2021 DOE Vehicle Technologies Office
Annual Merit Review

High-Performance Computing (HPC) and Big Data Solutions for Mobility (BDSM) Design and Planning

Jane Macfarlane
Lawrence Berkeley National Laboratory
June 22, 2021

Project ID: eems037

BERKELEY LAB
Argonne NATIONAL LABORATORY
NREL NATIONAL RENEWABLE ENERGY LABORATORY

Overview

TIMELINE

- Start: January 2021
- End: January 2023
- 10 % complete

BUDGET

- Total project funding
- $4 M / 2 years
- 3 Labs (LBL, ANL, NREL)

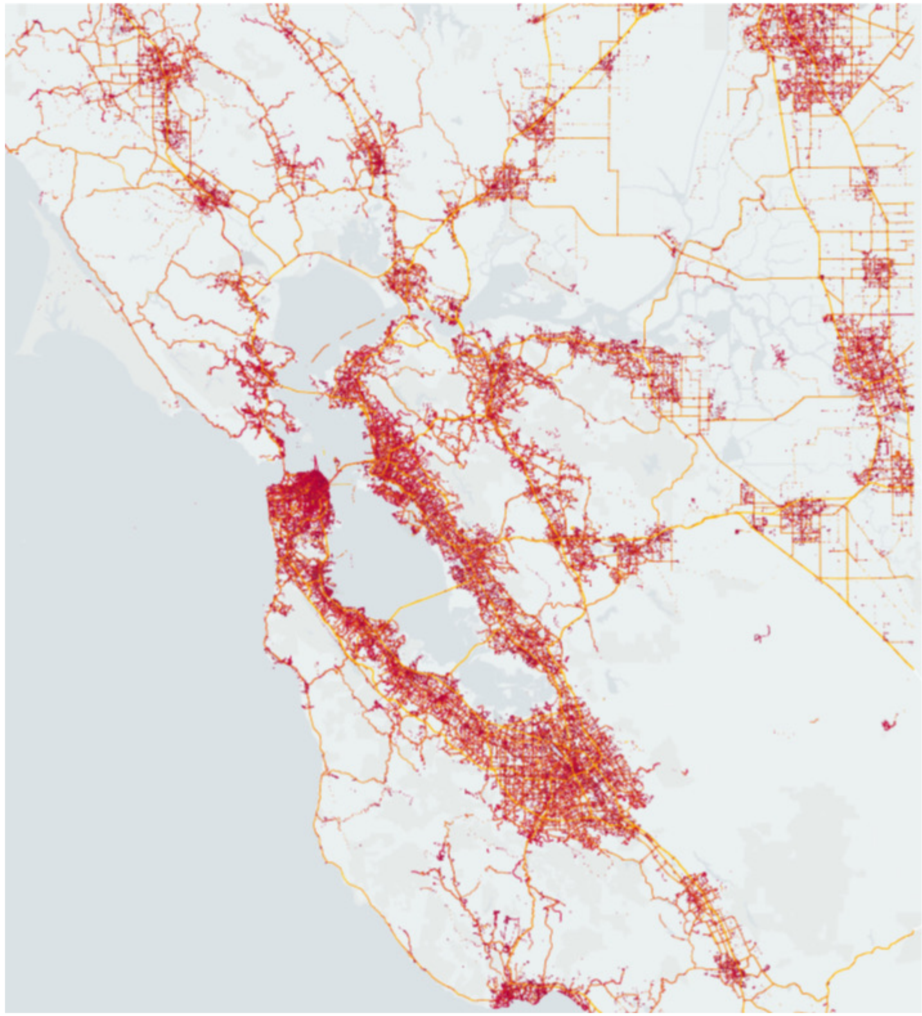
PARTNERS

- City of San Jose
- HERE Technologies
- Southern California Area Governments

BARRIERS

- Metropolitan scale networks are too complex to model in reasonable compute time.
- Sensors for capturing traffic dynamics provide a limited view and are difficult to mine for relevant information.
- Optimization of energy, travel time and mobility across complex networks has yet to be accomplished for real-world metropolitan scale networks.

Relevance and Project Objectives



Overall Objective:

- Develop tools to **rapidly model large-scale transportation networks** using real-world, near real-time data. Integrate energy, travel time and mobility measures to determine optimization opportunities. Develop new **active control ideas for connected vehicles** that will optimize energy, travel time and mobility for normal traffic conditions and networks under stress.

Objectives this Period:

- Integrate Diffusion Convolutional Recurrent Neural Network (DCRNN) models into Mobiliti Simulation Platform
- Develop simulation capability to allow modeling of infrastructure - signals and control.
- Improve energy modeling to capture link dynamics

Impact:

- Understand the impact of control variations on travel times, delay, and a variety of socially-aware city metrics.

Project Goals

- Extend and improve an ***integrated simulation and data analytics platform*** – Mobiliti - that can be used for solving transportation problems at scale.
- Use ***AI to build models that can be transferred*** to organizations that do not have access to HPC resources.
- Allow practitioners to ***run high-value scenarios of complex transportation solutions and develop next-generation planning and control solutions*** for transportation impacts in cities.

2021 Milestones

Milestone	Criteria	Q1	Q2	Q3	Q4
Data Agreement with new collaborator NREL		x			
GPS waypoint integration to provide integrated travel demand models	Validated with modelers		x		
Deep Learning uncertainty assessment in CA region Application of Uncertainty Quantification methods on similar region from BDSM 1.0	Similar prediction accuracy (vs non-UQ method) Robustness of new UQ method demonstrated			x	
Computational models of communication latencies for distributed control modeled in PDES Automated identification of traffic control locations and signal timing.	Initial integration of signal information into Mobiliti				x
ML and simulation methods providing similarly accurate predictions of traffic states. GO/NO GO	The trained CNN Surrogate Model from PDES ensemble demonstrates similar results				x

Approach

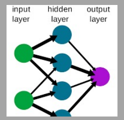
- Use data science and DOE National Laboratory high-performance computing resources and expertise to enhance understanding of urban-scale transportation dynamics.
- Develop strategies for significantly improved transportation system planning and operations.

Taking a data driven approach ***implementing large-scale machine learning/deep learning (ML/DL) methods for high-value applications as well as integrating real-world data into AI-enabled, high-speed, urban-scale transportation network parallel-discrete event simulations (PDES)***.

Focus on optimizing mobility, energy, productivity, regional economics, and quality of life in our cities by ***increasing mobility system efficiency, reducing cost, reducing fossil fuel use and increasing the effectiveness of transportation***.

Approach

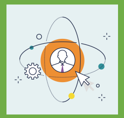
Task 1 Addresses a key task for all AI projects – data ingestion.



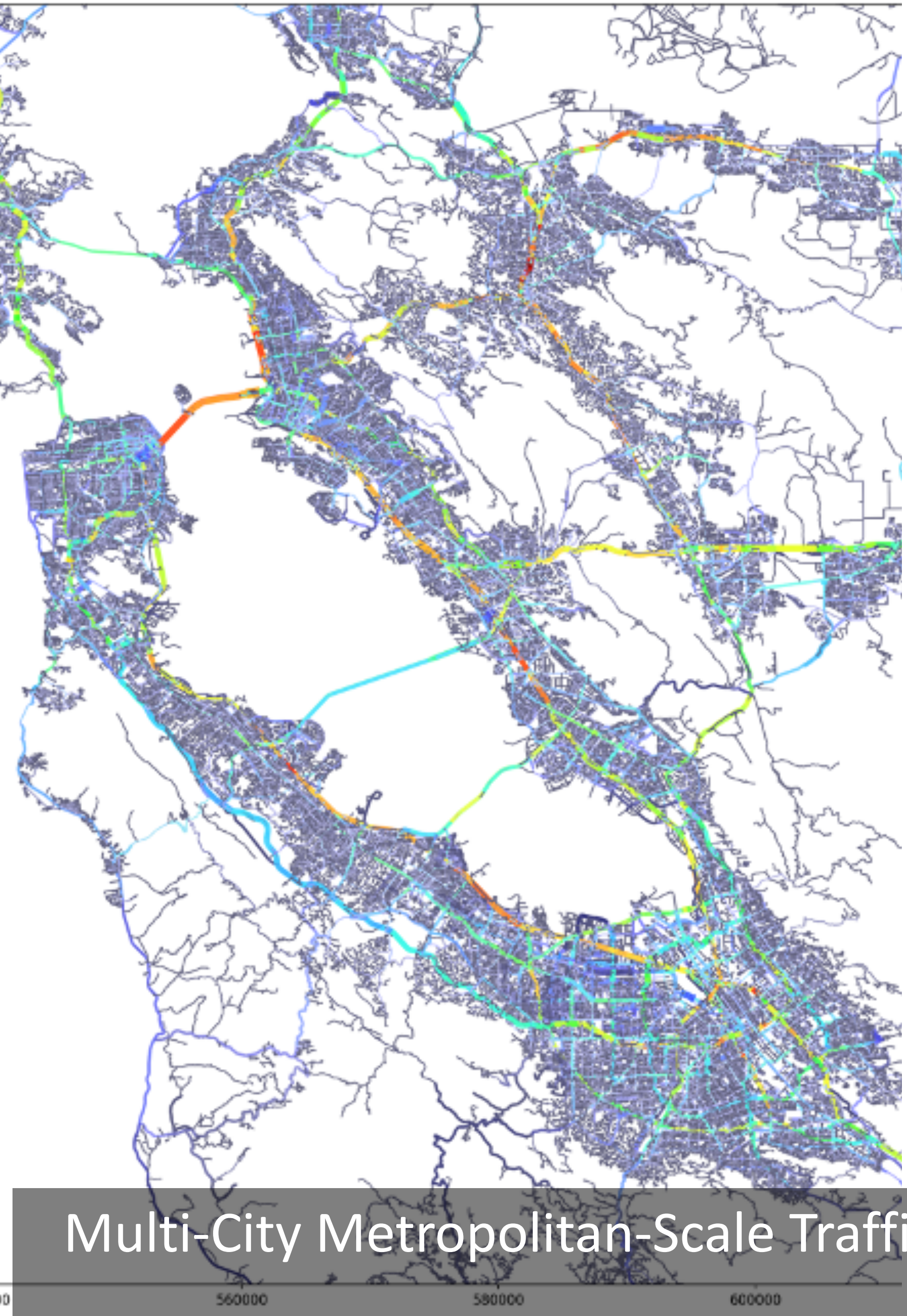
Task 2 Addresses improving our ML/DL models for system state prediction and characterization. (ANL, LBNL)

