

Modeling and Artificial Intelligence for Sustainable Transportation Solutions



Jeff Wojtowicz

Sr. Transportation Engineer
Civil and Environmental Engineering
Rensselaer Polytechnic Institute
wojtoj@rpi.edu



Ruimin Ke

Assistant Professor
Civil and Environmental Engineering
Rensselaer Polytechnic Institute
ker@rpi.edu



Sean He

Associate Professor
Civil and Environmental Engineering
Rensselaer Polytechnic Institute
hex6@rpi.edu

**NYSAMPO 2025 Planning
Conference
Syracuse, NY
Wednesday, May 14, 2025**

About this Session

- **Overview:**

- Presentation of emerging tools developed by RPI to support the MPO planning process through data integration, AI and simulation.

- **Summary of Presentations**

- Jeff Wojtowicz will present a dynamic freight database
- Ruimin Ke will present the integration of AI for transportation systems
- Sean He will present simulation modeling that can support scenario planning



Dynamic Freight Database Background

- **Project:** Dynamic Freight Database
- **Funded by:** FHWA SHRP2 Implementation Assistance Program: Innovative Local Freight Data
- **Partners:**



Rensselaer

- **Purpose:** Identify available freight data sources and fuse into a single database to support CRTC's freight plans, projects and programs.



Project Tasks

- **Task 1:** Identification of freight data sources and outreach activities
- **Task 2:** Data gathering and processing
- **Task 3:** New data collection and processing
- **Task 4:** Data fusing and integration
- **Task 5:** Build Capital Region's Dynamic Freight Database



Establishment of Dedicated Data Server

- Project FTP site setup for data sharing & warehousing

CapitalDistrictFreightData - SHRP2_RPI - WinSCP

File Commands Mark Session View Help

Address `/home/CapitalDistrictFreightData`

Find Files | Download | Edit | Properties

Transfer Settings Default

Name	Ext	Changed
0 CDTC Dynamic Database		2/24/2016 4:48:04 PM
511 NY data feed		2/24/2016 4:17:14 PM
Capital Region network data		2/22/2016 2:49:27 PM
Census of vehicles		1/31/2015 10:10:54 PM
Commodity Flow Survey (CFS)		2/22/2016 4:57:46 PM
Economic data (GIS format)		1/31/2015 8:49:15 PM
Employment Data (ZIP Code Business Patterns)		2/25/2016 8:56:28 AM
E-ZPass data		1/31/2015 9:00:05 PM
Freight Analysis Framework-3		1/31/2015 8:49:10 PM
HERE speed data		2/22/2016 2:52:52 PM
Oversize-Overweight Permitting Data		1/31/2015 9:00:08 PM
Smart Location Data		2/22/2016 2:25:37 PM
TRANSEARCH commodity flow data		2/22/2016 2:49:01 PM
Truck accidents data		1/31/2015 10:03:28 PM
Truck traffic counts 2010-2013		2/22/2016 2:48:45 PM
Weight-In-Motion (WIM) data		1/31/2015 9:39:02 PM

0 B of 0 B in 0 of 16

SFTP-3 0:01:05

NYSGIS

`/home/SHRPC20/NYSGIS`

Name	Ext
..	
Albany County Pa	
AlbanyCountyMur	
AlbanyCountyTran	
CityOfAlbanyZoni	
RensselaerCounty	
SaratogaCountyTa	

NYSTA

`/home/SHRPC20/NYSTA`

Name	Ext
..	
albany_2013.csv	
buffalo_2013.csv	
newyork_2013.csv	
plaza_list.csv	
syracuse_2013.csv	

NYSDOT_Class_2010_2013/NYS

Name	Ext
ContinuousClass2010.csv	
ContinuousClass2011.csv	
ContinuousClass2012.csv	
ContinuousClass2013.csv	
ShortClass2010.csv	
ShortClass2011.csv	
ShortClass2012.csv	
ShortClass2013.csv	

No.	Data Source	Parameter				Industry			Frequency				Mode						Aggregation					Policy								
		Freight trips	Freight generation	Ton-miles	Value	Service trips	NAICS	SCTG	STCC	5 years	Yearly	Quarterly	Hourly	Types of trucks	Trucks	Rail	Water	Air	Pipeline	Multi-modal	Import/Exports	State	County	ZIP Code	Census Tract	Along highways	Weight limit	Land-Use	Network	Traffic	Toll	Crashes
PUBLICLY AVAILABLE DATA																																
1	Commodity Flow Survey (CFS)	.	✓	✓	✓	.	✓	✓	.	✓			.	✓	✓	✓	✓	✓	✓	.	✓	.								✓		
2	FAF3 Origin-Destination	.	✓	✓	✓	.		✓	.	✓			.	✓	✓	✓		✓	✓	✓	✓	.							✓			
3	Smart location data		.				✓			2 nd version												✓	✓			✓	.	.	.			
4	Employment data		.	.						✓	.	.	.									✓	.									
DATA FROM OTHER AGENCIES/SOURCES																																
1	511 NY data feed											✓																✓	.	✓		
2	Capital Region updated network									2006-07																	✓	✓	.	✓		
3	Commercial Census data		✓				✓			Periodically			✓	✓									✓								✓	
4	HERE speed data										✓	.		✓									✓				✓	✓	.			
5	MIST speed and volume data									✓														✓			✓	✓	.			
6	TRANSEARCH commodity flow		✓	✓	✓	.		✓		✓												✓						✓				
7	Economic data (in GIS format)									✓											✓	✓				✓	✓					
8	E-ZPass data	✓										✓	✓											✓			✓	✓	✓			
9	Oversize/overweight permitting	✓	✓							✓			✓											✓		✓						
10	TRANSMIT data	✓	✓									✓												✓	✓		✓	✓	✓	.		
11	Truck crash data									✓			✓	.							✓			.	.				.	✓		✓
12	Truck traffic counts	✓								.		✓	✓								.	.		✓	.		✓	✓				
13	Weight-In-Motion (WIM)	✓										✓	✓								.	.		✓	✓			✓				
NEW DATASETS																																
1	GPS	✓					✓					✓	✓	✓									✓				✓	.				
2	In-Depth-Interviews	✓	✓				✓						✓	✓	✓	✓	✓				✓	✓	✓	✓	✓			
3	FTG, FG, STA models	✓	✓	.		✓	✓						✓				✓					✓				

- Describes each data source including:
 - Variables
 - Advantages
 - Limitations
 - Value for freight planning

Variables:

CFS data contain the following major variables aggregated for CFS geographic locations, states and country:

- Value of the commodity
- Tons of commodity
- Ton-miles of commodity
- Mode of transportation:
 - Single modes: Truck, Rail, Water, Air, Pipeline
 - Multiple modes: Parcel, Courier, Truck and Rail, Truck and Water, Rail and Water, Other multiple modes.
- Type of the commodity in two digit SCTG
- Type of the commodity in NAICS

Advantages:

- These data give valuable inputs on freight movement.
- It is an exclusive data on freight without interfering with passenger transportation.

Limitations:

- Since the data are based on surveys these are prone to sampling errors and low survey response rates.
- The 'S' mark in Table 1 and Table 2 refers to void data due to high sampling variability or low response rates whereas the 'Z' mark represents the data estimate is between zero and half so taken as zero. Because of such void data the O-D between all geographic areas cannot be found leading to unusability of the data for generating freight trip distribution patterns in CDTC region.
- The level of aggregation is CFS geographic area which is far aggregated for a state level freight planning. For state level freight planning the aggregation should be as low as possible, perhaps to city level or ZIP code level.
- The Ton-miles are calculated based on the shortest distances on the transportation network between origin and destinations which would not consider the freight trip pattern of these goods.
- Disaggregated data are highly confidential and the access for which requires a special sworn status from the census bureau.

Value for planning:

- Process the data for significant inputs for freight planning for CDTC region based on the CFS geographic locations.
- O-D matrices to be generated for CDTC region for studying freight trip generation patterns.



New Datasets

- Fleet GPS data
- In-depth interviews with major carriers, shippers and receivers
- Freight activity survey to local businesses, this was used to develop:
 - Freight trip attraction models
 - Freight trip production models



DO YOUR PART IN IMPROVING THE LOCAL FREIGHT NETWORK

The Capital District Transportation Committee (CDTC) and Rensselaer Polytechnic Institute (RPI) are working to ensure that the needs of the freight industry in upstate New York are taken into account. We are requesting your assistance to provide freight and travel data for upstate New York that could be used to identify freight related improvements in the area. This data could help identify and quantify the impacts of congestion, route restrictions and policies on the local freight network. This will allow CDTC to properly advocate that these freight improvements receive the proper funding in the future.

We are requesting data related to delivery patterns in the region. Data may include, but are not limited to:

- Information related to fleet size, main commodities hauled and main origins and destinations, etc.
- GPS data
 - 1-2 weeks of data past GPS would be sufficient. Although recent data would be ideal, we could
 - We have experience working with data provided by private sector onboard GPS systems. If you
- Sharing your thoughts, comments and experiences about the freight industry in the Capital District with the team. This insight would also be very useful and we invite you to contact us at the numbers or email addresses below.

Following our analysis, the team will be able to share general results with you. It is also important to note that any data you share with the team will be held strictly confidential. RPI has a long history of working with the freight community through the Center of Excellence for Sustainable Urban Freight Systems (CoE-SUFS). Any data shared with the team will be protected and will not be disclosed to anyone else. We are also prepared to sign non-disclosure agreements to further protect this data if you wish.

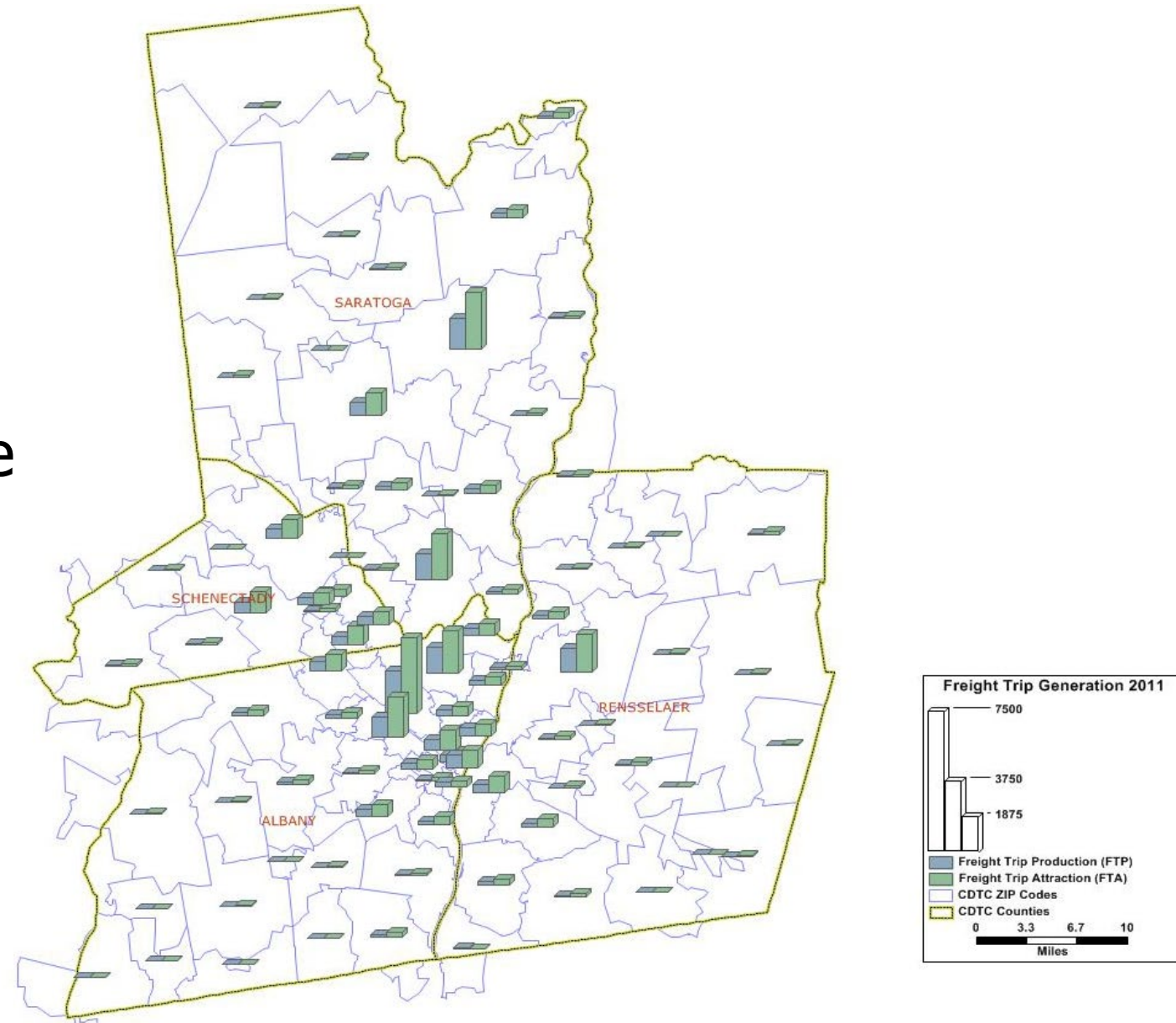
WE NEED YOUR HELP

If you are interested in sharing any freight data or learning more about this study please contact either Chris Bauer at CDTC (cbauer@cdtcmpo.org) or 458-2161) or Jeff Wojtowicz at RPI (wojtoj@rpi.edu or 276-2759).



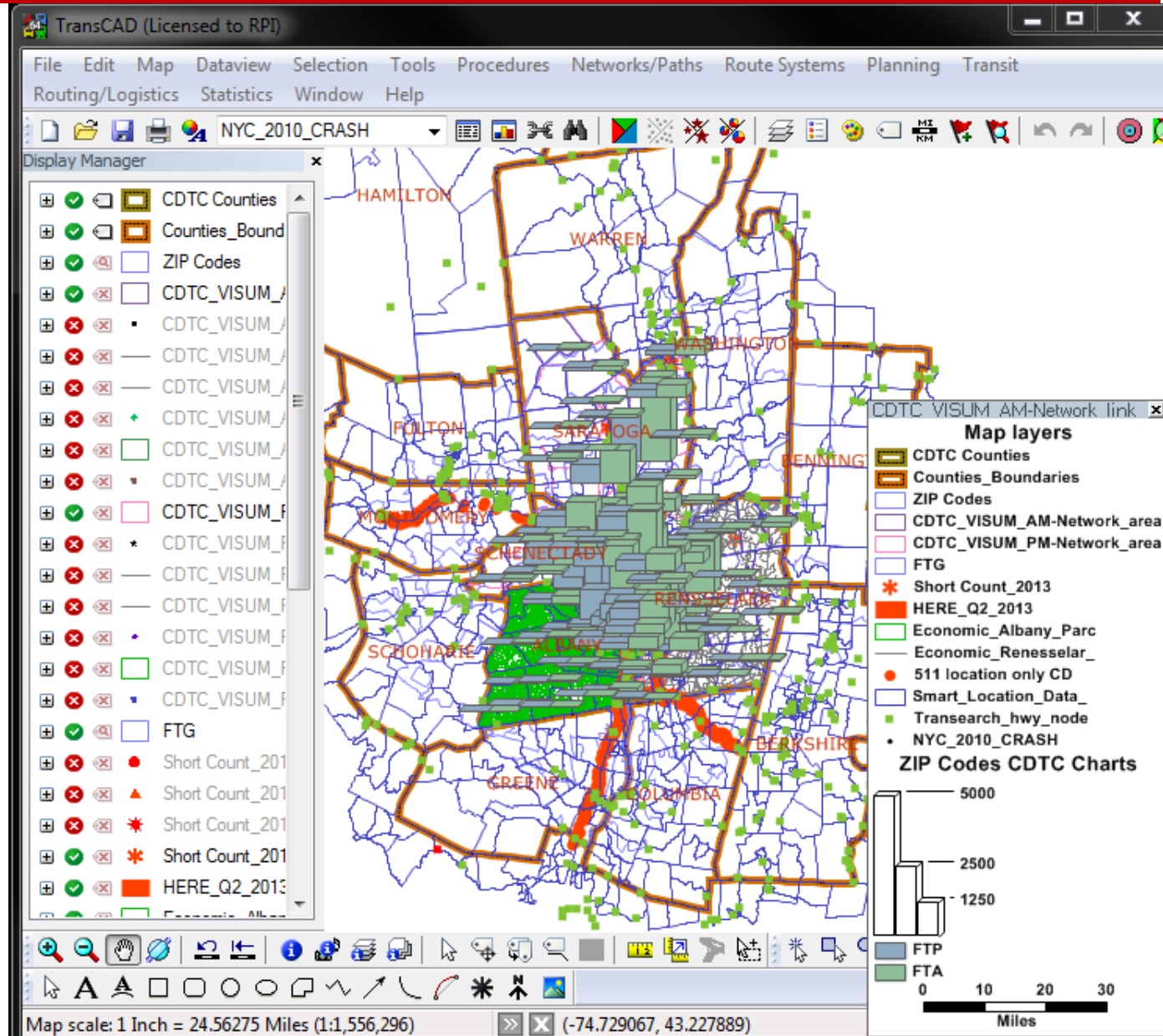
Freight Trip Generation for the Capital Region

- Freight activity survey supported the development of Upstate NY freight models
- 170 responses from freight intensive businesses
- Using RPI's Freight and Service Activity Software (FASTGS) software, freight trip generation (production & attraction) for the Capital Region were run at ZIP Code level



Dynamic Database

- Cleaned & fused all data sources
- Geo-located and processed data into ArcGIS and TransCAD
- Metadata was created to support the long-term use
- Database was kept current until ~2020



Sample Applications of the Database

- Bottleneck analysis
- Commodity flow (regional/local levels)
- Truck VMT estimates
- Forecasted truck flows
- Impacts of trucks on road capacity
- Temporal analysis of truck trips
- Analysis of special traffic generators
- Truck emission analysis
- Truck route evaluation
- Cost/benefit analysis of freight projects

No.	Database/Applications	Truck VMT estimates	Forecasted truck flows on networks	Bottleneck analyses	Impact of trucks on roadway capacity	Temporal analysis of truck trips	Analysis of special generators	Truck emissions inventory	Evaluating truck routes or restrictions	Cost-benefit analyses of freight projects	Commodity flows affecting region
PUBLICLY AVAILABLE DATA											
1	Commodity Flow Survey (CFS)	✓	✓							✓	✓
2	FAF3 Origin-Destination	✓	✓							✓	✓
3	Smart location data					✓	✓				
4	Employment data		✓								✓
DATA FROM OTHER AGENCIES / SOURCES											
1	511 NY data feed	✓	✓	✓	✓	✓					
2	Capital Region updated network		✓			✓			✓		
3	Commercial Census data							✓		✓	
4	HERE speed data	✓	✓		✓	✓				✓	
5	MIST speed and volume data	✓			✓	✓				✓	✓
6	TRANSEARCH commodity flow	✓			✓	✓				✓	✓
7	Economic data (in GIS format)			✓		✓	✓			✓	
8	E-ZPass data	✓	✓		✓	✓			✓	✓	✓
9	Oversize/overweight permitting			✓	✓				✓		
10	TRANSMIT data	✓	✓		✓	✓				✓	
11	Truck crash data										
12	Truck traffic counts	✓	✓	✓	✓	✓				✓	✓
13	Weight-In-Motion (WIM)				✓					✓	
NEW DATASETS											
1	GPS	✓	✓				✓			✓	
2	In-Depth-Interviews			✓			✓			✓	✓
3	FTG, FG, STA models	✓	✓	✓	✓	✓	✓			✓	✓



Questions?

Jeff Wojtowicz – wojtoj@rpi.edu

Leveraging AI and Data Integration for Practical Transportation Solutions

Dr. Ruimin Ke – Assistant Professor
Rensselaer Polytechnic Institute

May 14, 2025



Rensselaer

why not change the world?®

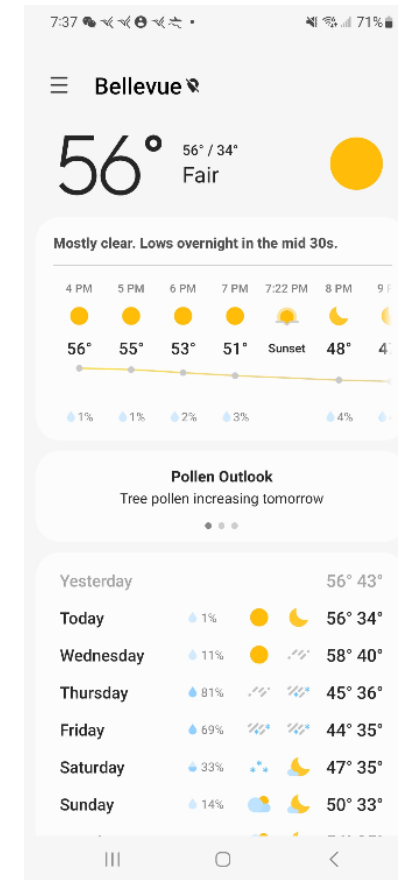
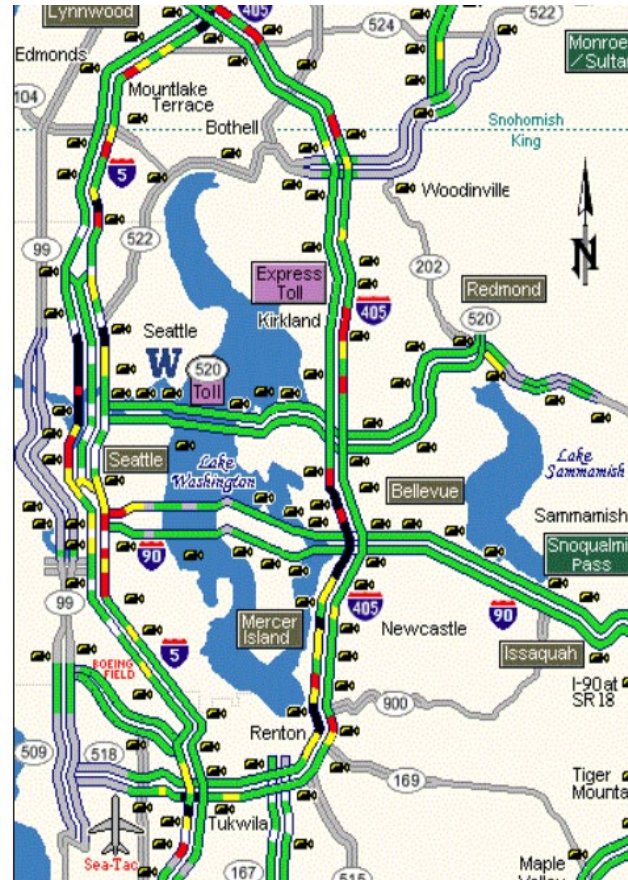
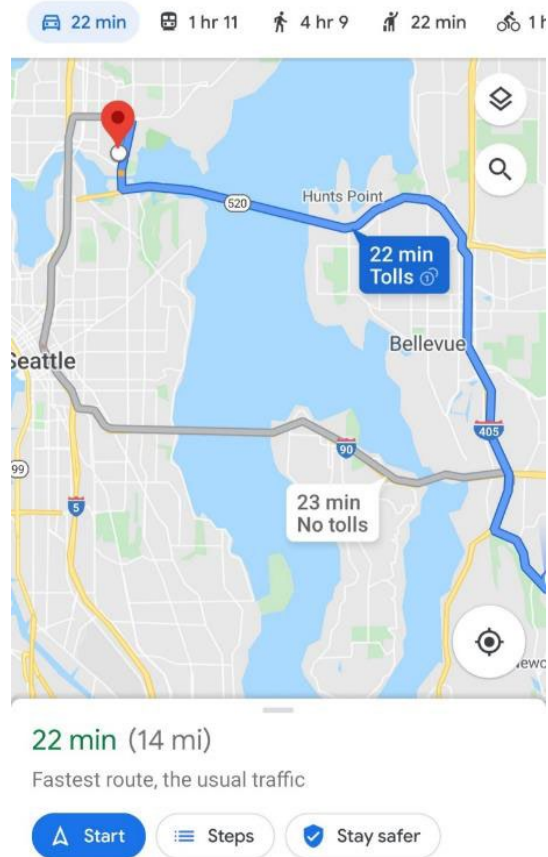
Background

