Machine learning-enhanced decision-making frameworks for sustainable mobility systems

Jimi Oke

Networks for Accessibility, Resilience and Sustainability Laboratory (NARS Lab)
Department of Civil and Environmental Engineering
University of Massachusetts Amherst

March 21, 2024

Outline

Academic journey

BA (2010)	Physics, Music, Williams College, MA
Teacher (2010–12)	Mathematics, Center for Learning, The Pennington School, NJ
PhD (2012–16), MSE (2012–14)	Civil [and Systems] Engineering, Johns Hopkins University
Postdoctoral Associate (2016–19)	Civil and Environmental Engineering, Massachusetts Institute of Technology
Asst. Prof. (2019–)	Civil and Environmental Engineering, University of Massachusetts Amherst

Research agenda

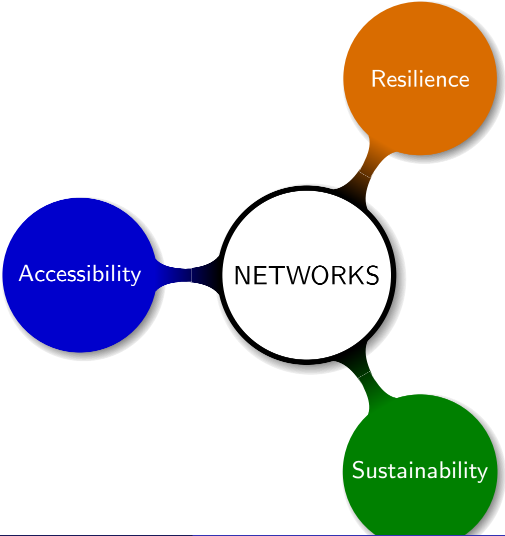
Networks for Accessibility, Resilience and Sustainability (NARS) Lab

Objective

- Advance sustainable and zero-emissions infrastructure systems
- Learn typology representations to guide decision-making
- Quantify and track emissions and energy metrics

Methods

- Machine learning
- Optimization (and simulation)



Team

Current members (status; date joined):

Zhuo Han
MS/PhD; Winter '20

Mohammed Abdalazeem
MS/PhD; Winter '21

Mahsa Arabi
PhD; Summer '21

Peiyao Zhao
PhD; Fall '23

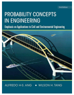
Jimi Oke
Fall '19

Alumni
Atanas Apostolov (MS '24); Alexa Weinman (BS '23); Hichul Chung (BS '22); Vivian Rost-Nasshan (NSF REU '23).

Teaching

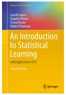
Undergraduate

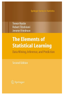
- Probability and Statistics in Engineering (Civil/Mechanical/Industrial); Fall 2019, '20, '21

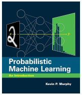


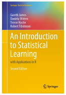
Graduate (new)

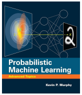
- Big Data and Machine Learning for Engineers (Spring 2020, '21, '22)
- Probabilistic Machine Learning (Spring 2023)
- Machine Learning Foundations and Applications (Fall 2023)
- Advanced Probabilistic Machine Learning (Spring 2024)

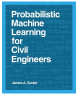












Overview

- How can we leverage machine learning (ML) for decision-making and sustainability planning in mobility systems?
- Four projects:
  - P1: Typology-informed trip chaining for boarding-only systems
  - P2: Road network typology at metropolitan statistical area level
  - P3: Tracking energy for optimal fleet transition pathways in bus transit network
  - P4: Robust decision-making framework for sustainable urban rail transit

ML-enhanced trip chaining

Supervised and unsupervised learning approaches can vastly improve trip chaining algorithms with boarding-only data.[1]

- Trip chaining connects dots between boardings and alightings to facilitate transit decision-making
- Traditional data collection methods are often resource-intensive and not always feasible
- Automated fare collection (AFC) systems offer detailed, high-frequency data
- Bus transit systems rely on mobile-based AFC but few trip chaining models exist for this type

Team

RA: Mohammed Abdalazeem

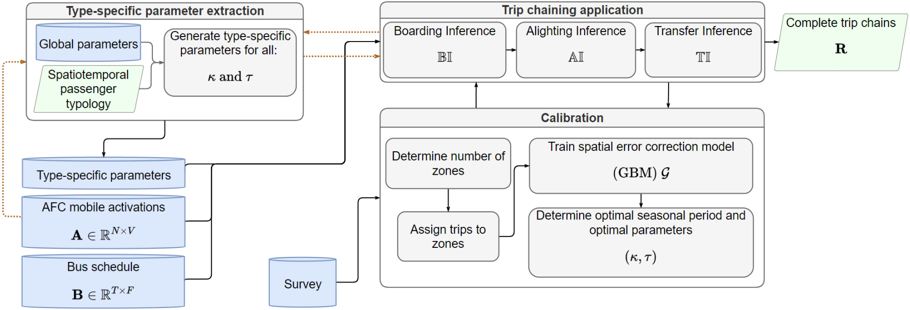
Sponsor: Pioneer Valley Transit Authority PVTA

Tolu Oke, Alex Forrest, Sandra Sheehan

[1]Mohammed and J. Oke 2023; Abdalazeem and J. Oke 2023.

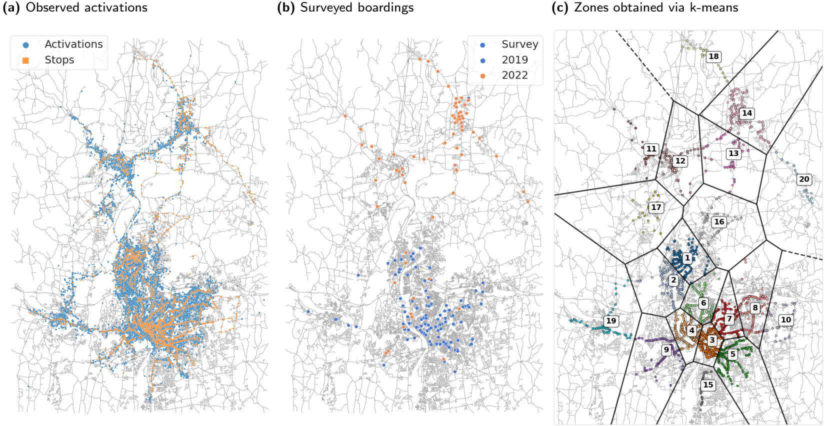
Research questions and framework

1. How can we leverage mobile ticketing data to infer complete passenger journeys within regional transit networks?
2. What role does typology information play in improving trip chaining accuracy?
3. How can we improve the framework's performance by using machine learning?