Task 3 Addresses the challenge of transferring the value from these powerful data driven analytics tools to the practitioner in understandable insights that can be applied to real-world transportation challenges. (LBNL, Siemens, NREL)

Task 4 Addresses the operational analytics for improving controls, including vehicle routing and distributed control using edge computing. (LBNL, Siemens, NREL)



Task 5 Addresses the stakeholders for whom the analytic insights will generate actionable improvements to their transportation challenges. (LBNL)

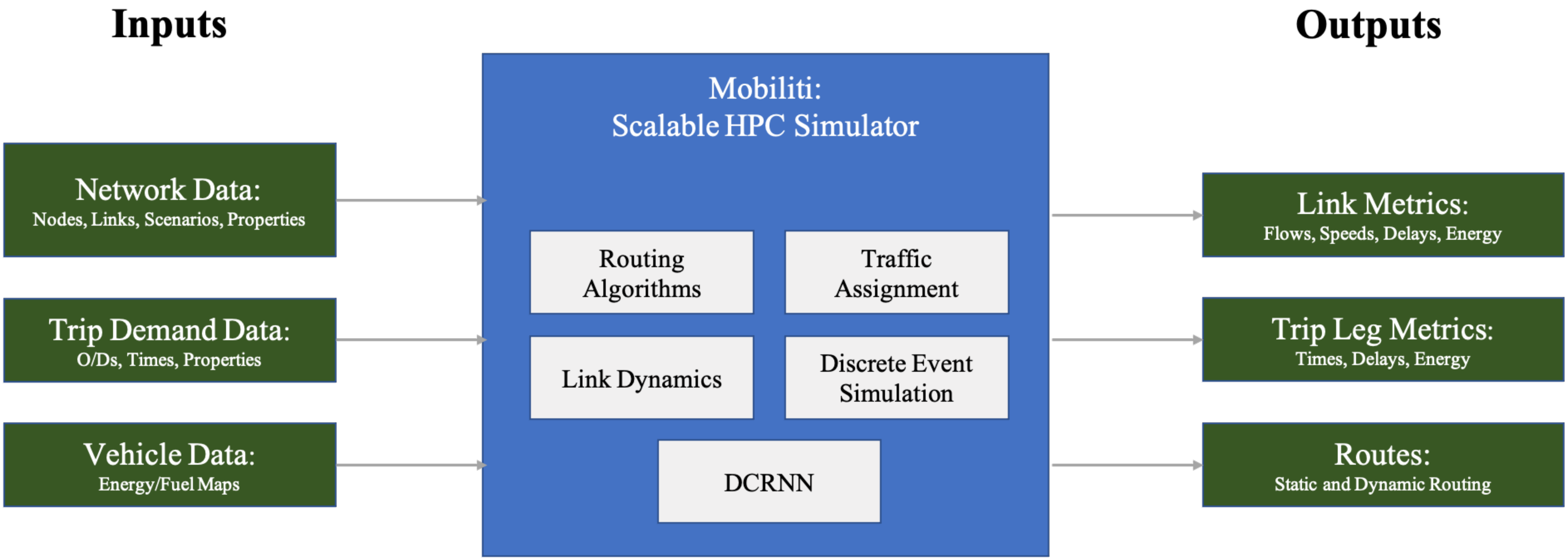


Multi-City Metropolitan-Scale Traffic Simulation

Mobiliti: An Urban-Scale Transportation System Modeling Platform

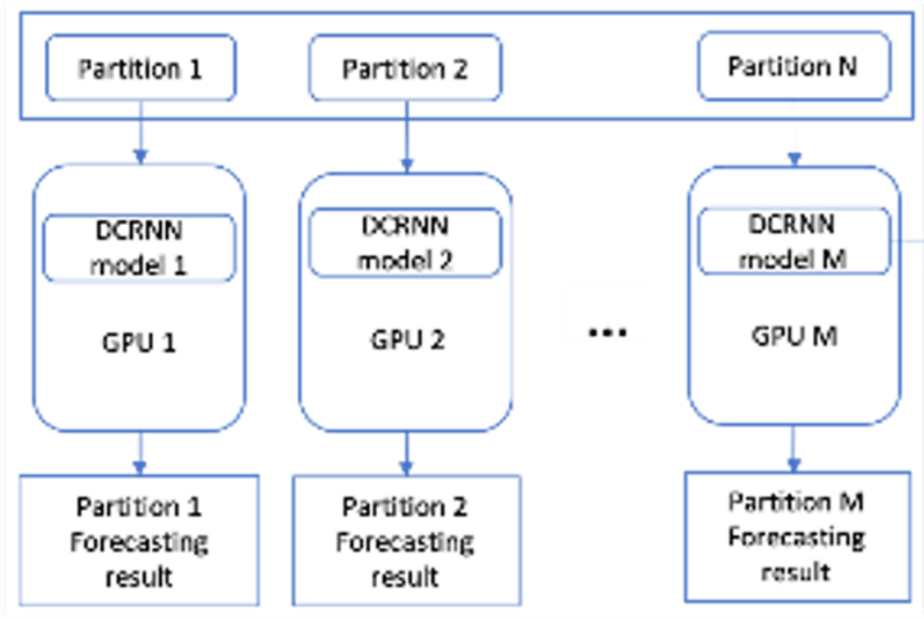
- Demonstrating that high performance computing technologies can enable efficient modeling and analysis for problems in energy and transportation that are too large for existing tools
  - Simulate vehicle congestion of San Francisco Bay Area (e.g. 19M vehicle trips, 1M link network) and LA Basin (e.g. 40Mvehicle trips, 2M link network) in minutes
  - Evaluate dynamic routing of vehicles
  - Integrate traffic assignment optimizations
  - Provide hybrid simulations for a variety of fleet profiles
    - Dynamic routing
    - Optimized routes
    - Fleet – vehicle types