- **Topic:** AI and Advanced Computing as Practical Solutions
- **Funded by:**



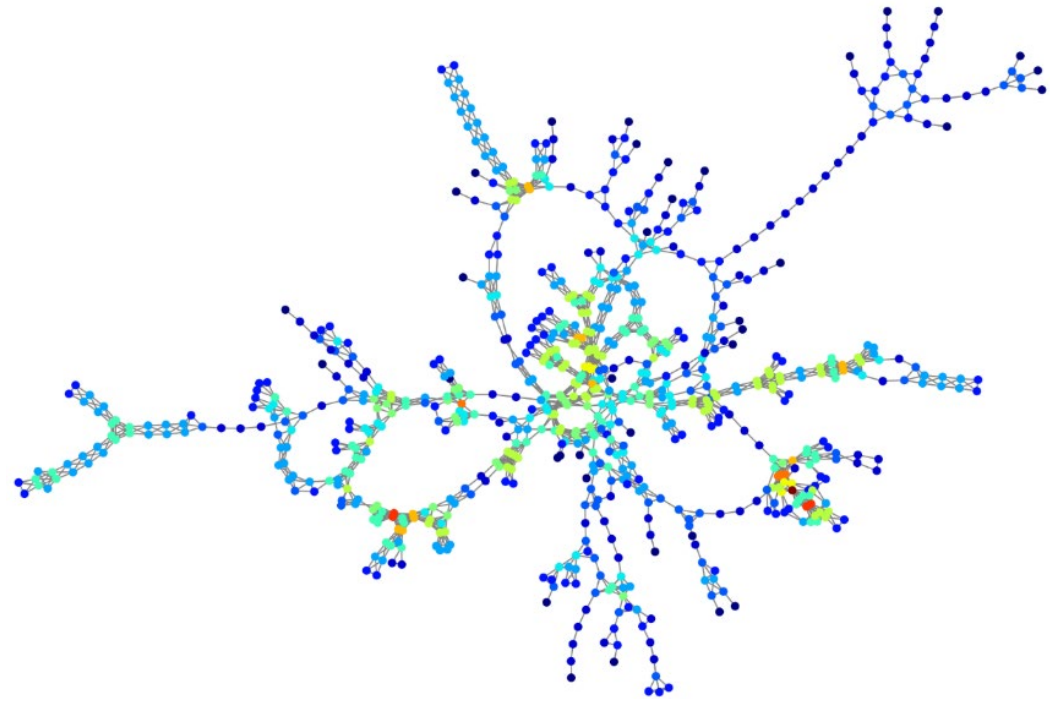
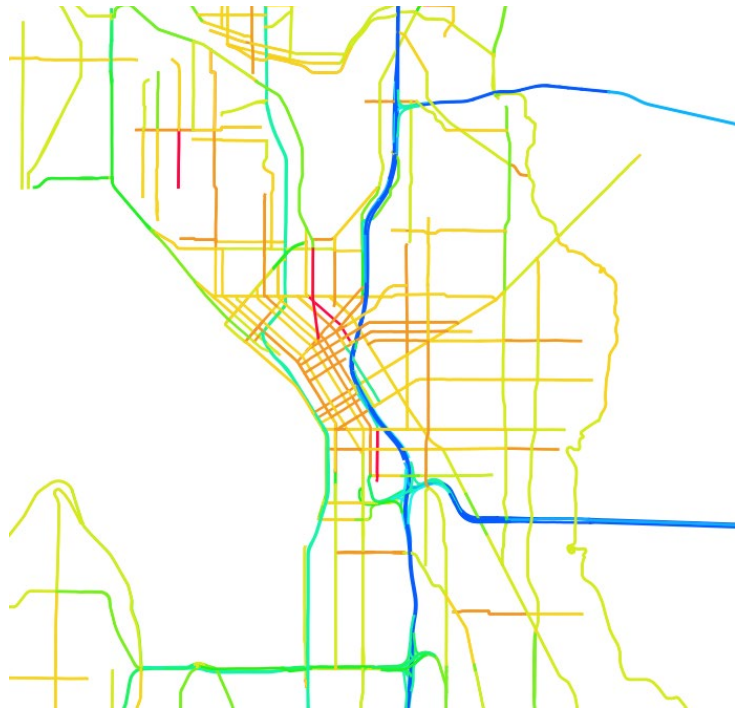
- **Applications:** (0) Traffic forecasting; (1) AI Transit: occlusion-resistant passenger re-identification for transit OD data collection; (2) ChatSUMO: Large language model for automating traffic simulation generation; (3) Realizing vehicle routing problem on IBM-RPI quantum computer.

Can We Forecast Traffic like Weather?



Can We Forecast Traffic like Weather?

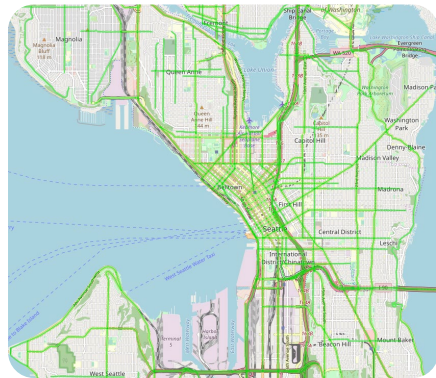
- Traffic modeled as graph



Can We Forecast Traffic like Weather?

Problem Definition

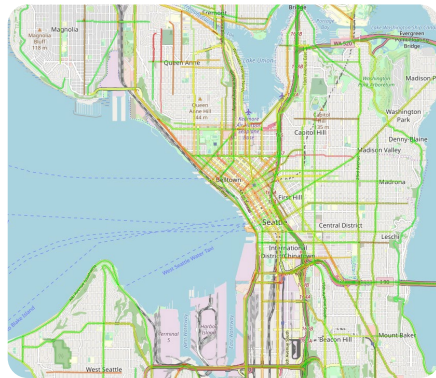
- Input: observed traffic states; Output: traffic states for future steps
- The traffic states of a traffic network with S links can be characterized as a T -step sequence $[x_1, x_2, \dots, x_t, \dots, x_T] \in \mathbb{R}^{T \times S}$



Time 1



x_1

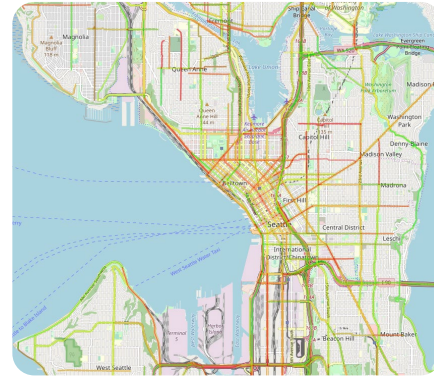


Time 2



x_2

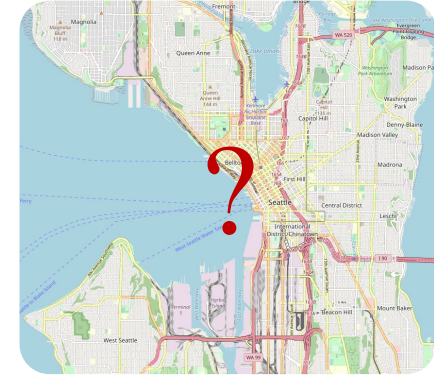
...



Time T



x_T



Time T+1



x_{T+1}

AI Transit: Transit Origin-Destination Data Collection through Occlusion-Resistant Passenger Re-Identification

Motivation

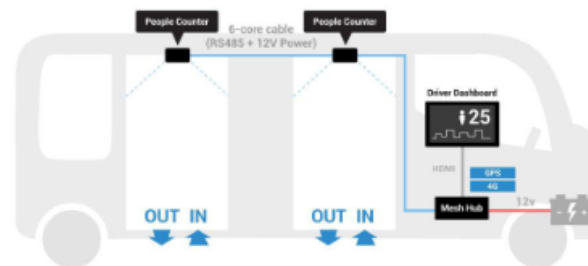
- Transit origin-destination (OD) data is useful in the design of transit systems for more efficient route designing, schedule planning, and ridership forecasting
- Automatic passenger counts (APC) do not tell the full OD story
- Existing approaches rely on manual surveys, Bluetooth or WiFi data
- Computer vision algorithms can remember features of persons entering and match them to exits
- Coordination with GPS data can allow the calculation of O-D pairs



Motivation

Current methods

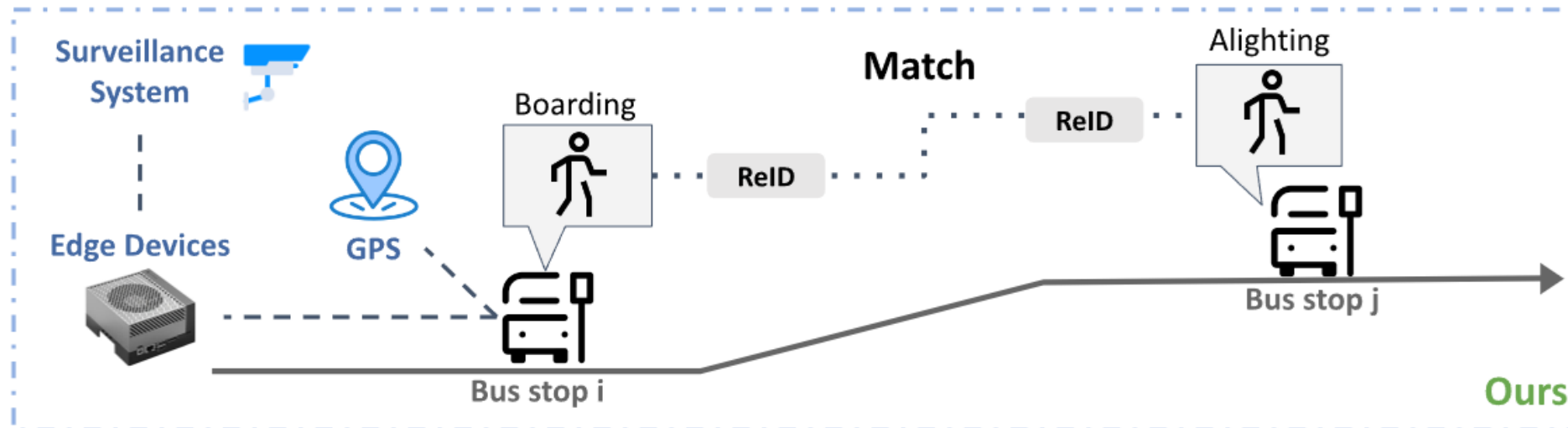
Surveys
Bluetooth data
Mobile phone data
Automated Passenger Counters
Automated Fare Collection Systems



Disadvantages

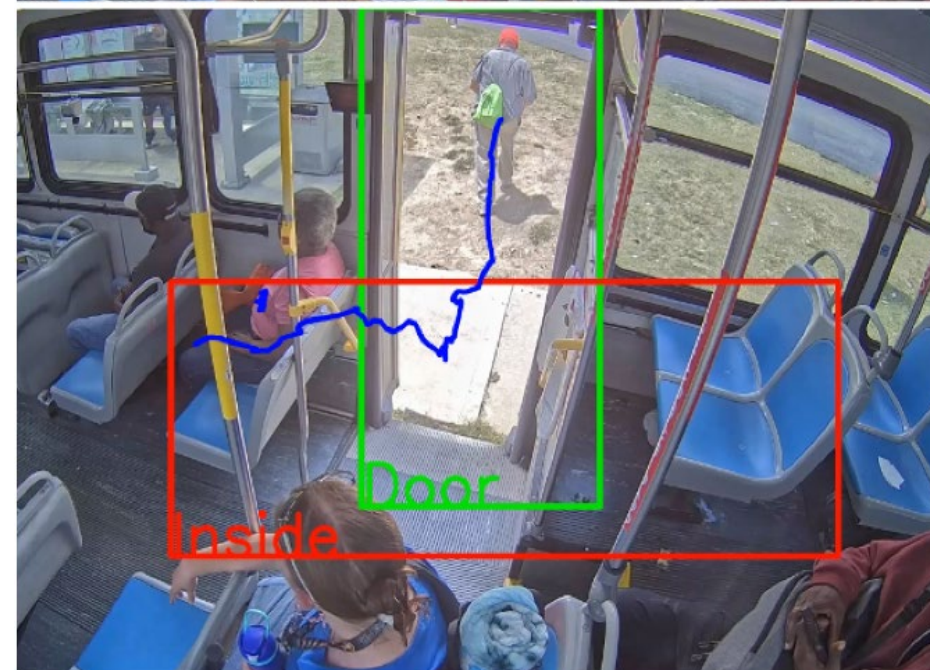
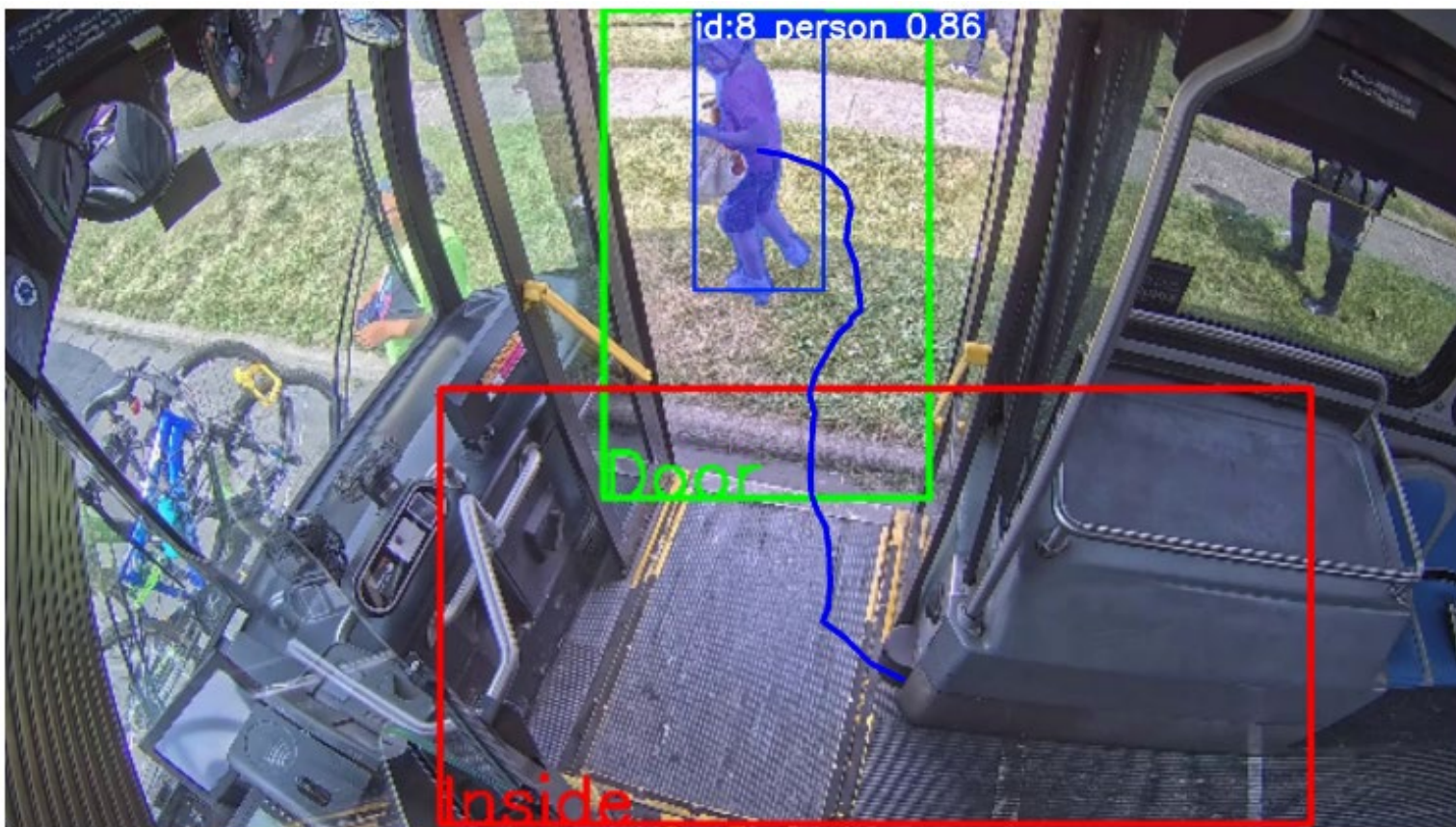
Labour-intensive
Need specific devices
Limited data coverage
Lack individual-level matching

Transit OD Data

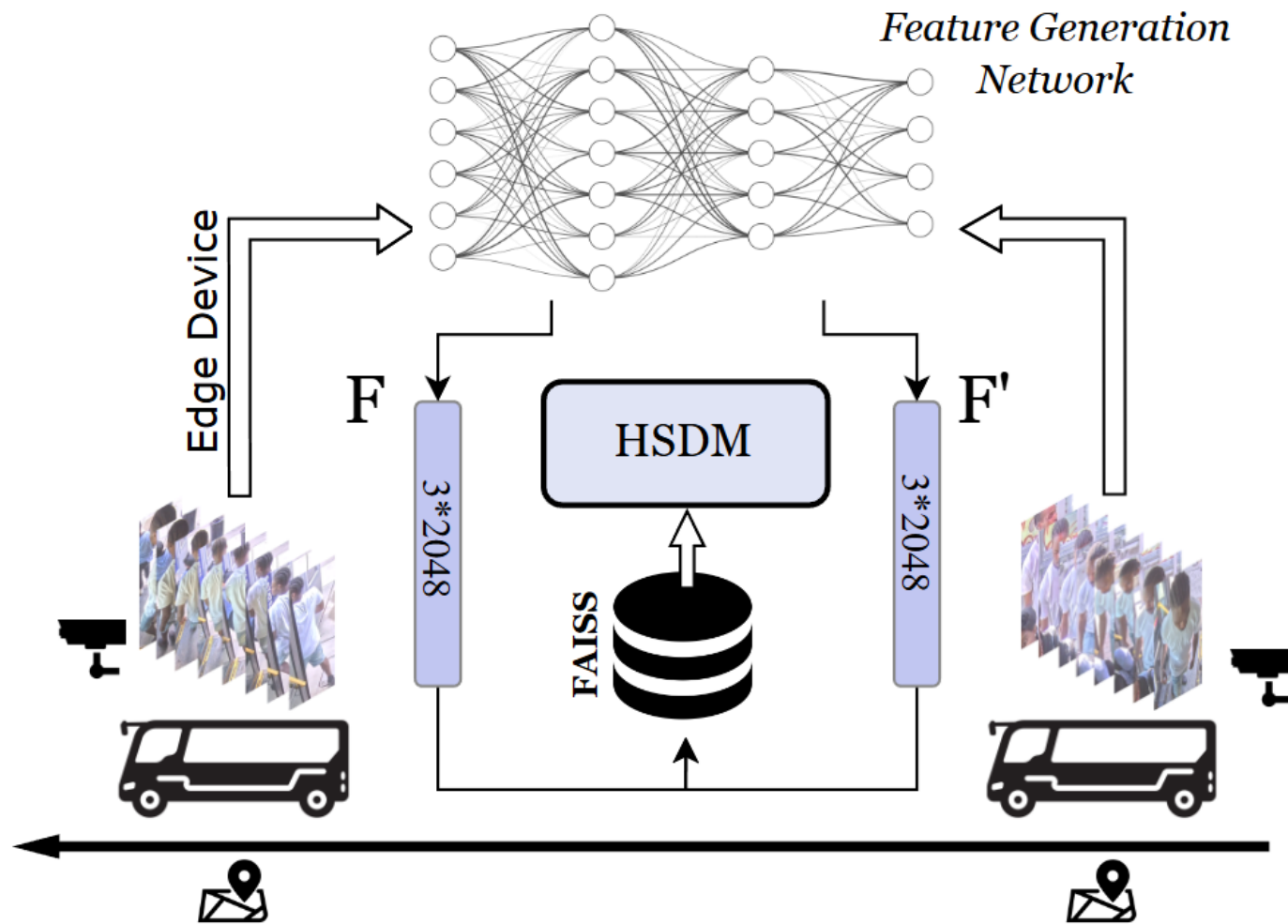


Motivation

- A person can board, alight, or move across the frame



Framework



**Features of
boarding**

F1
F2
F3
...
Fk

**Features of
Alighting**

F'1
F'2
F'3
...
F'k

Occlusion Challenge

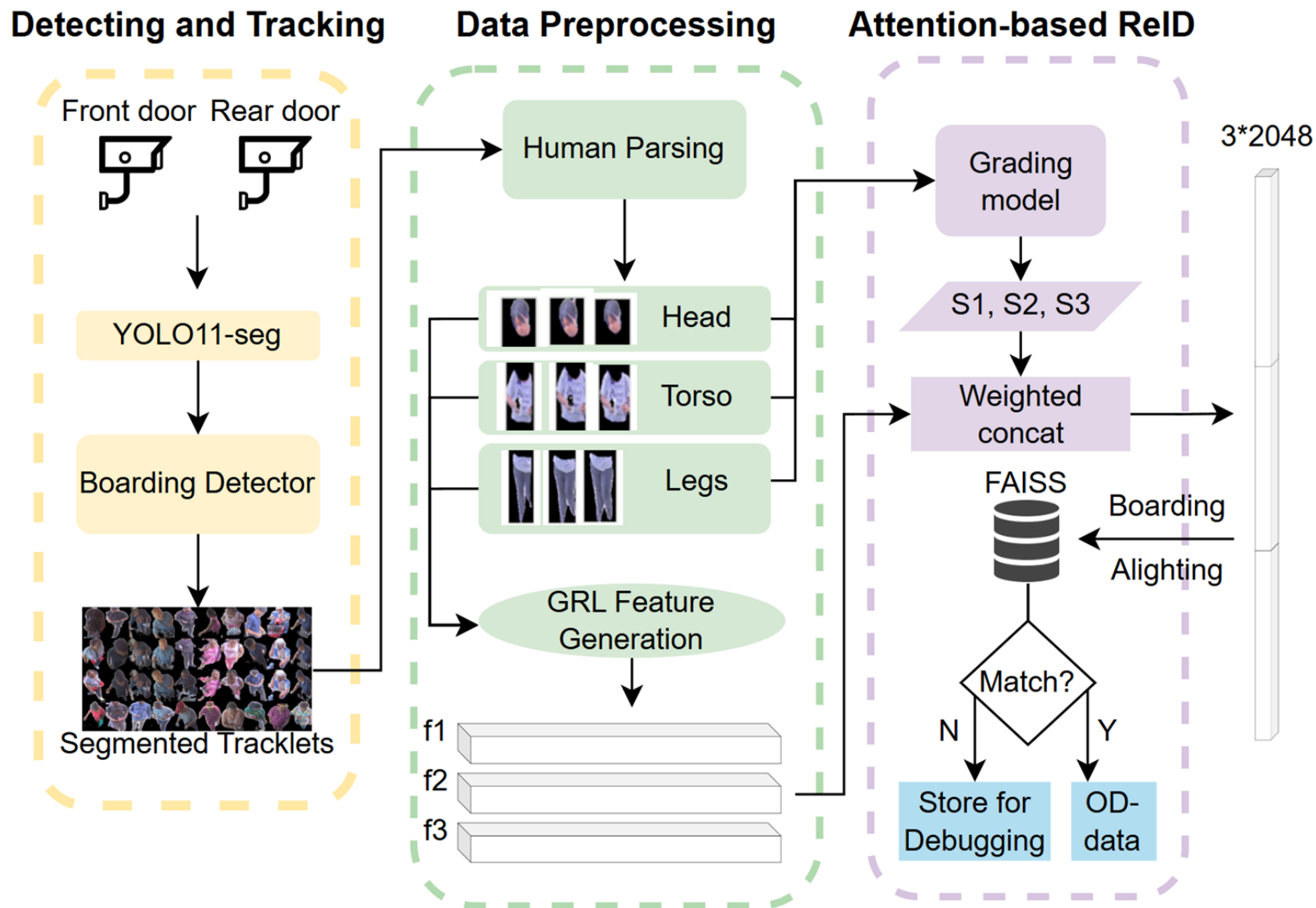


Loss of information will lead to matching errors!

