Estimating Highway Volumes Using Vehicle Probe Data – Proof of Concept

Preprint

Stanley E. Young and Yi Hou
National Renewable Energy Laboratory

Kaveh Sadabadi and Przemysław Sekuła
University of Maryland

Denise Markow
I95 Corridor Coalition

Presented at the ITS World Congress 2017
Montreal, Canada
October 29 – November 2, 2017

Suggested Citation
Young, Stanley E., Kaveh Sadabadi, Przemysław Sekuła, Yi Hou, and Denise Markow. 2018. "Estimating Highway Volumes Using Vehicle Probe Data – Proof of Concept: Preprint." Golden, CO: National Renewable Energy Laboratory. NREL/CP-5400-70938. URL.

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

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Conference Paper
NREL/CP-5400-70938
March 2018

Contract No. DE-AC36-08GO28308

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Available electronically at SciTech Connect http:/www.osti.gov/scitech

Available for a processing fee to U.S. Department of Energy and its contractors, in paper, from:

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
OSTI http://www.osti.gov
Phone: 865.576.8401
Fax: 865.576.5728
Email: reports@osti.gov

Available for sale to the public, in paper, from:

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Road
Alexandria, VA 22312
NTIS http://www.ntis.gov
Phone: 800.553.6847 or 703.605.6000
Fax: 703.605.6900
Email: orders@ntis.gov

Cover Photos by Dennis Schroeder: (left to right) NREL 26173, NREL 18302, NREL 19758, NREL 29642, NREL 19795.

NREL prints on paper that contains recycled content.

Paper ID # AM-SP 1300

Estimating Highway Volumes Using Vehicle Probe Data – Proof of Concept

Stanley E. Young[1], Kaveh Sadabadi[2], Przemysław Sekuła[2], Yi Hou[1*], Denise Markow[3]
1. National Renewable Energy Laboratory, 1530 Denver West Pkwy, Golden, CO, USA.
Tel: +1(303)384-7525 Email: Yi.Hou@nrel.gov
2. University of Maryland
3. I95 Corridor Coalition

Abstract
This paper examines the feasibility of using sampled commercial probe data in combination with validated continuous counter data to accurately estimate vehicle volume across the entire roadway network, for any hour during the year. Currently either real time or archive volume data for roadways at specific times are extremely sparse. Most volume data are average annual daily traffic (AADT) measures derived from the Highway Performance Monitoring System (HPMS). Although methods to factor the AADT to hourly averages for typical day of week exist, actual volume data is limited to a sparse collection of locations in which volumes are continuously recorded. This paper explores the use of commercial probe data to generate accurate volume measures that span the highway network providing ubiquitous coverage in space, and specific point-in-time measures for a specific date and time. The paper examines the need for the data, fundamental accuracy limitations based on a basic statistical model that takes into account the sampling nature of probe data, and early results from a proof of concept exercise revealing the potential of probe type data calibrated with public continuous count data to meet end user expectations in terms of accuracy of volume estimates.

KEYWORDS:
Volume estimation, Probe data, Neural network

Introduction
Real-time volume data remains the key missing dimension in operations data that would greatly improve the accuracy of assessing transportation system performance. Although agencies have invested in fixed sensors, volume data remains relatively sparse and of varying quality on the majority of the freeway and major arterial networks. Anticipated volume, a by-product from factoring of Highway Performance Monitoring System (HPMS) data, is currently the state-of-the-practice in assessing performance measures involving user cost, emissions, and energy efficiency network wide. However, anticipated volume does not reflect traffic volume fluctuations during weather events, major incidents, or even normal day to day fluctuations in traffic. Quality volume data is required to effectively assess user costs, assess extent of delay and congestion, detect real-time perturbations to the network, understand traffic density during major weather events such as blizzards and hurricanes, and to estimate congestion impact in terms of travel time and delay and their corresponding economic, environmental, and energy impacts. The cost of collecting accurate high-quality volume data with traditional infrastructure sensors would be prohibitively expensive. It requires sensors to be installed approximately every 1 to 5 miles on transportation networks. If successful, this method provides a cost-effective means to leverage existing continuous count stations with available commercial probe data for network-wide, 24x7x365 accurate volume estimates.

The I-95 Corridor Coalition put forth the proposition of providing traffic volumes through outsourced probe data as part of a 2013 Multistate Corridor Operations and Management Program (MCOMP) proposal, foreseeing that probe data will ultimately drive many of the operations and planning business processes. Probe data, with its ever evolving capability as more and more vehicles act as probe vehicles, is poised to deliver such information, and vendors are currently considering the technical challenges of providing volume flow information within the quality expectations of the industry. Since 2013 when the Coalition introduced the idea, other jurisdictions have inquired as to the viability of such a product, and some vendors have already taken internal efforts to explore product feasibility.

