Incorporating User-Generated Waze Data and Machine Learning into Traffic Analysis: A Case-Study in Louisville, Kentucky

Matthew D. Harris

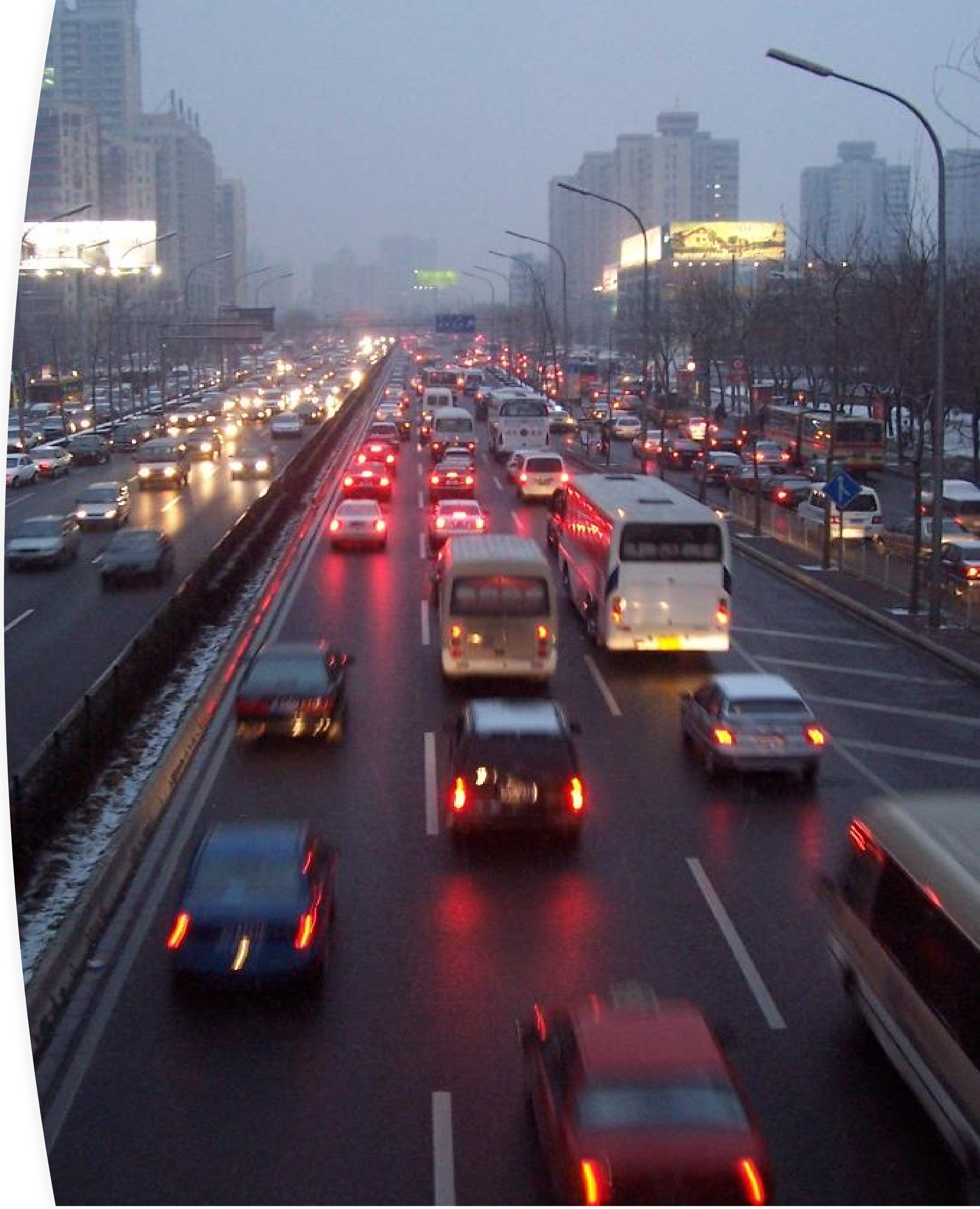
AECOM
Built to deliver a better world

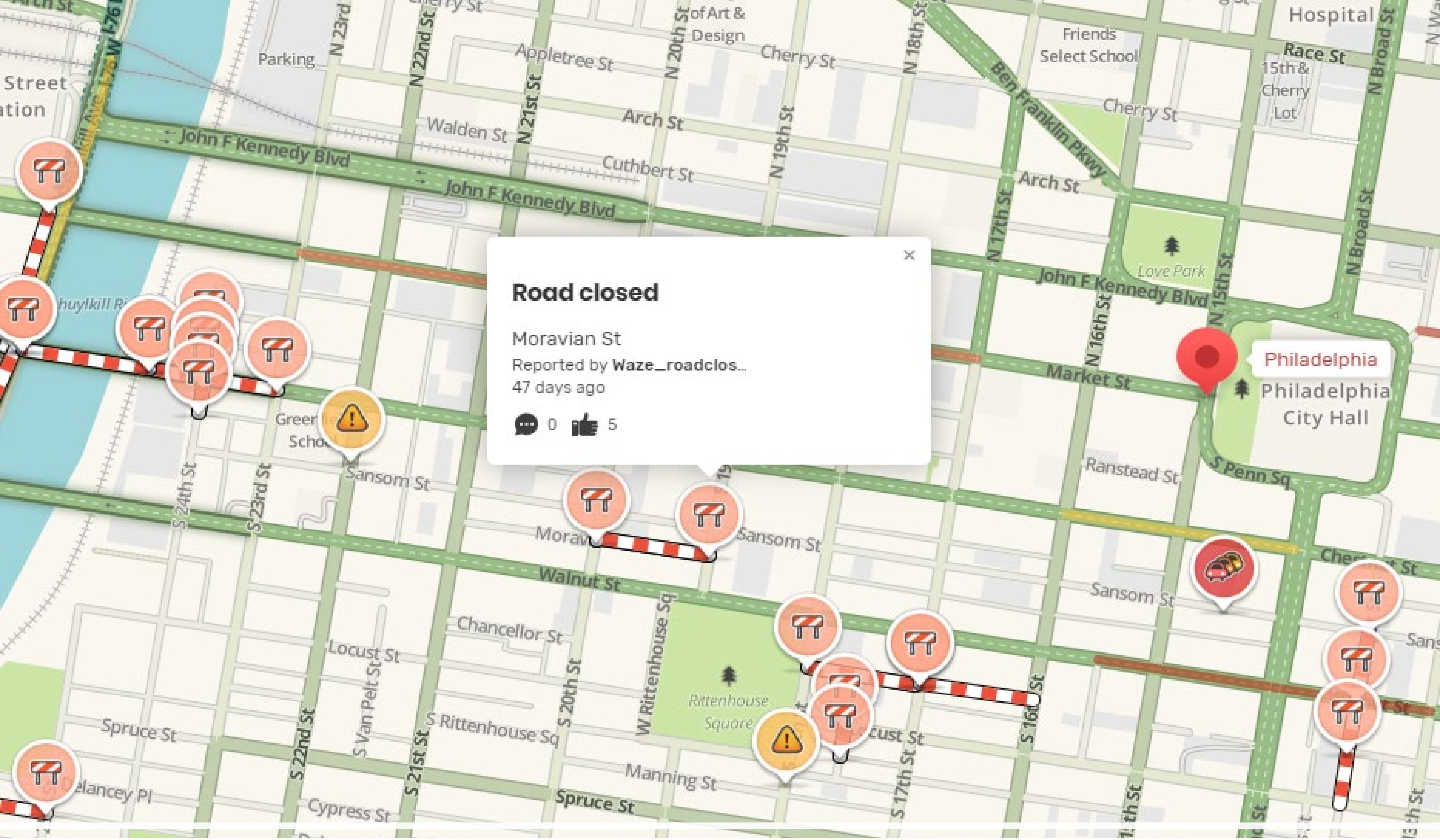
Weitzman School of Design
University of Pennsylvania

Let’s talk about…

1. Waze Warp - Louisville, KY OPI2
   - Cloud based Waze platform
   - Waze CCP program
   - Other similar analytics platforms

2. Use case – Louisville, KY
   - Penn MUSA Practicum
   - Congestion Prediction Model
   - Web-Application
   - Lessons Learned