Idea

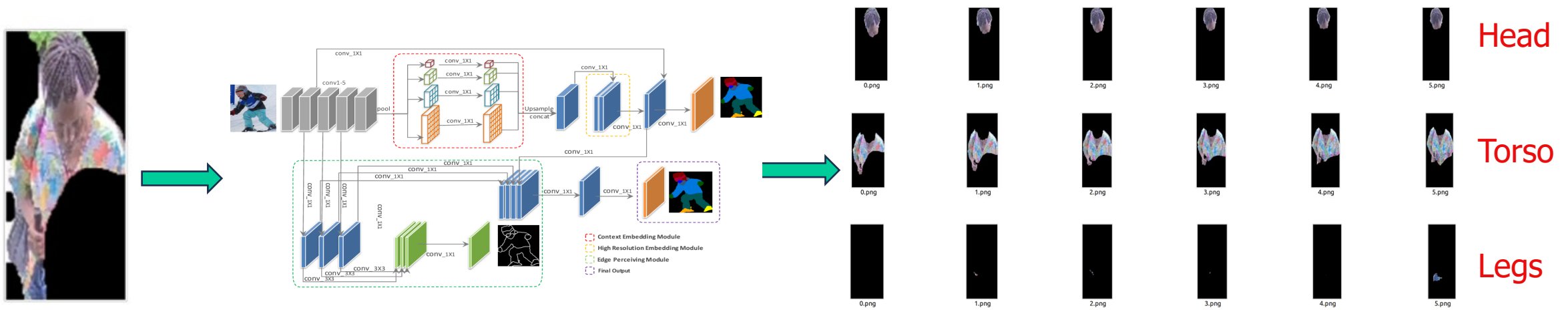
Give different attention to different regions.

Updated Framework



Data Preprocessing

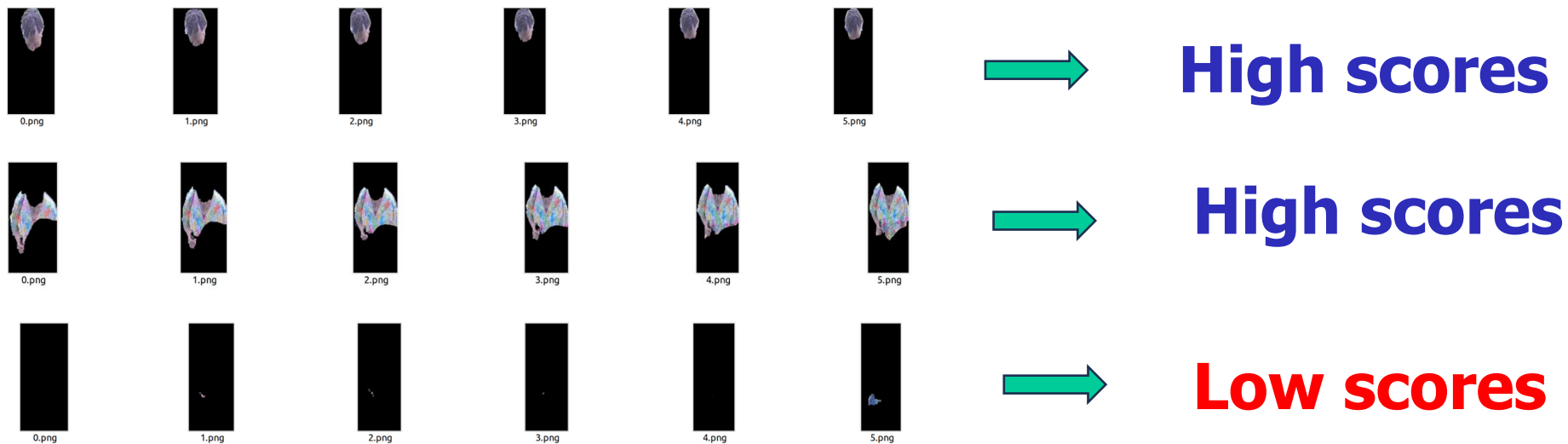
Human Parsing



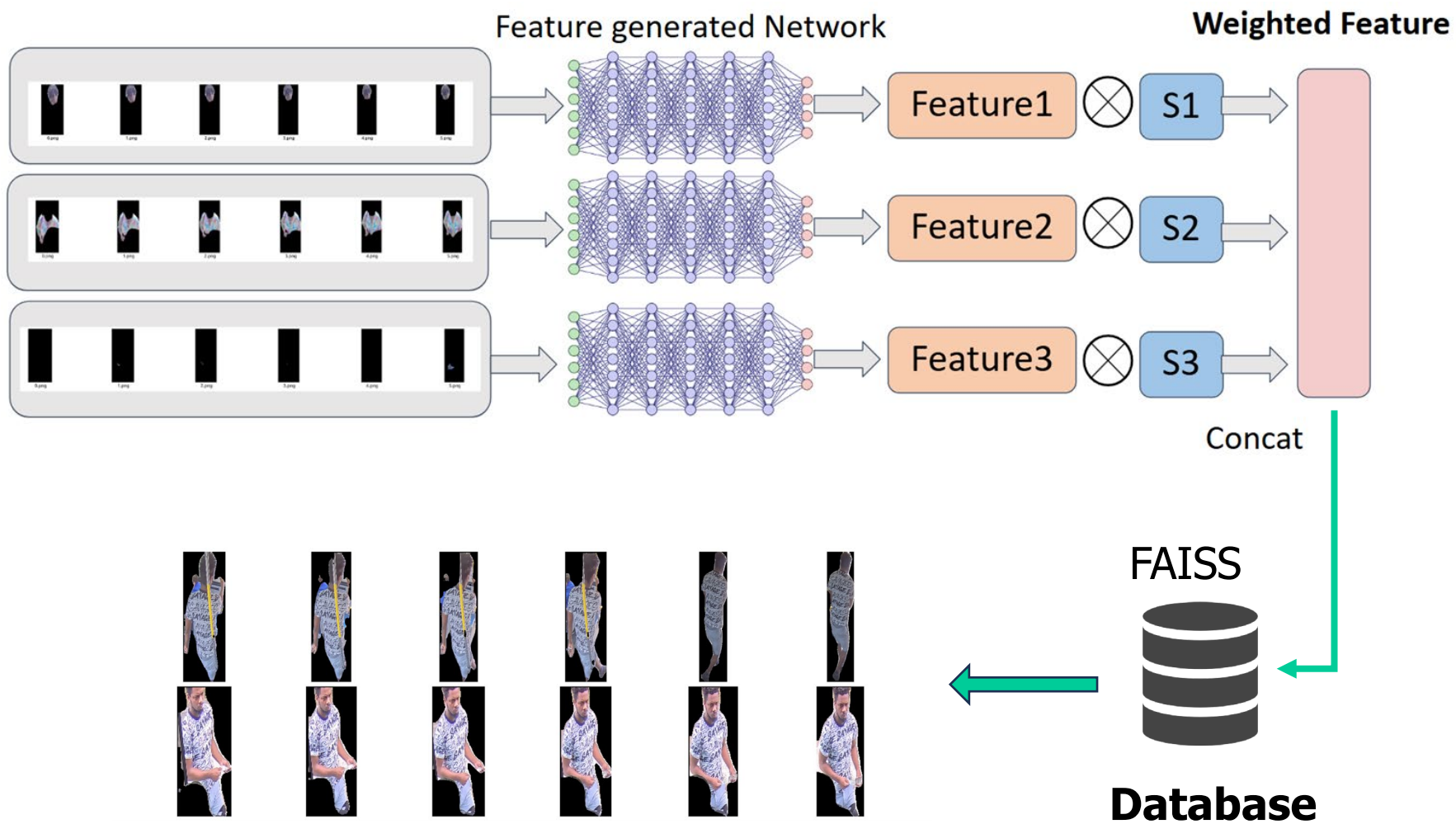
Data Preprocessing

Score Calculation

The scores **reflect the quality/occlusion condition** of each body part

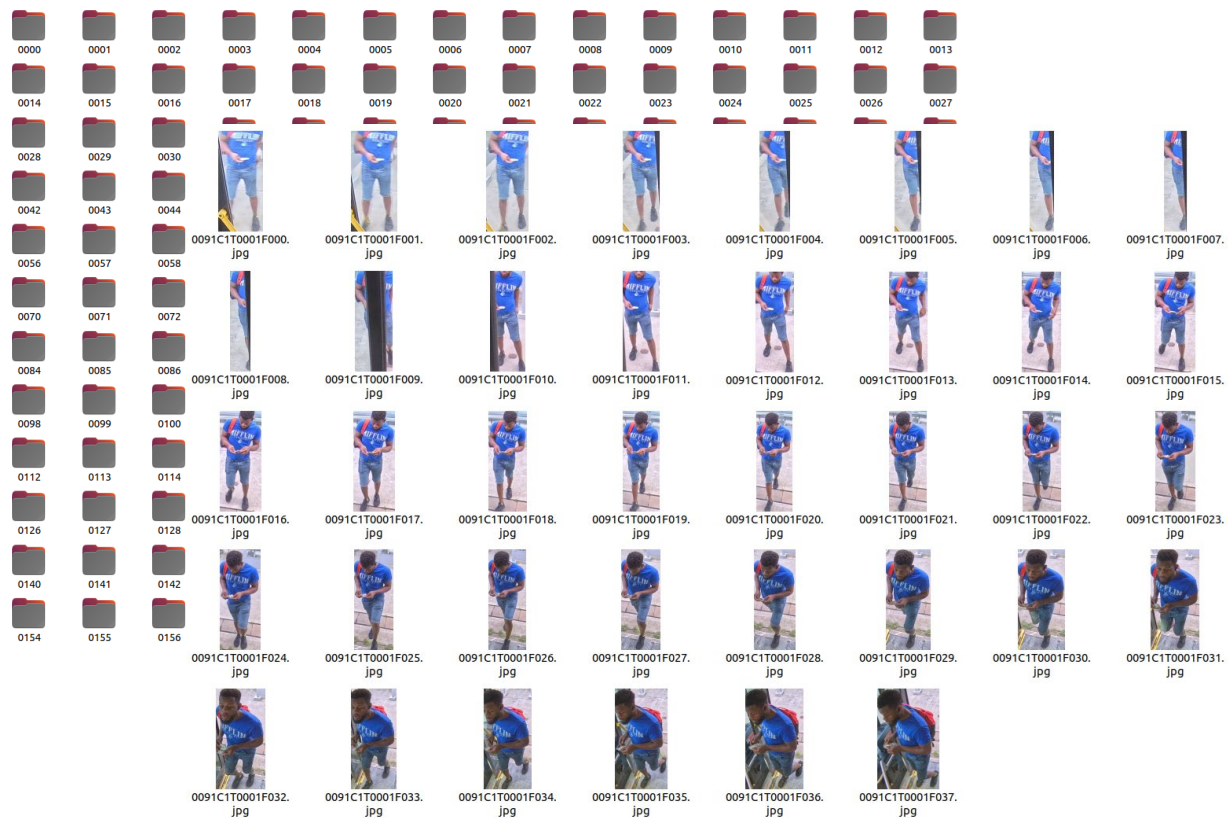


Attention-based Re-ID



The same person?

Experiment



Dataset

subset | # ids | # tracklets

train | 97 | 198
query | 61 | 63
gallery | 61 | 63

total | 158 | 324
number of images per tracklet:
6 ~ 443, average 55.8

Rank-5 retrieved results

ID: 13



ID: 18



ID: 21



ID: 41



ID: 42

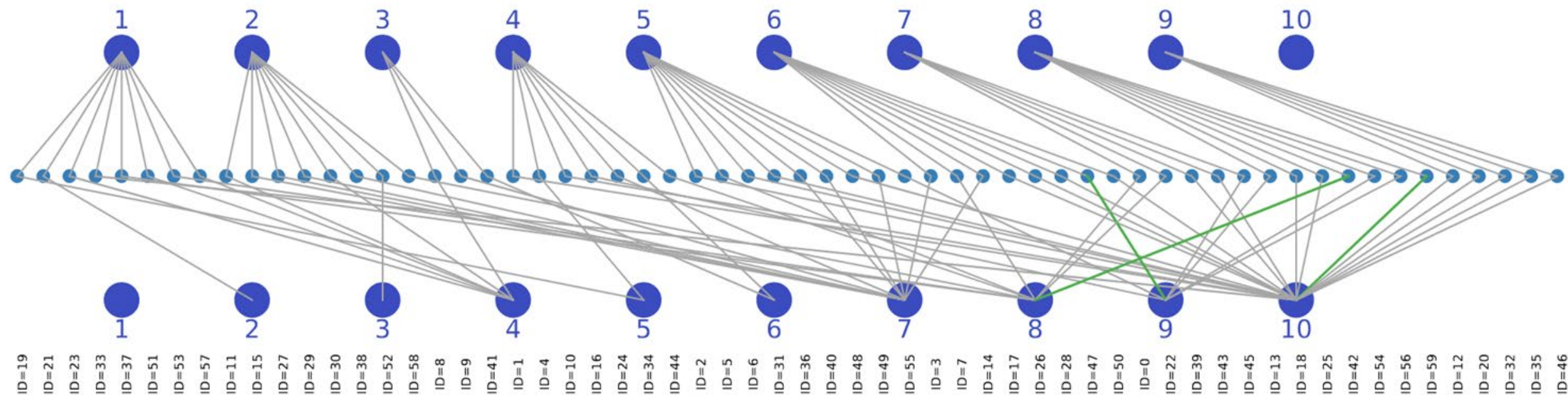


ID: 47



Experiment

O-D collection simulation with 10 bus stops





ChatSUMO: Large Language Model (GenAI) for Traffic Simulation Generation and Customization

Introduction

- Traffic congestion costs the U.S. economy over \$160 billion annually in wasted time and fuel, while also contributing to over 25% of greenhouse gas emissions from transportation.
- Despite the advancements in traffic simulation tools like SUMO and VISSIM, users still face a steep learning curve and significant time investment to create and customize simulations.
- This limits
 - The use of simulation to examine different traffic scenarios
 - The efficiency and effectiveness of finding optimal solutions to real-world problems

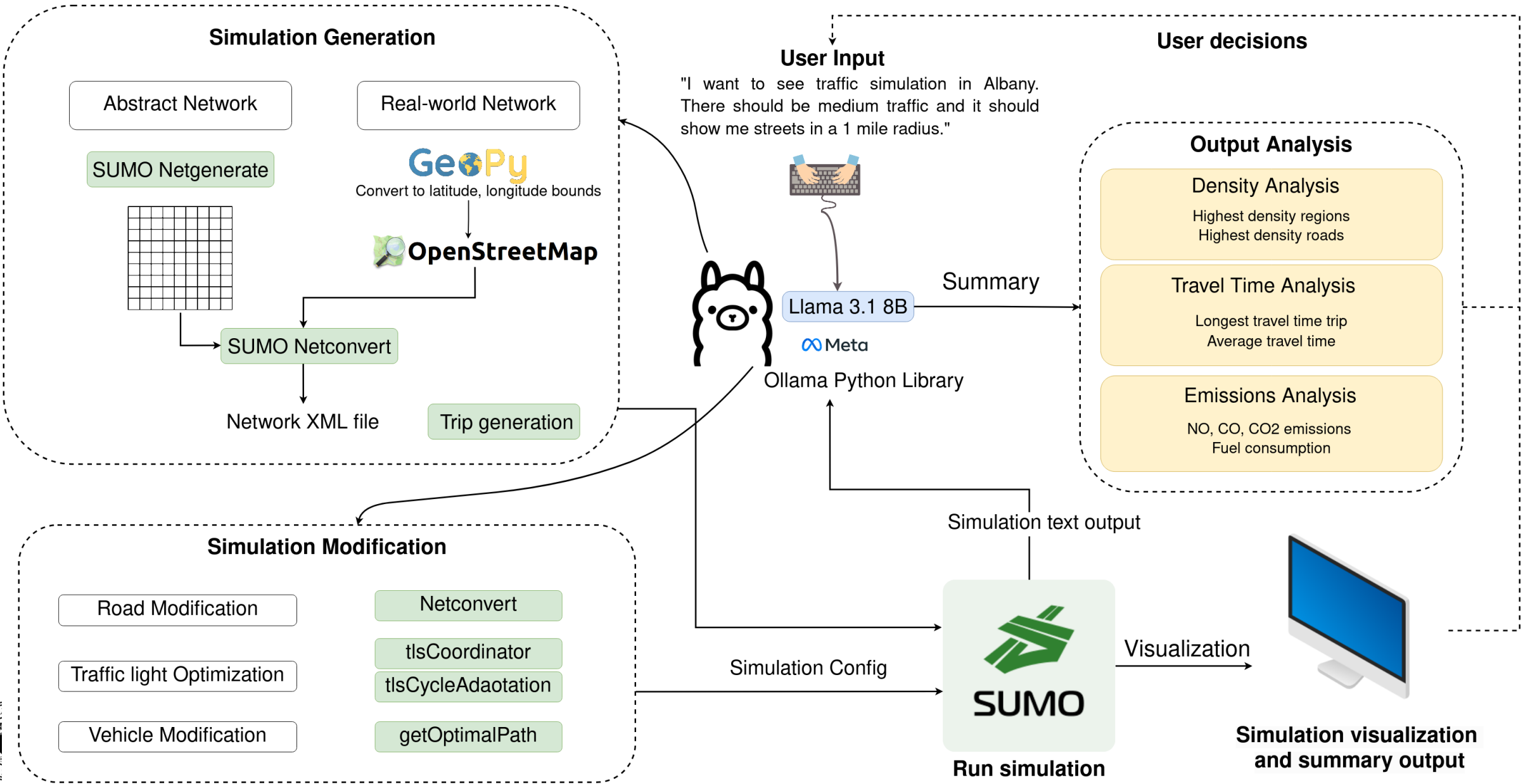


Introduction

- ChatSUMO integrates a Large Language Model (LLM) to simplify SUMO simulation creation
- Enabling users to design traffic scenarios using natural language
- Four-module framework: Input, Simulation Generation, Customization, and Analysis
- Uses OpenStreetMap to generate location-specific simulations
- Speeds up the initial setup for experienced users, and lowers the learning curve for new ones



Overall Framework



Simulation Generation

"""Show me a simulation inAlbany. There should be medium trafficand it should show me streets in a3 miles radius.
 """

Prompt generation



Llama 3.1 8B

```
{"City": "Albany", "radius": "3",
  "traffic_condition": "medium"}
```

Python Script

```
# Generate Network
generator.downm
load_osm_data(bbox)
generator.convert_osm_to_sumo(
)

#Generate Trips
generator.generate_trips

# Create sumocfg file
generator.create_config()

# Run simulation
simulation.run_simulation()
```



Road Segment Edit

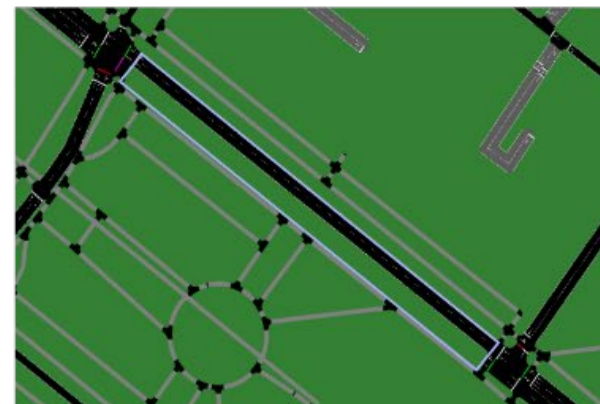
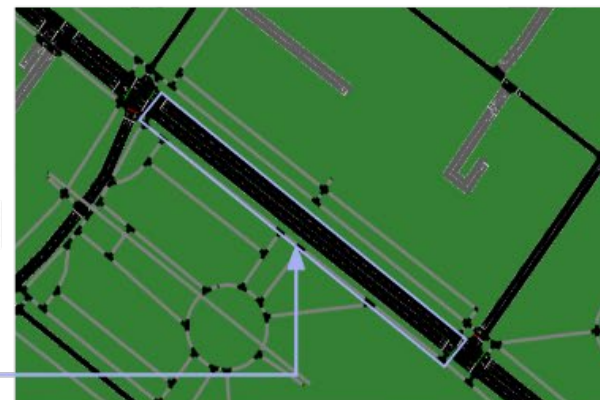
User Prompt: "I want to delete a section of Washington Avenue"



Llama 3.1

Identify the target road

```
# Read the net file
tree = ET.parse(net_file_path)
# Find the target edge
if edge.get("name") == edge_name
# Remove the edge
netconvert -s albany.net.rou --
remove-edges.explicit {edge.id}
```



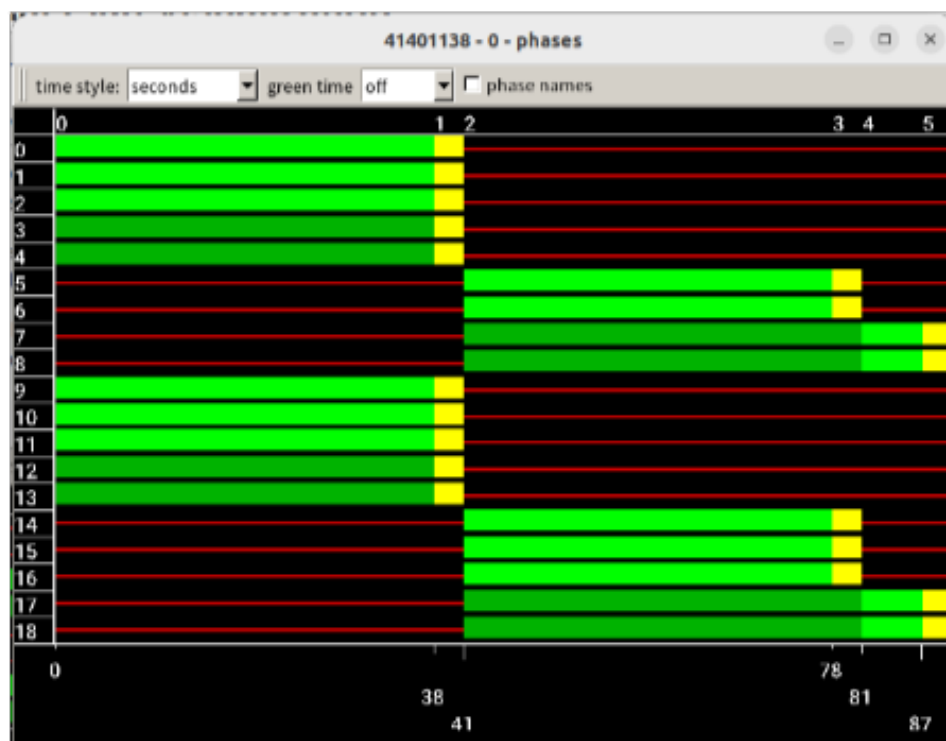
Traffic Signal Light Adaptation

Prompt: "I want to adapt the traffic lights with current traffic demand"



Llama 3.1

```
python tlsCycleAdaptation.py
-n Albany.net.xml
-r Albany.rou.xml
-o newTLS.add.xml
```



Websters Formula



Interface

ChatSUMO

C Hello, what type of network do you want to generate, abstract or real-world?

type of network

real world

Which city do you want to simulate and what kind of traffic do you want to see?

I want to see Paris with heavy traffic

Run Simulation

Here is a brief summary:

Traffic Simulation Results: Paris

- Average traffic density on major roads was around 341 vehicles per kilometer.
- Travel times were significant, with some trips taking 5483 seconds
- Average travel time across all trips was approximately 1048 seconds (~17.5 minutes).
- Emission output was moderate, with:
 - Total CO₂ emissions: 1.47 tons
 - Total CO emissions: 0.06 tons
 - Total NO_x emissions: 0.0006 tons
 - Total PM_x emissions: 29 μg
 - Total fuel consumption: 469 L
 - Total electricity consumption: 91128 kWh

Please enter the probability of gas vehicle and electric vehicle

Half of the vehicles should be electric and the other half should be gas |

Llama output:

Here's a brief summary:

Traffic Density: The average traffic density in Paris was 348.81 vehicles per kilometer, with some areas reaching up to 433.01 veh/km.

Travel Time: The average trip took around 1051 seconds (approximately 17.5 minutes), with the longest trip taking around 5484 seconds (around 91.4 minutes).