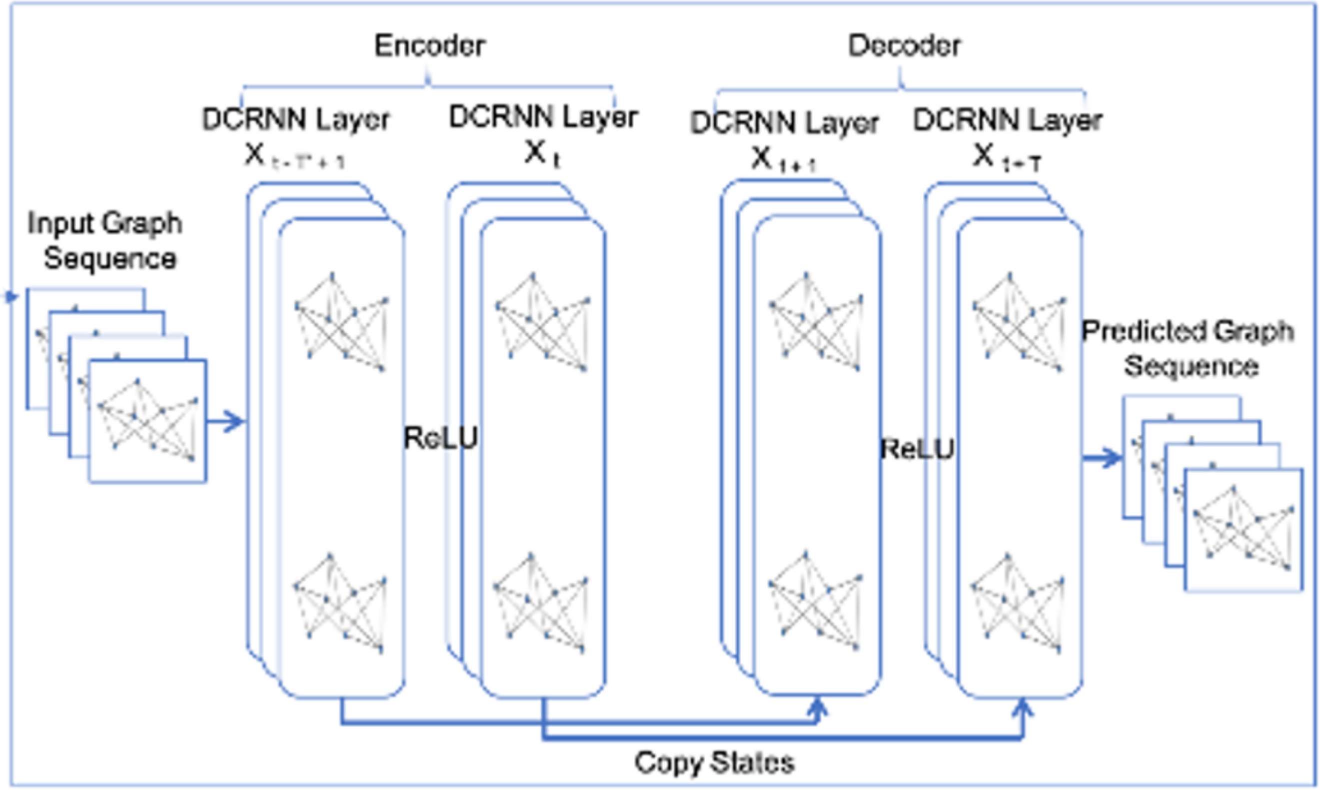
Mobiliti Platform



Leverage DCRNN

- Forecasting traffic dynamics
- Surrogate modeling using Transfer Learning

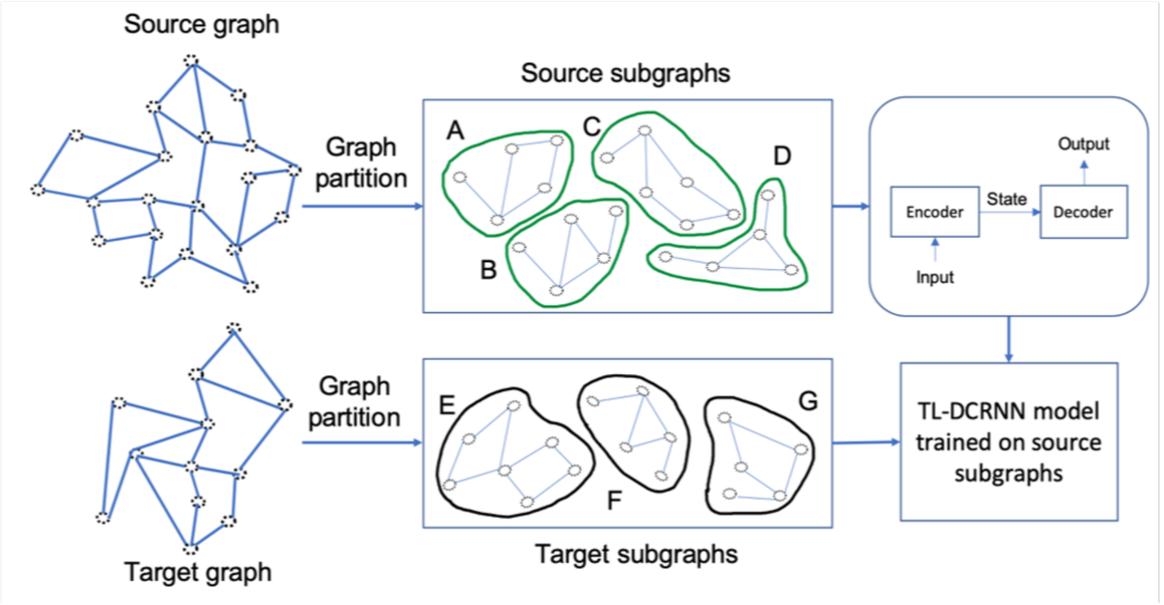


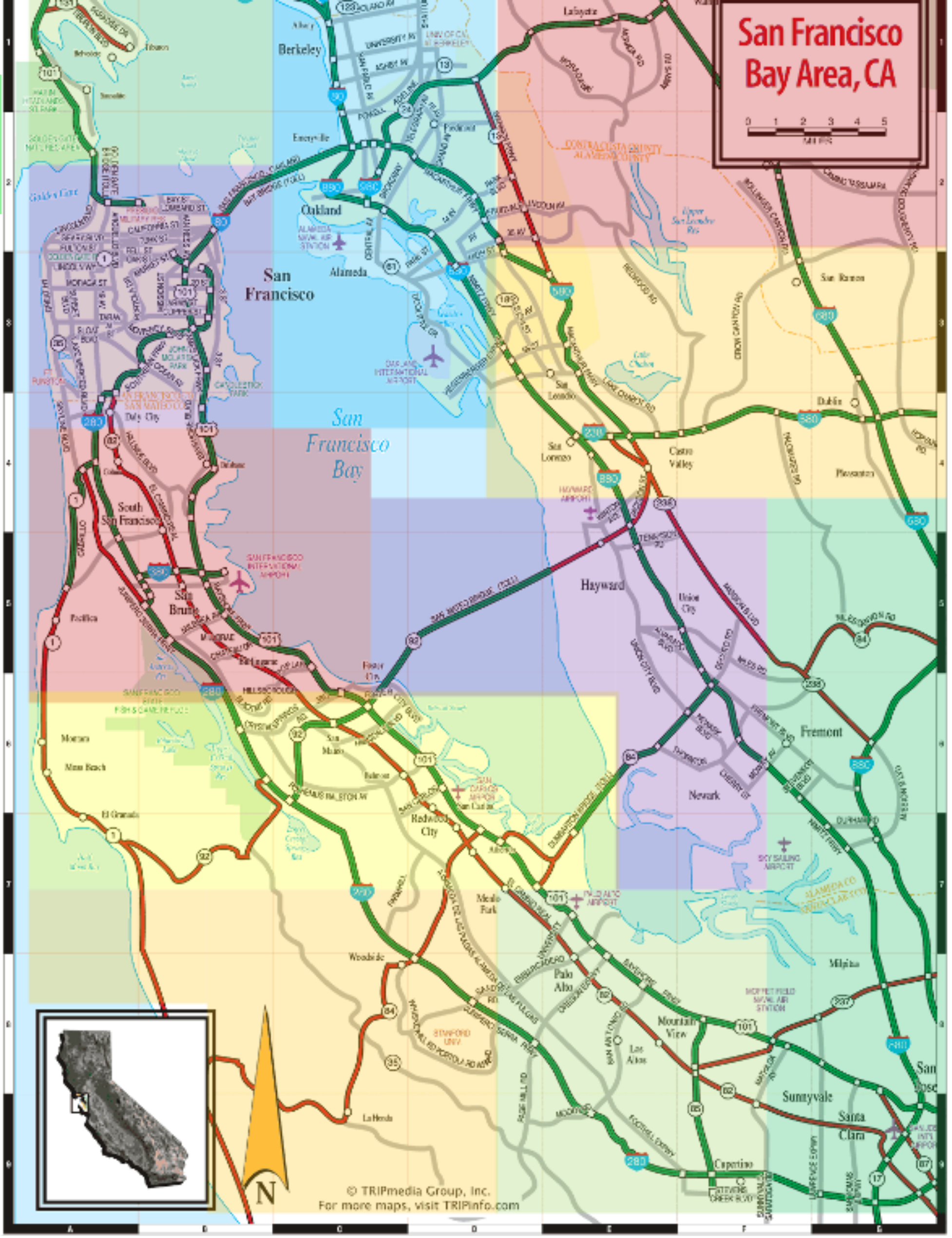


Transfer Learning with Graph Neural Networks (TL-DCRNN)

- TL-DCRNN predicted speeds and flows better than many models.
- It was trained with Los Angeles data and tested on a San Francisco network.

Method	MAE	RMSE	MAPE
Training and testing on PEMS-BAY			
ARIMA	3.38	6.50	8.30%
SVR	3.28	7.08	8.00%
FNN	2.46	4.98	5.89%
FC-LSTM	2.37	4.96	5.70%
STGCN	2.49	5.69	5.79%
DCRNN	2.07	4.74	4.90%
Graph Wevenet	1.95	4.52	4.63%
GMAN	1.86	4.32	4.31%
Training on LA and testing on PEMS-BAY			
TL-DCRNN	2.13 ± 1.09	5.23 ± 2.29	5.55 ± 4.34





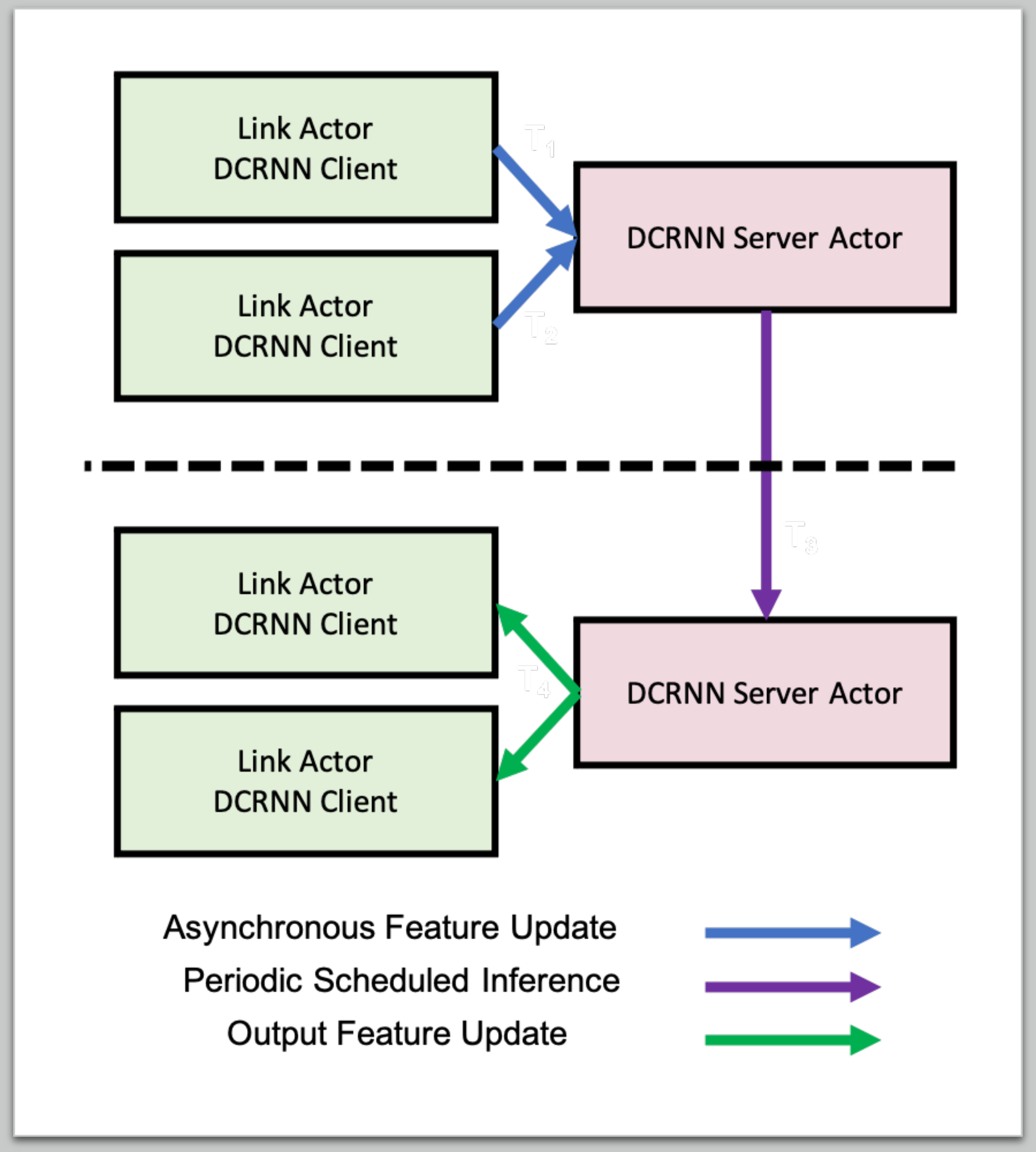
Example Partitioning of Freeways into Neighborhoods

Mobiliti DCRNN Model Integration – Design Considerations

- DCRNN inference is computationally expensive compared to simpler delay model function evaluations
- Since DCRNN inference happens at neighborhood granularity, Mobiliti instantiates an DCRNN Server actor per neighborhood of at most 200 links
- Each DCRNN Server aggregates input features from its links, periodically executes the model inference, and sends the outputs to its links
- This architecture amortizes the cost of inference over many links and permits tuning the inference frequency to balance model accuracy and speed