The University of Maryland Center for Advanced Transportation Technology (UMD CATT) in partnership with the National Renewable Energy Laboratory (NREL), and under the sponsorship of the I95 Corridor Coalition, are testing the feasibility of volume and turning movement information derived from probe sources to effectively meet the need of I-95 Coalition members (transportation agencies and their partners) to enhance planning, traveler information, operational activities and performance measures.

This paper presents: (1) a review of previous literature; (2) the necessity and use cases of broad-based volume; (3) the survey results of key personnel within the I-95 Corridor Coalition regarding the anticipated uses, criticality, and needed formats of such data; (4) anticipated accuracy limits based on a statistical model; (5) results from a proof of concept exercise to estimate volumes using archive probe data and machine learning techniques.

Literature Review

AASHTO regards the average daily traffic (ADT) volume as the most basic measure of traffic demand [1]. However, this measure does not reflect traffic volume variations during various months of the year, days of the week, and hour of the day. Different profile methods are proposed to breakdown the ADT to (sub)-hourly traffic (count) volume estimates. A method proposed by the Texas A&M Transportation Institute (TTI) [2] is widely used in practice to obtain 15 minute traffic count estimates during a typical week of the year. This method is based on a set of representative national level profiles and in part takes advantage of probe speed data to determine which profile should be applied to distribute a given segment's ADT.

Vlahogianni et al. [3] provide a comprehensive review of existing short-term traffic prediction methods. They categorize the documented approaches into time series analysis, function approximation, optimization, pattern recognition and clustering. Three of these methodological classes (function approximation, pattern recognition, and clustering) are in essence members of machine learning family of methods. Machine learning applications are very popular among the research community; at the same time, these methods are finding their way into data products offered by traffic data vendors.

Among the most notable machine learning techniques used in traffic estimation problems are linear regression, artificial neural networks (ANN) and support vector regression (SVR). Lin [4] provides a discussion of pros and cons of these techniques. Castro et al. [5] reports on the use of an SVR technique to predict AADT values.

Regression techniques are among the commonly reported machine learning techniques. For example, Zhao and Chung [6] employed linear regression and GIS technology, which allowed them to develop a variety of land use and accessibility measurements. Kingan and Westhuis [7] used robust regression techniques to remove outliers and forecast traffic growth factors.

Among the more recent works in this area Lv et al. [8], and Polson and Sokolov [9]—both report using ANNs. The former is interesting because they report using autoencoders and unsupervised learning. The latter however describes using a fully connected neural network with hyperbolical tangent activation function which also employs a dropout scheme.

Finally, while other works mainly focused on short-term traffic prediction, Islam [10] is one of the few works on AADT estimation using artificial intelligence. She focuses on ANN and SVR as two main methods to estimate AADT and uses road geometry, existing counts and local socio-economic data as inputs.

This paper is unique in taking advantage of hourly probe counts (among other inputs) and adopting a machine learning approach to estimate volumes in locations with no count stations based on training a model on locations and times where real-world measurements exist.

Section 1.0: Network Wide Volume Data Use Case Scenarios

The following are a collection of scenarios in which the availability of volume data may enable or enhance highway management operations. These were obtained from interaction with traffic professionals through the I95 Corridor Coalition.

Use case 1: Real-time volume flow during major events and incidents

Current volume information is available primarily in two forms. Volume sensors (in the forms of loops or radar) provide vehicle counts and occupancy at select locations, but are typically limited due to high cost of installation and maintenance. Although HPMS data provides sample counts to estimate hourly weekday and weekend volumes under normal demand, major events can dramatically alter usage patterns. Hurricanes, major snow storms, political events such as the inauguration, and major incidents are all examples of possible scenarios on the eastern portion of the United States that have occurred, and can occur in which anticipated volumes greatly differ from actual roadway volume/demand. During such events, anticipated volumes from HPMS have little value, and the density of most real-time traffic counters is too sparse to provide a comprehensive view of traffic flow. Whereas existing traffic monitoring systems based on commercial probe data can provide speeds and travel times, the demand characteristics on the network are absent.