Emission Output: The simulation generated a total of:

- 1.03 tons of CO₂
- 0.043 tons of CO
- 0.00044 tons of NO_x
- 0.0000203 tons of PM_x
- 0.329 tons of fuel consumed
- 153,602.84 kWh of electricity used

Would you like to make a comparison with a previous simulation, which one you want to compare?

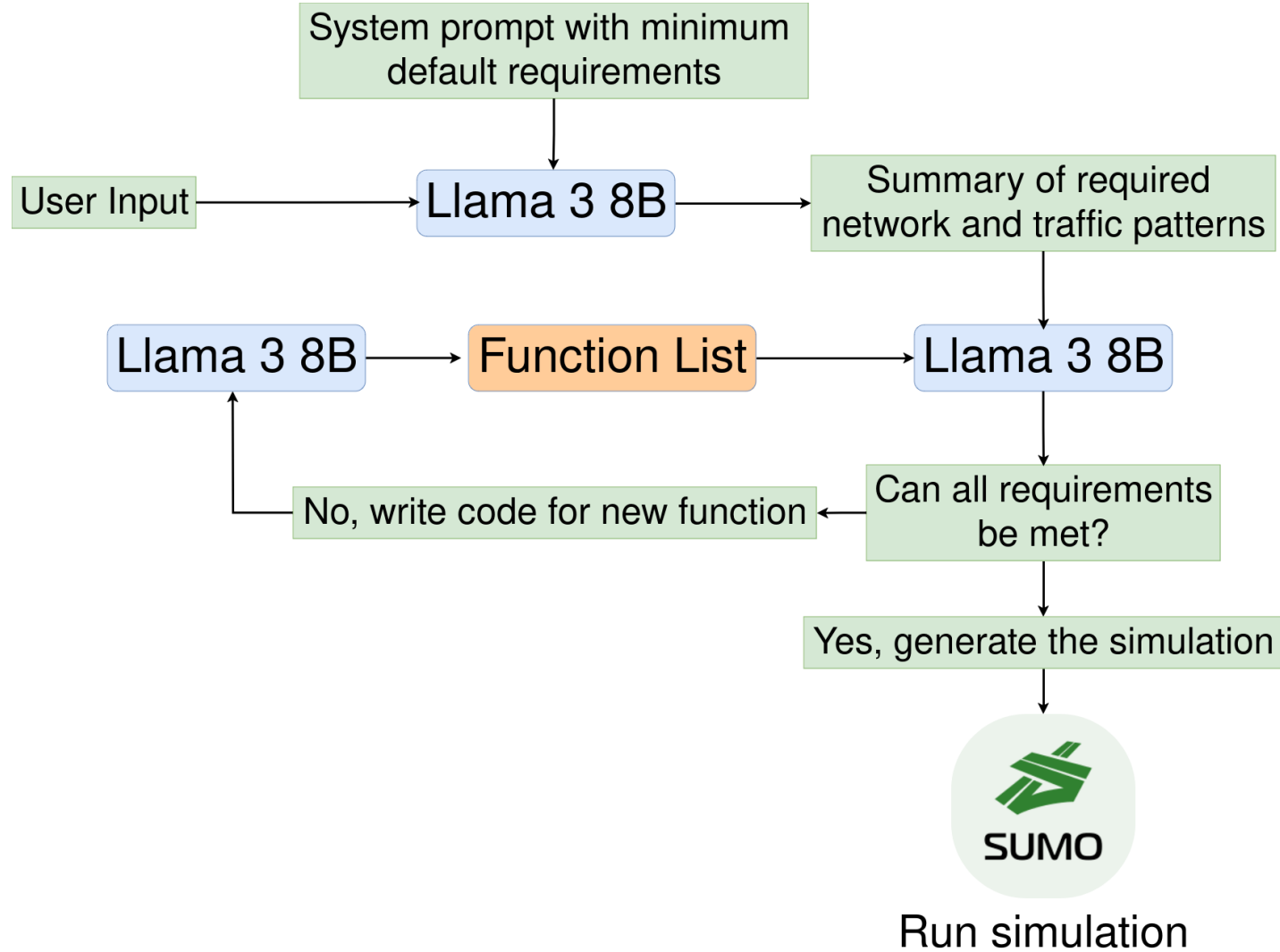
Value

Confirm

ChatSUMO Web Interface

Changing the distribution of electric and gas vehicles





Updated Methodology

ChatSUMO

Phase 2: Modification or Analysis

The simulation has been completed. You can now request modifications or view the analysis.

Choose an action:

- Request Modification
- View Analysis
- Restart Simulation Setup

Describe the modifications you want:

Available Functions:

- `generate_traffic_light_offsets`
- `generate_optimal_traffic_lights`

Your modifications:

edit the `sumocfg` file to use the `newTLS.add.xml` file

Press Ctrl+Enter to apply

Submit Modifications

Available Functions:

- `generate_traffic_light_offsets`
- `generate_optimal_traffic_lights`
- `edit_sumocfg`

Your modifications:

write code to remove a specified edge from the road network

Submit Modifications

Result: Function ``edit_sumocfg`

or

`def remove_edge_from_network(net_name, edge_id)`` is not available.

Function added successfully!

Modifications have been submitted. You can now re-run the simulation.

Quantum-Driven Vehicle Routing: Realizing QAOA on Gate-Based Quantum Computer

IBM-Rensselaer Quantum System One

Rensselaer Polytechnic Institute and IBM Unveil the World's first IBM Quantum System One on a University Campus

Now online, the IBM quantum system will expand the longstanding RPI and IBM partnership to accelerate quantum computing research, workforce development, and education in New York.

April 5, 2024

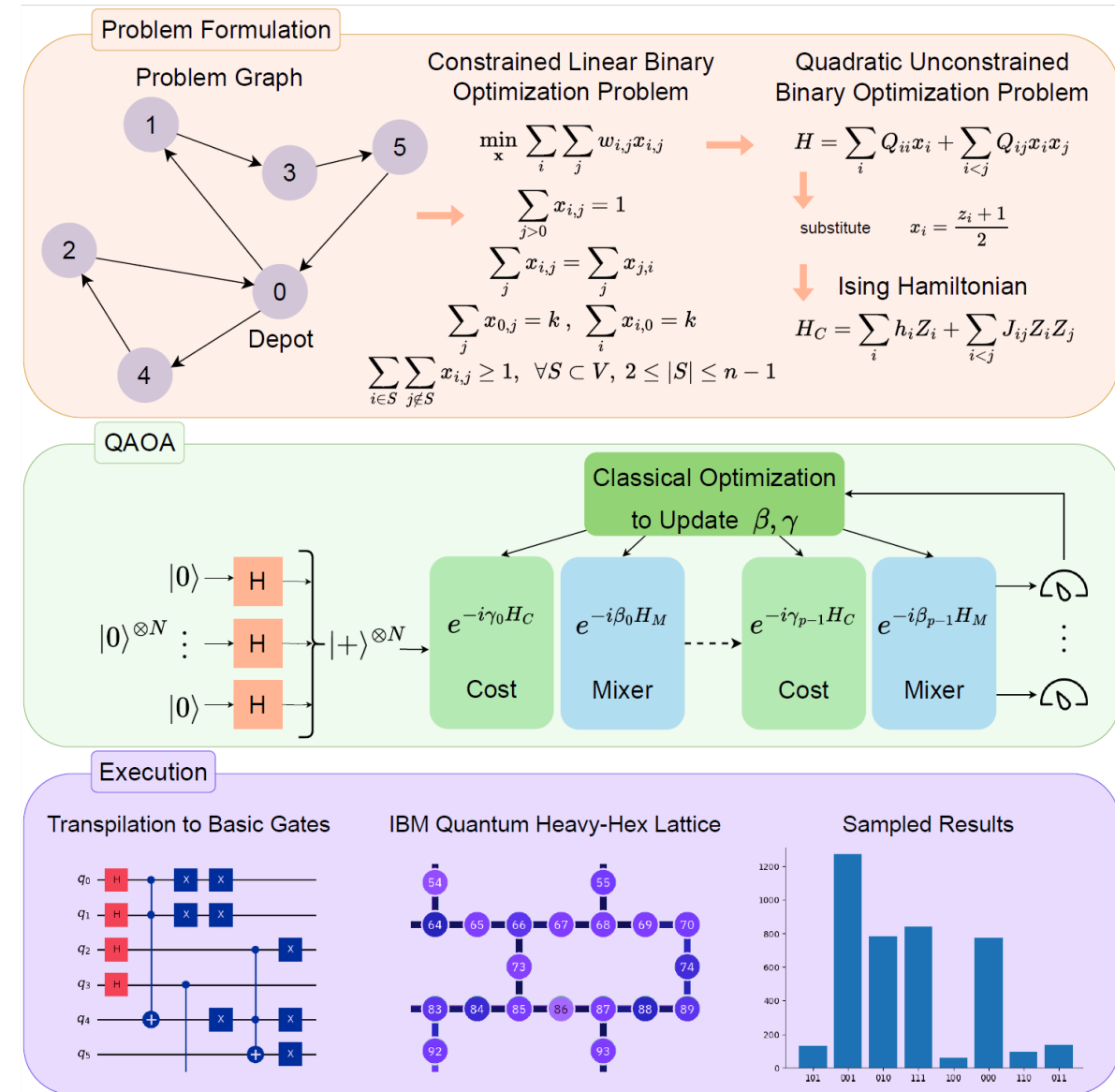


The IBM Quantum System One at Rensselaer Polytechnic Institute, unveiled on April 5, 2024. Credit: IBM

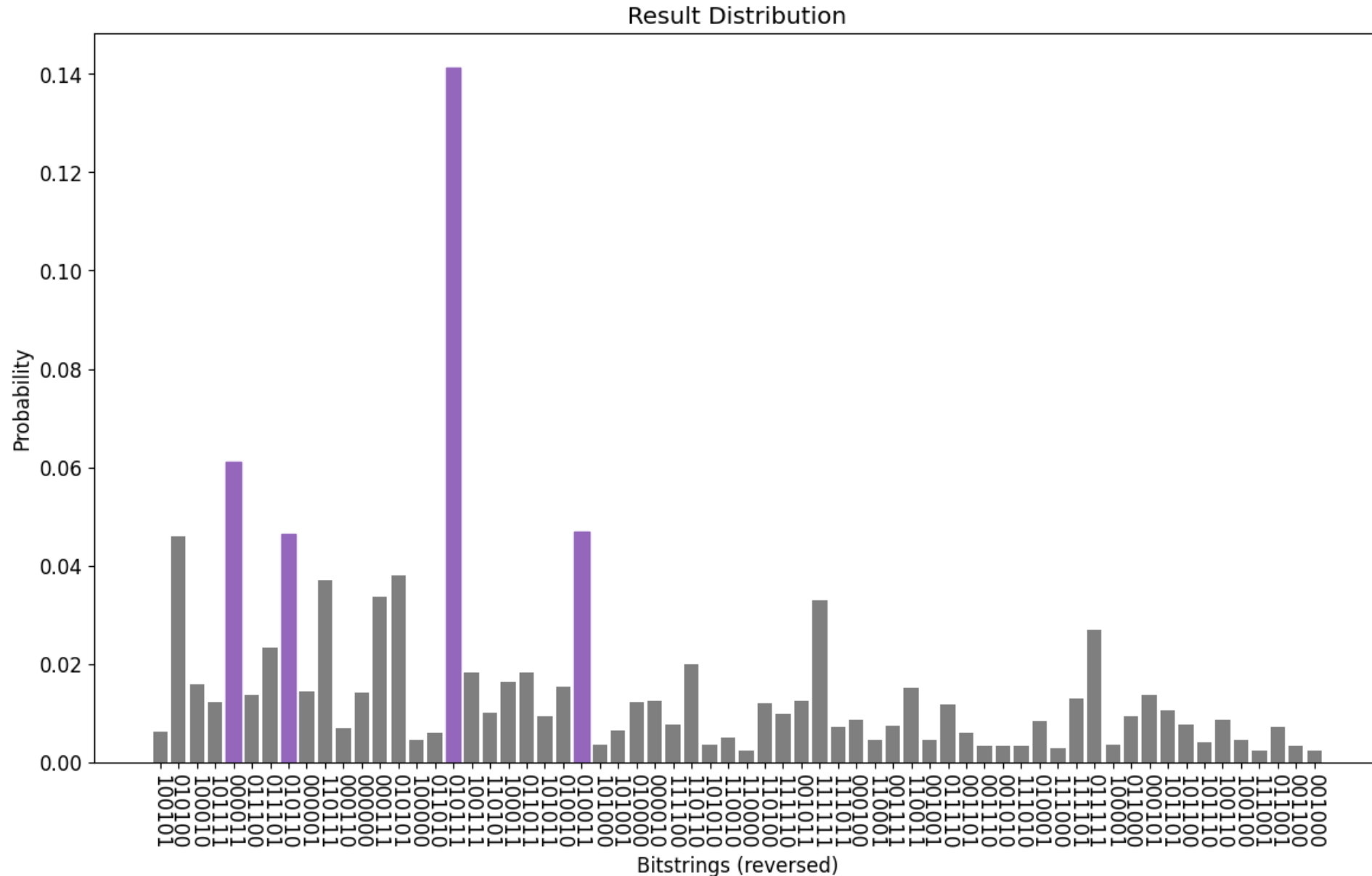


QAOA for Vehicle Routing Problem

- Three (3) nodes – One depot and two receivers
- Two (2) vehicles
- Six (6) links in total, each modeled as a qubit
- $C = \sum_{i,j} d_{ij} x_{ij}$
 - d_{ij} : Distance between nodes i and j
 - x_{ij} : Binary decision variable, 1 if route $i \rightarrow j$, 0 otherwise
- Solution: A final bit string with the largest number of measurements
- Compared with the classical solution from CPLEX - same cost



Vehicle Routing Problem Formulated and Implemented



Conclusions

AI Transportation Vision

- AI-Driven Smart Transportation
 - Autonomous vehicles with advanced AI perception and decision-making.
 - AI-powered traffic management for congestion reduction and optimization.
 - AI-enhanced mobility-as-a-service (MaaS) platforms integrating multiple transport modes.
- Advanced Computing in Transportation
 - Edge computing and federated learning for real-time decision-making.
 - Quantum computing for optimization problems in traffic routing, network modeling, shared mobility.
 - Digital twins and AI testbeds for simulating and improving transportation systems.
- Sustainability and Efficiency
 - AI-driven energy-efficient vehicle routing and speed planning.
 - Large language model for traffic simulation generation
 - Enhancing disaster resilience with UAV-assisted management of heterogeneous edge devices
- Human-Centric AI in Transportation
 - AI-driven adaptive driving assistance tailored to driver behavior.
 - AI-supported shared mobility and paratransit optimization for accessibility.
 - AI for enhancing safety through real-time hazard detection.



Questions?

Ruimin Ke – ker@rpi.edu

Advanced Simulation Framework for Enhancing Scenario Planning in Sustainable Transportation

Dr. Sean He – Associate Professor
Rensselaer Polytechnic Institute

May 14, 2025



Rensselaer

why not change the world?®

Background

- **Project:** AI-Powered Agent-Based Simulation

- **Funded by:**



Center for Social and Economic Mobility for
People And Communities through Transportation

SEMPACT

USDOT Region 2 University Transportation Center
New Jersey | New York | Puerto Rico | US Virgin Islands



- **Goal:** develop a comprehensive multiclass simulation platform by integrating real-world data with AI-powered modeling tools to enable policymakers to evaluate policies holistically, uncover hidden patterns, and make informed decisions

Simulation Modeling

- Explores "what-ifs"
- Supports data-driven insights
 - Emergency preparedness
 - Policy impact assessment
- Challenges in Practice
 - Transportation network complexities (Network scale and detail, Multimodal and multi-class Interactions, Spatiotemporal heterogeneity, Feedback loops)
 - Lack of panel data (e.g., passenger preferences, B2B delivery patterns)
 - High variability in transportation systems (stochastic demand, behavioral shifts)
- **AI/ML-enhanced simulation**: Uses synthetic data generation and behavior modeling to improve scalability, adaptability, and realism under complex, uncertain conditions



Framework

Scenario Generation



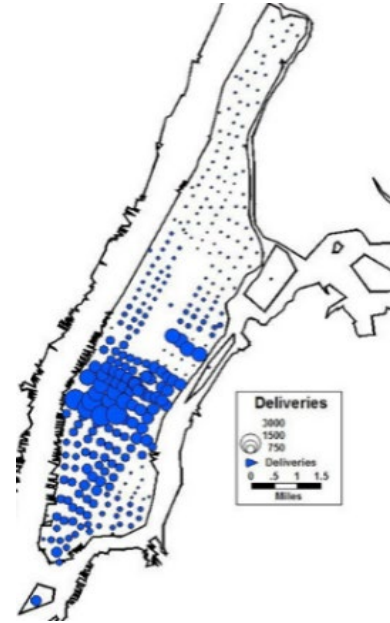
Supply-Side: Tech/Ops, policy constraints, etc.

Demand-Side: Behavior changes, freight demand, etc.

Multiclass Activities

Behavioral Micro-Simulation (BMS)

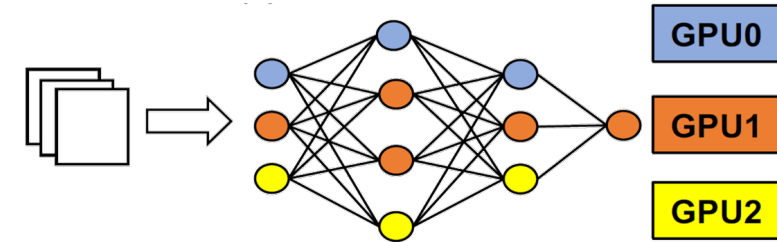
Parking Activity



Simulation Acceleration



MATSim
Multi-Agent Transport Simulation



GPU-Based Parallelism

Data

Census, Detectors, Camera, Radar, Ultrasonic, SafeGraph

Imputation:

GPS, ArcGIS, POIs, Mobile Phone, Trajectory, Inferred flows



NYC MATSim Network (Developed by C2SMART @ NYU)



#1: Multiclass Activities

Behavioral Micro-Simulation

- To produce a reasonable guess of freight traffic in metropolitan areas using:
 - Freight trip generation estimates (using NCFRP 25 models)
 - Known delivery patterns, such as tour length distributions by industry sectors (obtained from data collected by RPI from carriers and receivers)
 - Observed traffic counts at key corridors
- The BMS was originally developed to assess the impacts of alternative policies to foster off-hour deliveries (7PM to 6AM)

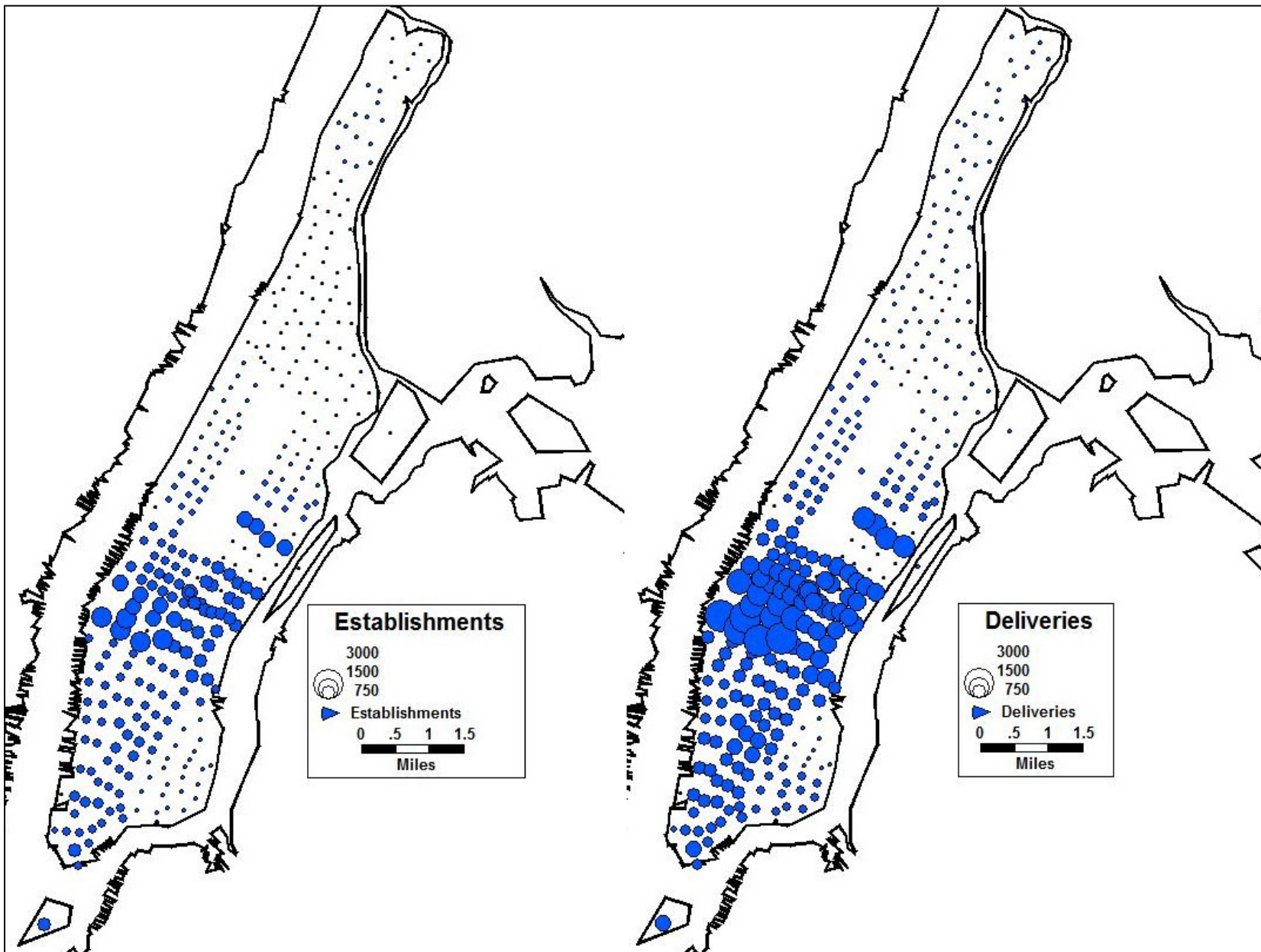


Key Components

- Freight trip generation (FTG): estimated using the NCFRP 25 models and Zip Code Business Pattern data
- Synthetic population of carriers (and receivers, if needed) is created
 - Using the data collected by RPI, the sample data is used to create the population of carriers needed to make all deliveries in the metro area
 - The origin of the deliveries are set to be the locations where warehouses and distribution centers are located
- Delivery tours are created:
 - Match the tour length (number of stops) by industry sector
 - Match the number of deliveries by ZIP code (or any other level of geography used)