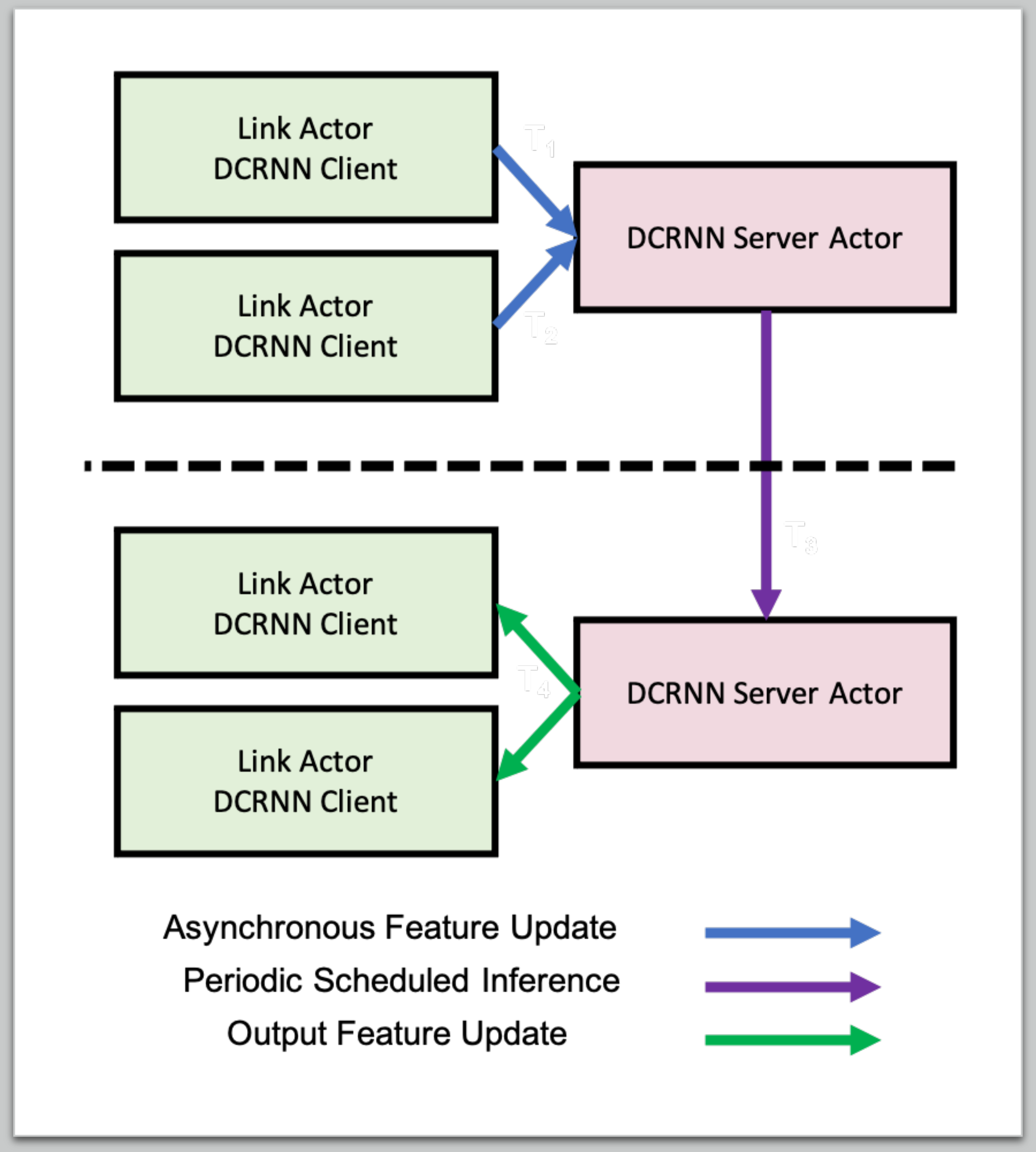
Mobiliti DCRNN Model Integration – Simulation

- Each link actor has a DCRNN Client component that interfaces with its assigned DCRNN Server
- Each participating link keeps track of its current DCRNN input variable values (e.g. vehicle speed or density)
- When the link's input variable(s) change more than a specified threshold, it asynchronously sends an update to its assigned DCRNN Server
- Each DCRNN Server aggregates model input features from its neighborhood of links
- The DCRNN Server periodically queries the DCRNN model with current link data and sends the output features (e.g. vehicle speeds) to the link actors in its neighborhood

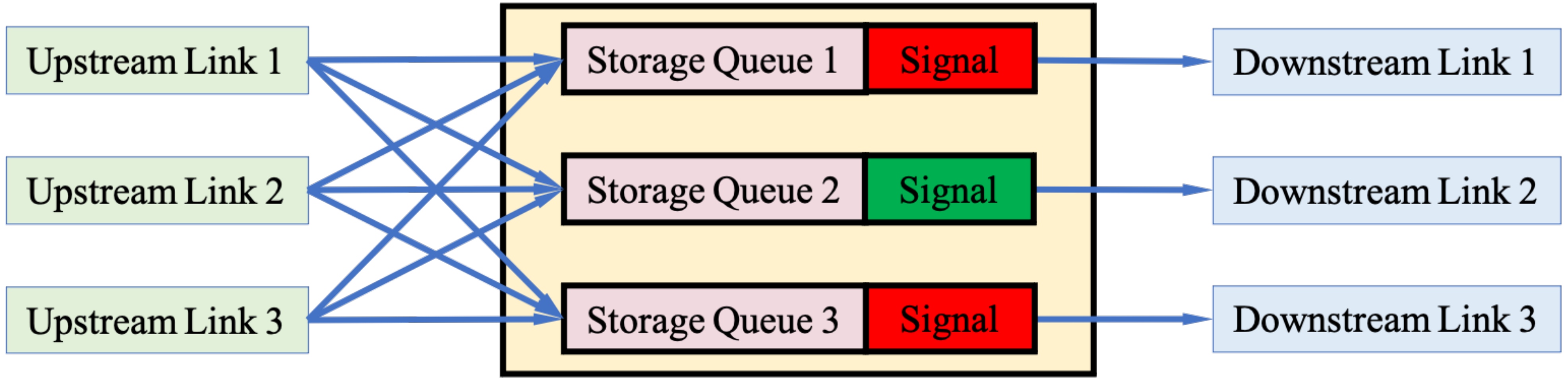


Mobiliti DCRNN Model Integration – Current Status

- TensorFlow build integration completed
- Network filters and DCRNN neighborhood partitioning completed
- Mapping of DCRNN Server actors to hardware ranks completed
- Most of DCRNN Client and Server class features are implemented
- Update message events between Client and Servers are implemented
- Need to determine the correct scaling parameters and units for model input/output to help with generalizability

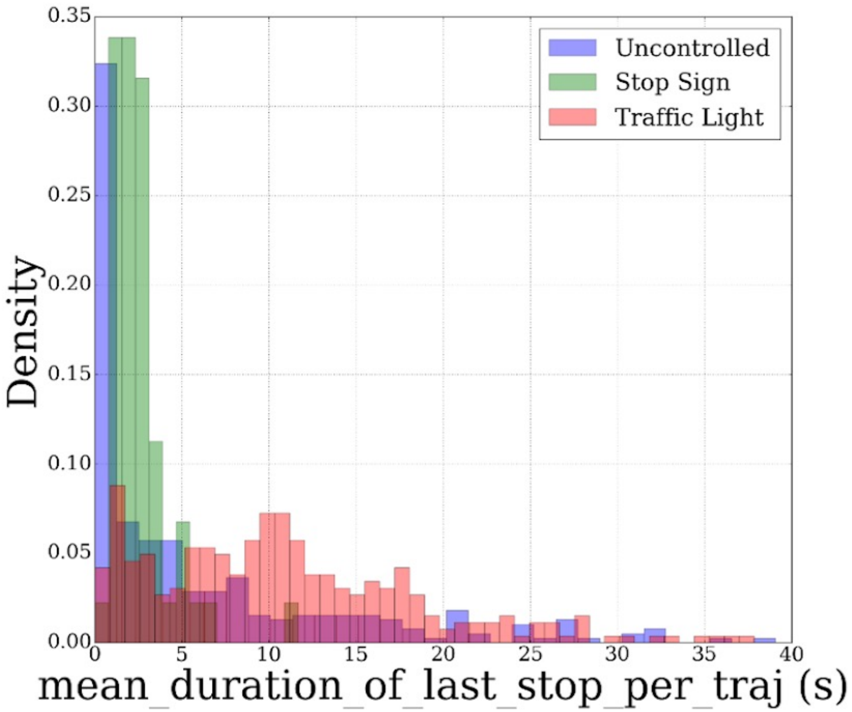


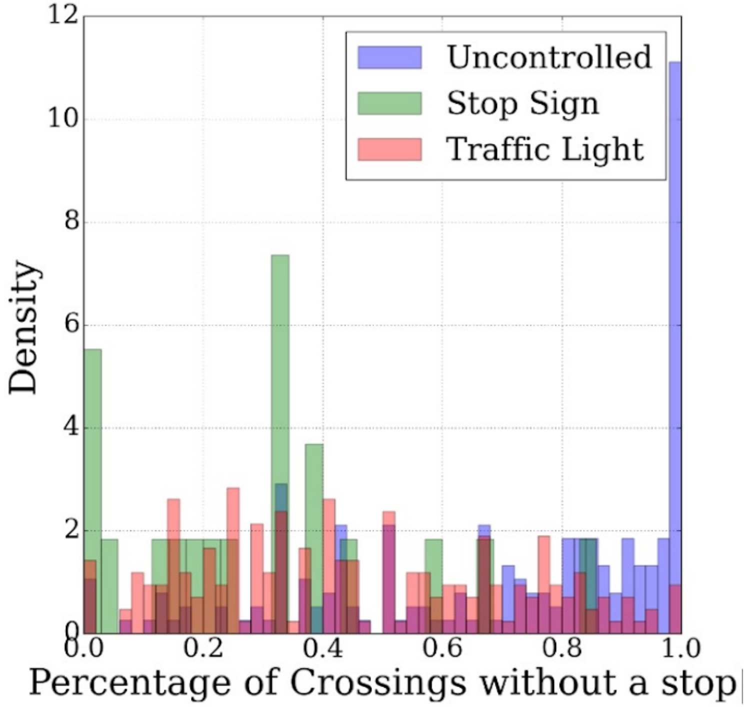
Updated Link Model Summary



- Computation steps when each vehicle arrives at link actor:
  - Compute congested link traversal time with link delay model (e.g. DCRNN)
  - Compute additional delay due to waiting for next available green signal phase and assign vehicle to a time slot
  - Send vehicle enqueue request with preliminary transition time to downstream link
  - Downstream link will reply with actual transition time when it has free capacity

Signal Detection Using GPS Trajectories





XGBoost				
Class	Precision	Recall	F1-score	Support
Uncontrolled	0.81	0.63	0.72	26
Stop Sign	0.86	0.75	0.80	8
Traffic Signal	0.93	0.98	0.95	128
Accuracy			0.91	162