Stops and zones

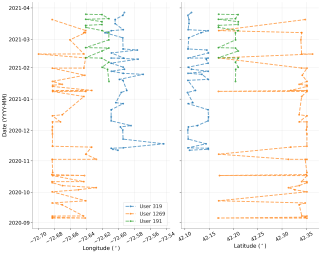
Study area: Pioneer Valley Transit Authority
Observation period: July 2020 – December 2021



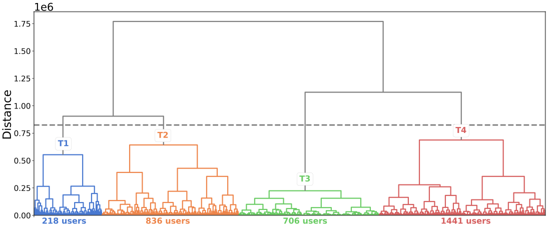
Clustering trajectories

- Aligned spatiotemporal passenger trajectories (encoded as binary 2D) using dynamic time warping[2]
- Obtained pairwise dissimilarity matrix
- Hierarchically clustered passengers using the Ward method

(a) Example trajectories



(b) Dendrogram

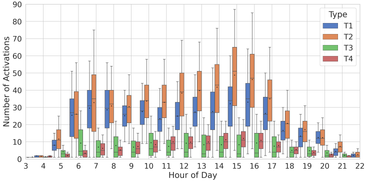


[2] Mueen et al. 2016.

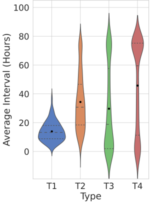
Typology identification and insights

T1: Commuters **T2:** Flex-Morning **T3:** Afternoon-Shift **T4:** Irregulars

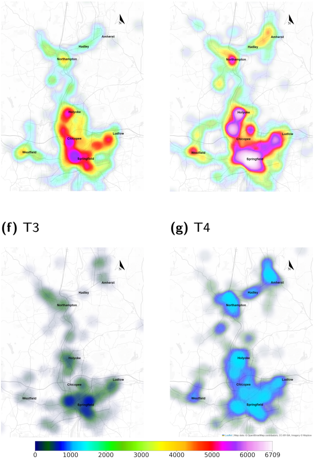
(a) Hourly user activations



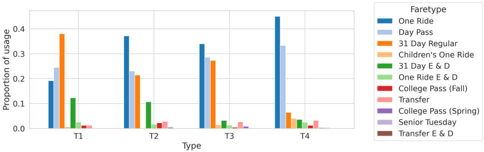
(b) Ave. interval between pax activations



(d) T1 **(e)** T2 **(f)** T3 **(g)** T4



(c) Faretype usage distribution



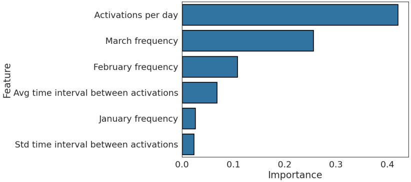
Provides type-specific parameters for trip chaining algorithm

Passenger type classification model

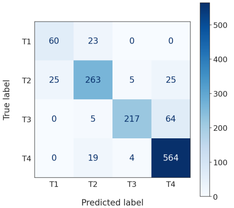
- Trained models to predict passenger type (60:40 train–test split)
- Compared several models: gradient boosting classifier was best
- Validates unsupervised typology and speeds up future type assignment

Model	Accuracy	Precision	Recall	F1 Score	AUC-ROC
Gradient boosting	0.8666	0.8713	0.8666	0.8650	0.9674
K-nearest neighbors	0.5714	0.5625	0.5714	0.5612	0.7972
Logistic regression	0.8140	0.8128	0.8140	0.8105	0.9433
Multilayer perceptron	0.6523	0.5238	0.6523	0.5656	0.8399
Naive bayes	0.3980	0.5413	0.3980	0.3748	0.7987
Random forests	0.8422	0.8473	0.8422	0.8391	0.9528

(a) Feature importance



(b) Confusion matrix



Spatial error correction

- Trip chaining algorithm infers transfer and destination zones (j, k) for each alighting zone (i) per passenger
- Error pattern identified in predicted trip probabilities

$$r_{ijk} = p_{ijk} - \hat{p}_{ijk} \tag{1}$$

  where r_{ijk}: residual; p_{ijk}: observed (survey) probs.; $\hat{p}_{ijk}$: predicted
- Fitted a gradient boosting model $\mathcal{G}$ (via to grid search) to learn error structure

$$\hat{r}_{ijk} = \mathcal{G}(i, j, k) \tag{2}$$

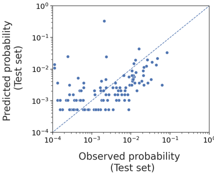
- Used predicted residuals $\hat{r}$ to

Overall improvements from enhancements

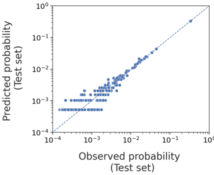
Typology, seasonality and spatial error correction (Model 6) yield 70% improvement in MAE (compared to Model 1)[3]

Model	Season-aware	Typology-informed	Spatial Error Corrected	Performance Metrics MAE ($\times 10^{-4}$)	MALE ($\times 10^{-4}$)	JSD
1	–	–	–	1.52	1.45	0.29
2	–	–	✓	0.65	0.65	0.04
3	–	✓	–	1.54	1.47	0.28
4	–	✓	✓	0.54	0.53	0.03
5	✓	✓	–	1.51	1.44	0.28
6	✓	✓	✓	0.42	0.42	0.02