Real-time flow information in terms of volume on a per road segment basis would greatly benefit situational awareness. For example, during major snow events in the Washington, D.C. metropolitan area the Federal Government has the latitude to dismiss government workers earlier, and has done so on occasion. On such occasions the prevailing evening rush hour patterns are completely disrupted, typically moving rush hour earlier in the day. Broad-based flow data on various facilities during such situations would provide more objective understanding of the impact during such events. Also, volumes could also be combined with incident information to better determine roadway incident rates to help identify high-incident segments/corridors. Ideally, as greater understanding is gained of these unusual off-peak traffic patterns, this data might also be used to support the Government's decisions regarding such early releases.

Use case 2: Monitoring road closure and clearance of major events

Probe-based traffic monitoring has succeeded in recent years to provide comprehensive speed and travel time situational awareness for urban networks, as well as rural highways. Low-volume conditions remain problematic. During events or weather that diminish the volume, it remains problematic to determine from probed based data if the roadway is closed, extremely slow, or simply absent of demand (volume of vehicles). Augmenting real-time speed and travel time with flow (volume or demand) estimates would provide the additional dimension to better resolve such situations.

One such example is determining clearance of rural incidents. Often the onset of an incident is reported to traffic management centers via normal emergency reporting. However, the clearance of incidents in rural portions of the network remains problematic. The corresponding 511 systems may inadvertently leave incorrect incident data long after the incident has been cleared. Volume information, integrated with travel time feeds would provide sufficient data to know when the road has been cleared and the road re-opened to traffic.

Applications are not limited to rural areas and could also apply to urban areas and/or any time when volume data are helpful to identify when traffic has returned to "normal" (historically normal) conditions. The time to "return to normal" conditions is an incident management performance measure that has been discussed but never put into practice due to the difficulty of determining normal flow based solely on speed or travel time data. Recent attempts to only use speed/travel time data to determine winter weather road clearance times suffer from similar issues [11]. Speed alone is often insufficient to conclusively determine return to normal conditions without accompanying volume data.

Use case 3: Accurate Operations Performance Measures

Current user costs for congestion (e.g., in terms of dollars, time and fuel wasted) are calculated based on HPMS anticipated volumes, factored for time of day, seasonal, and day of week. Actual volumes would provide more accurate measures of congestion impact. Also, identifying events in which volumes are drastically different from nominal weekday traffic would also prevent statistical outliers (such as major snow storms) that impact both travel time and volumes from incorrectly biasing performance measures. Although traffic speed is greatly reduced during such weather events, traffic volume may also be greatly reduced. If averages for time of day and day of week are applied to a major winter event (in which a large portion of normal traffic is off the roadway), the resulting user costs would be overly inflated.

Moreover, real-time volumes, together with speeds can be used to calculate real-time user delay costs which can be important to convey performance to management and elected officials.

Use case 4: Diversion Routes during Incidents
Traffic response to reported incidents, 511 or travel time information is difficult to monitor. Travel time and speed, though useful for traveler information, provides little data to understand driver reaction in terms of alternate routes taken. Most information on diversion is ad-hoc and/or subjective, as sensors are rarely positioned to provide objective information in changes of flow patterns. Volume flows and turning movements in the form of diversion paths from nominal would aid in understanding traffic diversion during incidents, predicting flows on adjoining roads, and possibly providing better traveler information on which alternate routes to use.

An example in the Maryland region is the impact of congestion on I-270, a radial interstate feeder to the Northwest of the DC metropolitan area, to traffic on Rockville Pike, a parallel arterial roadway. Another example is the balance of flow between I-95, Route 29, Route 1, the Baltimore-Washington Parkway, and to a lesser extent I-97, all of which provide parallel paths between the DC and Baltimore metropolitan regions. Do major incidents on I-95 impact expected volumes and associated diversion paths to other parallel facilities? Volume data may be able to resolve some of the diversion patterns. Volume and turning movements combined assist in the more complex networks, as well as to better understand the limits of diversion in some instances.

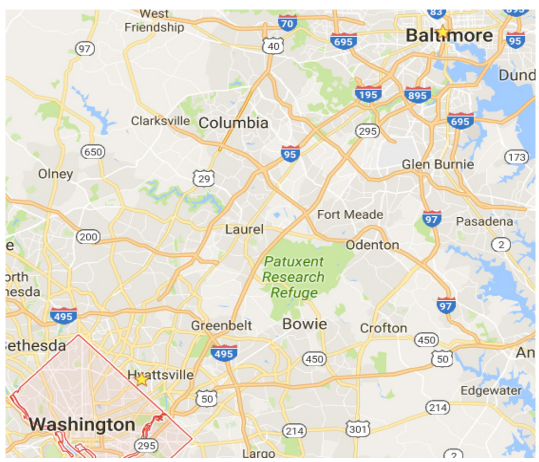

Figure 1 Parallel routes between Baltimore and Washington DC

Use case 5: Impact of Traveler Information on Diversion Routes
Most states have invested substantially in changeable message signs (CMS) to alert drivers of expected delay and for emergency situations. However, measuring the impact of information provided to the public (either due to posted travel times or in the event of emergencies that require evacuation) is problematic. Flow data provides a means of direct feedback on the effectiveness of CMS use. As a corollary, volume and turning movement data could also improve understanding of the relationship between message content and diversions.