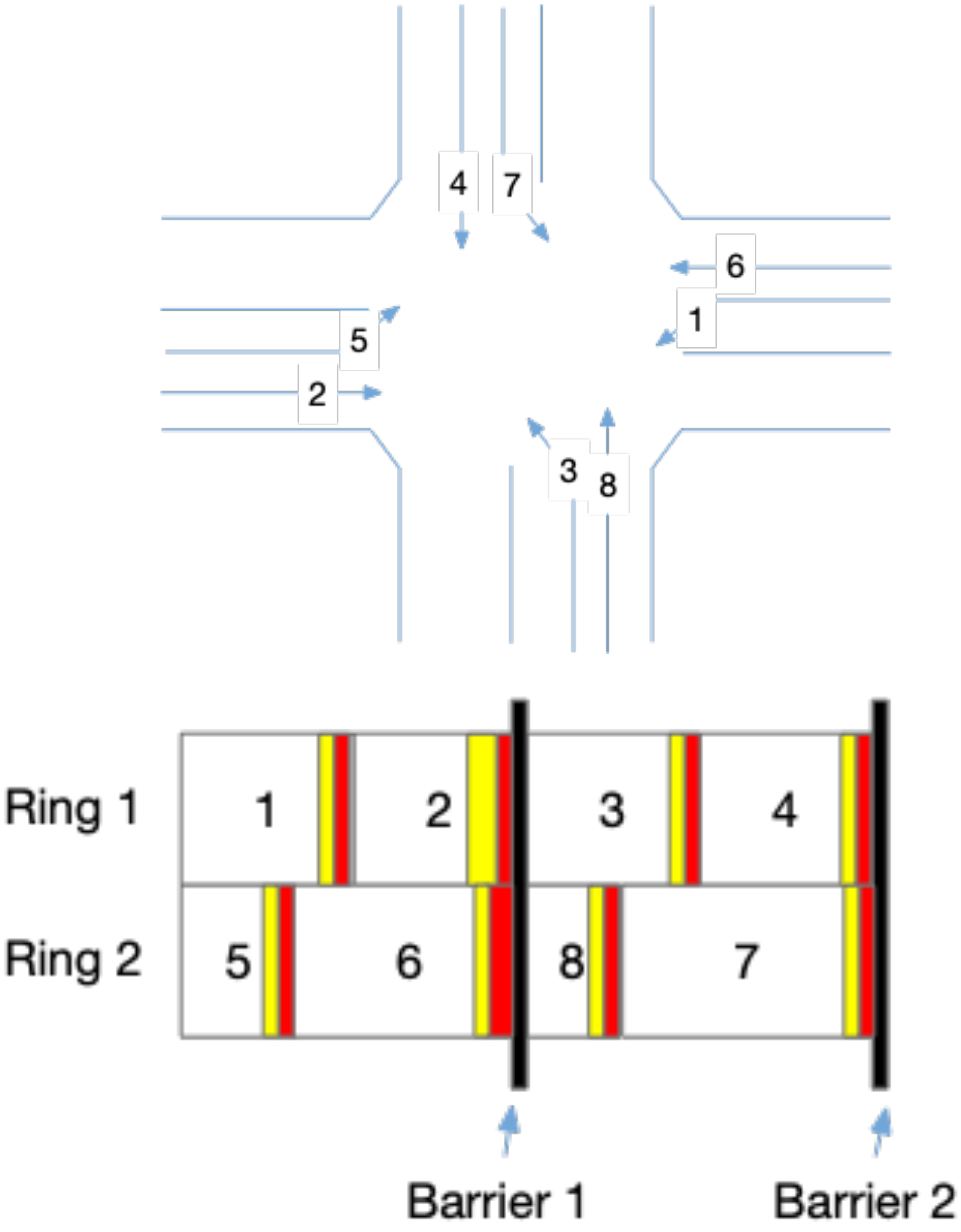
Bay Area Signal Map

- ~30k uncontrolled
- ~6.7k Stop Signs
- ~19k Traffic Lights

Signal Setups

- Signal phase and timing data for the areas of the Los Angeles (LA) County including Arcadia, Duarte, Temple City, Monrovia, Pasadena and San Marino.
- Mobiliti architected to represent ring control
- Mapping the phases of each signalized intersection to a movement represented by a sequence of nodes, generate the signal timing for all the intersections and estimate the offsets and phase sequence from probe data using machine learning.

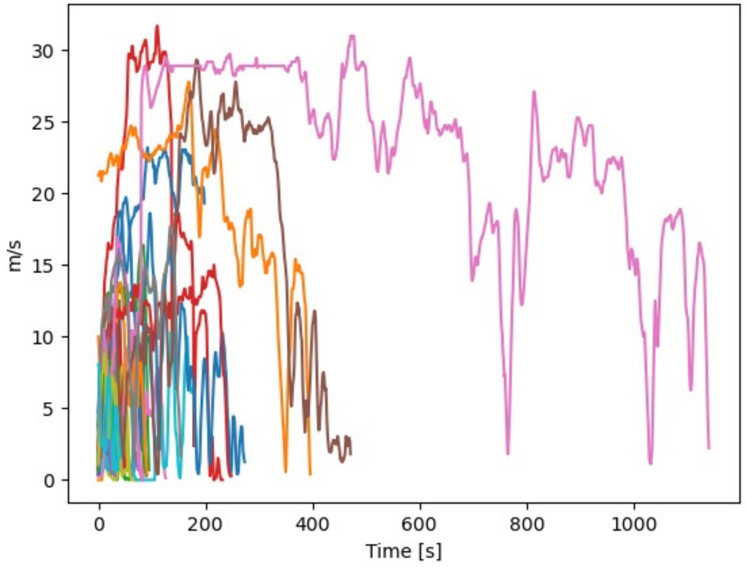
NEMA Representation of Signals



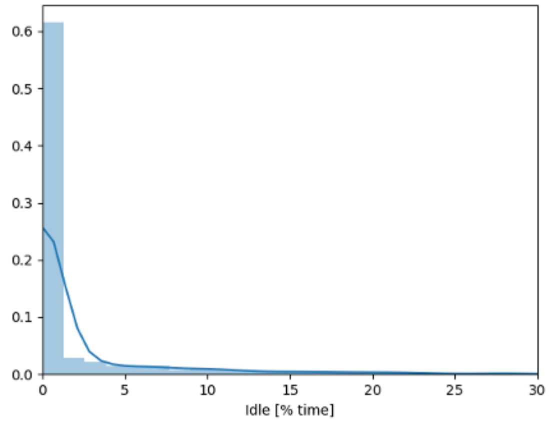
Drive Cycle Evaluation

- Velocity/time profiles used as input to predict consumption of EV and conventional vehicle
- Aggregate analysis of drive time:
  - Avg speed = 14.2 m/s
  - Relatively low average acceleration rates

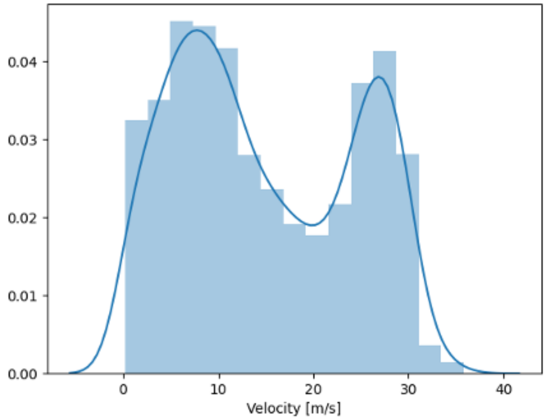
(+) First 15 file drive traces (velocity vs time)



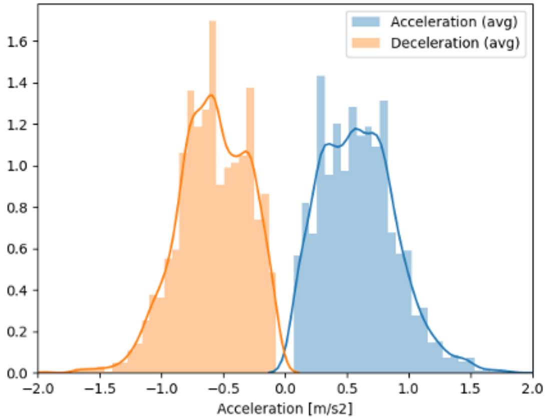
%Idle time
(speed = 0)



Speed (m/s)
avg = 14.2 m/s (32 mph)



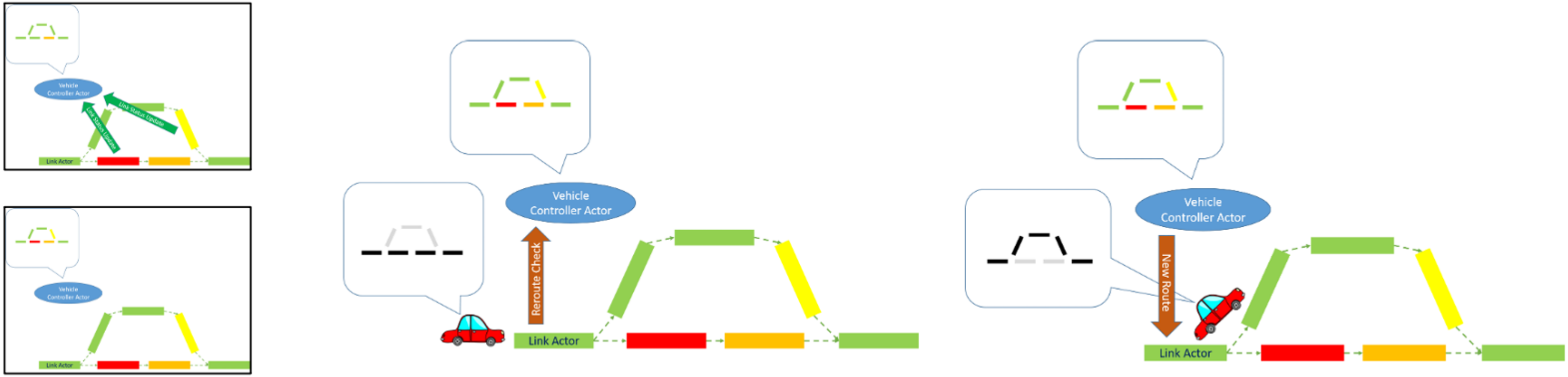
(+) and (-) acceleration
(+) and (-) avg = 0.6



Energy Modeling

- Completed the development of conventional and electric vehicle energy predictive models using supervised machine learning approaches on experimental vehicle data. These higher fidelity models will be compared to FASTSim and RouteE models to determine the model fidelity required to accurately assess the energy consumption of a heterogeneous mix of vehicles in Mobiliti.
- Exploring available vehicle powertrain simulation models, mesoscopic vehicle energy consumption models, and large-scale vehicle trajectory/drive cycle datasets.
- This approach will ultimately be represented in Mobiliti. ANL’s test vehicle data driven automated machine learning models and NREL’s FASTSim and RouteE models will all be used for microscopic and mesoscopic vehicle energy consumption modeling and RouteE models will ultimately be implemented in Mobiliti to represent the distribution of energy consumption over each link.