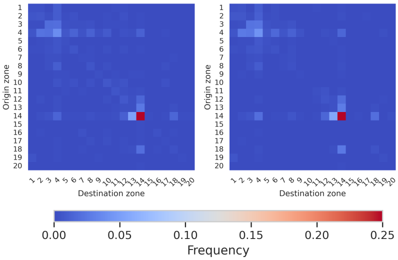
(a) Before enhancements



(b) After enhancements



(c) Predicted trip probabilities **(d)** Observed trip probabilities



[3] Paper in review: Abdalazeem and Oke

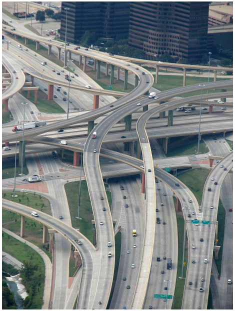
Key takeaways

- ML enhancements (typology, spatial error correction) and seasonality significantly improve trip chaining performance
- Typology offers insights into passenger travel patterns, facilitating more informed transit planning and resource allocation
- Scalable and cost-effective framework for regional transit system planning

Understanding our road networks

Effective and equitable deployment of shared autonomous electric vehicles, transit and other mobility initiatives require understanding of our road networks

- US emissions in 2021: 6.3 $GtCO_2e$[a]
  - Almost 25% of this was due to roadway emissions (gasoline and diesel consumption)
  - Globally, the roadway share is <15%[b]
- Earth is crisscrossed by 40 million miles of road; 16% of this in the US[c]
- Congestion is also a major issue with quality of life, socioeconomic and emissions impacts
  - From 1993-2017 (top 100 urban areas)[d]: freeway ↑ 40%, population ↑ 30%, congestion ↑ 140%
  - Demand for personal mobility continues to grow (25 mi/day in 2017)



High-Five Interchange, Dallas, TX[a]

[a]M. Bellis 2021.

[a]US EPA 2017.
[b]US EPA 2016.
[c]Raboteau 2023.
[d]R. Bellis et al. 2020.

Questions

- What are the latent factors underlying road network characteristics at the metropolitan statistical area (MSA) level in the US?
  - What are the prevailing MSA road network types?
- Can road network type help explain congestion and other outcomes, and ultimately inform mobility policy?

Objectives

- Learn a typology representation (via dimensionality reduction and clustering) of basic road network indicators observed from 372 MSAs
- Fit a model to predict the travel time index (a congestion measure) with and without the network type label as a predictor

Study fills gap for MSA road network typology.[4]

Team: Alexa Weinman, Vivan Rost-Nasshan; Mohammed Abdalazeem

Partial support: NSF

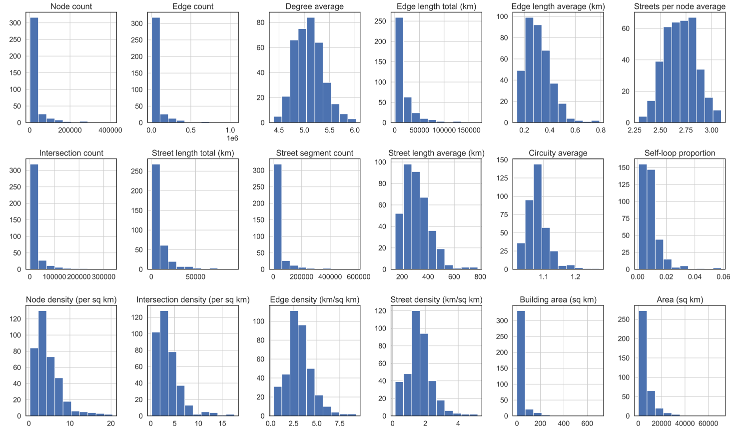
[4] Harris 1943; Bruce and Witt 1971; Louf and Barthelemy 2014; Hermosilla et al. 2014; J. B. Oke et al. 2019; Rath and Chow 2022.

Methods overview

- Dimensionality reduction of MSA network indicators: kernel principal components analysis (kPCA)
  - Analysis of extracted kernel principal components (kPCs) using random forests
- Clustering of MSA networks based on kPCs: Gaussian mixture model
- Fitting a model to predict congestion based on socioeconomic and mobility variables and network type factor: random forests
  - Target variable: travel time index in 2019 (average peak hour travel time by free-flow time) from Texas A&M Transportation Institute
  - Explanatory variables: population, income, modeshares (from American Community Surveys), and network type label

MSA road network indicators

- Obtained indicators from OpenStreetMap via the Python package: `osmnx`



Kernel principal components analysis (PCA)

- PCA: efficient lower-dimensional linear mapping of data matrix $\boldsymbol{X}$:

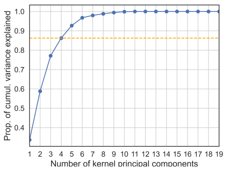
$$\boldsymbol{Z}_{N\times L} = \boldsymbol{X}_{N\times D}\boldsymbol{W}_{D\times L} \tag{4}$$

  where $\boldsymbol{W}$ is the weight matrix (eigenvectors for L largest eigenvalues of covariance matrix $\hat{\boldsymbol{\Sigma}} = \frac{1}{N}\boldsymbol{X}^\top\boldsymbol{X}$)
- **Kernel PCA:** projection in higher-dimensional feature space $\phi(\boldsymbol{x}_n)$ given by the kernel function:

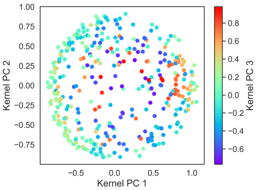
$$\mathcal{K}(\boldsymbol{x}, \boldsymbol{x}_n) = \phi(\boldsymbol{x}_n)^\top\phi(\boldsymbol{x}_n) \tag{5}$$

  - Weight matrix is given by the eigenvectors of the gram matrix $\boldsymbol{K} = \mathcal{K}(\boldsymbol{X}, \boldsymbol{X})$
  - The cosine kernel was used for this dataset