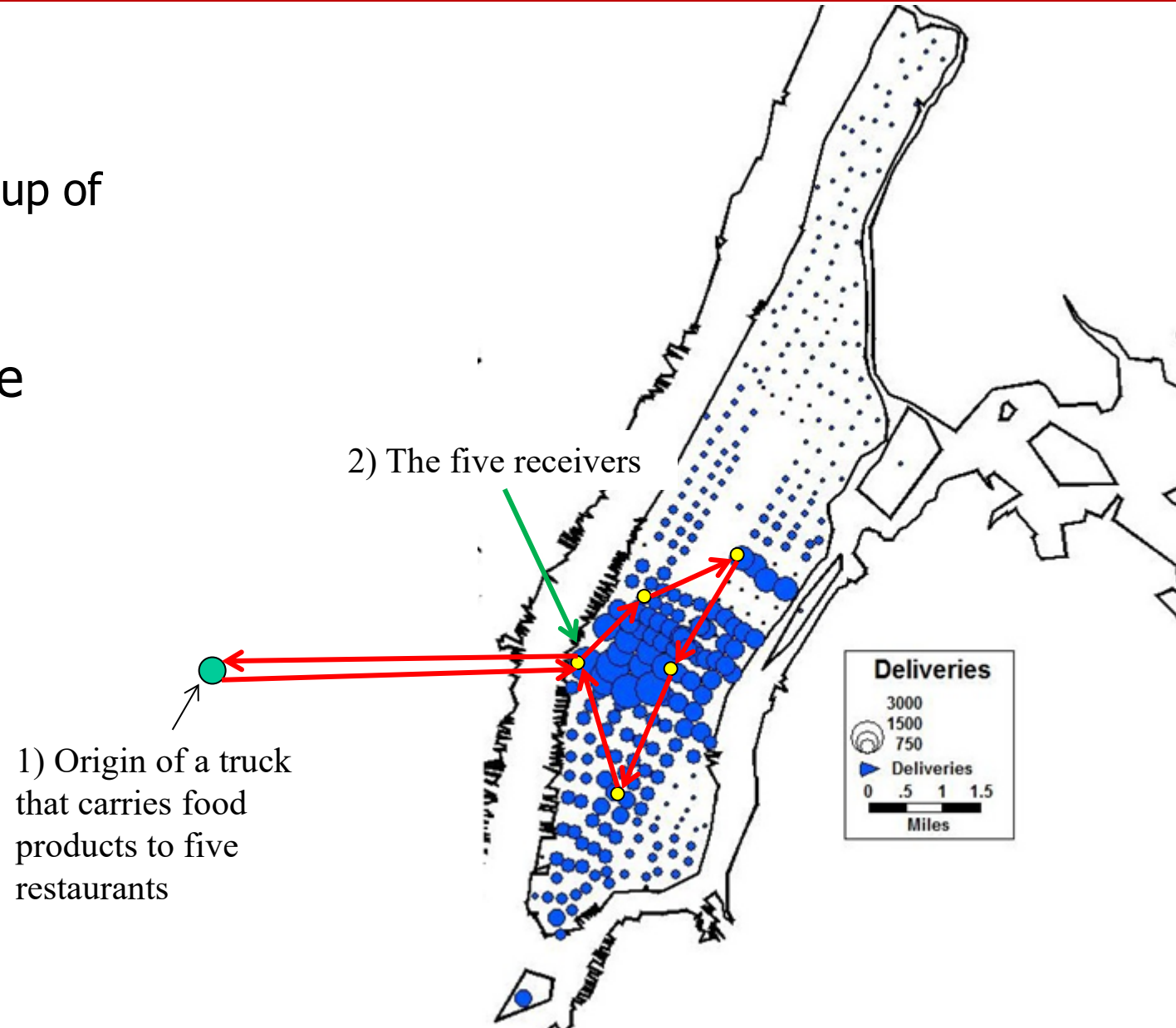


Freight Trip Generation



Tour Simulations

- Select a truck in an industry sector
 - Number of stops is randomly assigned
 - Select receivers at random from the group of receivers in that sector
 - Compute optimal tour and store it
- Repeat until delivery tours satisfy the FTG for the entire area



Parking Activities

- The conflict arising from increased parking demand and limited curbside space motivates the need for efficient on-street parking management

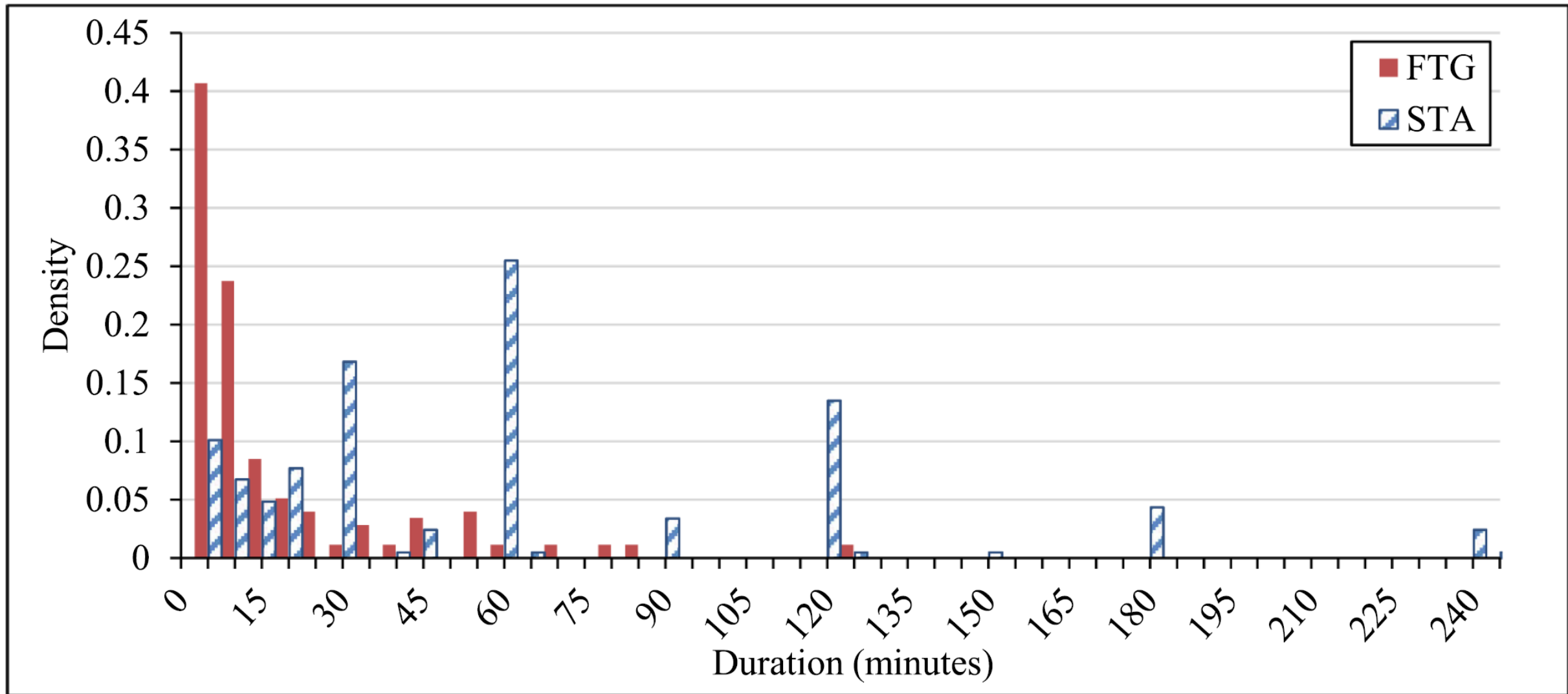


Double-parked freight vehicle

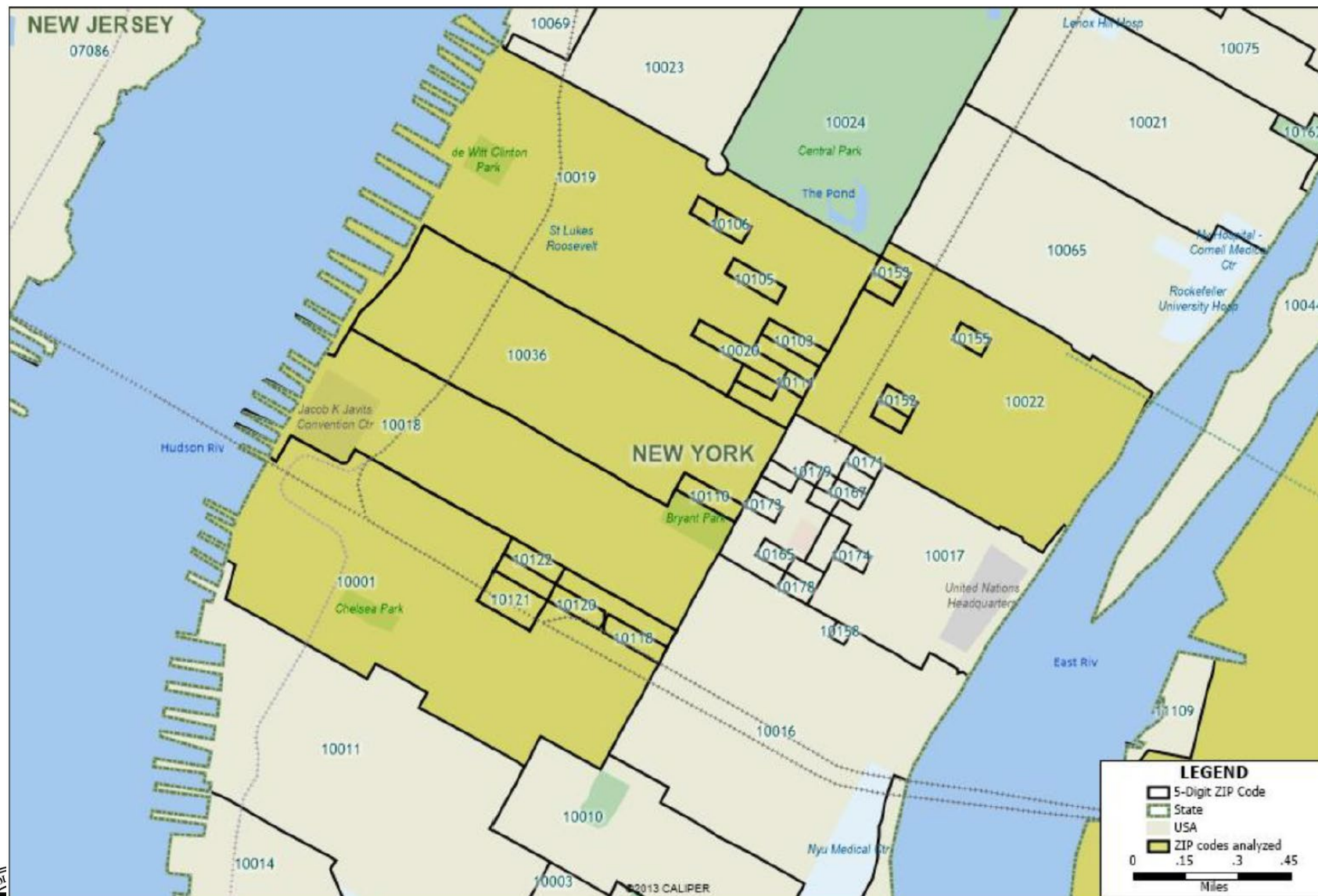


Service vehicle

Service (STA) and Freight (FTG) Duration Distribution

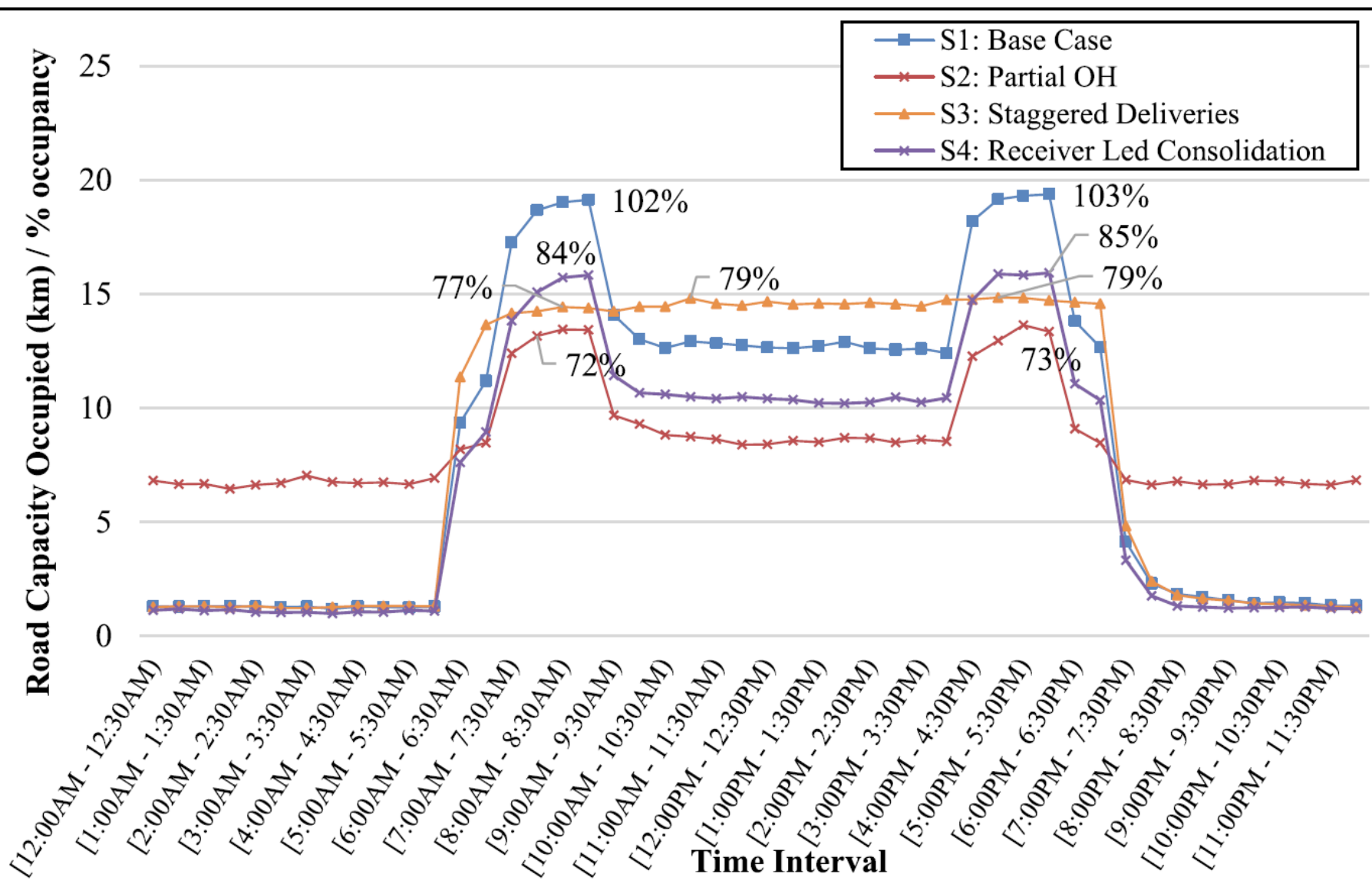


NYC's ZIP codes with highest freight and service activities



Ramirez-Rios et al. "On-street parking for freight, services, and e-commerce traffic in US cities: A simulation model incorporating demand and duration." *Transportation Research Part A: Policy and Practice* 169 (2023): 103590.

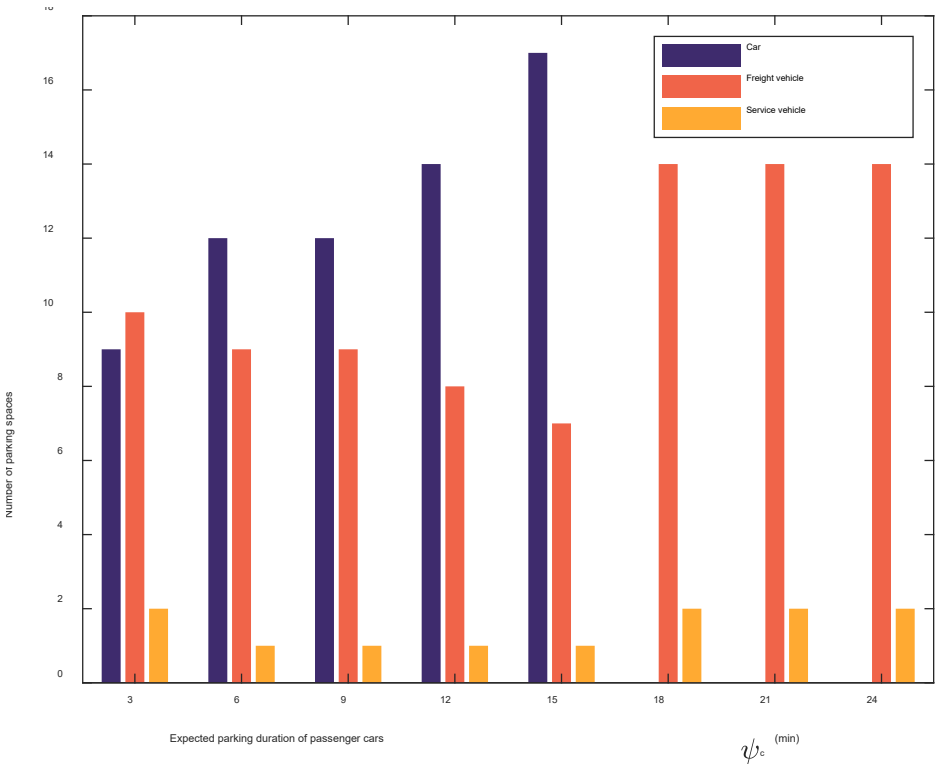
Parking Needs at ZIP 10018



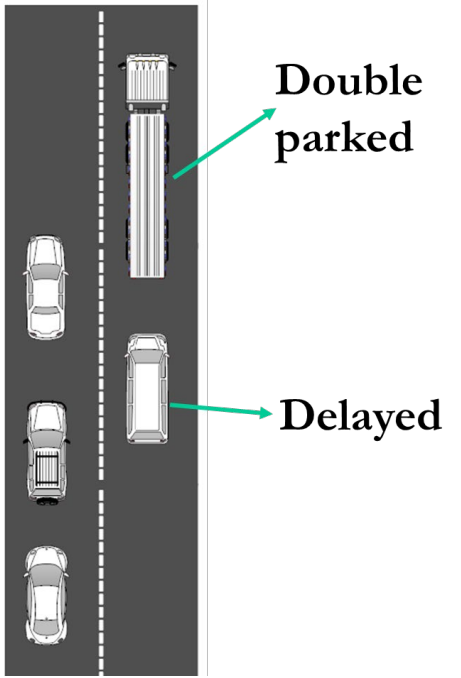
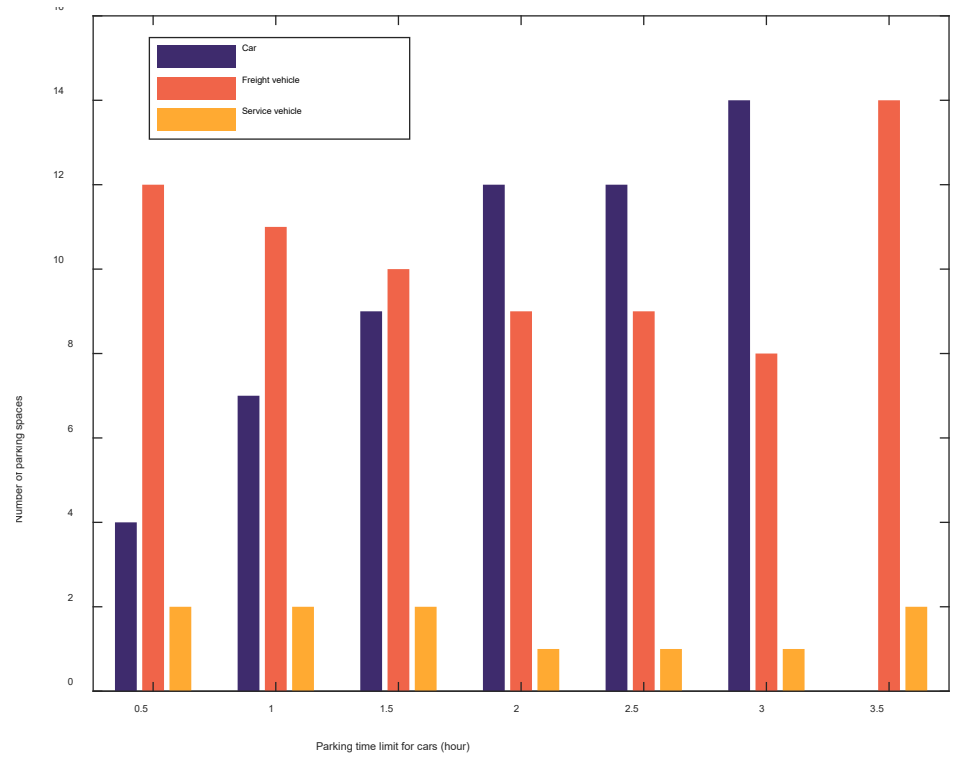
Ramirez-Rios et al. "On-street parking for freight, services, and e-commerce traffic in US cities: A simulation model incorporating demand and duration." *Transportation Research Part A: Policy and Practice* 169 (2023): 103590.

Impact of Double Parking

Without parking time limit

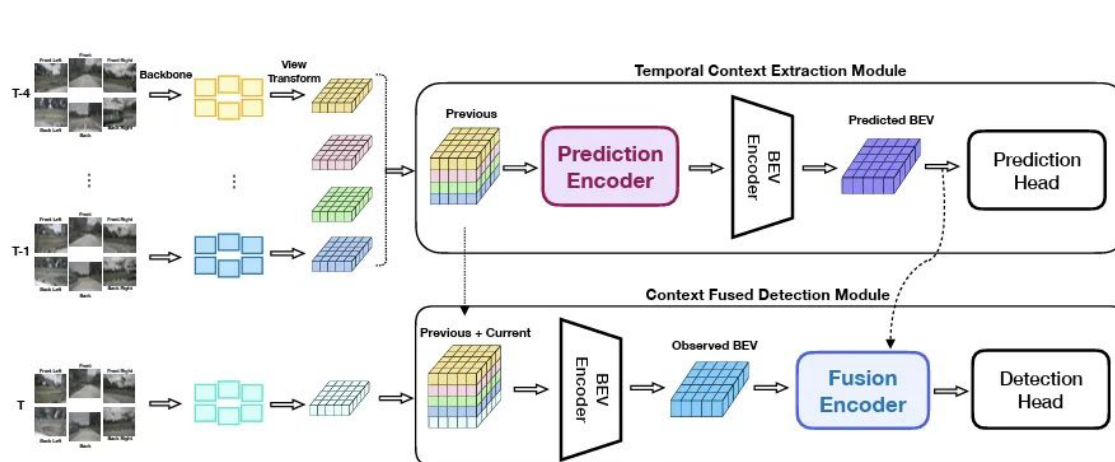
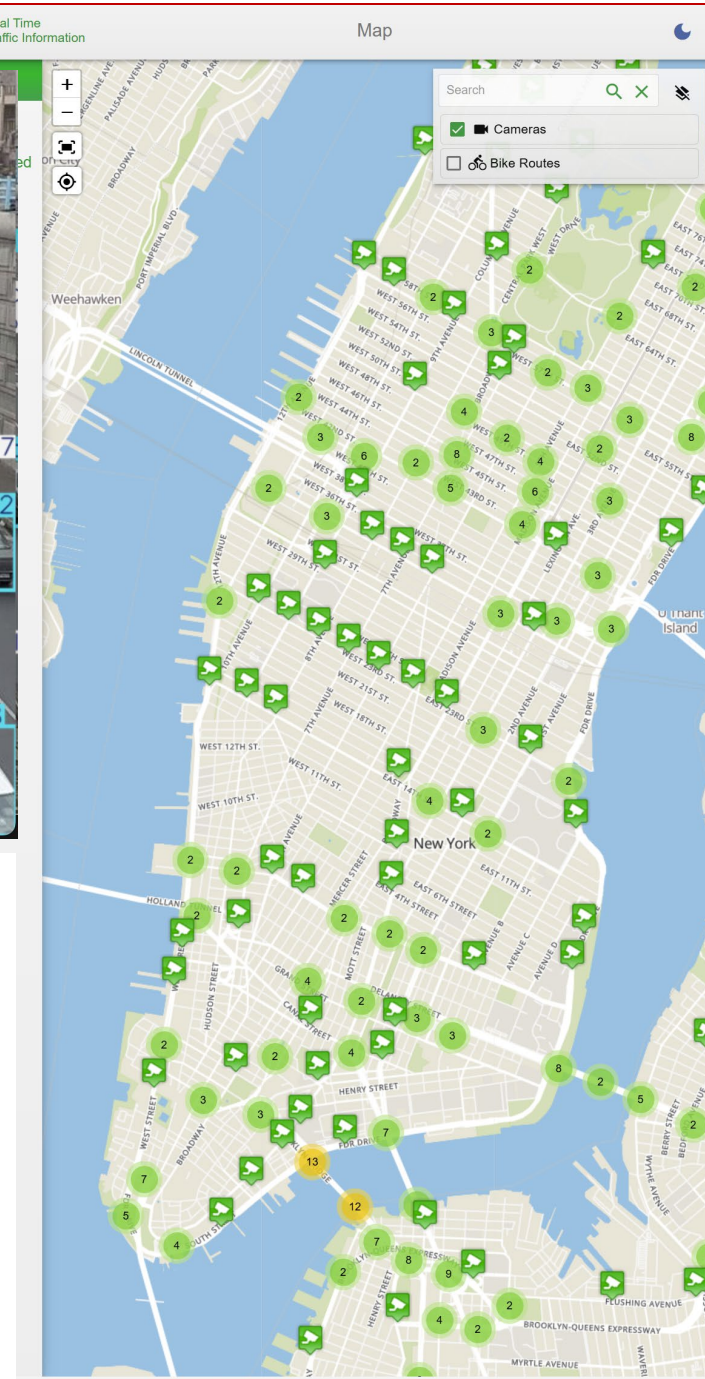
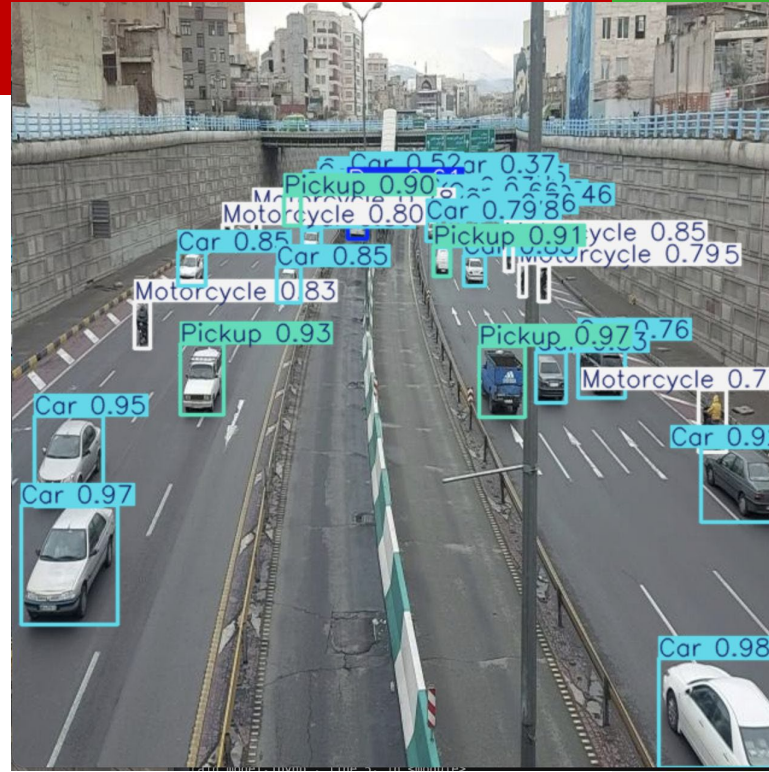