Use case 6: Volume Heat Maps, Understanding Utilization of Existing Capacity
Network monitoring consists primarily of speed and travel time at present. Although the theoretical relationship with volume is well understood, and many traffic management centers and transportation experts understand volume flow (tidal, peak, etc.) from experience, the knowledge of volume flow is less objectively monitored. Diversion of traffic from one facility that is experiencing an incident to another facility may simply overwhelm capacity, causing extreme slowdowns if traffic is diverted. Understanding capacity and under-utilized capacity is essential for further operations effectiveness.

Use case 7: Use of Travel Demand for Calculation of Signal Timing

Traditionally, signal timing is calculated based on traffic volumes measured using either machine counts or manual counts taken over a relatively short period of from 15 minutes to a single day. As a result, short term variations in demand occurring when the counts are taken will bias the results. Longer term measurements of traffic volumes would improve the quality of off-line signal timing. Similarly, signal timing patterns implemented for non-recurring congestion (incidents and special events) is typically developed using “seat-of-the-pants” estimates due to the difficulty of collecting demand data during these unpredictable conditions. The availability of reliable volume data during non-recurring conditions would be a valuable resource for calculation of signal timing to be used during non-recurrent congestion, which could reduce overall delay and/or fuel consumption for the vehicles traveling on the network.

Section 2.0: User Survey Results

To explore the user needs and requirements for broad-based volume and turning movement data as envisioned, the research team surveyed key operations and planning personnel within the I-95 Corridor Coalition regarding the anticipated uses, criticality, and needed formats of such data. This survey was administered starting in October 2016 using an on-line survey tool. The results of the survey are summarized in the following.

Respondents

The Coalition received 14 complete responses from the survey. Eleven were from agencies within the Coalition and another three were from agencies or entities outside the coalition The fourteen full responses revealed consistent feedback in terms of the data value and reporting preferences. A roster of participating organizations is provided below:

- Metropolitan Washington Council of Governments
- Federal Highway Administration
- Maryland State Highway Administration / Department of Transportation
- South Carolina Department of Transportation
- North Jersey Transportation Planning Authority
- Delaware Valley Regional Planning Commission
- Colorado Department of Transportation
- National Renewable Energy Laboratory
- Pennsylvania Department of Transportation
- New Hampshire Department of Transportation
- Virginia Department of Transportation

Summary of Survey Responses

Listed below are short summaries of the survey responses.

- ✓ There is great interest from a planning perspective for all aspects of this type of data.
- ✓ Real time volume data seems to have a higher perceived value for incident management monitoring than for traveler information.
- ✓ The preferred volume metric was vehicle flow (vehicles per hour) as opposed to percent capacity or vehicle density.
- ✓ The needed level of accuracy for flow data to begin to support anticipated applications was to within 10% of roadway capacity.
- ✓ The minimum time interval/aggregation that was recommended was 15 minute intervals for real time, however applications which utilize historical flow data could be longer intervals of 30 minutes to 1 hour.
- ✓ An overwhelming additional desirable attribute was the percentage (or volume) of heavy truck traffic.
- ✓ Particularly for planning and performance measures applications, the need for archived vehicle and turning movement (VTM) data is greater than for real-time VTM. There is a perceived need for

real-time VTM data primarily for detours and evacuation, more so than for day to day operations applications.

- ✓ With respect to turning movements, there was no clear preference for a defined metric. Either estimates of volume in each direction or percent of turning movements in each direction were acceptable. If percentage of turning vehicles was reported, accuracy (and precision) to within 10% is needed to begin to support applications. Similar to real-time volume, turning movement should be reported at 15-minute time aggregations.
- ✓ Lastly, if VTM data were broadly accessible like travel time and speed information through probe data sources, most agencies responded that it might take 1-2 years to implement applications.