Dynamic Routing



Link status updates

Vehicle checks with controller before traversing congested links

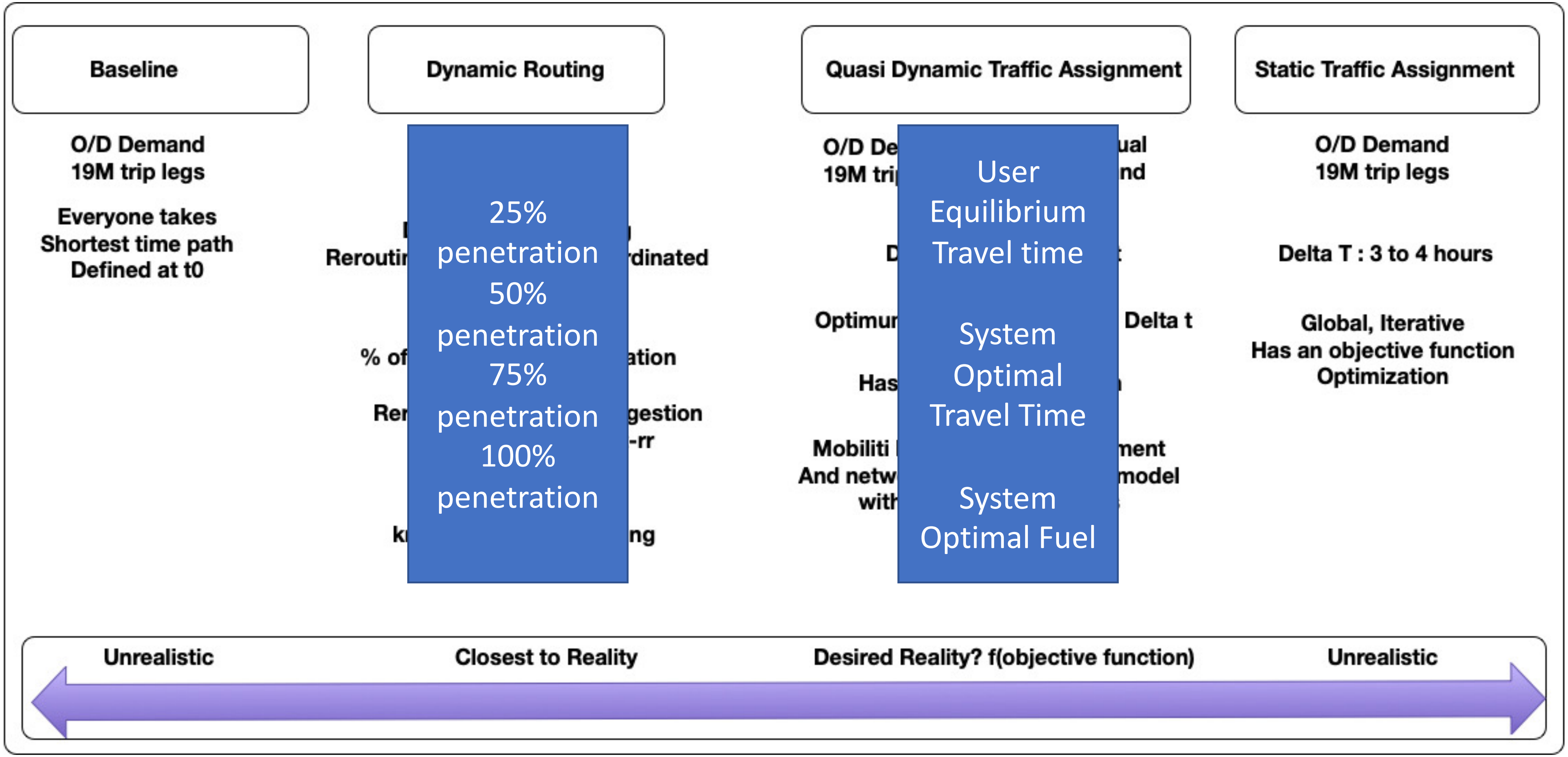
New route avoids congested links

Vehicle controllers provide localized routing capability. Each controller can have its own routing methodology.

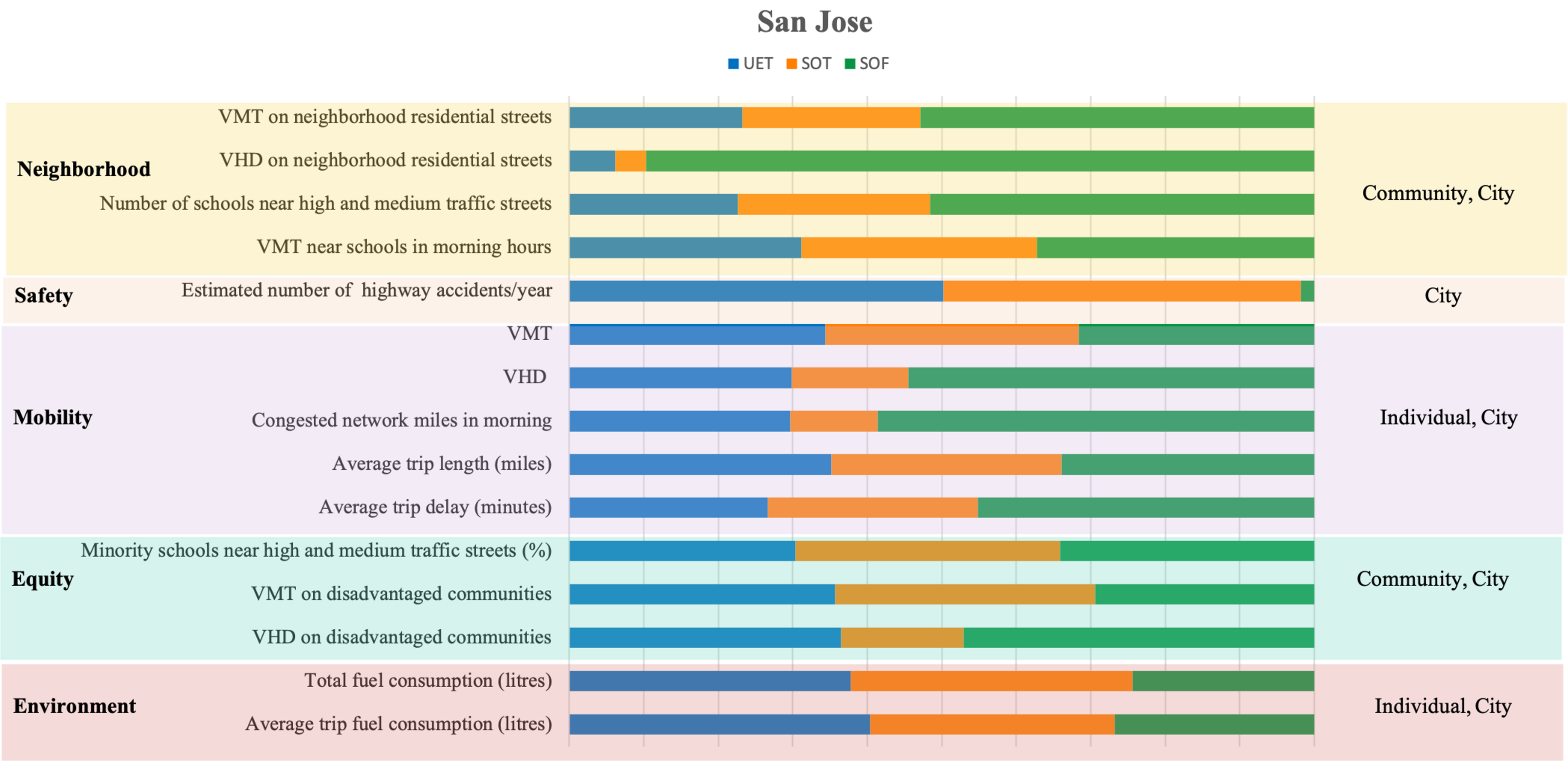
Mobiliti : Scenario Selection

Baseline	Dynamic Routing	Quasi Dynamic Traffic Assignment	Static Traffic Assignment
O/D Demand 19M trip legs	O/D Demand 19M trip legs	O/D Demand 19M trip legs + Residual Demand	O/D Demand 19M trip legs
Everyone takes Shortest time path Defined at t0	Delta t-rr for rerouting Rerouting is not globally coordinated	Delta T for assignment Mobiliti: 15 minutes	Delta T : 3 to 4 hours
	Selfish routing	Optimum determined at each Delta t	
	% of vehicle fleet participation	Global, Iterative Has an objective function Optimization	Global, Iterative Has an objective function Optimization
	Reroute is based on congestion In existing route at t-rr	Mobiliti PDES provides movement And network loading, queuing model with capacity constraints	
	% of vehicle fleet knowledge based routing - not modeled		
Unrealistic	Closest to Reality	Desired Reality? f(objective function)	Unrealistic