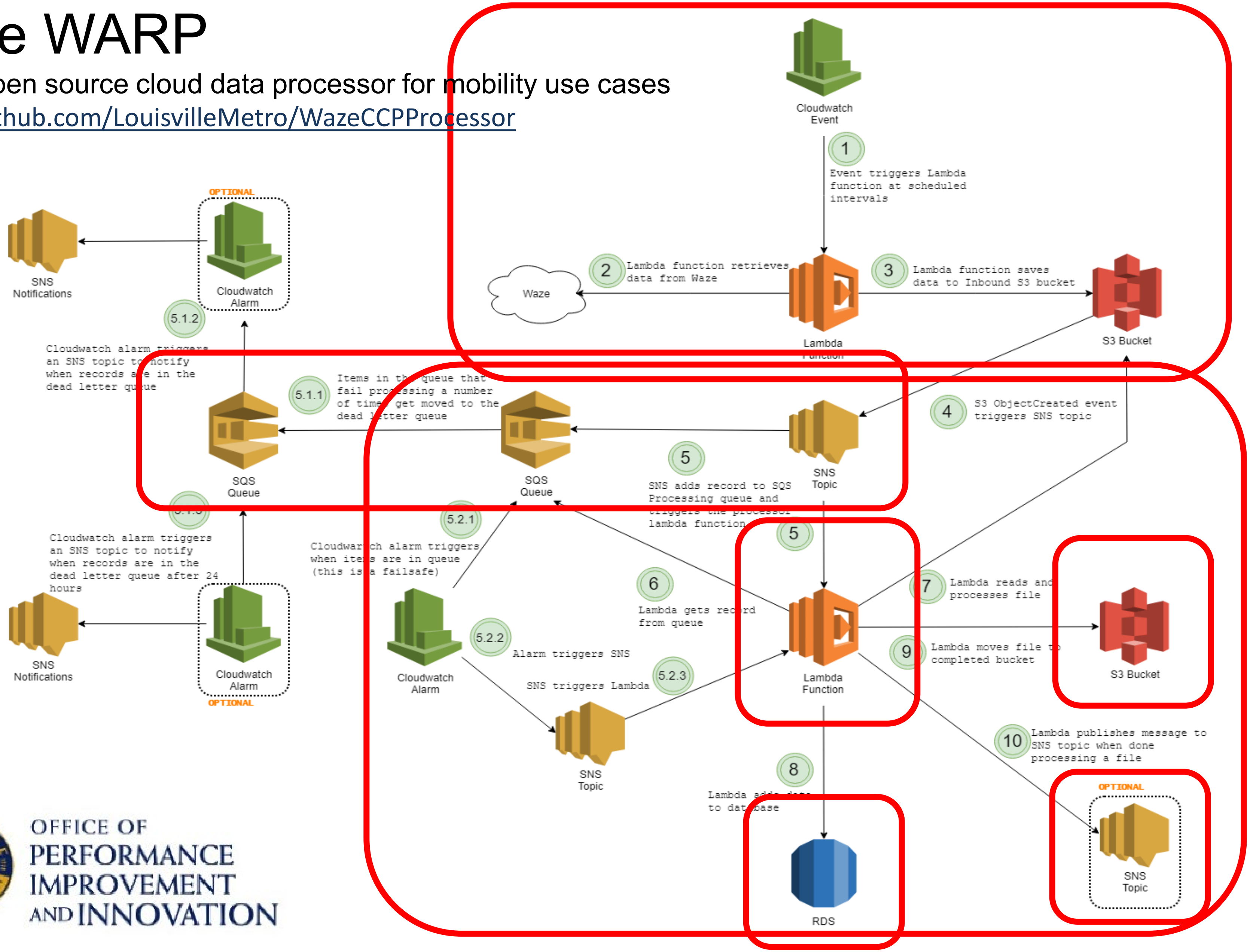
Waze Connected Citizens Program (CCP)

"What's happening and where?"

Waze WARP

A free, open source cloud data processor for mobility use cases

https://github.com/LouisvilleMetro/WazeCCPProcessor



OFFICE OF PERFORMANCE IMPROVEMENT AND INNOVATION

AECOM
Built to deliver a better world

Transportation Engineering and Safety Conference, December 12, 2019

Weitzman School of Design, University of Pennsylvania

Use Case: Louisville, KY

- Traffic Congestion Prediction
- Penn MUSA Team:
  - Sagari Datta
  - Dhruvi Kothari
  - Lufeng Lin
  - Andrew Renninger

