Average Annual Daily Traffic Estimation by Mobile Device Footprint Cardinality

Kentaro Iio
Zachry Department of Civil & Environmental Engineering
Texas A&M University, College Station, TX 77843-3136
Email: iio_kentaro@tamu.edu

Dominique Lord
Zachry Department of Civil & Environmental Engineering
Texas A&M University, College Station, TX 77843-3136
Email: d-lord@tamu.edu

Yunlong Zhang
Zachry Department of Civil & Environmental Engineering
Texas A&M University, College Station, TX 77843-3136
Email: yzhang@civil.tamu.edu

Word Count: 3,615 words + 3 tables = 4,365 words

Submitted [August 1, 2019]

ABSTRACT

Knowing average annual daily traffic (AADT) is important because it can be used for many activities including transportation planning, safety assessment, and even marketing. Today, AADT is estimated based on traffic volume counts at fixed locations. Although the current AADT estimation method achieves small measurement errors, the cost associated with traffic counts is not small. In addition, the traffic volume at locations that do not have any traffic counting stations remains unknown. In recent years, there have been attempts to estimate AADT based on vehicle trajectory reconstruction. Because such a line of data may not always be available for practical use, this paper proposes a new method to estimate AADT based on point data gathered with mobile devices but without reconstructing vehicle trajectories. The method uses the concept of cardinality as the relative likelihood of vehicle existence. Using datasets provided by Safe2Save and the base AADT provided by the Texas Department of Transportation, the accuracy of the methodology was tested in two locations in Texas. The results showed good overall predictability (r = 0.955). Mean signed difference, mean absolute difference, mean absolute percent error, and root mean square error were as small as those in methods that required trajectory reconstruction. Since this methodology is still at its infancy, there may be room for improvements including the potential application on hourly volume estimation. It may be desirable to initiate a larger-scale research to investigate the effects of the sampling rate, seasonal factors, and weather on the estimation of AADT.

Keywords: Traffic Volume, AADT, Mobile Device, Remote Sensing, Global Positioning System

INTRODUCTION

Vehicular traffic volume is a fundamental variable of transportation in a motorized society. It not only indicates the level of traveler activity but also represents vehicle occupants' exposure to a transportation system at an aggregated level. Depending on the purpose, vehicular traffic volume can be measured for different durations, such as hourly, daily, weekly, monthly, and yearly. In the United States, commonly-used traffic volume measures include average daily traffic (ADT), average annual daily traffic (AADT) (*1*) or annual average daily traffic (AADT) (*2*), average weekday traffic (AWDT), annual average weekday traffic (AAWT), average weekend daily traffic (AWET), annual average weekend traffic (AAWET), and seasonal average daily traffic (SADT) (*3*). Among these measures, AADT is widely used to determine transportation planning, operation, and safety applications such as the calculation of crash rates (*4*).

Perfectly-accurate traffic counts are only achievable when all volumes during a coverage period are counted. However, it is not always practical nor meaningful to count all traffic in terms of cost, especially when such counts are conducted at many locations. Therefore, AADT is often a subject to be estimated. In the United States, the Federal Highway Administration (FHWA) recommends estimating AADT based on 24-hour or 48-hour short-term counts (*5*). Traffic volumes are affected by multiple factors, including weather and special events (*6*). Short-term counts are multiplied by adjustment factors, such as seasonal factor and axle factor (*7*) to get estimated AADT (**Equation 1**).

$$AADT = (Axles) \times (Axle\ Factor) \times (Seasonal\ Factor) \qquad (1)$$

Other counting methods include manual counts, inductive loops, weigh-in-motion, and video counts (*3*).

Point Data and Line Data

Although fixed-location counts can be a source of relatively accurate AADT estimations, such counting methods are often accompanied by installation cost, maintenance cost as well as labor costs. Also, counts are only available at the locations of counting facilities. However, the recent development of the global positioning system (GPS) and the widespread use of mobile devices has opened a new opportunity to collect traffic data remotely. With GPS-enabled devices, geocoordinates and other attributes, such as device identification number and speed, are often recorded as granular "point data" at a fixed interval (*e.g.*, every 10 seconds) (*8*). Because point data are like animals' footprints, lines chronologically connecting points can reconstruct original trajectories (*9*). When recording intervals are short enough, it is possible to reconstruct trajectories of the objects with high accuracy.

If an AADT estimation method does not require special counting devices or laborforce at the scene, it could eliminate spatial constraints, significantly reduce associated cost, and even enable faster decision making (*e.g.*, crash hot spot identification) (*10*). Actually, StreetLight Data (*11*, *12*) reconstructed vehicle trajectories from waypoints obtained by smartphones. They trained the data with factors including U.S. Census data, speed limits, number of lanes, road classification, precipitation, and air temperature. Although the company claims that they achieved mean absolute percent errors (MAPE) of as low as 14.9 percent (for AADT larger than 50,000), the exact methodology has not been disclosed to the public. Turner and Koeneman (*13*) investigated the accuracy of AADT estimation based on vehicle waypoints provided by

StreetLight data for Minnesota Department of Transportation (MnDOT). The researchers compared estimations with counts at 69 permanent counting locations in Minnesota and found 29 percent (for 20,000 to 50,000 AADT) to 68 percent (300 to 5,000 AADT) MAPE. The researchers concluded the method needs to be improved to be practical for agencies to use. The StreetLight Data reports (*11*, *12*) did not include the estimation procedure, so the multiplicity of its use is limited.