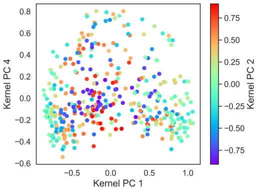
(a) Scree plot



(b) kPC2 vs. kPC1

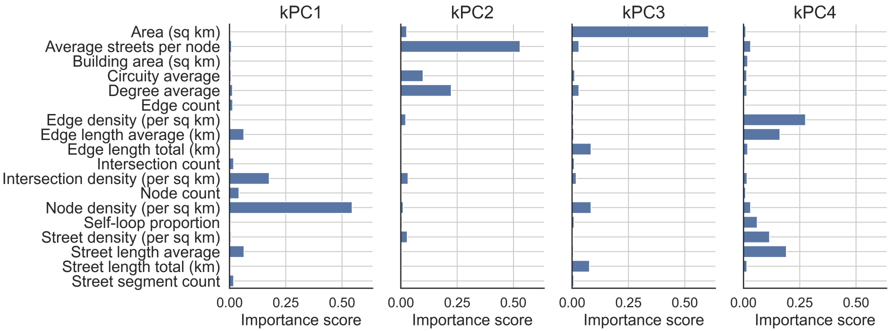


(c) kPC4 vs. kPC1



Interpreting the kernel principal components

- Trained random forests models to predict each kPC based on the 18 network indicators
- Characterizations:
  - kPC1: node/intersection density
  - kPC2: degree (average connections/streets per node)
  - kPC3: size/area
  - kPC4: edge/street density and length

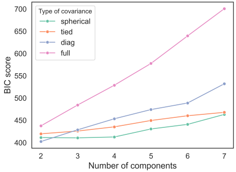


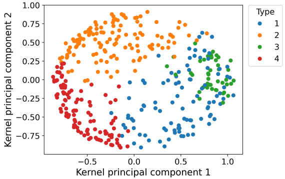
Gaussian mixture model (GMM)

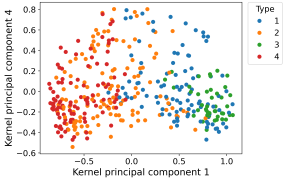
- A GMM for a multivariate observation $\boldsymbol{x}$ is specified as

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{6}$$

  - π_k is the mixture weight for cluster k
  - Parameters $\boldsymbol{\theta} = \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}$ estimated via expectation-maximization
  - Clusters assigned by maximizing posterior membership probability $p(z_n = k|\boldsymbol{x}_n, \boldsymbol{\theta})$
  - Hyperparameters: number of clusters K; covariance form: full, diagonal, tied, spherical
- Best fit: $K = 4$ clusters obtained; covariance structure: spherical

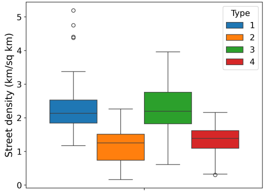


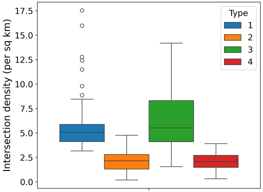


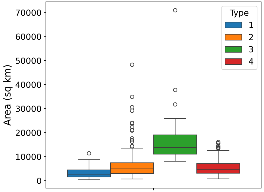


Characterizing the typology

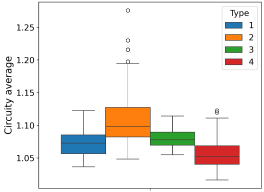
- **Types 1 and 3**: street- and intersection-dense; **Type 3**: largest in size

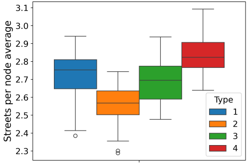






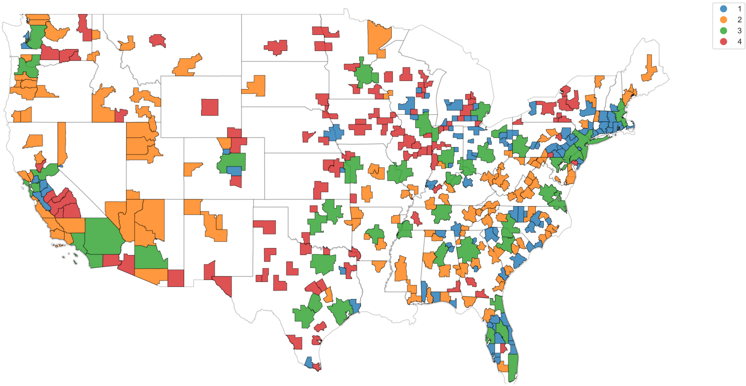
- **Types 2 and 4**: smallest size
- **Type 2**: most circuitous and lowest degree
- **Type 4**: highest degree average





Typology map and examples

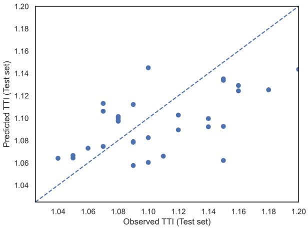
- Type 1: Akron, OH; Flint, MI; Worcester, MA-CT
- Type 2: Albuquerque, NM; Durham-Chapel Hill, NC; Portland-South Portland, ME
- Type 3: Autstin, Atlanta, Baltimore, Boston, LA, NY metro areas
- Type 4: Des Moines, IA; Fresno, CA; Wichita Falls, TX



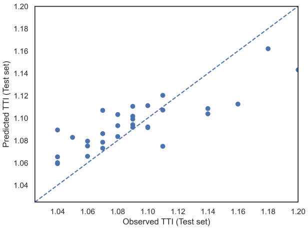
Preliminary results: predicting congestion

- Fitted random forests (RF) model to predict the 2021 travel time index (TTI) based on modeshares, income and household population ($N = 153$)
- Including type label yields mild improvement in performance

(a) Mean absolute % error: 1.8% without type label



(b) Mean absolute % error: 2.9% with type label



- Efficient latent representation of basic MSA road network indicators
- Four network types found via GMM
- Potential for typology to inform models for congestion, emissions, etc.