With parking time limit

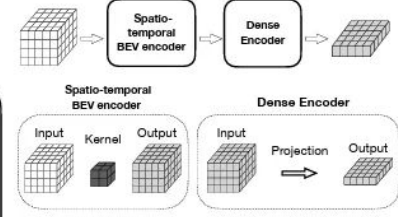


Data Sources

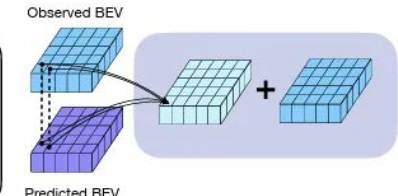
- Detection
 - Camera
 - Radar/Lidar
- Localization
 - GPS
 - Connected vehicles
 - Ultra-Wideband (UWB)



(a) Overall architecture



(b) Prediction Encoder



(c) Fusion Encoder



#2: Scenario Generation

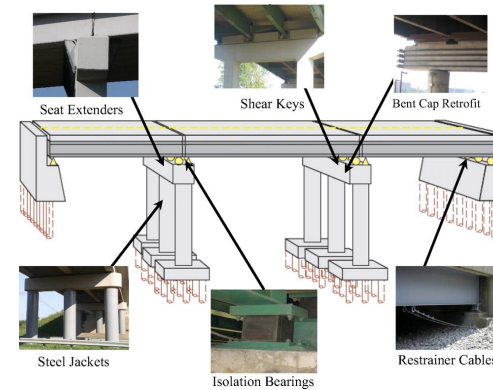
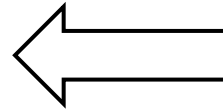
Real-World Data Limitations

- Data gaps
 - Key behavioral and operational data—such as passenger preferences, freight delivery schedules, and mode-shift responses—are often missing or inconsistent
- Limited Longitudinal Data
 - Few panel datasets exist to track changes in traveler/shipper behavior over time
- Sparse Coverage
 - Data is often geographically biased or temporally limited
- Infrequent Updates:
 - Rapidly evolving mobility technologies and market trends outpace traditional data collection cycles
- Privacy and Access Constraints
 - B2B and personal travel data are often proprietary or protected



Pre-disaster Mitigation

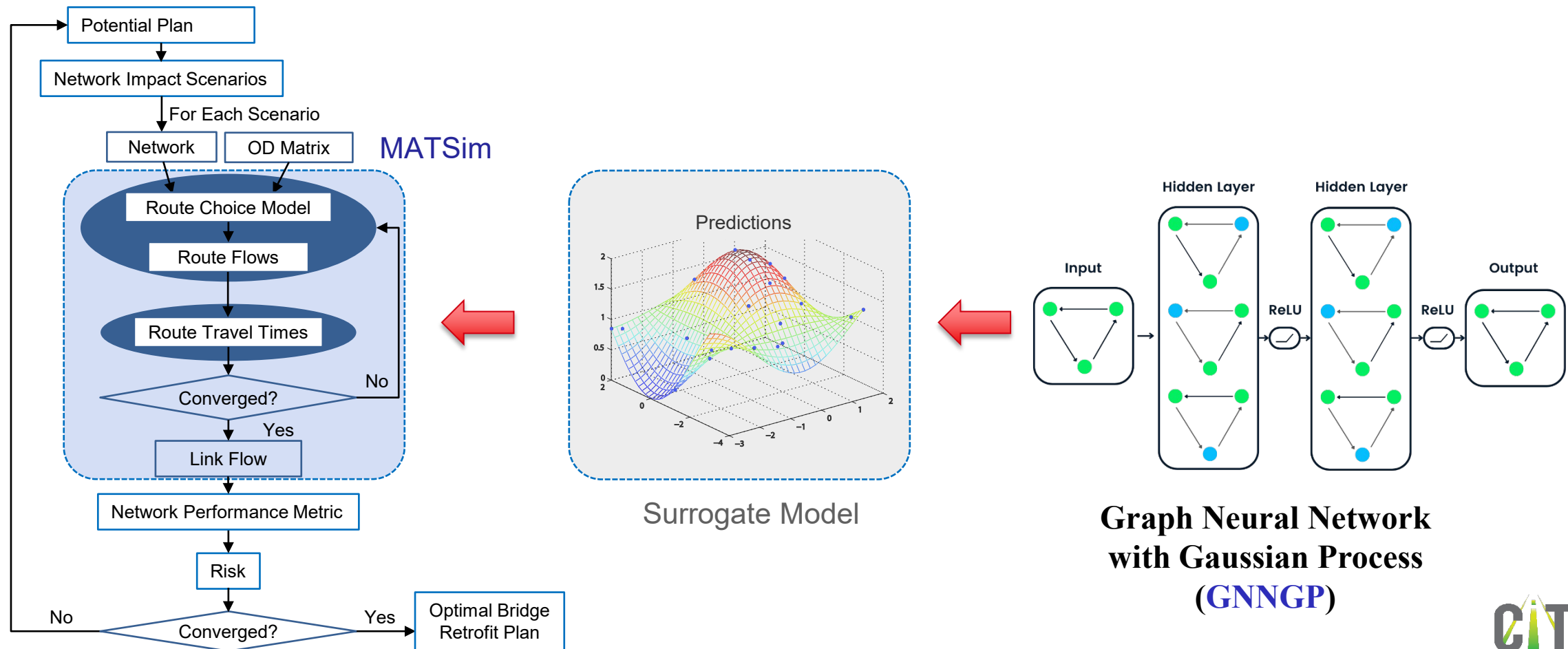
- The retrofitting of critical components can enhance network resilience and significantly reduce the impact of disastrous events



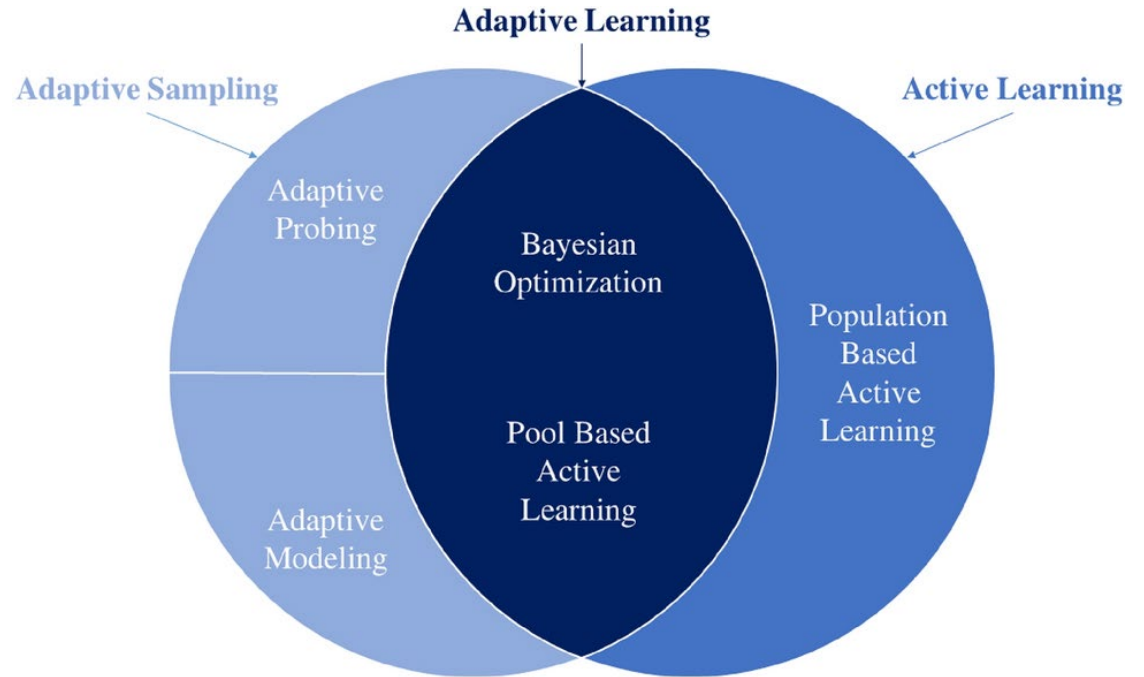
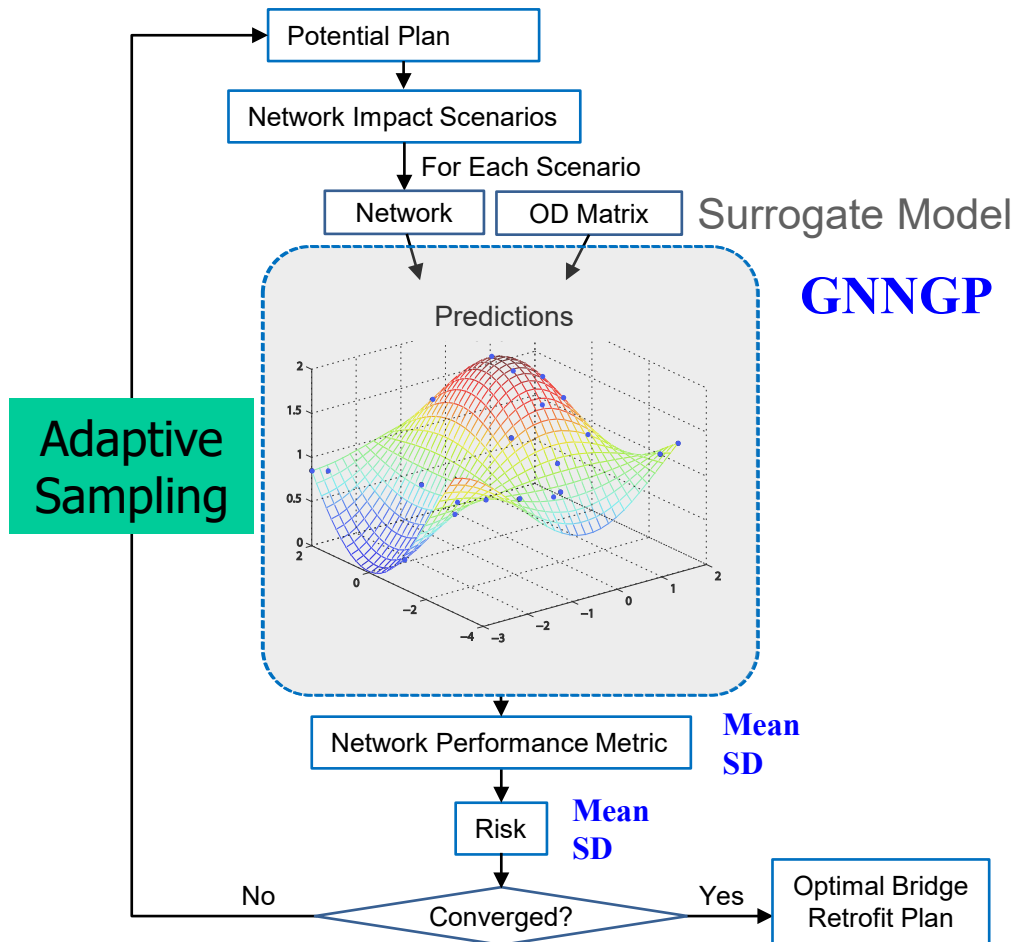
- Challenges in:
 - Modeling disruptive events realistically
 - Capturing dynamic system responses
 - Integrating stakeholder priorities and constraints
 - Cross-sector and network interdependencies

Machine Learning Techniques for Scenario Generation

- An approximation of the input-output relationship from simulation
- Apply the surrogate model to support scenario generation

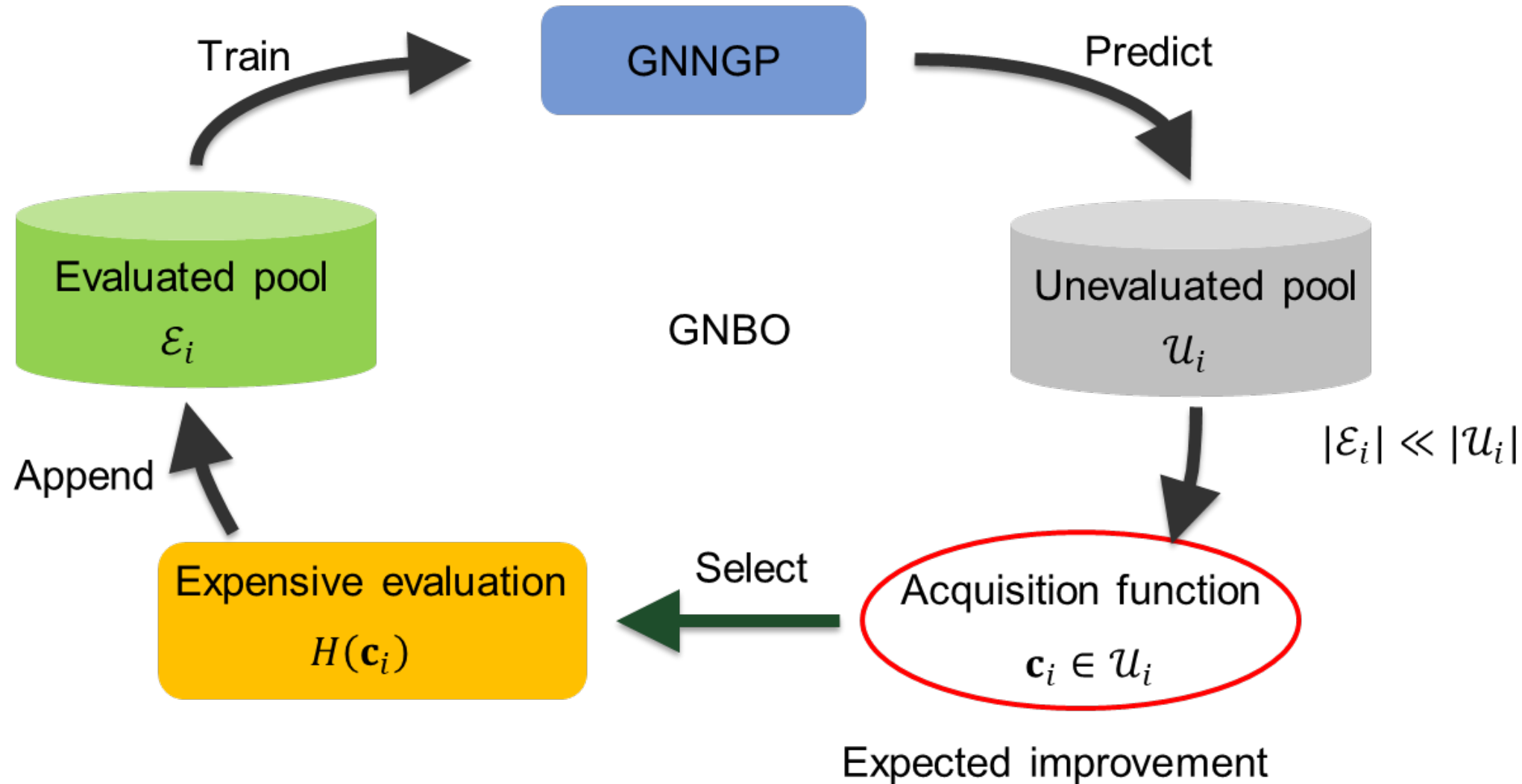


Adaptive Sampling Techniques

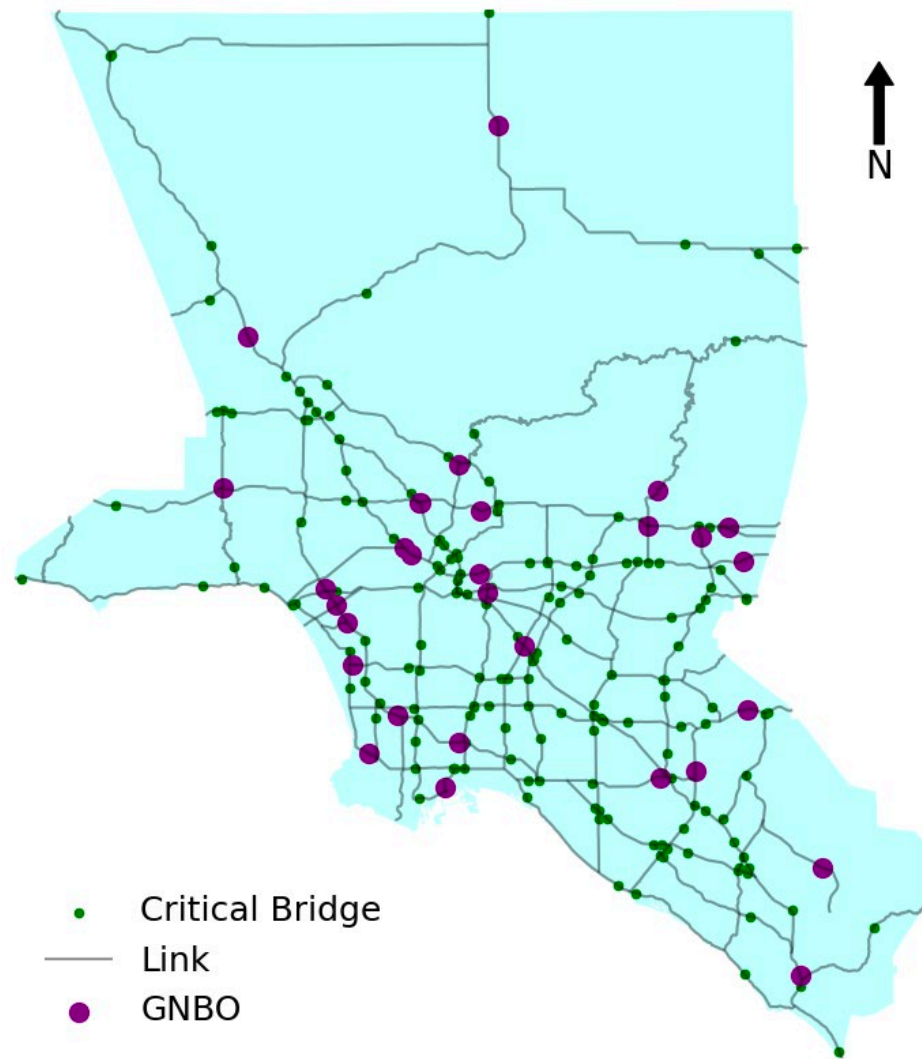
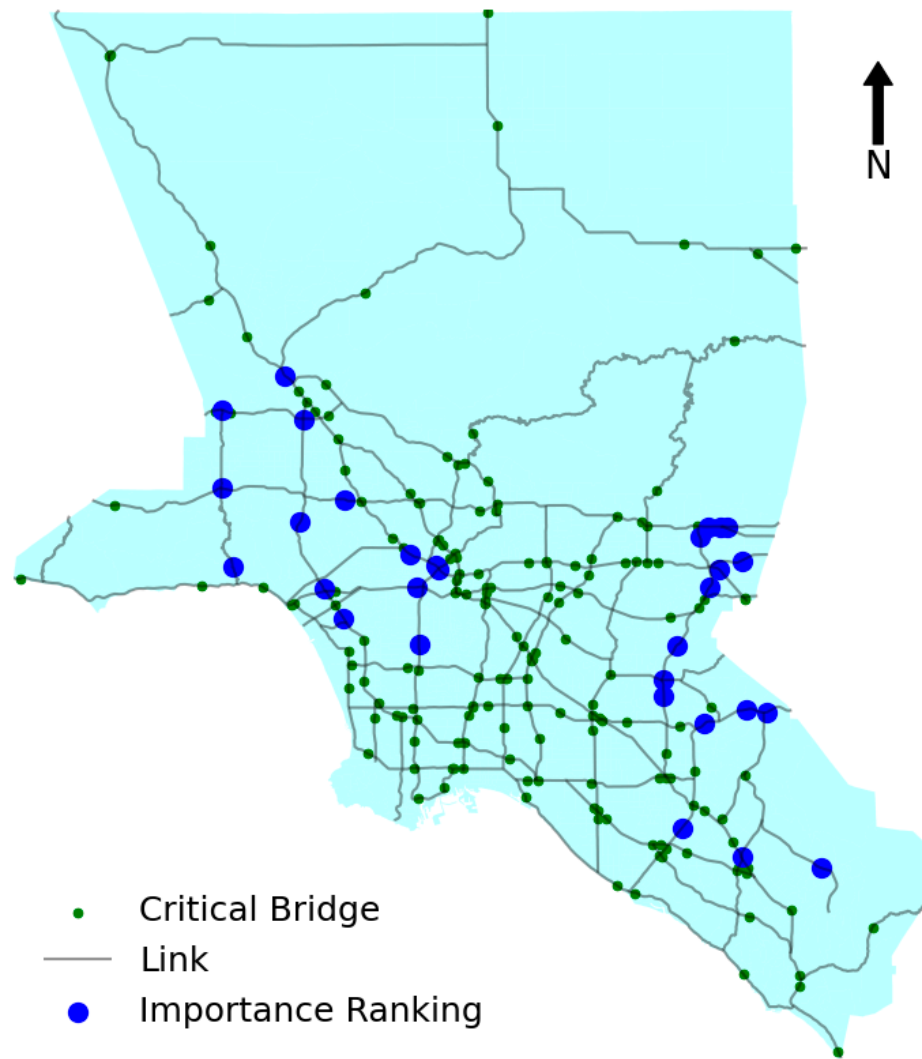


- Bayesian sampling techniques
 - Upper confidence bound
 - Entropy reduction / Mutual Information
 - Markov Chain Monte Carlo (MCMC)
 - Co-Kriging or hierarchical GPs

Graph Neural Bayesian Optimization (GNBO)