https://pennmusa.github.io/MUSA_801.io/project_8/index.html

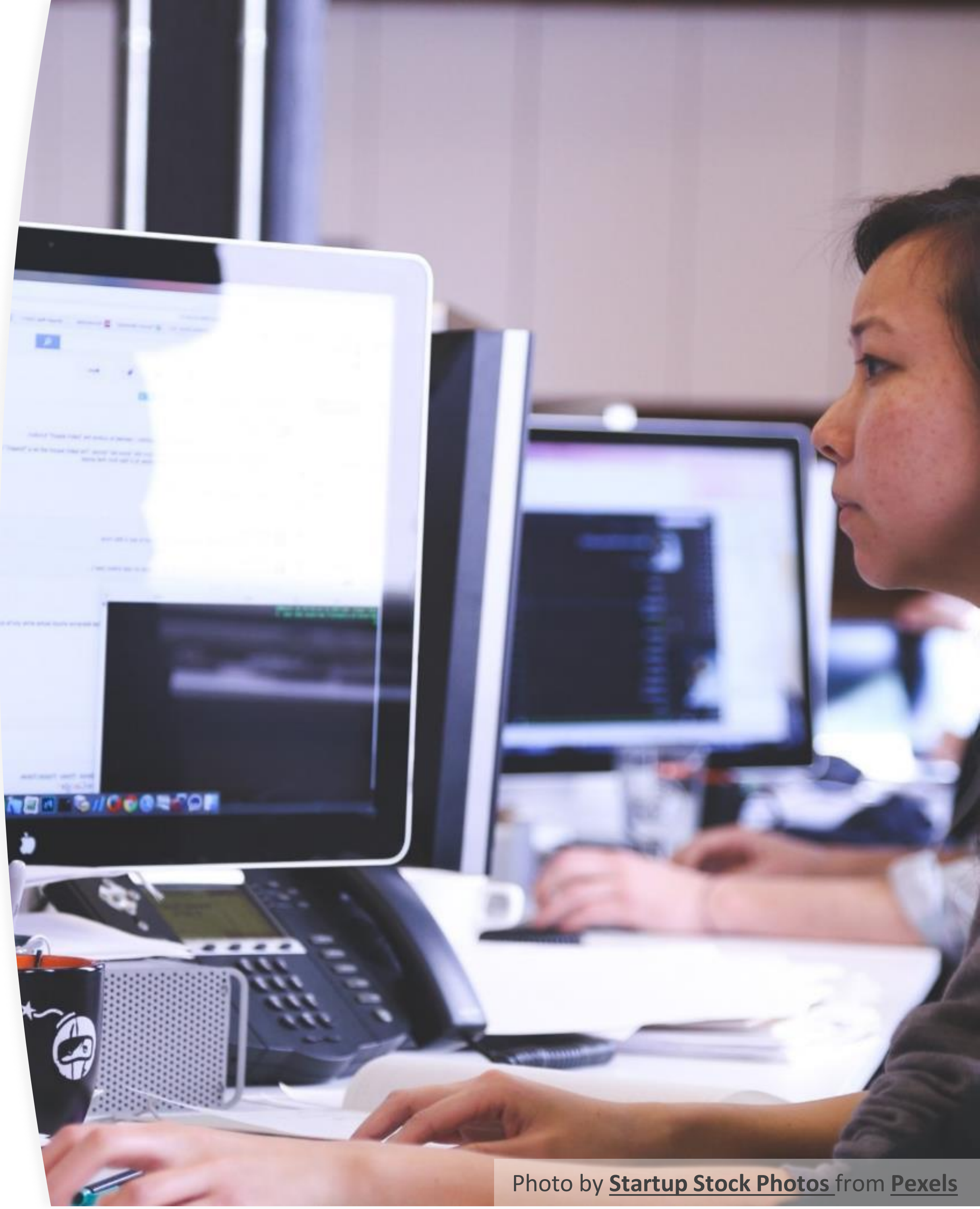


Photo by Startup Stock Photos from Pexels

Objective

Operations Planning - Predict city-wide average traffic intensity for any one-hour period under different conditions

Data

All 2018 Waze data for City; ~14.5 million reports, Weather, OpenStreetMaps, Holidays, and Build Environment. Covering 2,075 grid cells (500-meters)

Model

Mixed-Effects Linear model
Prophet – Bayesian Timeseries model
Mixed Random Forest Ensemble model

