Spatial Transferability of Machine Learning based Volume Estimation Models

Kevin Kasundra, Venu Garikapati, Chris Hoehne, Yi Hou, and Stan Young

National Renewable Energy Laboratory, 15013 Denver West Parkway, Golden, Colorado 80401

TRBAM-21-02017

TRB 2021 Annual Meeting

January 2021

INTRODUCTION

Motivation

Recent advancements in probe-based vehicle data could provide a cost-effective solution to increase the observability of traffic volumes. There are numerous potential benefits to ubiquitous traffic volumes:

- Operation
  - Detect real-time traffic volume in the network
  - Detect during bad weather and special events
- Performance measures
  - Assess user costs
  - Utilization of existing capacity
- Economic and energy assessment
  - Estimate economic impact of congestion
  - Quantify vehicle miles travelled and energy use

Project Background

- Probe-based data offers a cost-effective solution for increasing the observability of traffic volumes.
- Ample ground truth traffic data is a prerequisite for developing robust volume estimation models.
- States with scarce continuous count station data could benefit from robust volume estimation models developed in (adjacent) data-rich states.
- This paper explores spatial transferability of volume estimation models from three states (Colorado, North Carolina, and Pennsylvania).

Key Research Question

Are probe-based volume estimation models spatially transferable? If so, which factors ensure a good spatial transfer? What are the critical pitfalls to look for while spatially transferring volume estimation models?

Data

Probe data is combined with other data sources for use in predicting underlying spatiotemporal trends in vehicle traffic.

- Continuous-count station traffic volume data (DOTs)
  - Hourly volume, road class, number of lanes
- Weather Underground
  - Temperature, precipitation, visibility, fog, rain, snow
- TomTom GPS Data *(1)*
  - Probe counts (key ingredient) and average speed
- Temporal information
  - Month, day of week, hour of day

METHODOLOGY

Model Specification

Average hourly volumes are modeled as a function of commercial probe vehicle data, weather information, and roadway characteristics (**Figure 1**). Tree-based ensemble learning method Extreme Boost Machine (XGBoost) is chosen for its established quick and accurate performance *(2)*:

$$F_K(x) = \sum_{k=1}^{K} h(x; a_k),$$

where h(·) is a tree model and a_k is set of the parameters for the k^{th} tree model. In each iteration, XGBoost is trained in a forward stagewise manner to minimize the loss function specified as:

$$\hat{a}_k = \arg\min_{a_k} \sum_{i=1}^{N} L(y_i, F_{k-1}(x_i) + h(x_i; a_k)) + \Omega(h_k),$$

where L(·) is the loss function, N is the number of training data points, y is the true value, and $\Omega(h_k)$ is the regularization term.

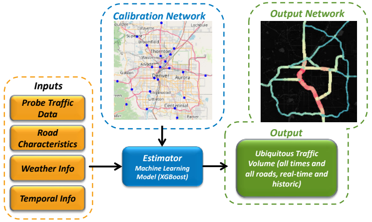

Figure 1. Modeling framework.

Model Training and Validation

The models are trained and validated using the cross-validation technique; in each iteration, all but one continuous count station is used for training. The held-out station is used for validation. This is repeated for N stations.

Baseline Volume Estimation Models

Table 1. Baseline Model Results.

State	R^2	MAE (veh/hr)	EMFR (%)
Colorado	0.91	357	5.3
North Carolina	0.92	257	6.4
Pennsylvania	0.91	194	5.7

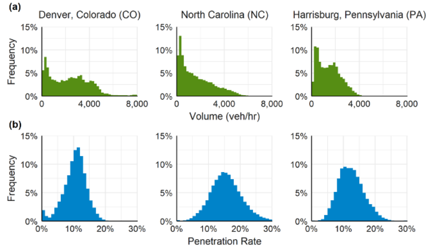

Figure 2. Distribution of (a) hourly road volumes and (b) penetration rates.

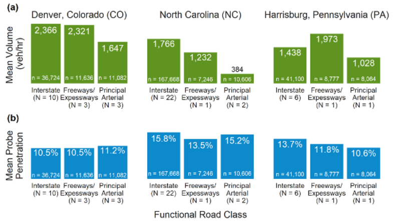

Figure 3. Mean volume (a) and mean probe penetration rate (b) by road class.