Energy modeling and optimization for transition planning

- What is the best pathway to transition bus fleet from diesel to electric powertrains given:
  - Budget constraints
  - Charging capacity constraints
- Can we track vehicle-specific energy consumption at high resolutions for better decision-making?

Team

RA: Mahsa Arabi
Sponsor: Pioneer Valley Transit Authority PVTA
Tolu Oke, Alex Forrest, Sandra Sheehan

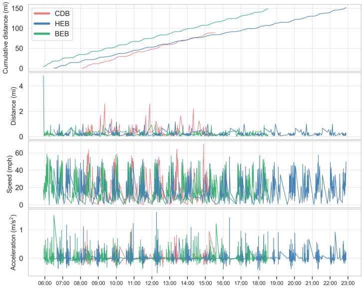
Tractive energy modeling

- Computed bus trajectories from year-long data (Oct '21 – Sep '22)
- Calibrated physics-based tractive energy models for conventional diesel (CDB), hybrid electric (HEB) and battery electric (BEB) buses
  - Example: diesel in gallons ($p \in \{C, H\}$):

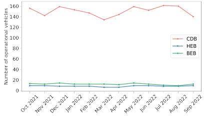
$$D_{p,i,t} = 3.79\Delta_t \left[\alpha_{0,p} + \left(\alpha_{1,p}P_{p,i,t} + \alpha_2 P_{p,i,t}^2\right)\mathbb{I}(P_{p,i,t} \geq 0)\right] \qquad (7)$$

  - $P_{p,i,t}$: tractive power function (speed, acceleration, mass); α_i: coefficients

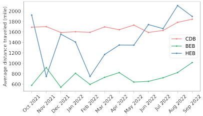
(a) Day-long trajectories of 3 sample buses (Aug 2022)



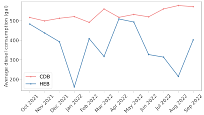
(b) Operational vehicles



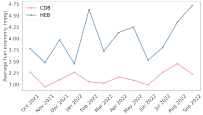
(c) Monthly avg. bus distance



(d) Monthly avg. diesel

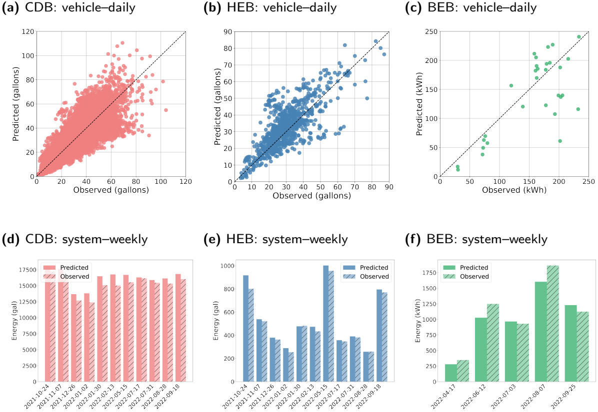


(e) Average fuel economy



Calibration and validation

- Used grid search to calibrate 8 physical model parameters
- 2×10^4 points (CDB, HEB)
- 10^5 points (BEB)
- Performance could be improved by including other variables and using ML



Powertrain	Vehicle-level RMSE	Vehicle-level MAPE	System-level RMSE	System-level MAPE
CDB	9.49 gal	22%	1029 gal	6%
HEB	8.98 gal	24%	64 gal	8%
BEB	51.75 kWh	26%	164 kWh	13%

Transition strategy

Goal: minimize diesel $\rightarrow$ minimize emissions over planning horizon.

$$\min_{x_{p,\tau,y}} \sum_{p\in\{C,H\}} \sum_{\tau\in\mathcal{T}} \sum_{y\in\mathcal{Y}} \hat{D}_{p,\tau,y,d^*} x_{p,\tau,y} \tag{8}$$

where:

- p: powertrain (C: conventional diesel; H: hybrid; B: battery-electric)
- $x_{p,\tau,y}$: distance of trip τ assigned to bus of powertrain p in year y
- $\hat{D}$: predicted diesel consumption obtained from tractive energy model
- d^*: day with highest trip volume

subject to: investment caps, charging capacity, battery electric bus range, etc.

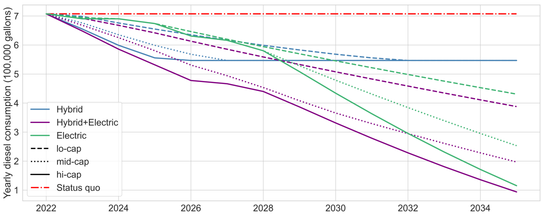
10 futures (scenario under given strategy) evaluated

			Scenario	
Strategy	**Description**	*lo-cap* \$10M	*mid-cap* \$20M	*hi-cap* \$30M
Status Quo	Maintain existing powertrain distribution	✓	-	-
Hybrid	All future bus purchases hybrids	✓	✓	✓
Hybrid+Electric	Future bus purchases hybrid or electric	✓	✓	✓
Electric	future bus purchases electric	✓	✓	✓

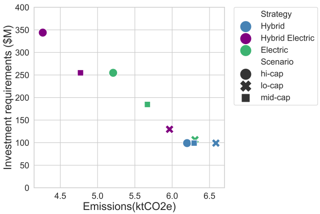
Results highlights

Mixed hybrid+electric strategy most effective for maximizing emissions reduction (-45%) in the medium term.