Outcome

Overall cross-validation error (MAE) of 1.71-meters

AECOM
Built to deliver a better world

Weitzman
SCHOOL OF DESIGN
UNIVERSITY of PENNSYLVANIA

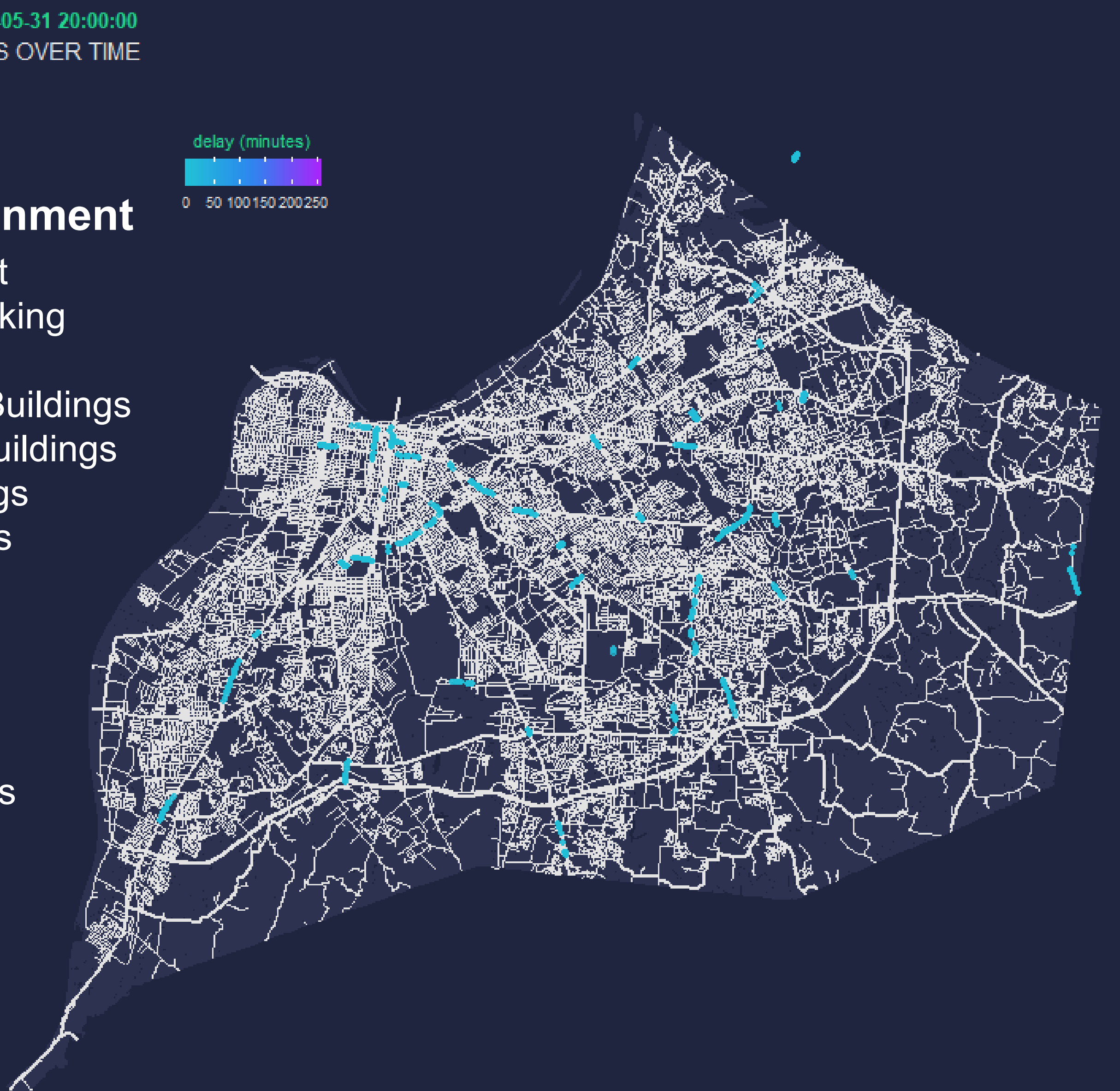
Model Features

Temporal
Local time
Hour
Peak
Day
Weekday
Weekend
Month
Year
Holiday

Built Environment
Parking Count
Off Street Parking
Incidents
Commercial Buildings
Residential Buildings
Retail Buildings
Total Buildings

Weather
Precip. Probability
Temperature
Humidity
Pressure
Wind Speed
Snow
Heavy Rain
Fog
Hurricane

Roadways
Freeway
Count of Turns
Roundabouts
Stop Signs
Crossways
Tolls
Traffic Signs
Intersections

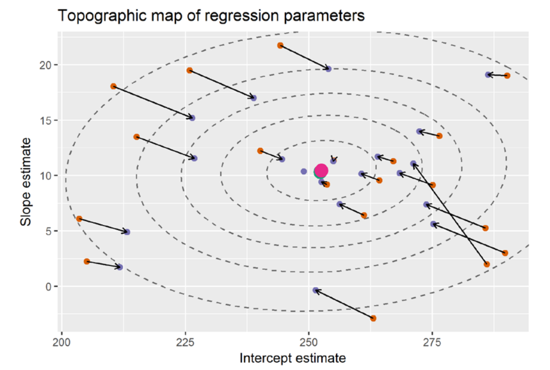


Modeling Approaches

Conditional on previous traffic, time, weather, and the build environment, what is the predicted length (meters) of a traffic jam for any given place and time?