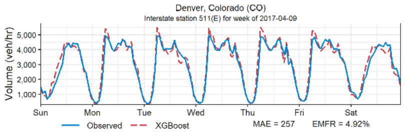

Figure 4. Observed vs. predicted volumes in the Denver baseline model.

SPATIAL TRANSFERABILITY

Region-to-Region

Spatial transferability exercises were carried out by applying the model estimated in one state to predict hourly traffic volumes (on the same type of roadway segments) in a different state. The model performance between NC and PA spatial transfer is much better than that of CO-NC spatial transfer. Upper bounds of observed traffic volumes was the most impactful factor in poor transferability model performances.

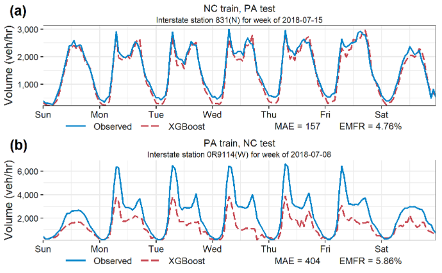

Figure 5. Predicted versus observed hourly volumes for a typical week for (a) the model trained on NC and tested on PA (thus displayed are PA predicted volumes), and (b) the model trained on PA and tested on NC (thus displayed are NC predicted volumes).

Meta Model

A meta model was constructed and evaluated to see if this approach could bolster the shortcomings (e.g. bounding issues in XGBoost, differing local probe penetrations) from single region transferability models.

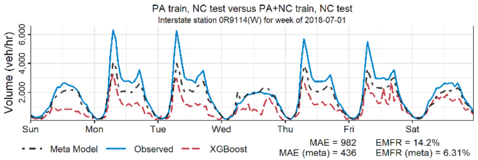

Figure 7. Predicted versus observed hourly volumes for a typical week for the model trained on PA and tested on NC (thus displayed are NC predicted volumes) with the meta model (PA+NC train) and test on NC. The displayed station and week are the same as Figure 5b.

Table 2. Results from spatial transferability exercises.

Type of Model	Train → Test	R^2	MAE (veh/hr)	EMFR (%)
CO ⇔ NC Spatial Transferability	CO → NC	0.71	577	15.6
	NC → CO	0.67	704	13.6
NC ⇔ PA Spatial Transferability	NC → PA	0.89	217	6.1
	PA → NC	0.79	403	10.3
NC + PA Meta Model	**NC + PA**	**0.91**	**266**	**5.5**

EMFR: Error to Maximum Flow Ratio **MAE**: Mean Absolute Error

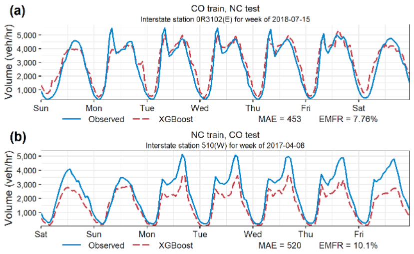

Figure 6. Predicted versus observed hourly volumes for a typical week for (a) the model trained on CO and tested on NC (thus displayed are NC predicted volumes), and (b) the model trained on NC and tested on CO (thus displayed are CO predicted volumes).

CONCLUSIONS & FUTURE WORK

- When transferring a volume estimation model from one region to another, it is extremely important to have temporal consistency.
- Transferring a model between states with similar weather and traffic characteristics will result in a better spatial transfer.
- The meta-model was observed to have the best model performance and is shown to overcome the bounding issue identified with spatial transfers.
- Future will consider focus on identifying state typologies based on traffic volume patterns, weather information, and roadway characteristics to build meta-models.

REFRENCES

1. TomTom. Real-Time Traffic. https://www.tomtom.com/products/real-time-traffic/.
2. Chen, T., and C. Guestrin. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Vol. 13, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy Efficiency and Renewable Energy Vehicle Technologies Office. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

NREL/PO-5400-78829

christopher.hoehne@nrel.gov