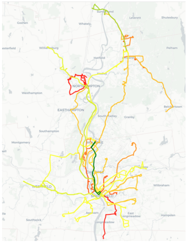
(a) Yearly diesel consumption of simulated futures



(b) Investment–emissions tradeoff

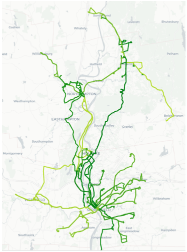


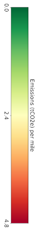
(c) Status Quo



(d) Hi-cap Hybrid+Electric

(e) Hi-cap Electric



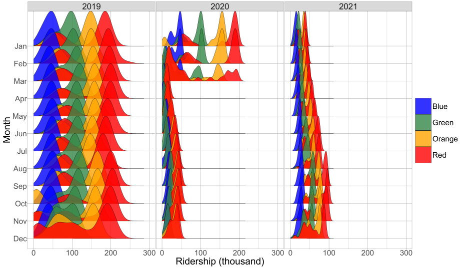


Next steps

- Use ML to predict vehicle energy for improved tracking and fleet assignment

Motivation

- Rail transit systems consume significant amounts of energy
  - Massachusetts Bay Transportation Authority (MBTA) serving Boston spends \$38M annually on electricity (422 GWh on average)
  - 2019 stats: 1.5M veh-hrs, 10.5M veh-mi, 150M riders
- How can agencies improve efficiency, reduce energy expenditure and respond to disruption sustainably?



Team

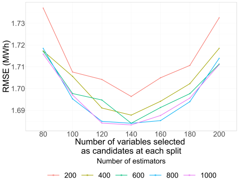
RA: Zhuo Han; Co-PIs: Dr. Eleni Christofa, Dr. Eric Gonzales

Sponsor: Mass DOT/MBTA, Sean Donaghy

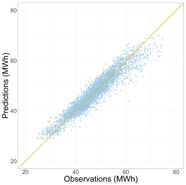
Key energy modeling results

- Trained ridge regression model ($R^2 = .91$) to predict system-level hourly energy consumption (used random forests for feature selection)[5]

(a) Random forests feature selection

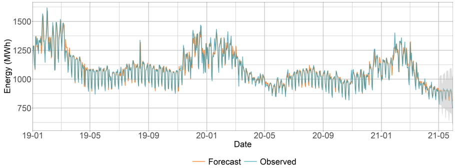


(b) Ridge regression performance

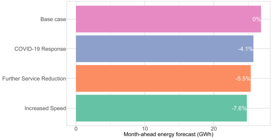


- Developed ARIMAX model for daily forecasts and limited scenario planning

(c) Daily forecasting model



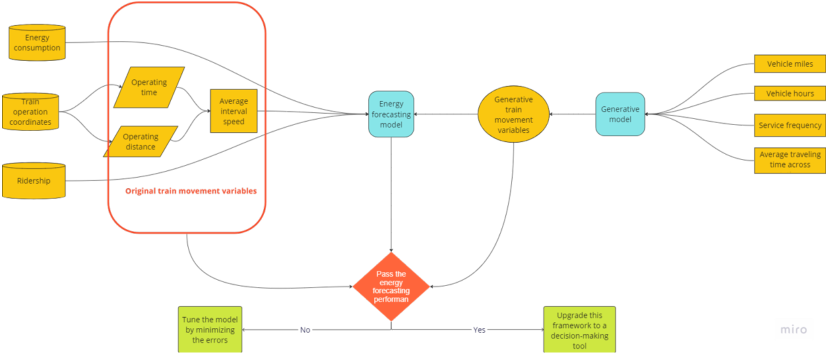
(d) Scenario outcomes



[5] Han et al. 2022; Han et al. 2023.

Decision-making framework

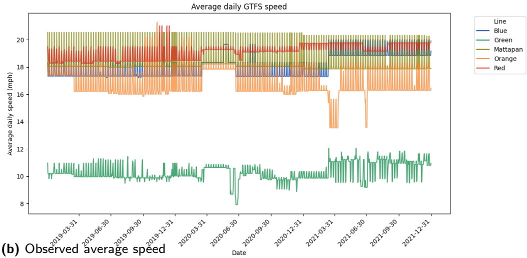
- Identify high-level planning metric (HPMs)
- Learn mapping to low-level movement variables
- Thus, we can generate sequences for specified plan (decision) and forecast energy accordingly



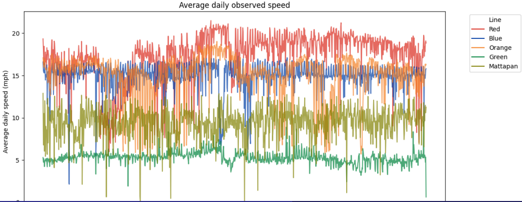
Planning metrics

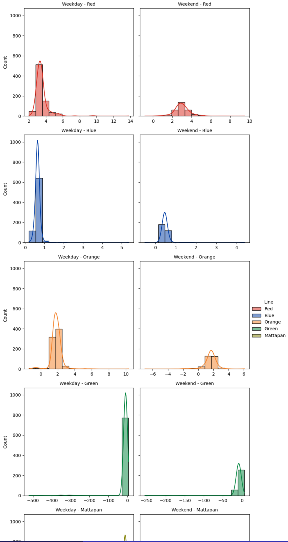
Three planning metrics: average speed, number of trips, trip distance
Goal: model noise and generate samples for training

(a) Planned average speed



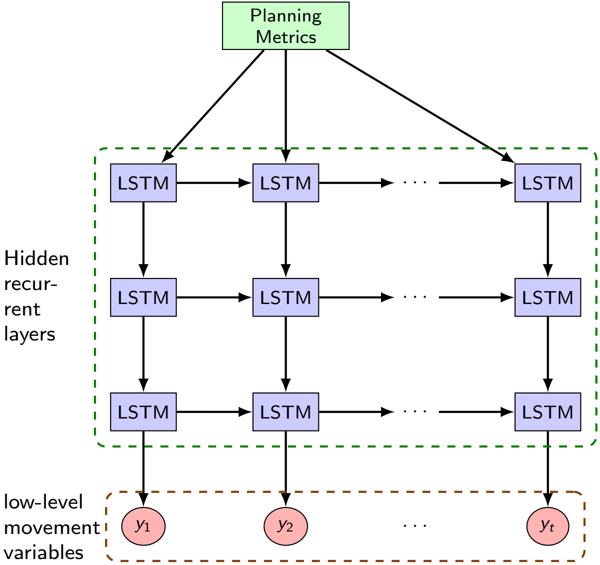
(b) Observed average speed





Map planning metrics to movement variables

- Train vector-to-sequence recurrent neural network to generate low-level trajectory variables for given day
- Use to model to generate planning scenario-specific trajectories
- Forecast energy



Summary

Supervised and unsupervised learning approaches can improve decision-making and provide insights for sustainable systems

- Bus transit network planning[6]
- Energy modeling and fleet transition optimization[7]
- Rail transit energy management and planning[8]
- Regional road network characterization

[6] Abdalazeem and J. Oke 2023.
[7] Arabi and J. Oke 2023; Arabi, T. Oke, et al. 2024.
[8] Han et al. 2022; Han et al. 2023.

