoast-2,0)	1.3351	☐	☑
Bike: Miles greater than 5 to coast from origin MGRA	Max(df.DistanceToCoast-5,0)	0.0796	☐	☑
Bike: ASC Zero Auto HH	df.is_indiv and df.auto_ownership = 0	-7.6023	☐	☑
Bike: ASC Auto Insuff HH	df.is_indiv and (df.auto_ownership < df.num_adults) and (df.auto_owners	-6.6419	☐	☑
Bike: ASC Auto Suff HH	df.is_indiv and (df.auto_ownership >= df.num_adults) and (df.auto_owne	-3.6053	☐	☑
Bike: Pro Bike District	(df.district27=8) or (df.district27=9)	1.5524	☐	☑

DISTRICT FLOWS DASHBOARD

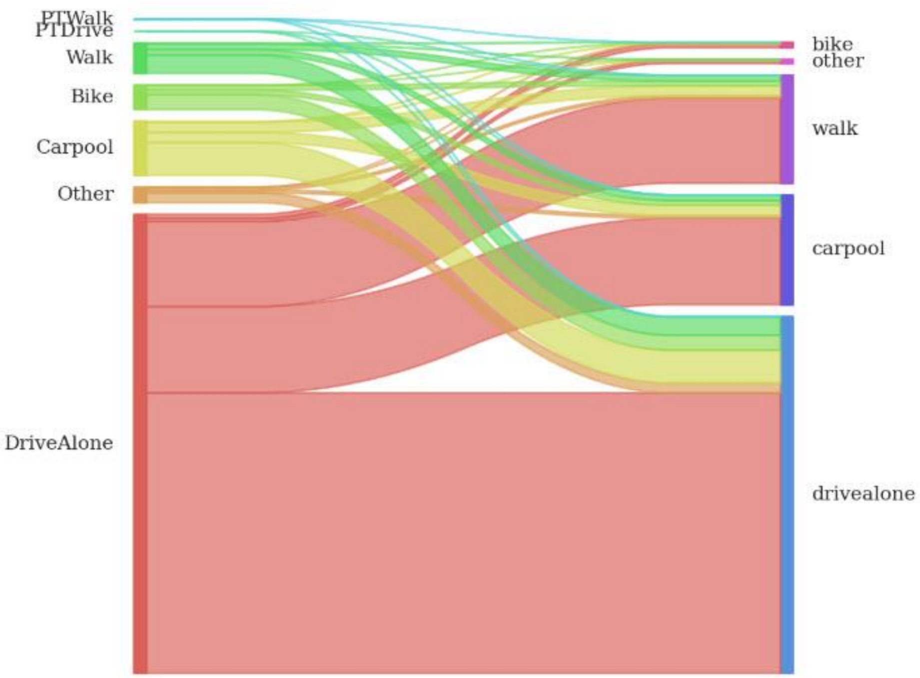




MODE SUMMARY DASHBOARD



Mandatory Tour Mode Shares



ASSIGNMENT DASHBOARD

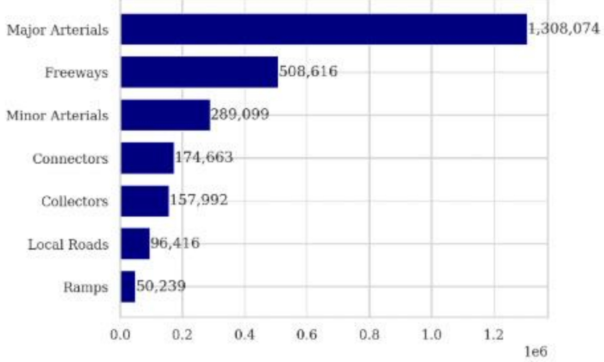
EA | **AM** (selected) | MD | PM | EV | NT

