NREL
NATIONAL RENEWABLE ENERGY LABORATORY

NREL is a national laboratory of the U. S. Department of Energy, Office of Energy Efficiency and Renewable Energy, operated by the Alliance for Sustainable Energy, LLC.

Network-Scale Ubiquitous Volume Estimation Using Tree-Based Ensemble Learning Methods

Authors: Yi Hou, Stanley Young, Anuj Dimri, Nicholas Cohn

January 7–11, 2018, Washington D.C. • Transportation Research Board Annual Meeting (18-01542)

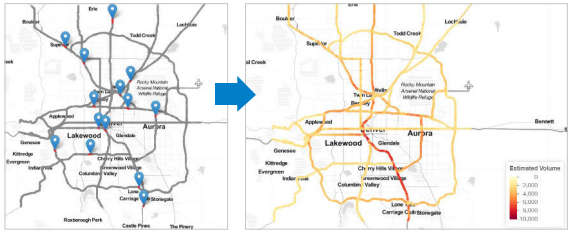
A PIONEERING PROJECT

The I-95 Corridor Coalition was the first to put forth the proposition of providing traffic volumes through outsourced probe data as part of a 2013 Multistate Corridor Operations and Management Program proposal, foreseeing that probe data will ultimately drive many of the operations and planning associated with business processes.

The University of Maryland and NREL recognize that the success of this project is critical to broader national initiatives that require quality data to operate and model the transportation system with the goal of optimizing it for safety, mobility, and energy efficiency.

GOAL

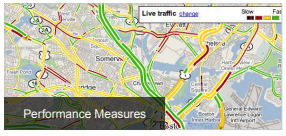
Ubiquitous network observability	Ideal but expensive to achieve
Best alternative	Utilize and fuse existing "sparse" data to predict traffic volumes on each and every link of the road network



WHY DO WE NEED UBIQUITOUS VOLUME?

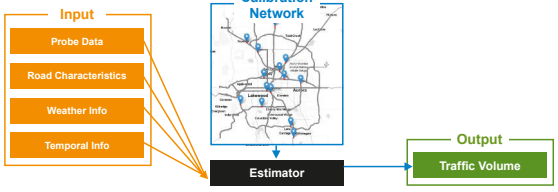

Planning


Operations


Performance Measures

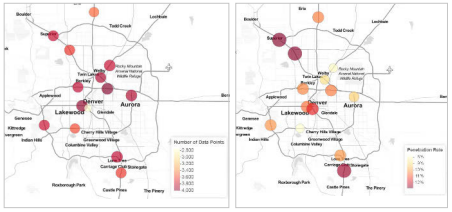
Forecasts (Special Events)

PROPOSED SOLUTIONS



DATA

- Data sources
  - Colorado Department of Transportation
  - TomTom
  - Federal Highway Administration's Travel Monitoring Analysis System
  - Weather Underground
- February 1, 2017 – April 30, 2017
- A total of 14 automated traffic controller locations with 52,092 data points



INPUT VARIABLES

- TomTom probe data: hourly average speed and probe count
- Weather: temperature, visibility, snow, and precipitation
- Road characteristics: road type, speed limit, number of lanes, and capacity
- Temporal information: month, day of week, hour of day

TREE-BASED ENSEMBLE LEARNING MODELS

- Random Forest (RF)
- Gradient Boost Machine (GBM)
- Extreme Gradient Boost (XGBoost)

MODEL VALIDATION

In each iteration, 13 stations are used for training and one station is used for validation. Repeat this 14 times and report validation results for all 14 locations.


1st iteration 2nd iteration 3rd iteration 4th iteration

MODEL EVALUATION CRITERIA

- Mean Absolute Percentage Error: $\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|V_i - \widehat{V_i}|}{V_i}$
- Error to Theoretical Capacity Ratio: $\text{ETCR} = \frac{1}{N}\sum_{i=1}^{N}\frac{|V_i - \widehat{V_i}|}{C_i}$
- Coefficient of Determination: $R^2 = 1 - \frac{(\widehat{V_i} - V_i)^2}{(V_i - \overline{V})^2}$

MODEL RESULTS

Model	Overall MAPE	Overall ETCR	Median R^2	Training Time
RF	17.8%	5.2%	0.92	73s
GBM	18.3%	4.8%	0.93	124s
XGBoost	17.7%	5.3%	0.91	13s

MODEL COMPARISON

Compare with Texas A&M Transportation Institute method
- MAPE: **~50%** reduction
- ETCR: **~30%** reduction
- R2 : **~10%** increase

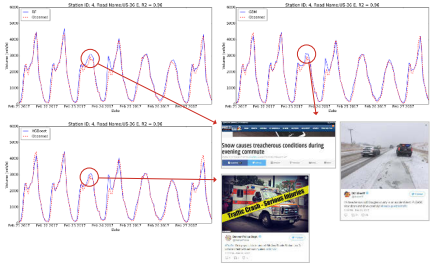
Compare with linear regression
- MAPE: **~60%** reduction
- ETCR: **~30%** reduction
- R2 : **~10%** increase

In the TTI method, 16 typical time-of-day traffic volume distributions are produced from 713 ATR stations in urban areas of 37 states based on four variables—functional class, weekday or weekend, congestion level, and dual directionality. Then, hourly volume is calculated from AADT supplied by HPMS.

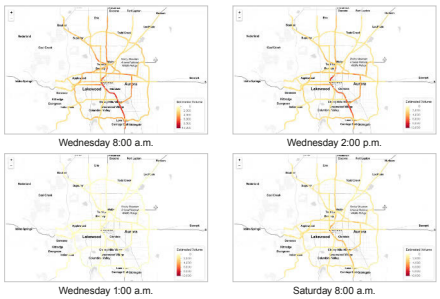
CONTRIBUTION OF PROBE DATA

	Overall MAPE	Overall ETCR	Median R^2
Without Probe Data	39.4%	12.4%	0.65
With Probe Data	**17.7%**	**5.3%**	**0.91**

ESTIMATIONS VS. OBSERVATIONS



ESTIMATED TRAFFIC FOR DIFFERENT TIME PERIODS


Wednesday 8:00 a.m. Wednesday 2:00 p.m. Wednesday 1:00 a.m. Saturday 8:00 a.m.

PROJECT VALUE

For many agencies, network-wide volume and turning movement data remain key missing dimensions for complete and actionable situational awareness, accurately assessing transportation system performance, and developing targeted, cost-effective mobility projects and programs. Having the ability to easily access and leverage these data (both real-time and historic) along with probe speed and travel time data offers these substantial benefits:

- Improve incident management monitoring and action
- Improve work zone monitoring, impact analysis, and safety
- Anticipate and verify "jam" conditions
- More accurate user-delay cost reporting for weather, sports, or other events
- Comprehensive special event management with the ability to monitor roadway utilization in the event of an emergency evacuation
- Improve traffic signal system timing management, enabling more cost-effective, timely, and accurate updates to signal timing plans
- More complete after-action reviews
- Better problem identification, root cause analysis, and project development
- More accurate system performance evaluation
- Enhance project/program assessment
- Improve travel demand modeling accuracy
- Better address air quality and emissions requirements, enabling more accurate transportation energy analysis

ACKNOWLEDGEMENTS

This work was funded by the I-95 Corridor Coalition.

Transportation Research Board 97th Annual Meeting
January 7-11, 2018, Washington, D.C.
NREL/PO-5400-70733