Ongoing and future work

- Developing framework for regional inventory estimation and forecasting[9]
- Train ML models for vehicle-level energy/emissions prediction in bus transit networks
- Spatiotemporal multimodal crash typology[10]

Thank you! Questions?

[9] RA: Peiyao Zhao; Sponsors: Connecticut Councils of Government (EPA)

[10] RA: Mohammed Abdalazeem; Sponsor: New England UTC/US DOT

Bibliography I

Abdalazeem, M. and J. Oke (2023). "Extracting Spatiotemporal Bus Passenger Trip Typologies from Noisy Mobile Ticketing Boarding Data". In: *Data Science for Transportation* 5.3, p. 20. ISSN: 2948-1368. DOI: 10.1007/s42421-023-00082-x.

Arabi, M. and J. Oke (2023). "Comprehensive Energy Modeling Framework for Multi-Powertrain Bus Transit Systems". In: *Transportation Research Record*, p. 03611981231172502. ISSN: 0361-1981. DOI: 10.1177/03611981231172502.

Arabi, M., T. Oke, and J. Oke (2024). "Toward Zero-Emission Transit Futures: Energy Modeling and Optimization for Fleet Transition Planning". In: *Submitted to Transportation Research Part D: Transport and Environment*.

Bellis, M. (2021). "History of Roads". In: *ThoughtCo*.

Bellis, R., S. L. Davis, C. McCahill, E. Mangan, and B. Osborne (2020). *The Congestion Con*. Transportation for America.

Boeing, G. (2020). "A Multi-Scale Analysis of 27,000 Urban Street Networks: Every US City, Town, Urbanized Area, and Zillow Neighborhood". In: *Environment and Planning B: Urban Analytics and City Science* 47.4, pp. 590–608. ISSN: 2399-8083. DOI: 10.1177/2399808318784595.

Bruce, G. D. and R. E. Witt (1971). "Developing Empirically Derived City Typologies: An Application of Cluster Analysis". In: *The Sociological Quarterly* 12.2, pp. 238–246. ISSN: 0038-0253.

Han, Z., E. Gonzales, E. Christofa, and J. Oke (2022). "Modeling System-Wide Urban Rail Transit Energy Consumption: A Case Study of Boston". In: *Transportation Research Record* 2676.12, pp. 627–640. ISSN: 0361-1981. DOI: 10.1177/03611981221096442.

Bibliography II

Han, Z., E. Gonzales, E. Christofa, and J. Oke (2023). "Line-Specific Energy Modeling Framework for Urban Rail Transit Systems: A Case Study of Boston". In: *Transportation Research Record*, p. 03611981231170181. ISSN: 0361-1981. DOI: 10.1177/03611981231170181.

Harris, C. D. (1943). "A Functional Classification of Cities in the United States". In: *Geographical Review* 33.1, pp. 86–99. ISSN: 0016-7428. DOI: 10.2307/210620.

Hermosilla, T., J. Palomar-Vázquez, Á. Balaguer-Beser, J. Balsa-Barreiro, and L. A. Ruiz (2014). "Using Street Based Metrics to Characterize Urban Typologies". In: *Computers, Environment and Urban Systems* 44, pp. 68–79. ISSN: 0198-9715. DOI: 10.1016/j.compenvurbsys.2013.12.002.

Louf, R. and M. Barthelemy (2014). "A Typology of Street Patterns". In: *Journal of The Royal Society Interface* 11.101, p. 20140924. DOI: 10.1098/rsif.2014.0924.

Mohammed, M. and J. Oke (2023). "Origin-Destination Inference in Public Transportation Systems: A Comprehensive Review". In: *International Journal of Transportation Science and Technology* 12.1, pp. 315–328. ISSN: 2046-0430. DOI: 10.1016/j.ijtst.2022.03.002.

Mueen, A., N. Chavoshi, N. Abu-El-Rub, H. Hamooni, and A. Minnich (2016). "AWarp: Fast Warping Distance for Sparse Time Series". In: *2016 IEEE 16th International Conference on Data Mining (ICDM)*. 2016 IEEE 16th International Conference on Data Mining (ICDM), pp. 350–359. DOI: 10.1109/ICDM.2016.0046.

Oke, J. B., Y. M. Aboutaleb, A. Akkinepally, C. L. Azevedo, Y. Han, P. C. Zegras, J. Ferreira, and M. E. Ben-Akiva (2019). "A Novel Global Urban Typology Framework for Sustainable Mobility Futures". In: *Environmental Research Letters* 14.9, p. 095006. ISSN: 1748-9326. DOI: 10.1088/1748-9326/ab22c7.

Bibliography III

Raboteau, E. (2023). "How Our Roads Hurt Us and Everything Around Us". In: *The New York Times. Books*. ISSN: 0362-4331.

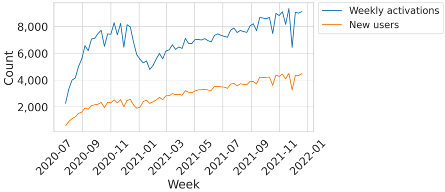
Rath, S. and J. Y. J. Chow (2022). "Worldwide City Transport Typology Prediction with Sentence-BERT Based Supervised Learning via Wikipedia". In: *Transportation Research Part C: Emerging Technologies* 139, p. 103661. ISSN: 0968-090X. DOI: 10.1016/j.trc.2022.103661.