Total Vehicle Miles Traveled (VMT)
2,585,099

Total Vehicle Hours Traveled (VHT)
68,969

Total Link Flow
9,118,435

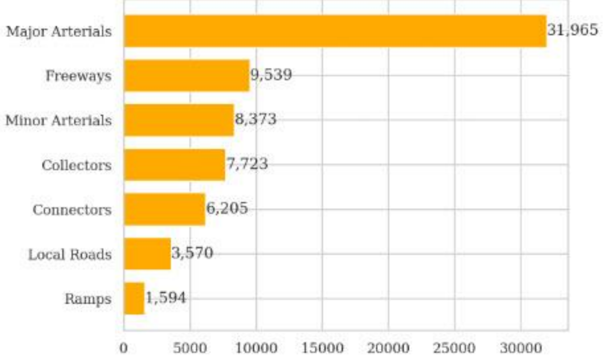
VMT by Class



Average Speed by Class

Class	AvgSpeed	Tot_Flow
Collectors	20.5	938,557
Connectors	28.1	479,231
Freeways	53.3	1,178,572
Local Roads	27.0	823,234
Major Arterials	40.9	3,817,819
Minor Arterials	34.5	1,652,444
Ramps	31.5	228,578

VHT by Class



Medium-High congestion mileage: VOC [0.75, 0.90)
86.4 miles (1.2%)

High congestion mileage: VOC [0.90, 1.00)
29.5 miles (0.4%)

Severe congestion mileage: VOC ≥ 1
25.2 miles (0.4%)

REPRESENTATIVE REALIZATION

SIMULATION VARIATION

- How many simulations are needed to ensure Monte Carlo results are stable and converged to the analytic solution?

SIMULATION VARIATION

- How many simulations are needed to ensure Monte Carlo results are stable and converged to the analytic solution?
- The answer is given by Hoeffding's inequality

SIMULATION VARIATION

- How many simulations are needed to ensure Monte Carlo results are stable and converged to the analytic solution?
- The answer is given by Hoeffding's inequality
- Let's say for the change in destinations to be within 1% error in a DC choice model with 2,000 zones?

SIMULATION VARIATION

- How many simulations are needed to ensure Monte Carlo results are stable and converged to the analytic solution?
- The answer is given by Hoeffding's inequality
- Let's say for the change in destinations to be within 1% error in a DC choice model with 2,000 zones?

226,000 runs

AN INCONVENIENT TRUTH

- The Law of Large Numbers requires LARGE NUMBERS
- Fixing a random seed DOES NOT reduce simulation variation, only hides it

- We do not have the computational resources to do real Monte Carlo simulation of an activity-based model.
- So, what should we do?

REPRESENTATIVE REALIZATION

- The most likely single set of choices
  with the correct shares chosen of each alternative.

REPRESENTATIVE REALIZATION

- The most likely single set of choices
  with the correct shares chosen of each alternative.
- In other words, most-likely, least biased realization

REPRESENTATIVE REALIZATION

- The most likely single set of choices
  with the correct shares chosen of each alternative.
- In other words, most-likely, least biased realization
- In math,

$$Max\ LL = \sum_{i,j,k} ln\left(P_{ij}\delta_{ijk}\right)$$

Subject to

$$Min\ Bias = \sum_{j}\left(\sum_{i} P_{ij} - \sum_{i}\frac{\sum_k \delta_{ijk}}{K}\right)^2$$

BOTH CRITERIA ARE NECESSARY

- Not simply the most likely realization
  - The most likely realization is where everyone chooses their most likely alternative (i.e., everyone drives)
- There are many, many unbiased realizations
  - Where the share of people choosing each alternative is equal to the shares given by the analytic solution
- But the most likely unbiased realization is practically unique

ON-GOING WORK

- Testing a variety of heuristic algorithms
  - Rigorous Myopic
  - Shadow-Pricing
  - Top-Down
- Exploring the mathematical properties

MOST LIKELY OUTCOME

- You can entirely eliminate random number draws and have 100% confidence that a single run of your activity-based model gives you exactly the right solution

- It will probably take marginally longer (compared to current, improper practice)
  - Probably less than twice as long
  - Hopefully much less than twice as long

CONTACTS

Vince Bernardin, PhD | Vice-President

vince@caliper.com | +1 812-459-3500

Andrew Rohne | Sr. Planner/Engineer

andrew@caliper.com | +1 513-314-9901

Caliper