Example: Optimal Bridge Retrofit Plan



#3: Simulation Acceleration

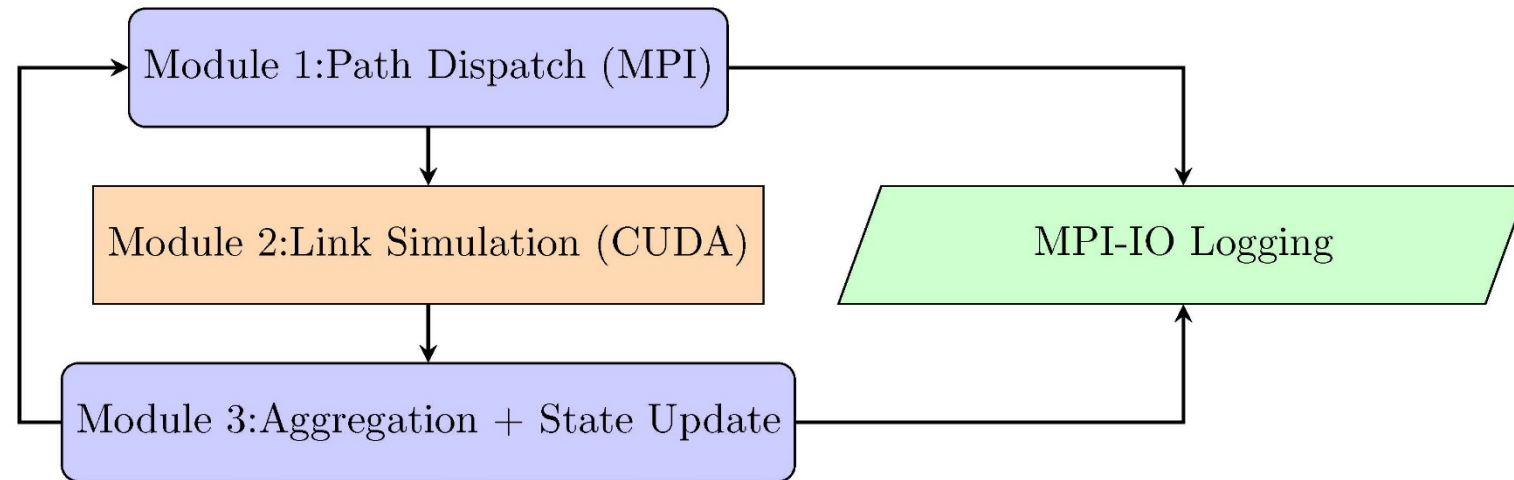
Simulation Bottlenecks

- Time-consuming process to develop the simulation model
 - LLM-based solutions (ChatSUMO)
 - Model calibration and validation
- Computational cost of large-scale simulations
 - **Parallelization**
 - Surrogate models
 - Adaptive sampling or use of reinforcement learning
- Tradeoff between fidelity and efficiency
 - Choice between macro-, meso-, and micro-models



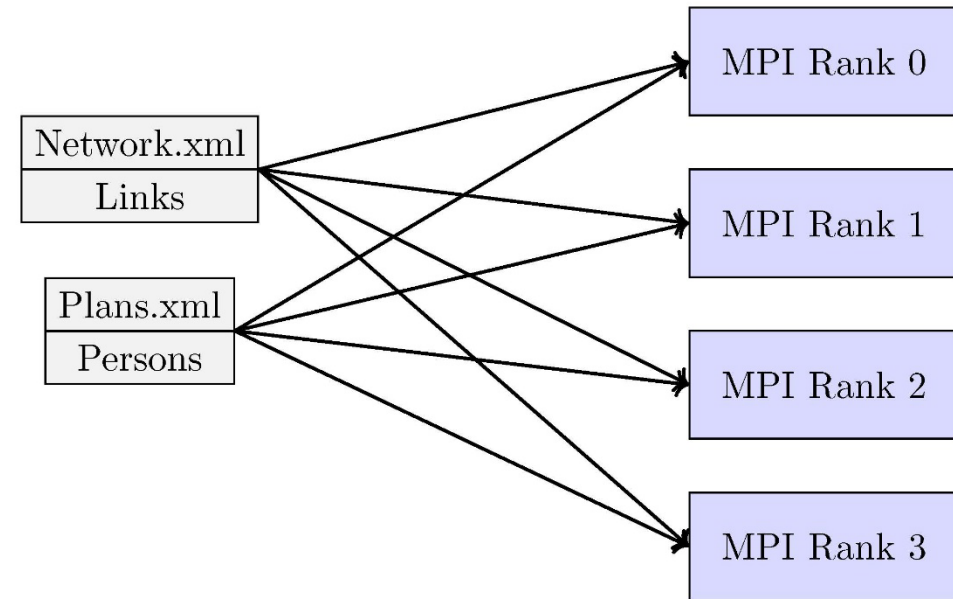
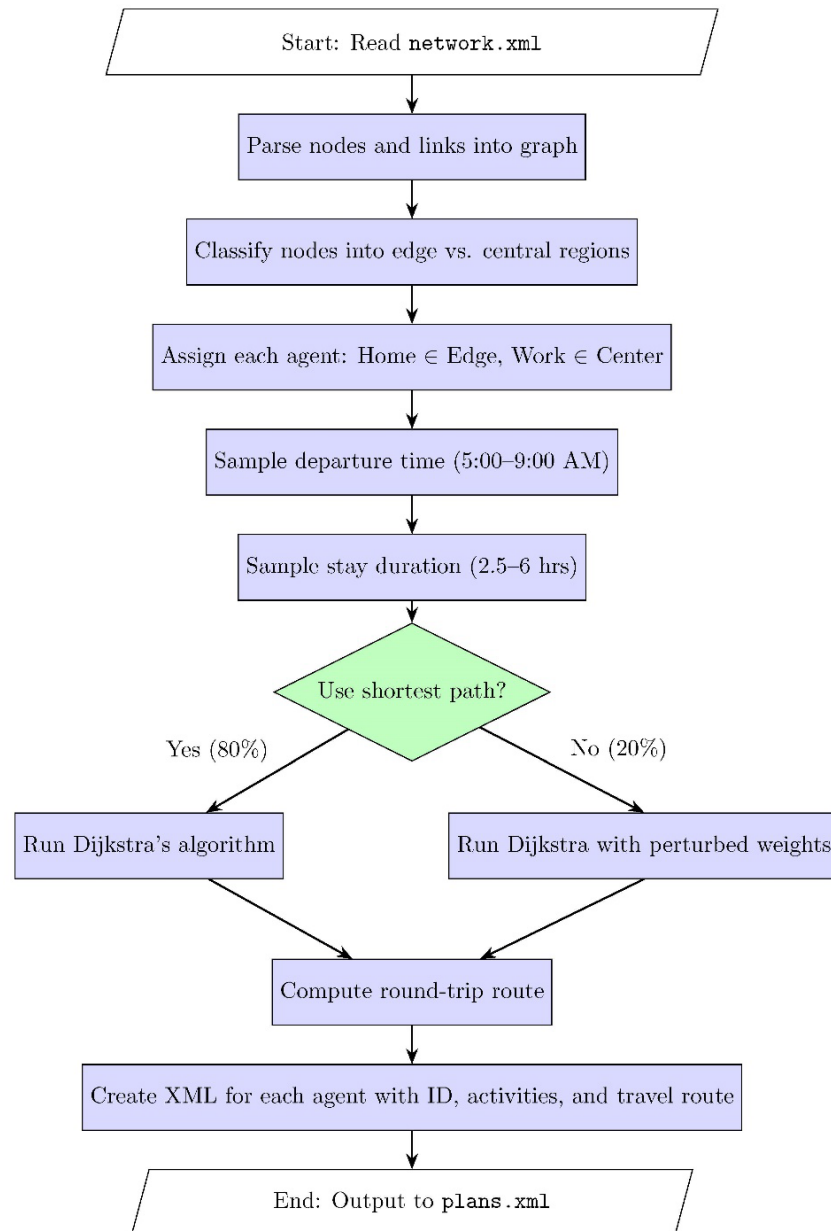
GPU-Based Parallelism

- Combining Message Passing Interface (MPI) with CUDA for intra-node parallelization strategy
 - Decompose the problem at the link level to capture traffic queuing behavior, with each link assigned to a GPU node
 - Extending GPU-only approaches with MPI-based distributed-memory techniques to span multiple compute nodes

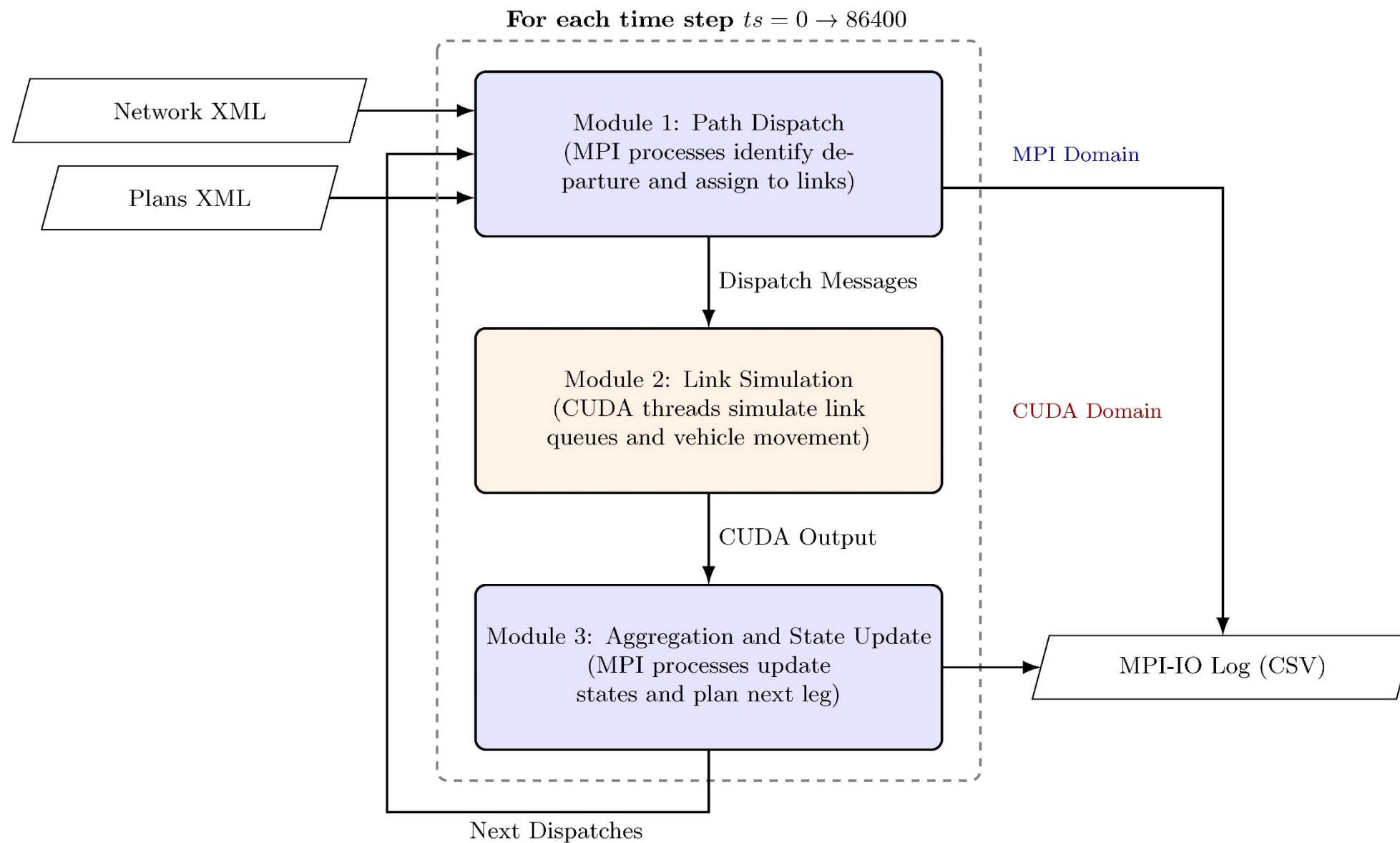


Overall Simulation Architecture

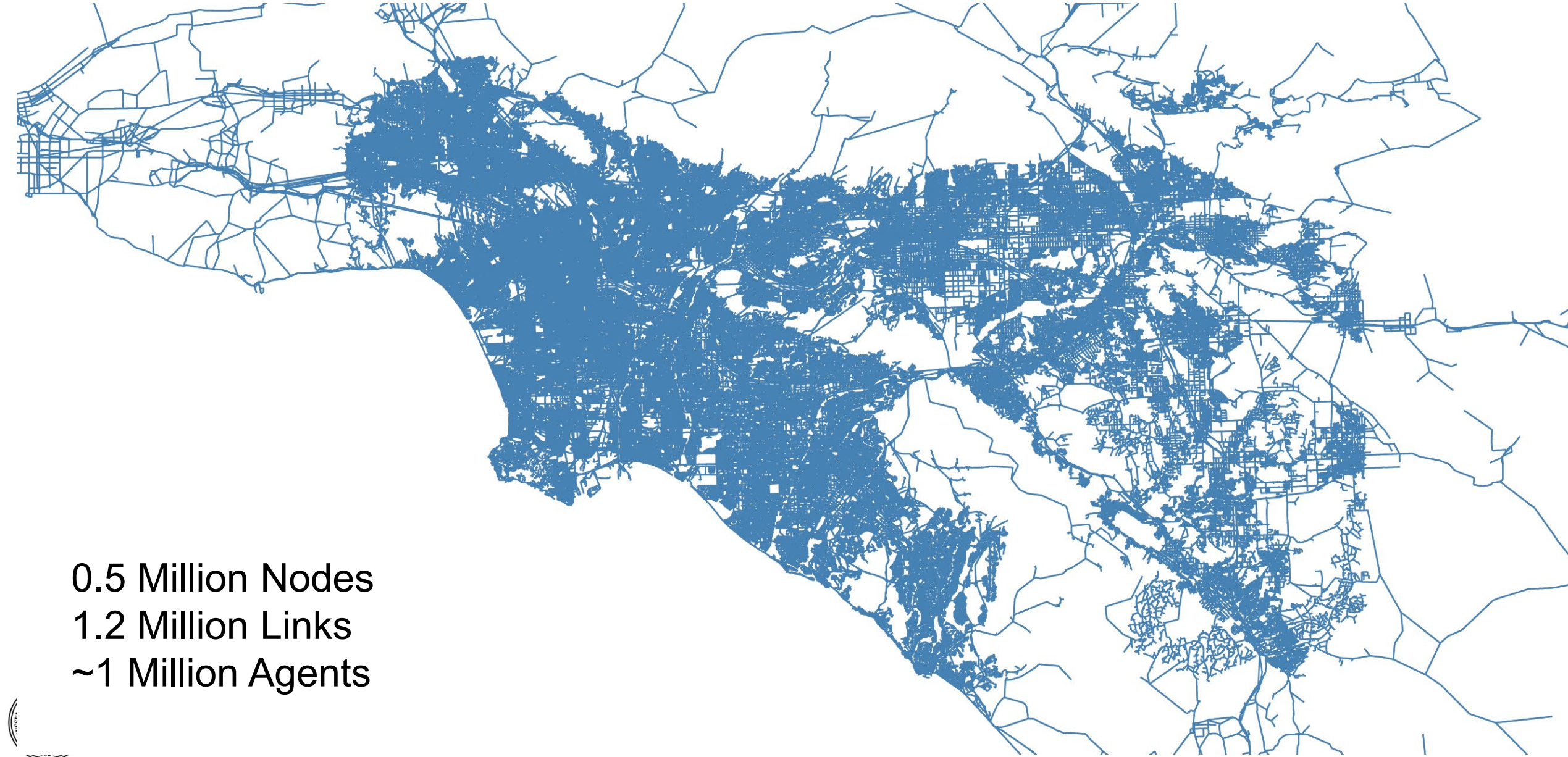
Parallel Agent Routing and Data Partitioning