Section 3.0: Accuracy Limitations Based on Statistical Modeling

The accuracy of volume estimates based on sampling data is explored in the following four charts derived based on a statistical model. As the sampling rate (also referred to as penetration rate of probe vehicles in some literature) increases, the percentage error decreases. Also, as the volume on the roadway increases, the percentage error also decreases. The percentage error is defined here as the ratio of the binomial standard deviation (based on flow rate, reporting interval and sampling percentage) to the flow rate itself. Results are shown for sampling ratios of 5, 10, 25 and 50 percent. Note that these error metrics are with respect to roadway volume, not roadway capacity as discussed in the survey.

Table 1 Estimated volume error as a function of sampling percentage, time aggregation, and volume

	Time Aggregation (minutes)	Volume (vehicles/hour)						
		100	500	1000	1500	2000	3000	5000
5% Sampling	15	87%	39%	28%	23%	19%	16%	12%
	30	62%	28%	19%	16%	14%	11%	9%
	60	44%	19%	14%	11%	10%	8%	6%
10% Sampling	15	60%	27%	19%	15%	13%	11%	8%
	30	42%	19%	13%	11%	9%	8%	6%
	60	30%	13%	9%	8%	7%	5%	4%
25% Sampling	15	35%	15%	11%	9%	8%	6%	5%
	30	24%	11%	8%	6%	5%	4%	3%
	60	17%	8%	5%	4%	4%	3%	2%
50% Sampling	15	20%	9%	6%	5%	4%	4%	3%
	30	14%	6%	4%	4%	3%	3%	2%
	60	10%	4%	3%	3%	2%	2%	1%

Section 4.0: Initial Results of Volume Estimation from Probe Data

Initial estimates of roadway volume were conducted based on a dataset of INRIX Trip Records dating from 2015 which were licensed by the Maryland State Highway Administration (SHA) spanning four months. These data sets contain GPS traces of a sample of vehicles in the traffic stream. The GPS trace data, combined with other data sources are modeled to provide a volume estimate. Traffic volume data obtained from the Maryland SHA from their system of automated traffic recorders (ATRs) were used as both training data for machine learning and as a validation data set. All data were aggregated to an hourly basis. Data from twelve automated traffic recorders (ATRs) were selected for use in the initial analysis. The GPS trajectory data was from February, June, July, and October of 2016. The trajectory data were processed to determine the number of probed vehicles which passed each ATR station.

The analysis indicated that the average penetration rates of GPS probe vehicle traces at the 12 reference count locations varied from 0.18% to 0.72% with median of 0.57%. The corresponding average hourly observed GPS volumes (probe vehicle traces that passed the 12 reference stations every hour) varied from 22.3 to 62.3 vehicles with median of 37.3 vehicles, as displayed in Figure 2. Additional data such as speed data estimated from GPS traces, road characteristics (i.e., type of the road, number of lanes, speed limit), incident reports (i.e., work zones, collisions) and weather information (i.e., temperature, humidity, visibility, precipitation) were also included for machine learning modeling.

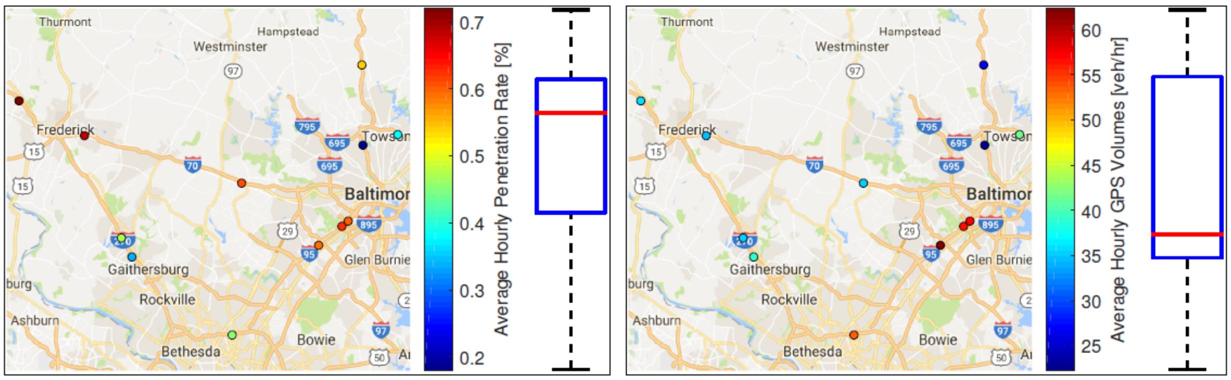

Figure 2 Average hourly penetration rates of GPS traces and average hourly GPS volumes.