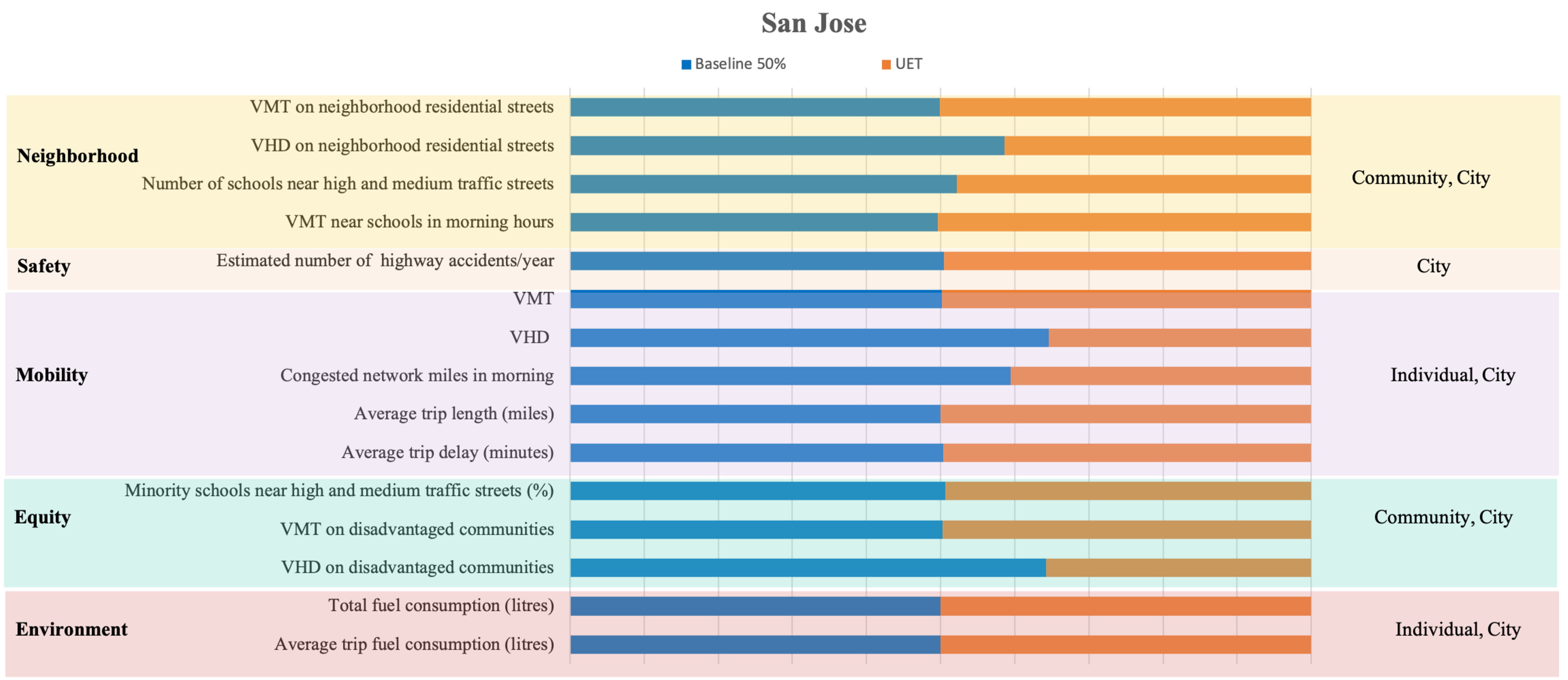
Mobiliti : Hybrid Selection



Understanding the Impacts on a City



Understanding the Impacts on a City



Collaboration and Coordination

Challenges and Proposed Future Research

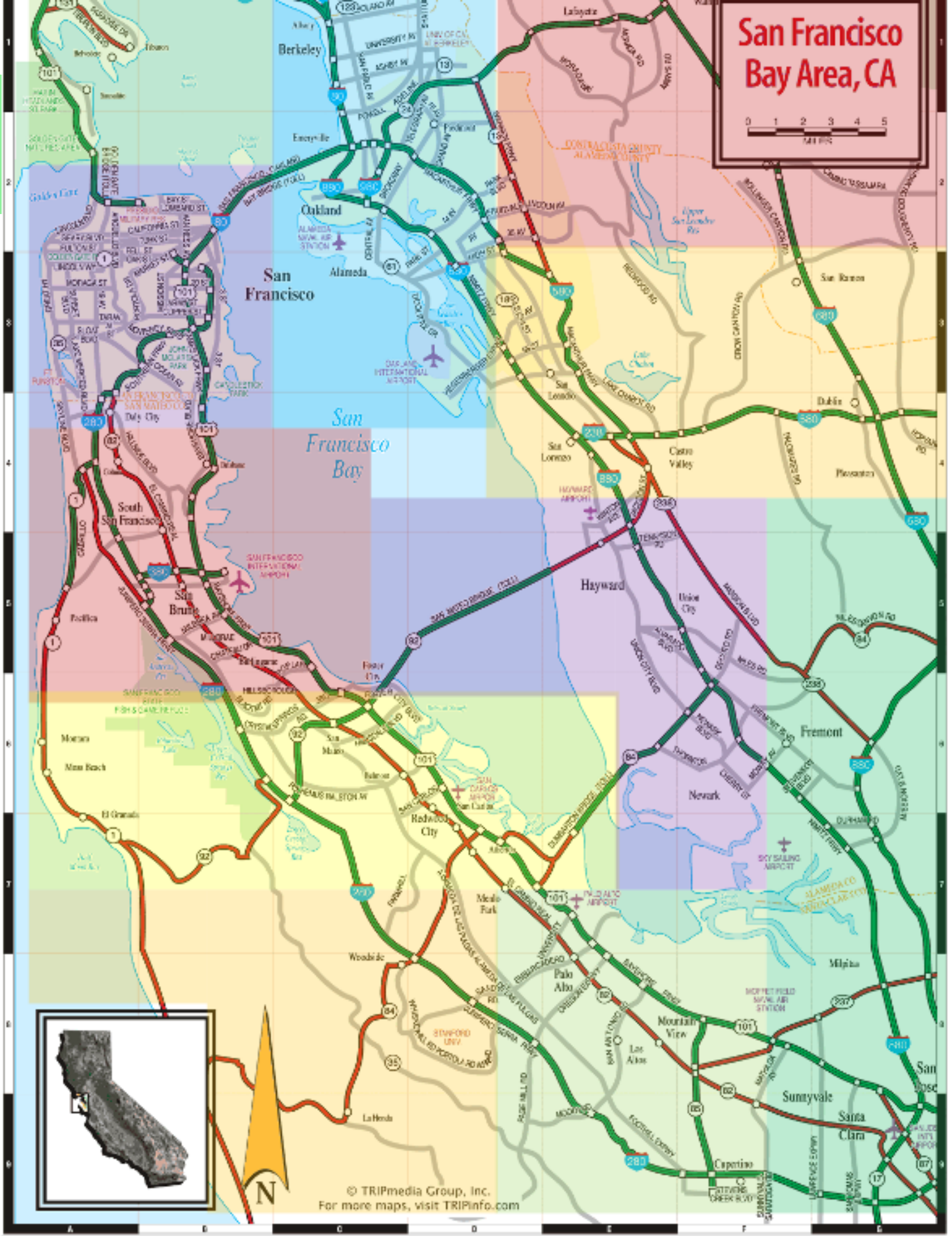
- Finding appropriate data to improve our data driven approach
- Using GPS trajectory data as input to DCRNN for neighborhood coverage
- Generating simulation results that are comprehensive enough to cover the full network for training the surrogate model
- HPC compute time allotment for training surrogate model
- Signal control using reinforcement learning on a large grid
- Determining fuel model fidelity appropriate for urban scale models
- Determining appropriate traffic control partitioning for future edge solutions

Any proposed future work is subject to change based on funding levels.

Summary

- Integrated the DCRNN into the Mobiliti architecture for expanding it predictive capabilities into arterials
- Expanded the travel demand model to include GPS derived locations of interest
- Evaluated 1Hz trajectories for designing better fuel use approximations
- Generated city level metrics for San Jose to include in Decision Support System

Technical Backup

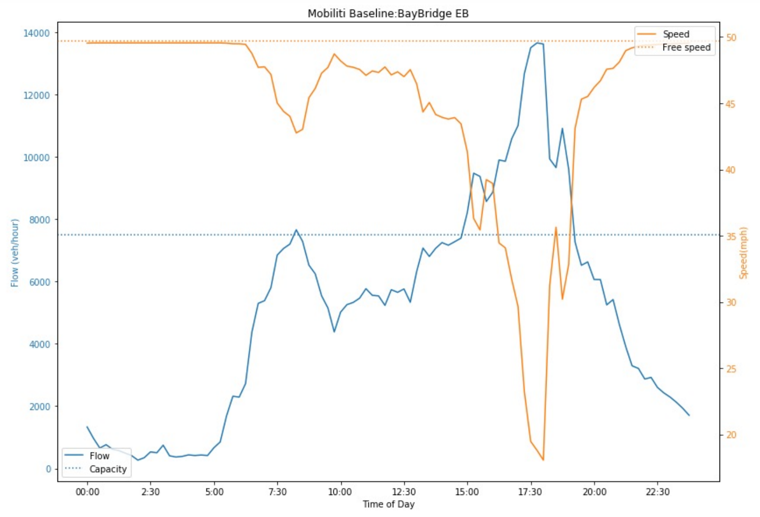


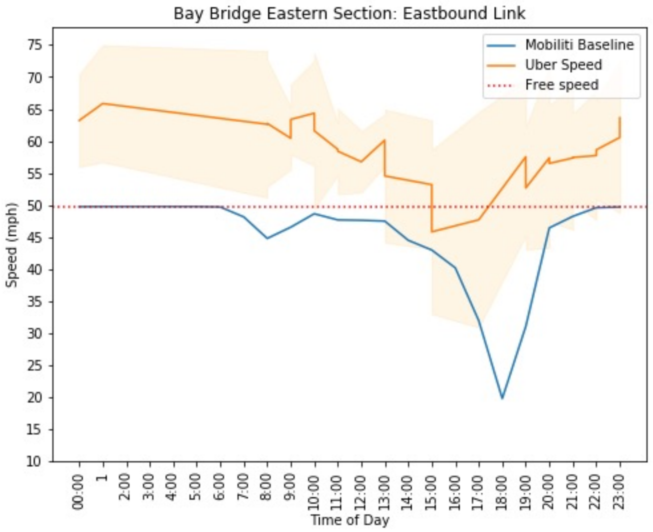
Example Partitioning of Freeways into Neighborhoods

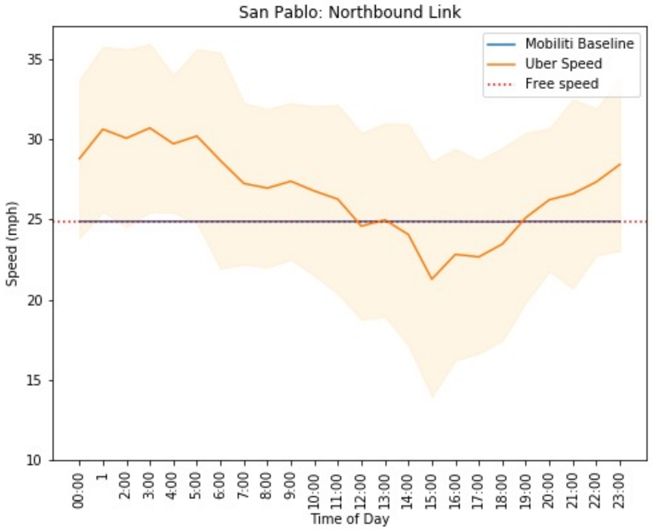
Mobiliti DCRNN Model Integration – Build and Initialization