US EPA, O. A. R. (2016). *Global Greenhouse Gas Emissions Data*. URL: https://www.epa.gov/ghgemissions/global-greenhouse-gas-emissions-data (visited on 03/19/2024).

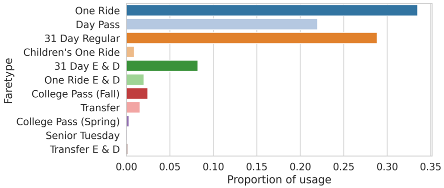
— (2017). *Inventory of U.S. Greenhouse Gas Emissions and Sinks*. URL: https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks (visited on 03/19/2024).

Pioneer valley mobile and ticketing data

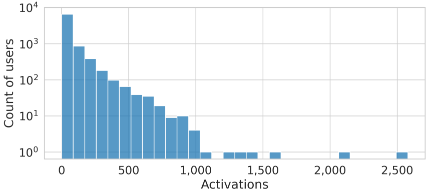
(a) Weekly ticket activations and new users



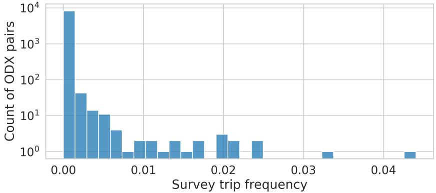
(b) Faretype usage by all



(c) Ticket activations per user



(d) Observed survey trip frequencies by zone



Dynamic Time Warping (DTW)

- DTW is used to compute pairwise distances between passenger trajectories, considering both longitude and latitude dimensions
- It accounts for temporal shifts, making it ideal for comparing spatiotemporal trajectories with varying lengths and time intervals

$$D_{ij} = \sum_{l=1}^{L} \frac{1}{C} \sum_{k=1}^{M} c_k \min_{\phi_k} d(\phi_k(\boldsymbol{y}_{n,l}, \boldsymbol{y}_{n',l})) \tag{9}$$

- D_{ij} represents the DTW distance between passengers n and n'
- C is the normalization constant to adjust for different sequence lengths, ensuring comparability
- c_k is the cost associated with each step in the path, with ϕ_k mapping the alignment between points in two sequences
- $d(\cdot)$ denotes the Euclidean distance, capturing the spatial discrepancy between points matched by ϕ_k

Hierarchical clustering

HAC is a bottom-up clustering method that builds a cluster hierarchy based on distance metrics. In our study, we use HAC with the Ward method to minimize the total within-cluster variance at each step, leading to more homogeneous clusters.

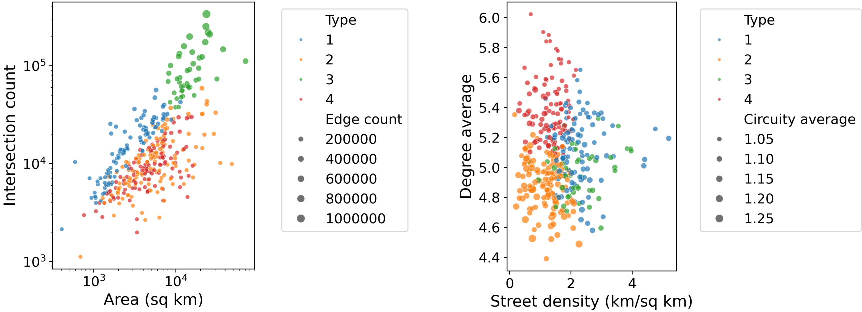
$$(\mathcal{C}, \mathcal{C}')^* = \arg\min_{\mathcal{C}, \mathcal{C}'} \frac{|\mathcal{C}||\mathcal{C}'|}{|\mathcal{C}| + |\mathcal{C}'|} \|\boldsymbol{c}_{\mathcal{C}} - \boldsymbol{c}_{\mathcal{C}'}\|^2 \tag{10}$$

- $\mathcal{C}, \mathcal{C}'$ are clusters being considered for merging.
- $|\mathcal{C}|$ and $|\mathcal{C}'|$ represent the sizes of the clusters, and $\boldsymbol{c}_{\mathcal{C}}, \boldsymbol{c}_{\mathcal{C}'}$ their centroids.
- The goal is to find the pair of clusters $\mathcal{C}, \mathcal{C}'$ that, when merged, result in the minimum increase in total within-cluster variance.

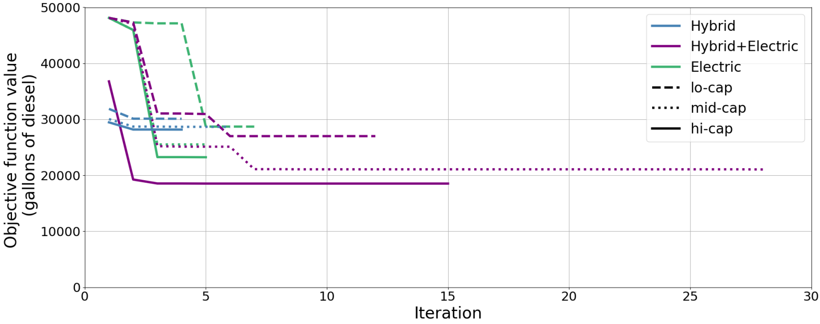
Through HAC, we obtain distinct passenger types by their spatiotemporal characteristics, facilitating targeted analysis and understanding of travel behavior across different user groups.

Relationships between network indicators

- Good separation among the types



Optimization convergence



Powertrain distributions

Figure: Powertrain distribution in each future

Related work: network typology

- Typologizing cities in terms of form and function is well established within urban planning and related fields[11]
  - These have been based on socioeconomic, geographic, and more recently, mobility indicators
- Network analyses have tended to be at the *intra*-urban scale[12]
- Notably, Boeing 2020 analyzed network indicators at various scales but did not derive a typology
- A recent global urban typology study and related work[13] included network indicators, but these were combined with other socioeconomic, mobility and sustainability indicators

[11] Harris 1943; Bruce and Witt 1971.
[12] Louf and Barthelemy 2014; Hermosilla et al. 2014.
[13] J. B. Oke et al. 2019; Rath and Chow 2022.