Volume Estimation Methodology

A neural network was used for modeling which took the GPS trace counts as inputs, along with calibrated volumes (from ATRs), as well as speeds, road characteristics, incident reports, weather information and temporal information. The output of the model was estimated volumes. For each of the twelve reference locations, a neural network model was trained using the data from the other eleven reference locations, and validated by data from the reference location of focus. The process was repeated 12 times for all 12 reference count locations. Each reference count location had approximately 6,000 data points (approximately one data point for each hour of each day in each direction of travel). Test results for all 12 locations are evaluated in terms of R^2 and mean absolute percentage error (MAPE). MAPE is defined as

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{V_i - \widehat{V_i}}{V_i}$$

where V_i is the observed volume, $\widehat{V_i}$ is the predicted or forecasted volume, and *N* is the total number of observations.

Modeling Results

The model performance is presented in Figure 3. The R^2 varies from 0.61 to 0.94, with median of 0.82. MAPE varies from 14% to 48% with a median of 27%. The prediction results for the typical, the worst, and the best cases for a randomly selected one-week period from June 3rd to 9th, 2016 are plotted in Figure 8. The red line indicates predicted volume based on GPS traces. The blue line represents the actual volume measured by ATR at that location. Figure 4 provides a visual reference for actual and predicted traffic volumes indexed to the R^2 value.

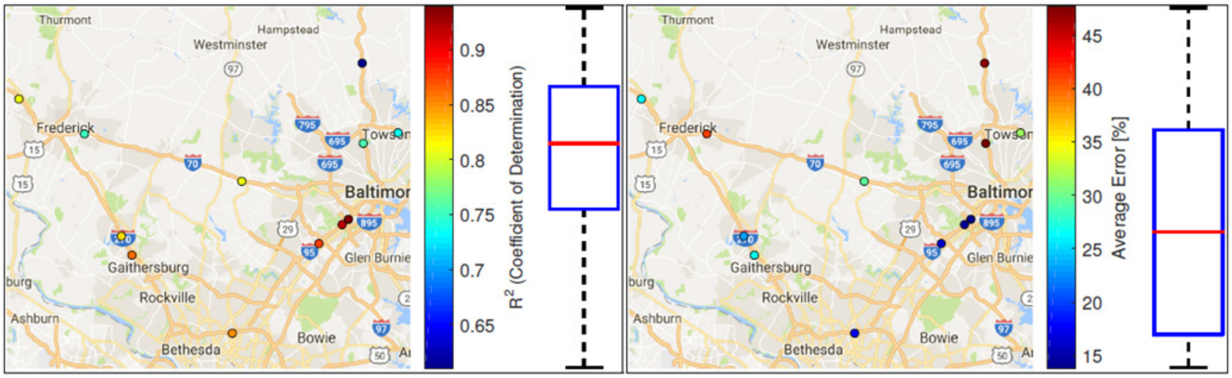

Figure 3 Model accuracy.