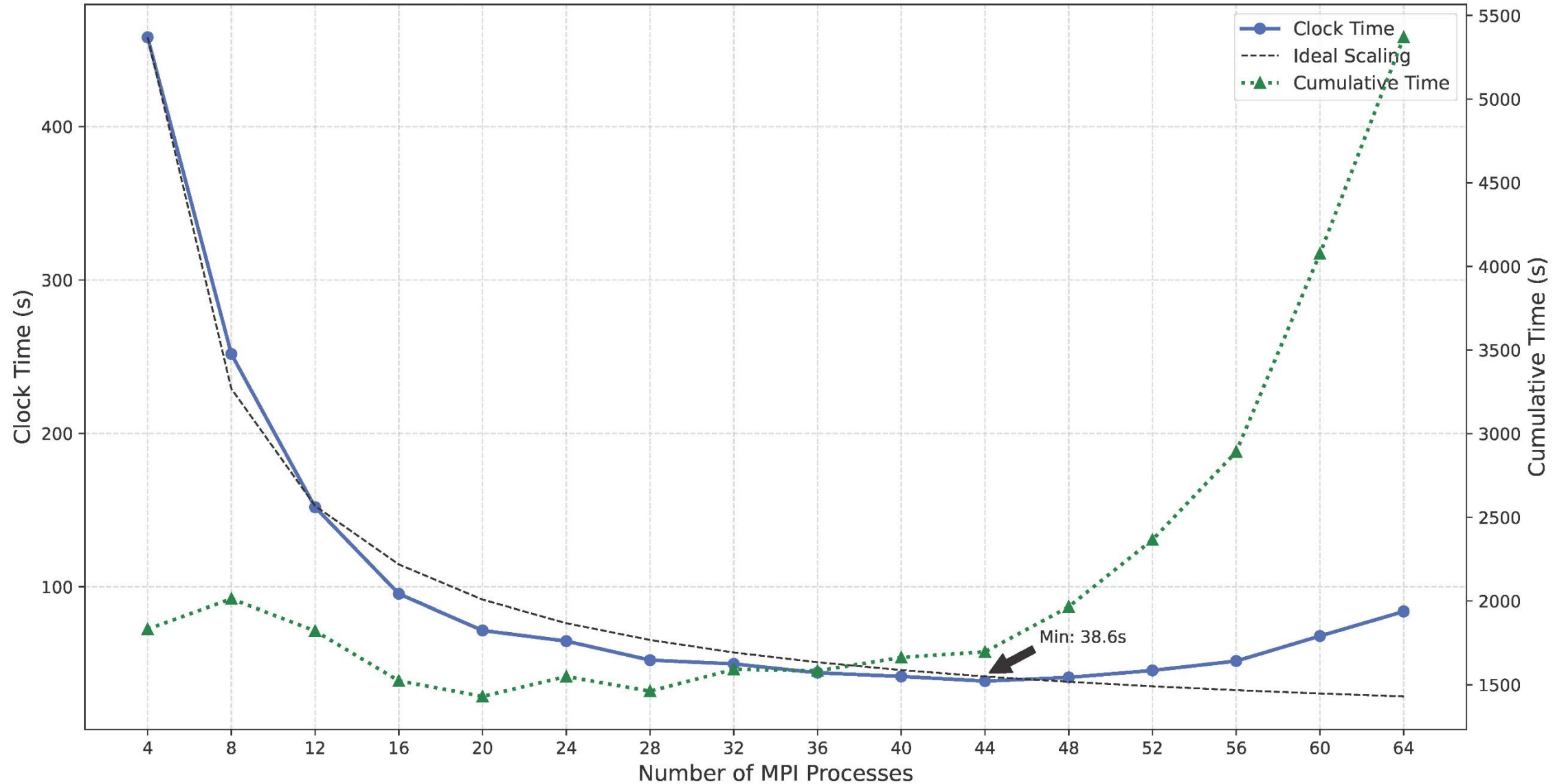
Hybrid Simulation Framework



Los Angeles Network

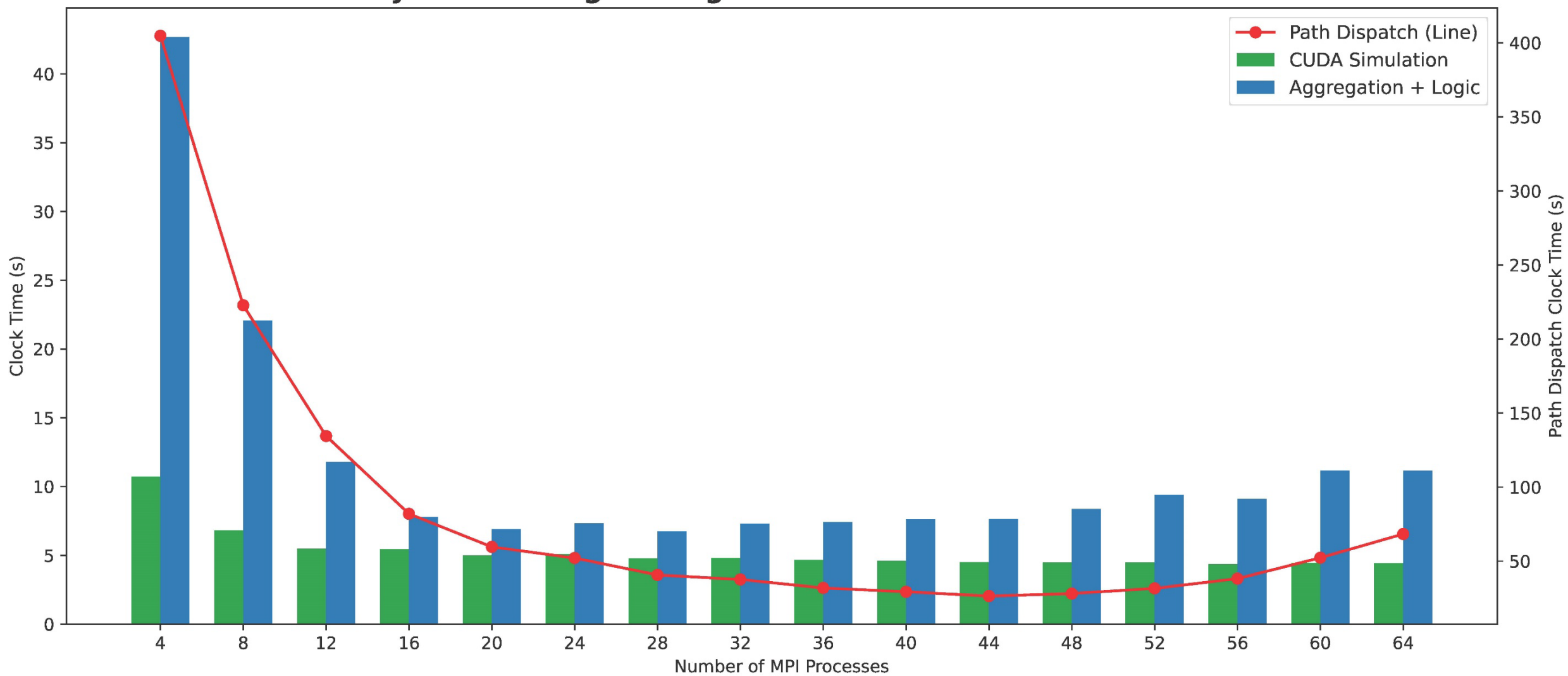


Strong Scaling Analysis with 1000k Simulation Modules: Clock vs Cumulative Time

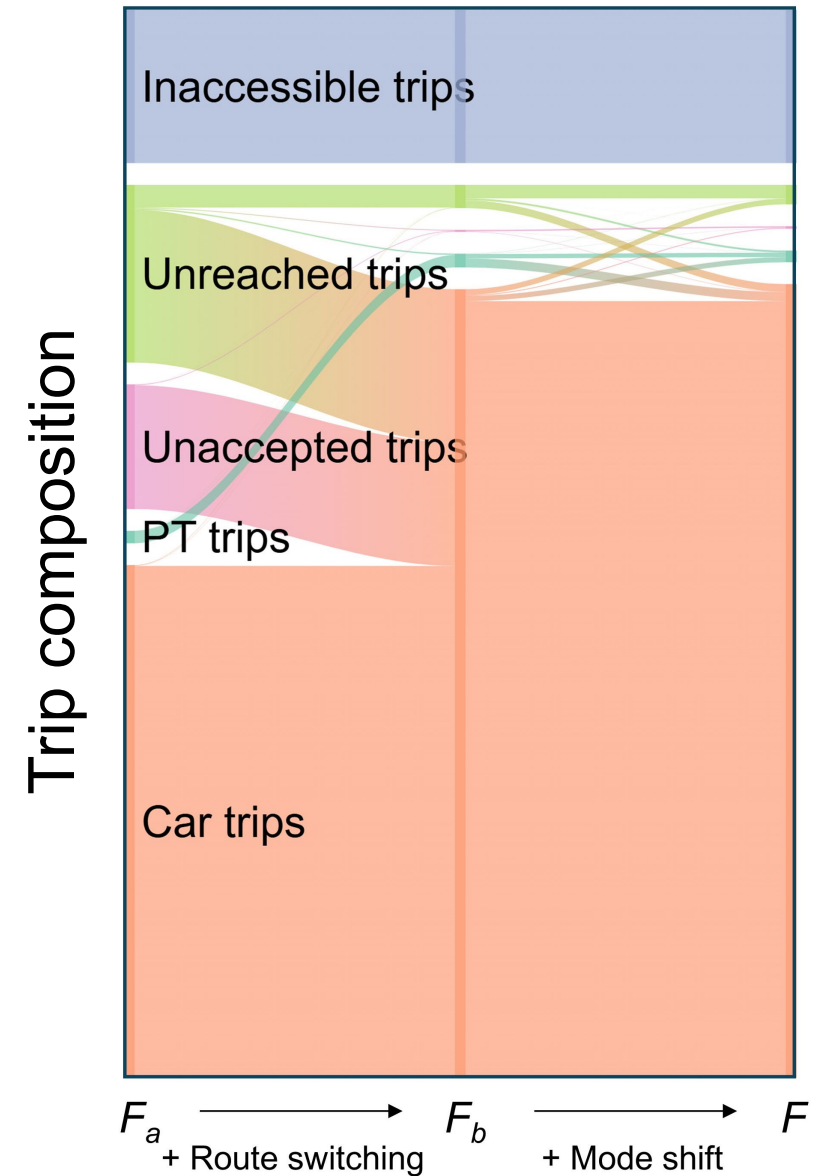
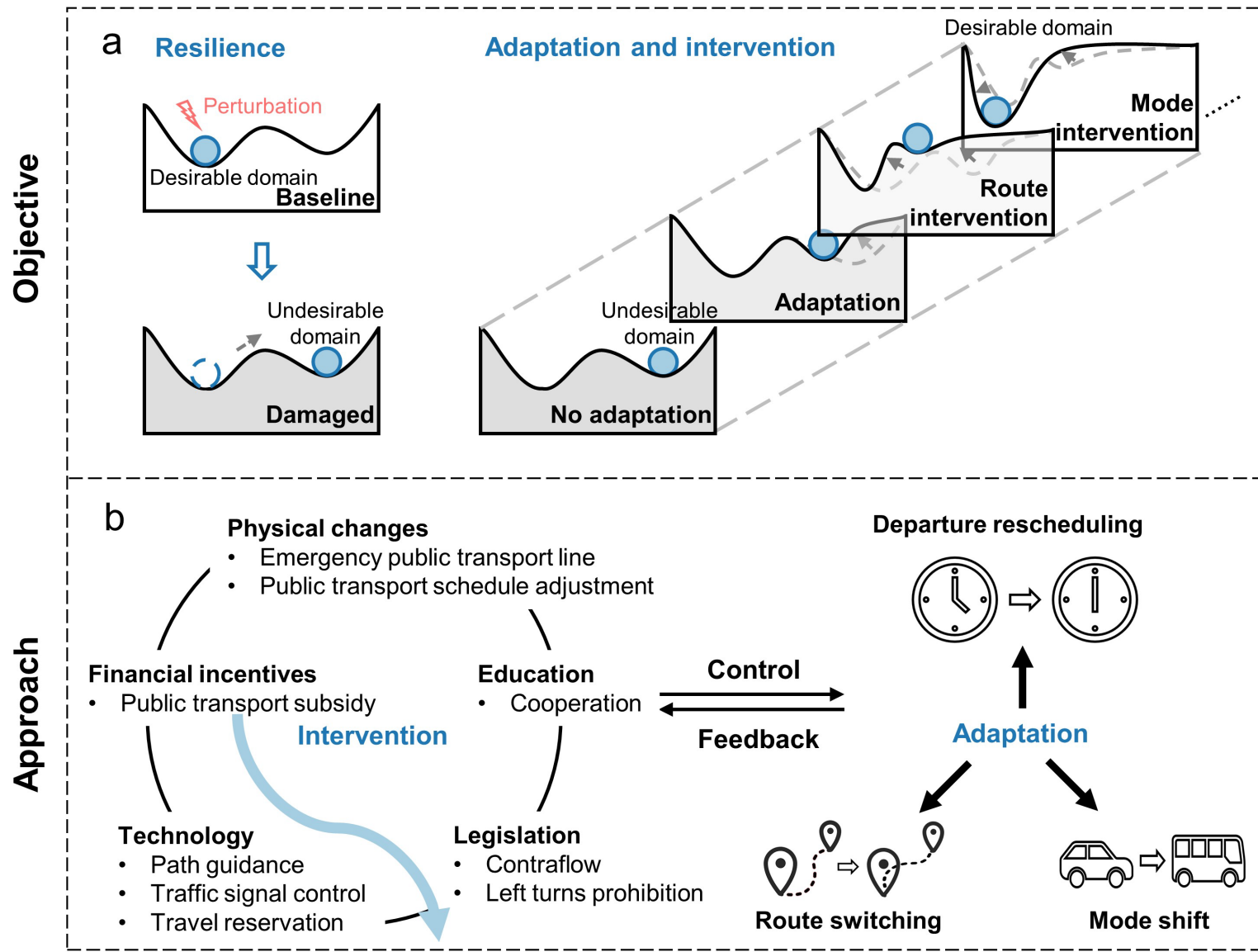


Performance (2)

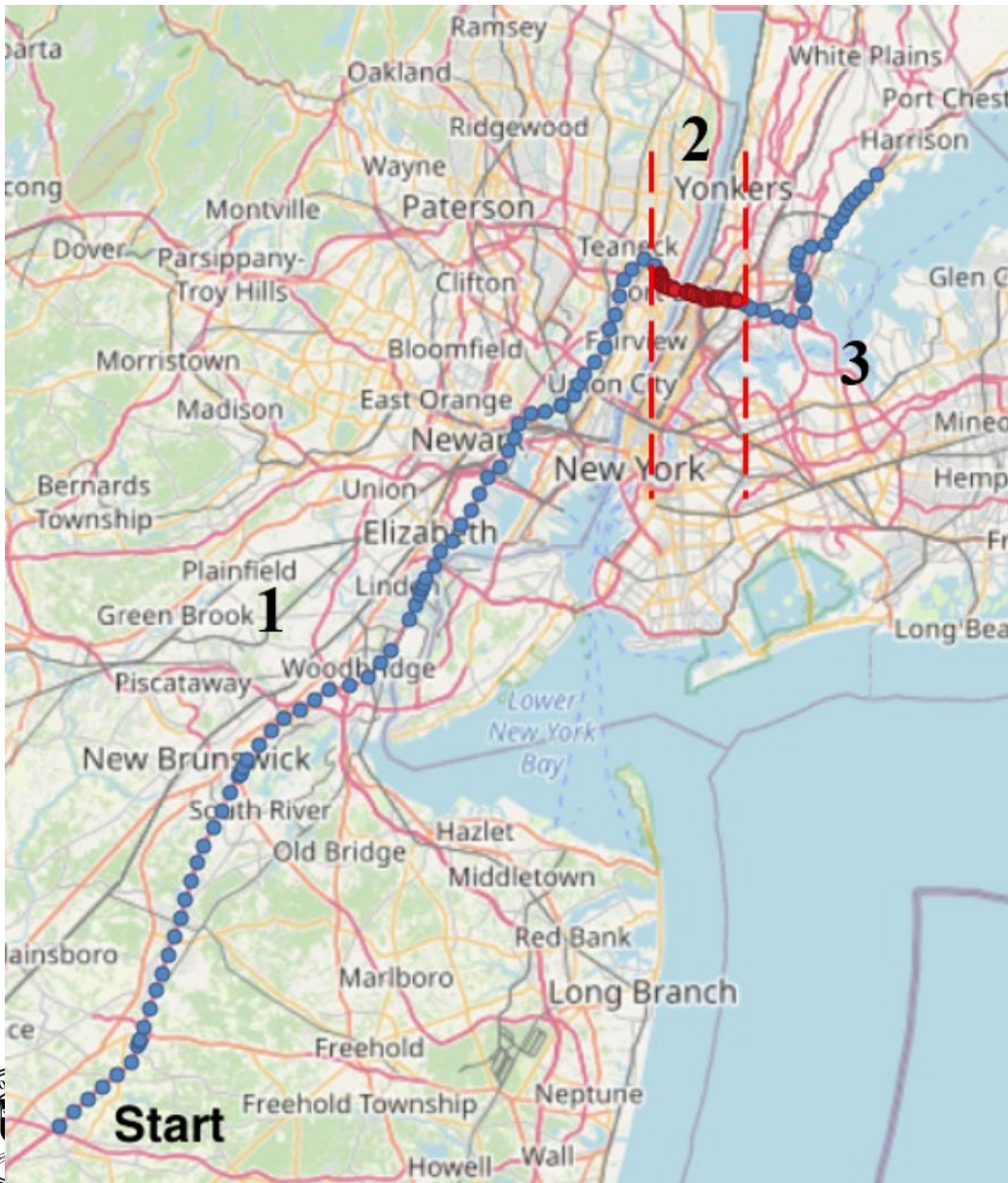
Breakdown Analysis of Strong Scaling Performance with 1000k Simulation Modules



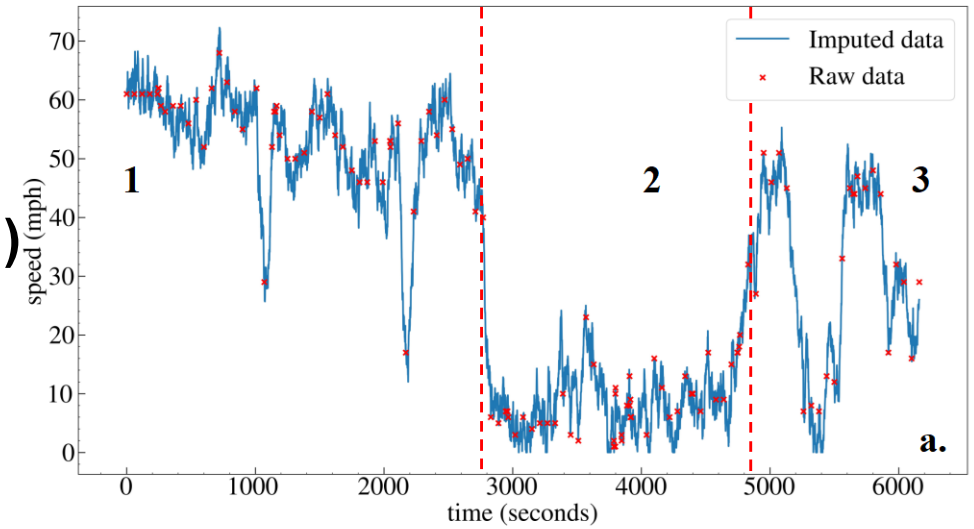
Application—Multimodal Network Response



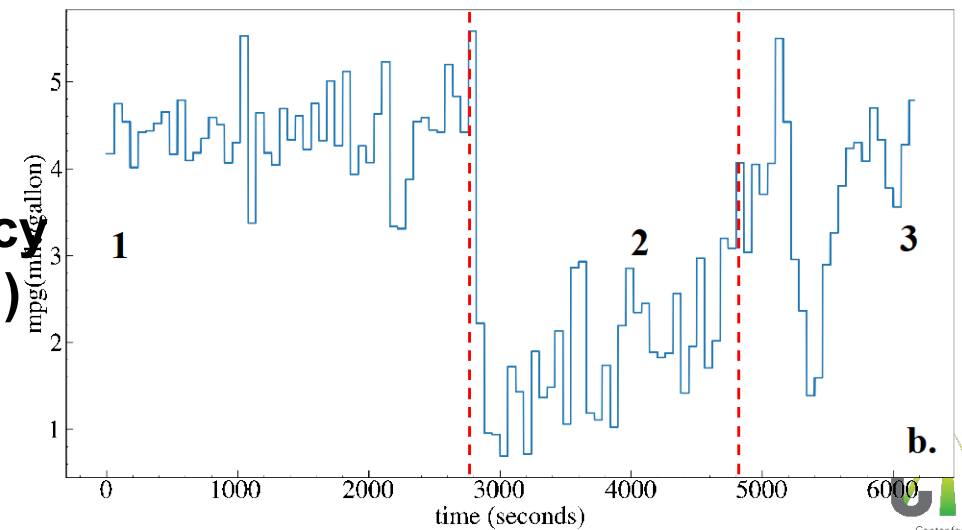
Application—Energy-Efficient Fleet Operation



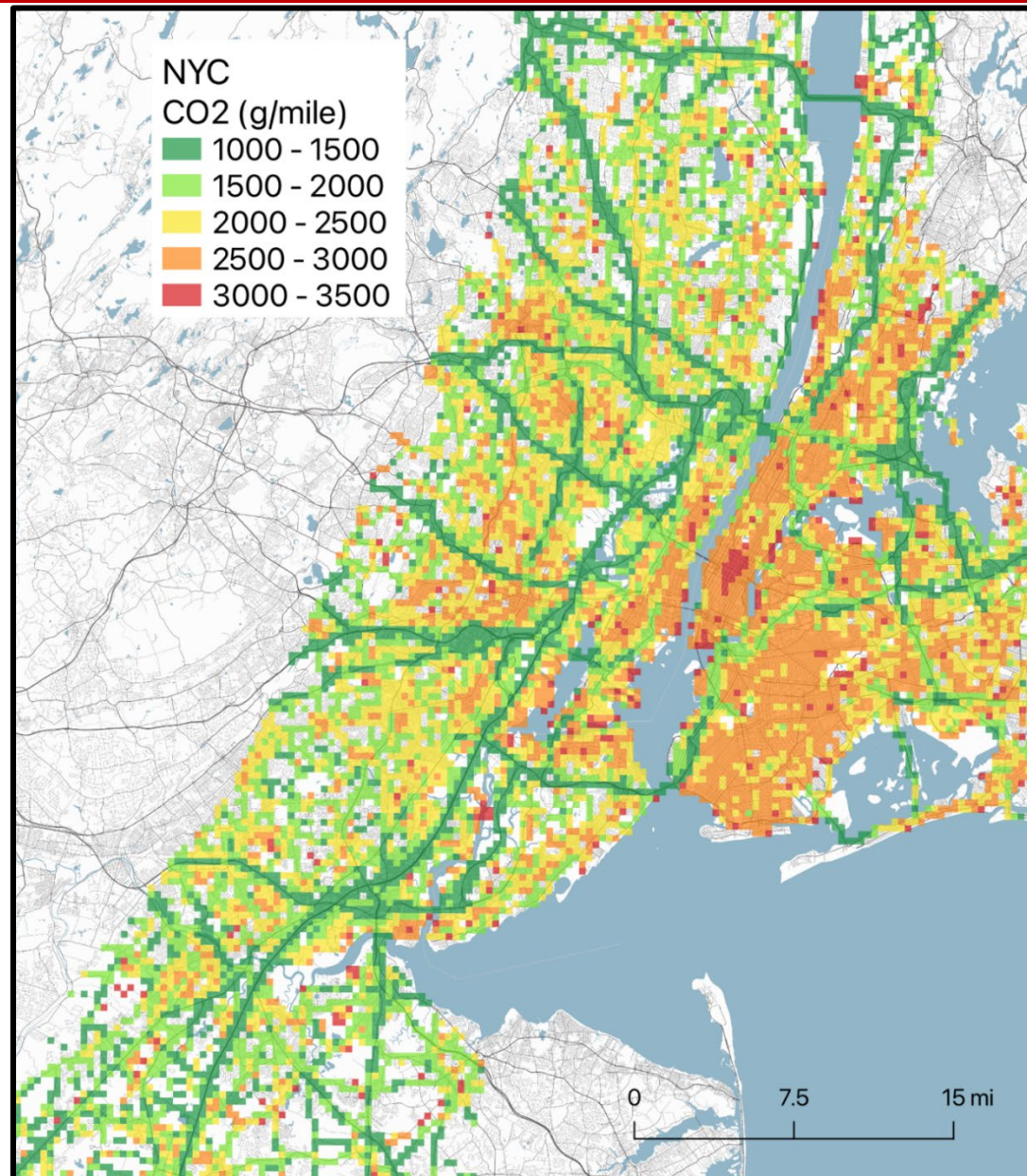
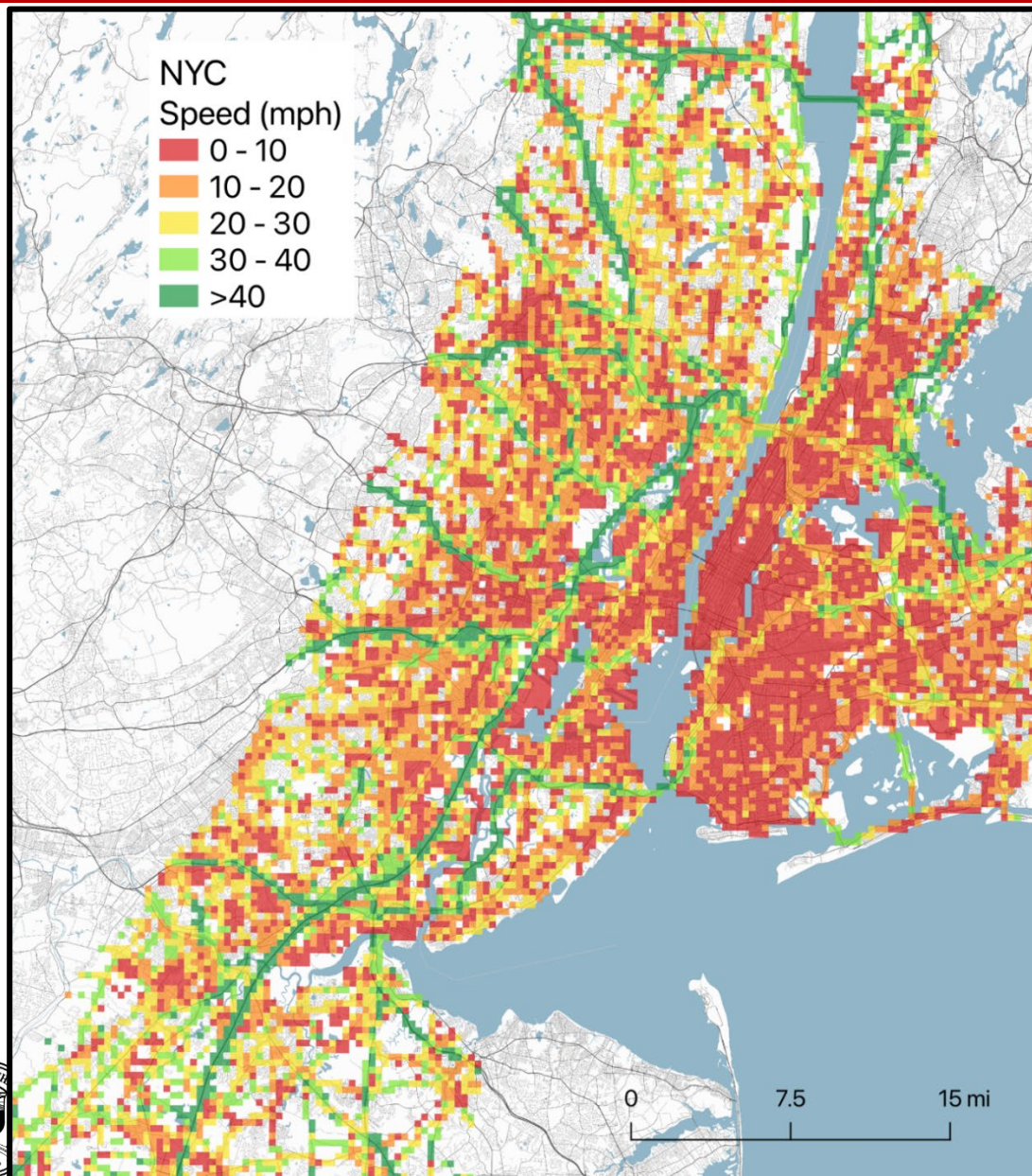
Speed (mph)



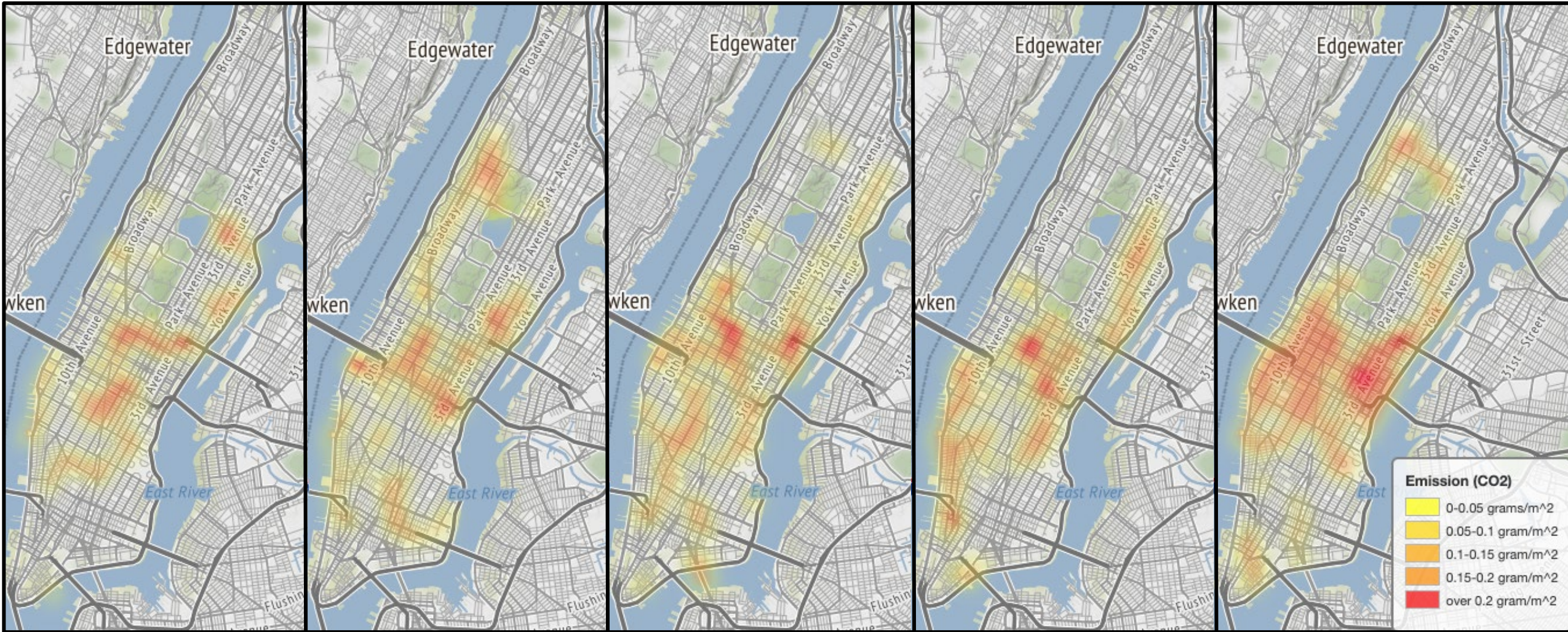
Fuel efficiency (miles/gallon)



Application—Generating Baseline Conditions—NYC



Application—Emission Dynamics



July 16 07:12-07:17 AM

07:27-07:32 AM

07:42-07:47 AM

07:57-08:02 AM

08:12-08:17 AM



Conclusions

Summary & Key Takeaways

- **Multiclass Activities**
 - Enables realistic modeling of diverse travel behaviors and vehicle classes
 - Enhances simulation fidelity across passenger, freight, and emerging modes
- **Scenario Generation**
 - Fills critical data gaps with synthetic, behaviorally plausible data
 - Supports robust testing of policies under uncertainty and disruption
- **Simulation Acceleration**
 - Leverages LLMs to automate and streamline simulation model development
 - Utilizes GPU-based parallelism to significantly speed up high-fidelity simulations
- **Real-world data collection remains foundational**
 - High-quality, representative real-world data is essential for calibrating and validating AI-enhanced simulation models



Questions?

Sean He – hex6@rpi.edu