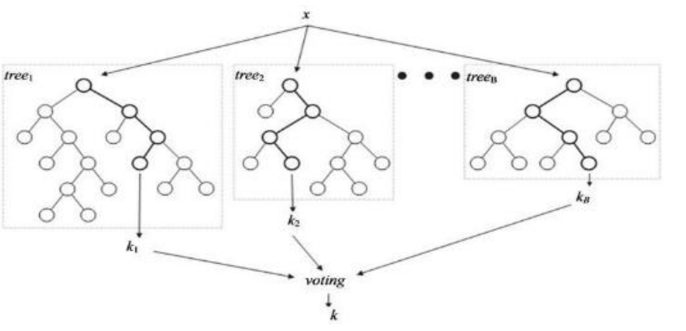
Mixed Effect Linear Model (lme4)

- Control for space and time
- Linear and additive functions
- Partial Pooling
- Specify error distribution



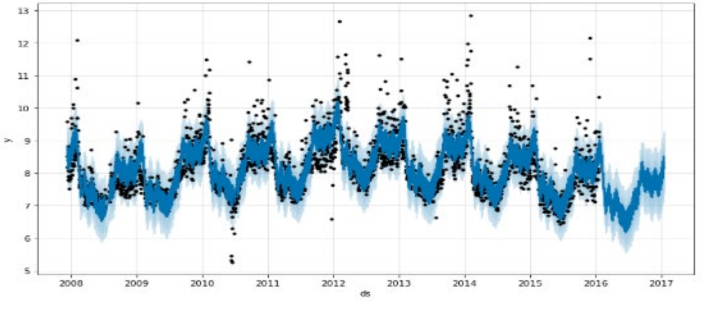
Mixed Random Forest (MixRF)

- Control for space and time
- Piecewise linear functions
- Lowers variance via ensemble



Bayesian Timeseries Forecasting (Prophet)

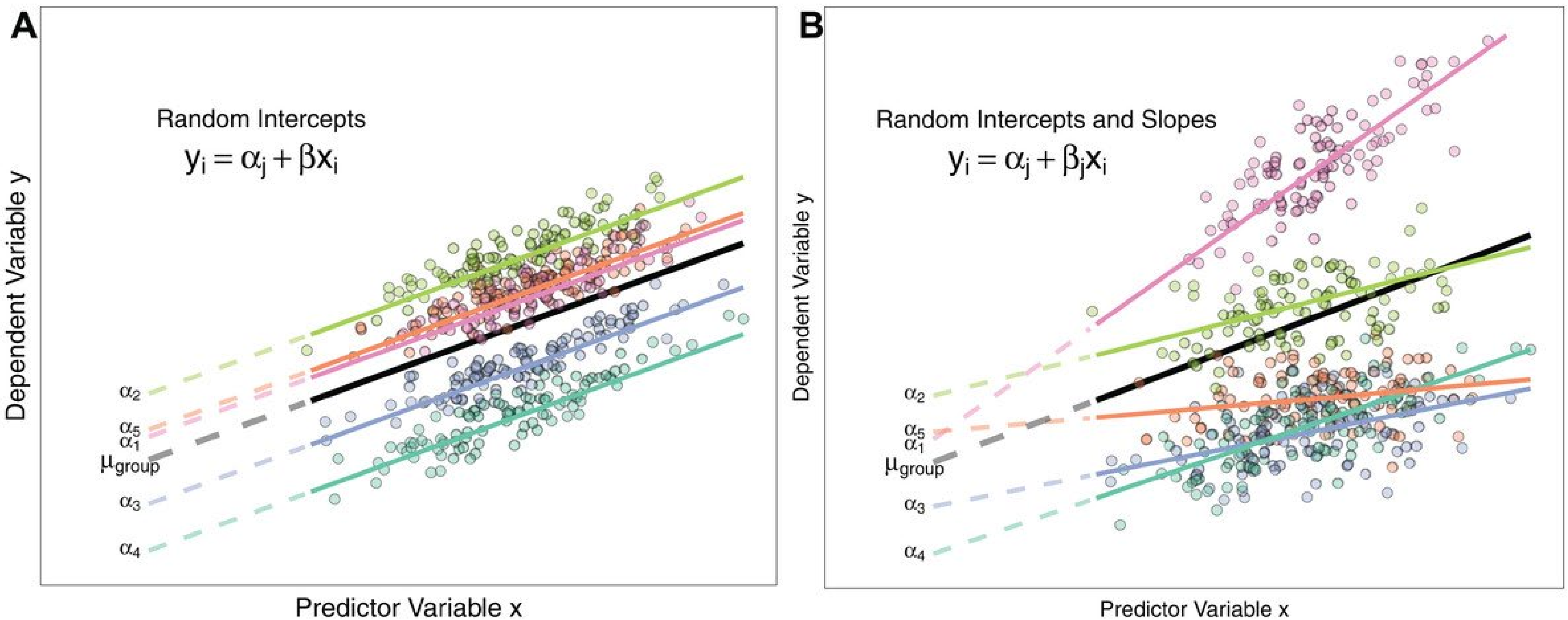
- Control for time
- Piecewise non-linear functions
- Robust to cycles and trends