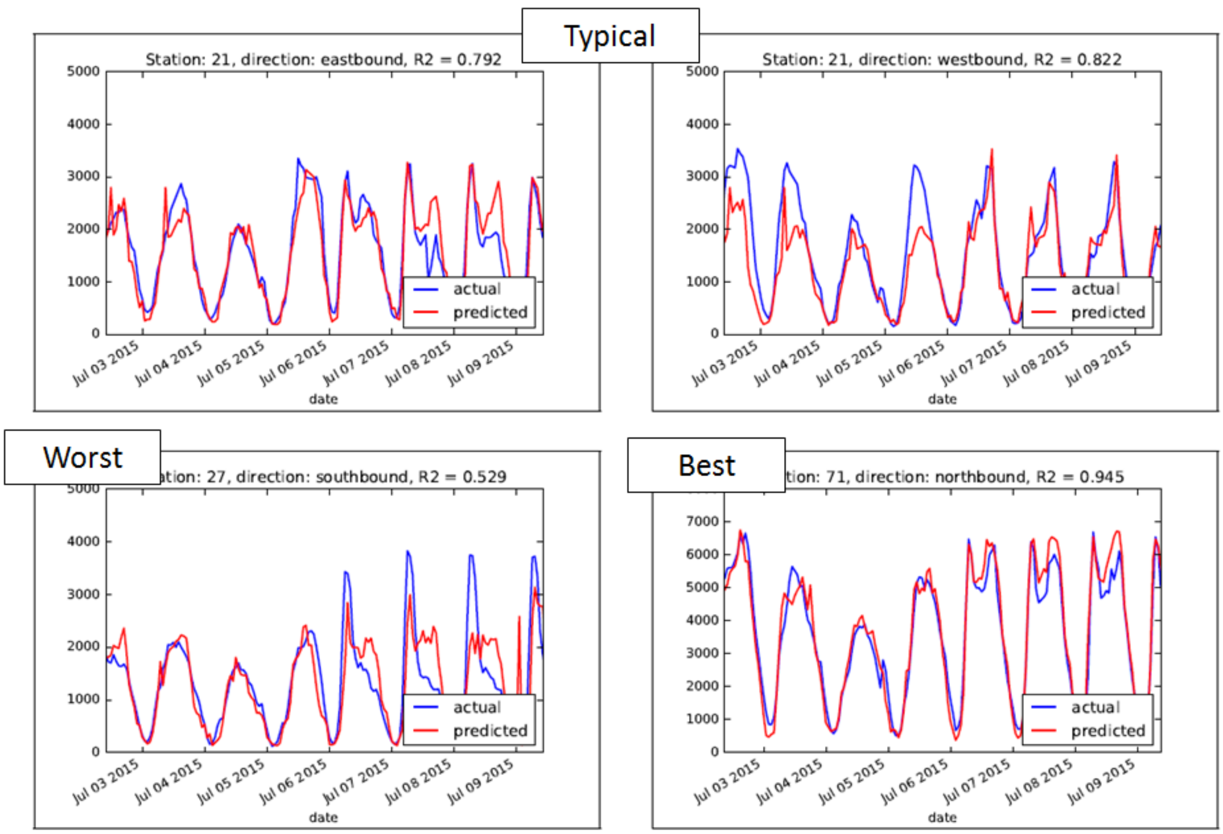

Figure 4 One-week temporal variation of predicted and actual traffic volume (in vehicles per hour).

To demonstrate the contribution of GPS probe vehicle traces in the modeling process, neural network models were re-trained and re-tested without including GPS traces as an input variable. The models operated only on auxiliary data such as number of lanes, weather, time of day, day of week, etc. The results showed that R^2 varies from 0.49 to 0.90, with the median of 0.73 compared to 0.82 with GPS traces included in the model. Similarly, the MAPE metric varies from 16% to 54% with a median of 37% without the GPS traces, compared to 27% with GPS traces. The difference in R^2 and MAPE with and without GPS trajectory data reinforces the value of probe vehicle information for improved volume estimates.

Model performance from the perspective of capacity (rather than absolute volume) was further examined. For each 10 percentile of road capacity, MAPE with respect to both volume and capacity was calculated.

The results were plotted in Figure 5. Figure 5 demonstrates that with the increase of traffic volume the estimation error percentage with respect to volume varies from 64.2% to 14.5%. This is similar to the trend results from the simple statistical model shown in Table 1. The estimation error percentage with respect to capacity increases from 3.5% to 13.7%, with an average of 9.5%. The accuracy in this initial model thus meets the stated preferred level of accuracy by respondents in the survey (i.e., within 10% of road capacity). Note, however, that the metric of accuracy within the survey was not strictly defined, but also that modeling accuracy is anticipated to improve as probe vehicle penetration increases.

% Capacity (from)	% Capacity (to)	Number of samples	Error	
			% Volume	% Capacity
0%	10%	8348	64.23%	3.46%
10%	20%	7430	39.31%	5.57%
20%	30%	4985	32.31%	8.08%
30%	40%	4899	28.47%	9.92%
40%	50%	5497	23.15%	10.43%
50%	60%	6125	19.64%	10.76%
60%	70%	5818	16.04%	10.41%
70%	80%	6372	14.28%	10.71%
80%	90%	5912	14.17%	12.01%
90%	100%	5394	14.49%	13.74%

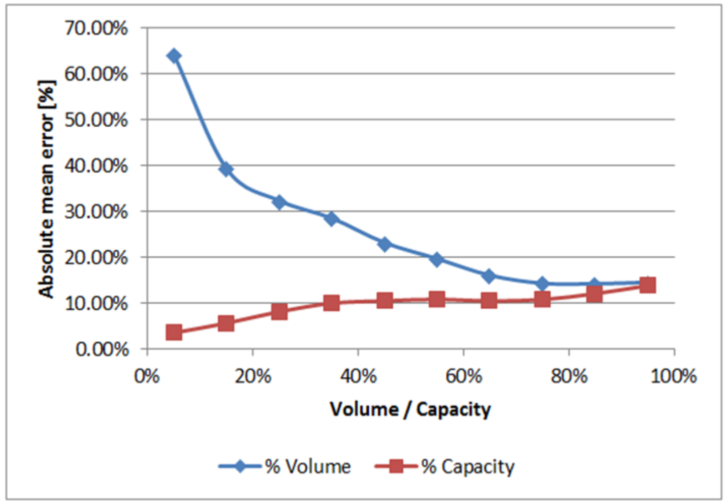

Figure 5 Estimation error percentage with respect to volume and capacity.