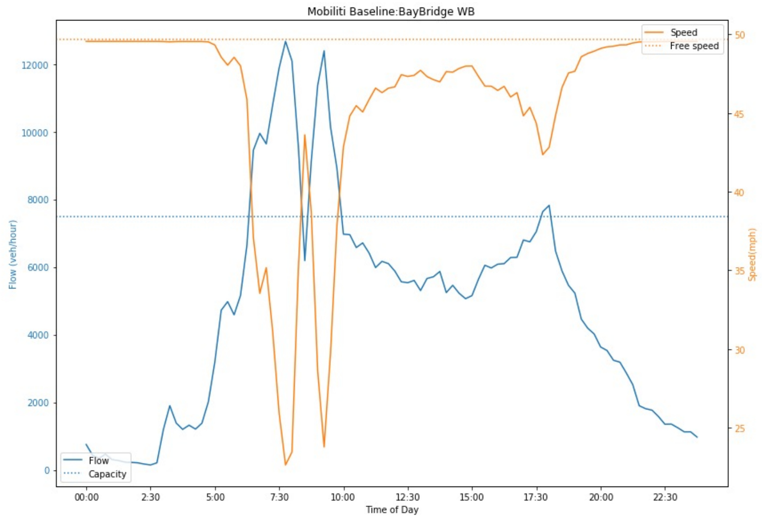
- Build process:
  - Augmented Mobiliti build to link external TensorFlow C API library and C++ interface (CPPFlow)
- During program initialization:
  - Identify which links will participate in the DCRNN model (e.g. FC 1/2 links only)
  - Partition selected links into neighborhoods of at most 200 links (about 100 groups for SF) using METIS to compute balanced partitions with low edge cuts
  - Instantiate DCRNN Server for each neighborhood partition, close (in hardware) to constituent link actors
  - Each server loads pre-trained TF model from disk and performs a warm-up inference computation

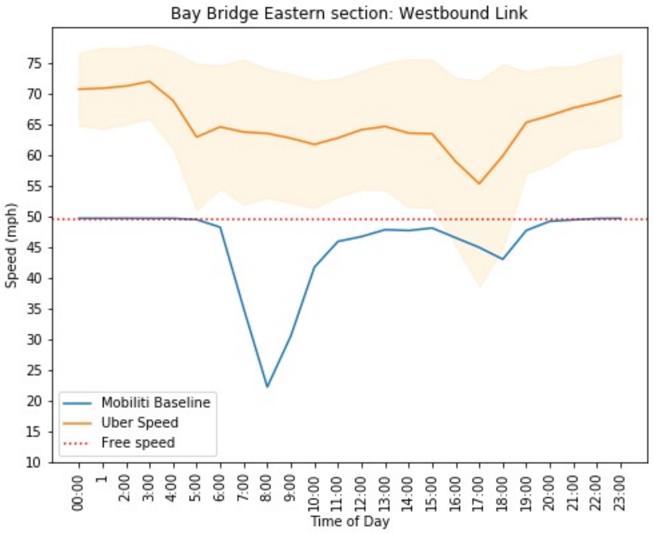
Uber Validation : Bridge Speeds and Flows

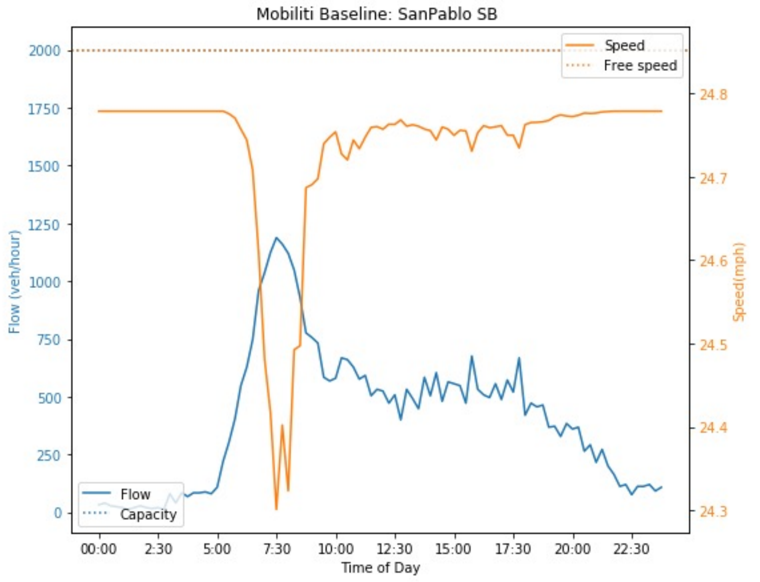












Validation : Daily Bridge Counts

Mobiliti

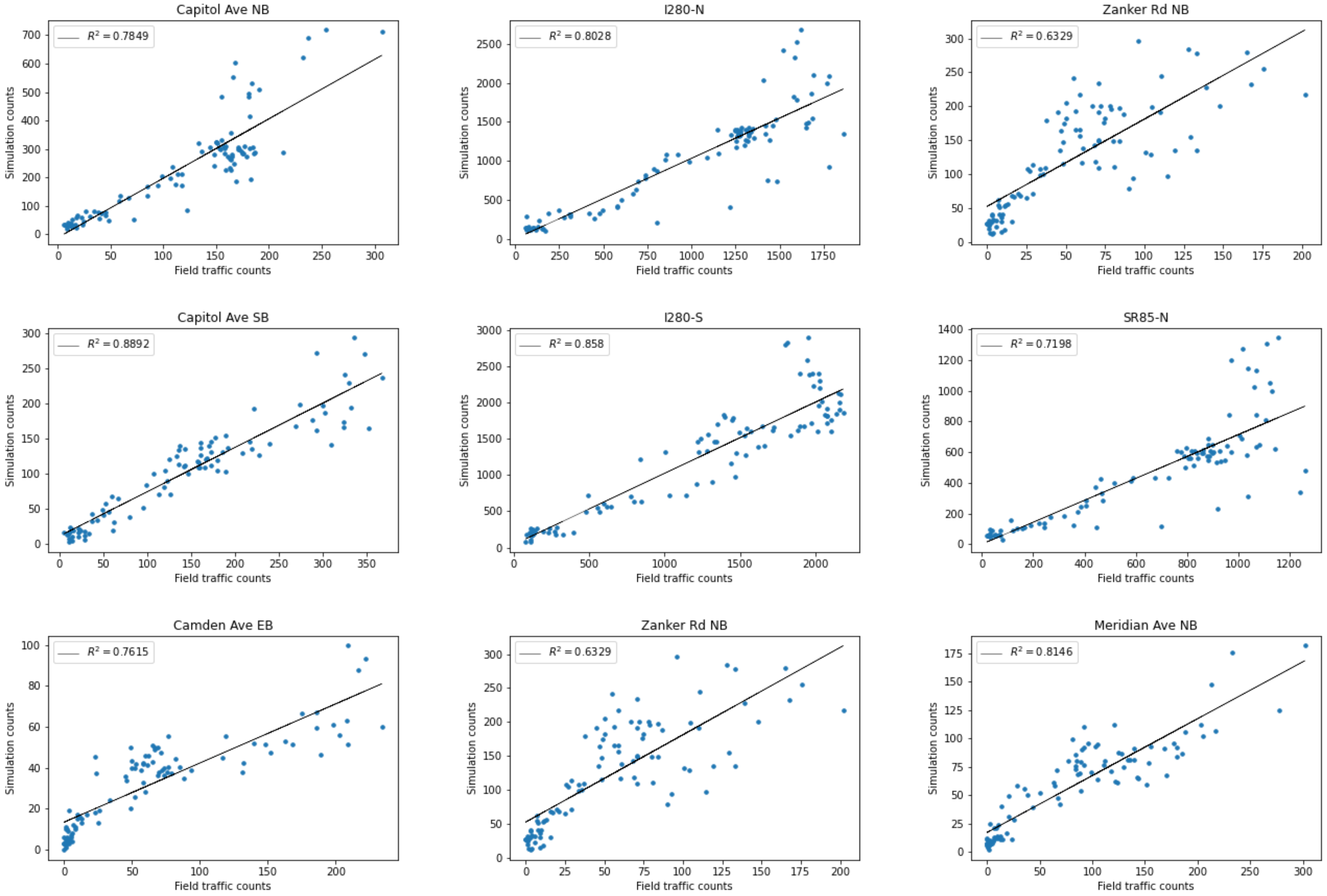
Bridge	Observed Count (Literature) [1]	Estimated Count (Mobiliti Baseline)	Relative Error
Bay Bridge	247,500	238,648	-3.58%
San Mateo	97,000	94,081	-3.01%
Dumbarton CA 84	81,800	97,218	18.85
Richmond	79,200	79,610	0.52
Golden Gate	112,000	94,230	-15.87%

Dynamic Traffic Assignment

Bridge	Observed Count (Literature) [1]	Estimated Count (DTA)	Relative Error
Bay Bridge	247,500	257,956	-3.58%
San Mateo	97,000	108,643	-12%
Dumbarton CA 84	81,800	105,921	-30%
Richmond	79,200	89,500	-13%
Golden Gate	112,000	104,144	-6%

Source: [1] Current and Projected Conditions Report San Francisco Bay Crossings Study Update Prepared for the Bay Area Toll Authority, 2010

Validation – San Jose Vehicle Counts, PeMS



Project Publications

Graph-partitioning-based diffusion convolutional recurrent neural network for large-scale traffic forecasting. T. Mallick, P. Balaprakash, E. Rask, and J. Macfarlane. *Transportation Research Record 2674 (9), 473-488, 2020*

Transfer learning with graph neural networks for short-term highway traffic forecasting. T. Mallick, P. Balaprakash, E. Rask, and J. Macfarlane. *International Conference on Pattern Recognition (ICPR) 2020.*

Scalable Diffusion Convolution Recurrent Neural Network for Large-Scale Traffic Forecasting. T. Mallick, P. Balaprakash, E. Rask, and J. Macfarlane. *Poster presentation at WiML workshop, NeurIPS, 2019*.

P. Balaprakash. Artificial intelligence for science. TRB ExComm A.I. Policy Session, Invited talk and panelist, January 2020.

Cy Chan, Anu Kuncheria, and Jane Macfarlane, "Understanding the Impact of Dynamic Rerouting on Urban-Scale Traffic Dynamics Using Parallel Discrete Event Simulation," Submitted to IEEE Transactions on Intelligent Transportation Systems, 2021.

To Be Submitted:
Anu Kuncheria, Colin Laurence, Cy Chan, Jane Macfarlane, "Evaluating Major Transportation Events in Urban Environments Using Real-World Movement Data and Simulation"