Mixed Effects Model – Partial Pooling

- Each grouping is assumed to have separate, but correlated regression parameters
- Groups with lots of information share strength to groups lacking information
- All groups regress to the global mean

$$Jam\ Length_i \sim \alpha_{grid\_cell,\ freeway} + \beta_{hour,\ incidents} x_i + \epsilon$$



Graphic from: Harrison XA, Donaldson L, Correa-Cano ME, Evans J, Fisher DN, Goodwin CED, Robinson BS, Hodgson DJ, Inger R. 2018. A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ* 6:e4794https://doi.org/10.7717/peerj.4794

Prediction Errors – How does it generalize?

Conventional CV

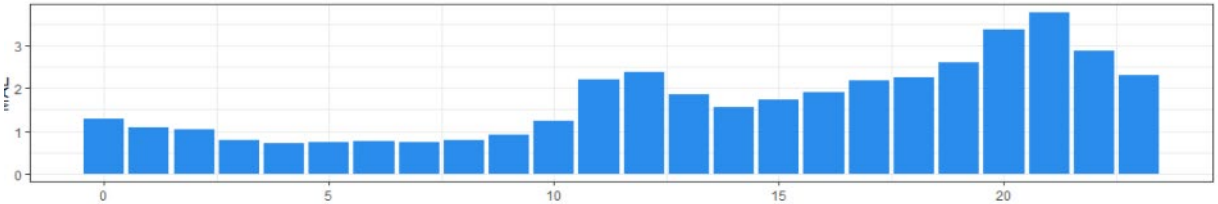
Data set	Mean Absolute Error
Training	1.70 meters
Test	1.71 meters

Spatial CV

Neighborhood	Mean Absolute Error
Central Business District	4.34 meters
Clifton	2.12 meters
Southside	0.97 meters