Conclusions

The lack of ubiquitous traffic volume counts available to support either real-time or planning application motivated this study to examine the potential of commercial probe data to provide volume estimates of adequate accuracy. This initial research effort explored uses that broad-based volume and turning movement data could enable. Seven use cases of real-time traffic volume data were examined that would enable transportation agencies to enhance operations, improve accuracy of performance metrics, and support planning activities. A survey of operations and planning personnel within the I-95 Corridor Coalition regarding the anticipated uses, criticality, and needed formats demonstrated that DOTs show great interest in ubiquitous traffic volume and turning movement data both for planning and incident management. The preferred volume metric was vehicles per hour reported at least hourly for planning applications, and at least every 15 minutes for operations applications. The level of accuracy needed to begin to support such applications was to within 10% of roadway capacity.

An initial analysis tested the used of GPS probe vehicle trajectory data to estimate traffic volumes in conjunction with other available roadway data attributes. A neural network was used to model the relationship between GPS traces and real-time traffic volume at 12 reference stations. Additional data such as speed data estimated from GPS traces, road characteristics, and weather information was also included as inputs for modeling. The resulting accuracy of the neural network model had an R^2 varying from 0.61 to 0.94 with a median of 0.82, and a MAPE varying from 14% to 48% with a median of 27%.

When GPS traces were not included as an input variable, the R^2 dropped to 0.73, and the MAPE rose to 37% indicating that GPS traces played a significant role in estimating traffic volume. Further analysis shows that the estimation error percentage with respect to capacity (as opposed to volume) averaged 9.5% which is within the preferred level of accuracy specified by respondents in the survey to begin to support applications.

The research in this paper only used data collected at 12 ATR locations in the state of Maryland. In the future, additional and more geographically diverse data will be used to gain further insights. Other machine learning methods and data sources may also be explored.

References

1. American Association of State Highway and Transportation Officials (2011). *A Policy on Geometric Design of Highways and Streets*, *6th Edition*.
2. Texas Transportation Institute (2015). Appendix A: Methodology for the 2015 Urban Mobility Scorecard, Texas A&M University: https://static.tti.tamu.edu/tti.tamu.edu/documents/mobility-scorecard-2015-appx-a.pdf
3. Vlahogianni, E., M. Karlaftis, J. Golias (2014). Short-term traffic forecasting: where we are and where we're going, *Transport. Res. Part C: Emerg. Technol.*, No. 43, pp. 3–19.
4. Lin, W. (2004). Case Study on Support Vector Machine versus Artificial Neural Networks. University of Pittsburgh, Pittsburgh, Pa.
5. Castro-Neto, M., Y. Jeong, M. Jeong, L. Han (2009). AADT prediction using support vector regression with data-dependent parameters. *Expert Systems with Applications*, No. 36, pp. 2979-2986.
6. Zhao, F., S. Chung (2001). Estimation of annual average daily traffic in a Florida county using GIS and regression. In proceedings Transportation Research Board Annual Meeting, Washington, DC, USA, Paper No. 01-3440
7. Kingan R., T. Westhuis (2006). Robust regression methods for traffic growth forecasting. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1957: 51-55.
8. Lv Y., Y. Duan, W. Kang, Li Z., F.-Y. Wang (2015). Traffic flow prediction with big data: a deep learning approach, *IEEE Trans. Intell. Transport. Syst.*, No. 16 (2), pp. 865–873.
9. Polson, N. G., , V. O. Sokolov (2017). Deep learning for short-term traffic flow prediction. *Transportation Research Part C: Emerging Technologies*, No. 79, 1-17.
10. Islam, S. (2016). Estimation of Annual Average Daily Traffic (AADT) and missing hourly volume using artificial intelligence, MS thesis, Clemson University.
11. Sharifi, E., S. E. Young, T. H. Jacobs, and S. M. Rochon (2014). Estimating Winter Weather Road Restoration Time using Outsourced Traffic Data: Three Case Studies in Maryland. In proceeding 93nd TRB annual meeting, Washington, D.C., January 12-16.