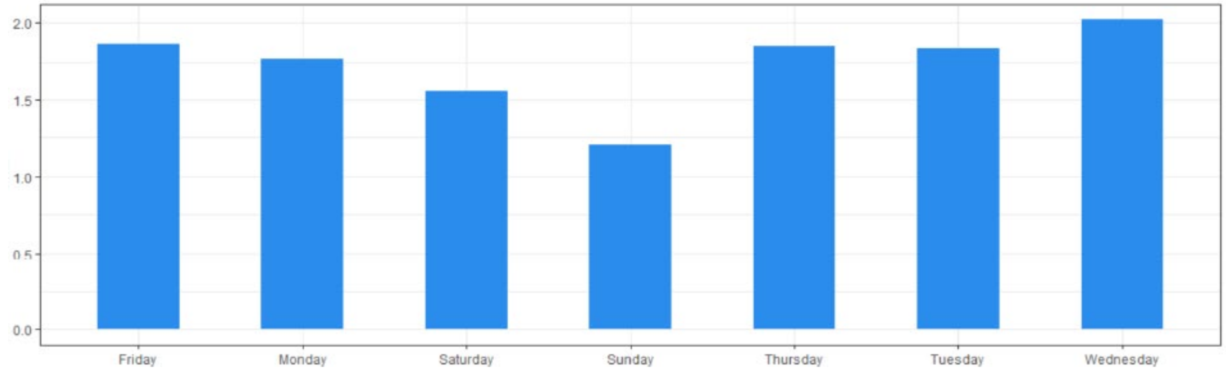
Temporal CV

MAE



Hour of the Day

MAE

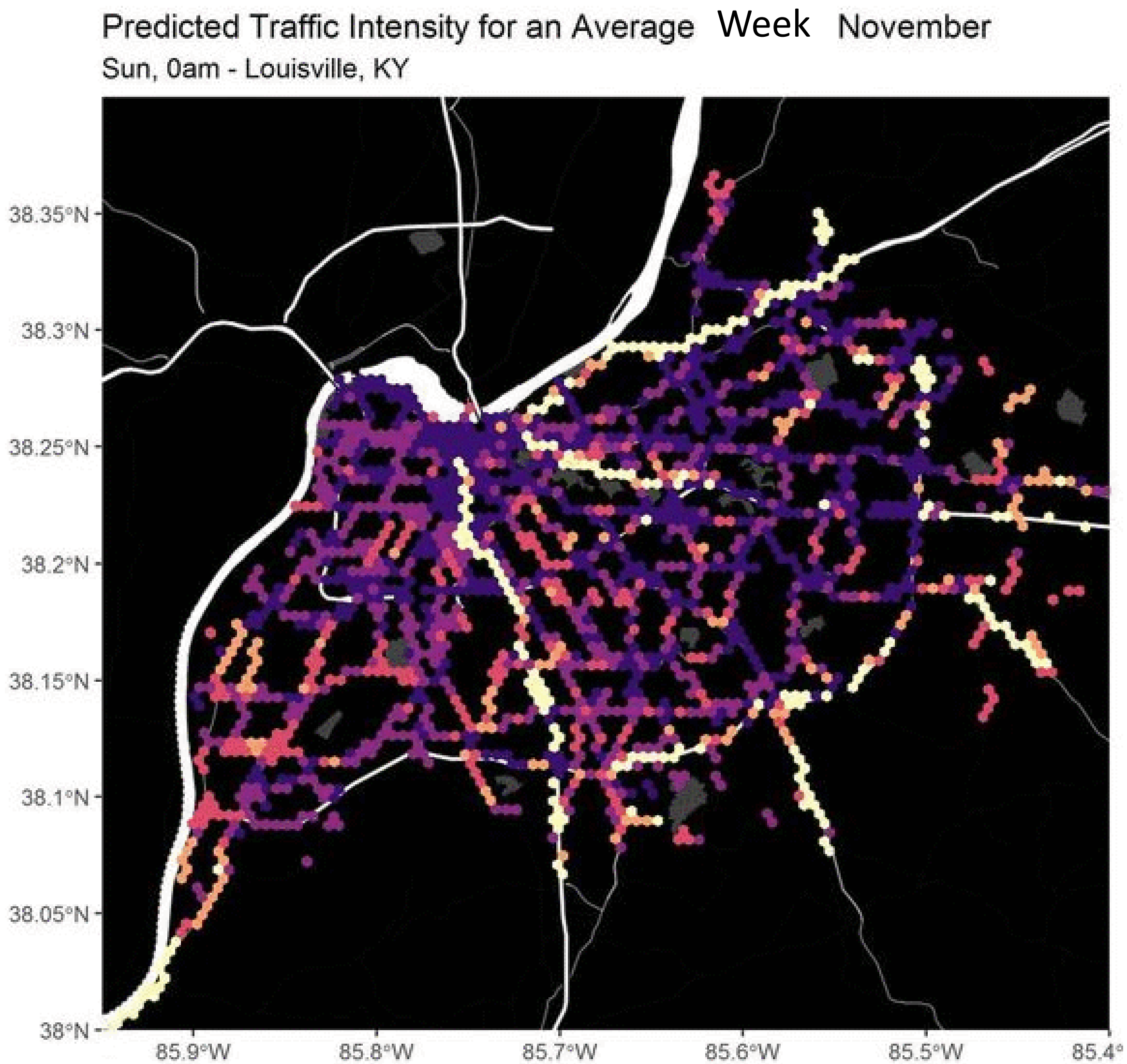


Day of the Week

AECOM
Built to deliver a better world

Weitzman
SCHOOL OF DESIGN
UNIVERSITY of PENNSYLVANIA

Predicting Average Under New Conditions



Opportunities and Challenges of this Model:

<u>Opportunities</u>:

- Methods to address large volume of user data
- Model Endogenous and Exogenous factors
- Address space and time correlation
- Measure errors across space and time
- New way to approach traditional problems

<u>Challenges</u>:

- Large data volume problematic in spatial joins and aggregation
- Need to include road network topology & spatially explicit relationships
- Dimensionality of model grows with each feature
- Need to connect to metrics for social and economic impacts of traffic

Operationalizing Model

https://msdakot.github.io/Congestion-Prediction-in-Louisville-KY/index.html#

Thank you!

Any question?

Matthew D. Harris
AECOM & Penn MUSA

matthew.d.harris@aecom.com

md_harris