Although trajectory reconstruction can be useful for AADT estimation, it has some disadvantages. One of them is that mass conversion of point data to line data requires large computing power. In addition, the reliability of the reconstructed line data decreases when the recording interval is not short enough. For these reasons, vehicle trajectory (or line data) may not always be available for practitioners, so it would be valuable if there was a method to estimate AADT based on granular point data. In this paper, the authors propose a method to estimate AADT based on point data associated with vehicle speed, test its accuracy, and discuss its applicability.

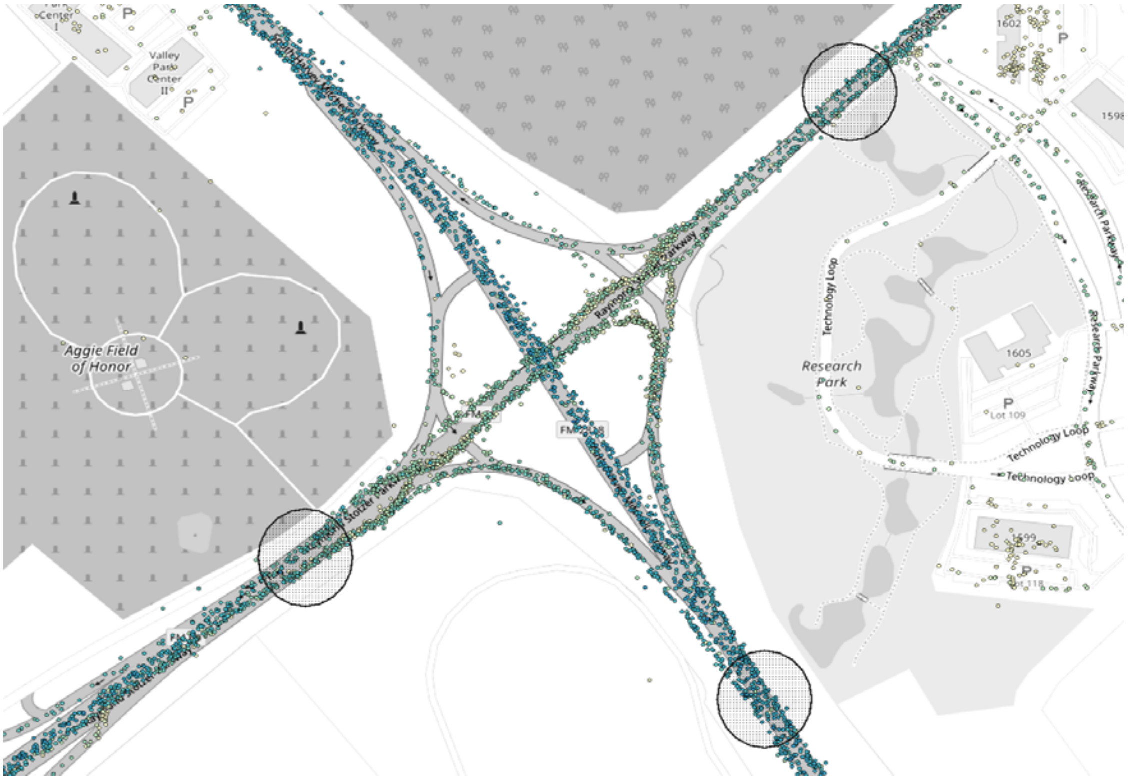
METHODOLOGY

In this research, the term "cardinality" is introduced and defined as the relative likelihood of the number of vehicles passing through a unit area in a unit time. It is the morphism of true traffic volume and conceptually close to cardinality in set theory, representing "how thick" the value is. When the cordon *i* has *k* data points that represent vehicle occupants' locations, its cardinality can be expressed as **Equation 2**.

$$\mathfrak{c}_i \coloneqq \frac{1}{a_i}\sum_{j=1}^{k}\left(SMS_j \prod_{l=1}^{m} AF_{j,l}\right) \tag{2}$$

where: $\mathfrak{c}_i$ = cardinality of cordon i
a_i = area of cordon i (ft^2) on the lanes of measurement
SMS_j = space mean speed of the jth vehicle footprint in area i (mph)
k = number of total vehicle footprints in the cordon i
$AF_{j,l}$ = the lth adjustment factor for the jth vehicle
m = number of adjustment factors

Cardinality is the sum of the number of observed data points weighted by the likelihood (or "rarity") of their existence at the location. To calculate cardinality, each data point representing a vehicle location is multiplied by the space mean speed. This is because cardinality is similar to probabilistic density, meaning the likelihood that a vehicle leaves a data point in a cordon in a unit time is inversely proportional to its space mean speed in a cordon. For example, a vehicle moving at 30 mph is twice more likely to leave its footprint than a vehicle traveling at 60 mph in a given segment in unit time when point data were recorded at a fixed interval. Like traditional AADT estimation, each speed can also be multiplied by adjustment factors (*e.g.*, seasonal factors) when such factors are available. **Figure 1** shows an example of cardinality cordons.



Note: Greener colors indicate lower speed.

FIGURE 1 Vehicle footprints and cardinality cordons in College Station, TX.

The unit of cardinality can be "vehicle miles per hour," but cardinality can be treated as a unitless value because it is only usable as a relative value. Therefore, this method requires known AADT for calibration. In this paper, AADT was calibrated in the following manner by highway functional class (*14*).

$$Estimated\ AADT_i = c_i\ \widetilde{F}_{\bar{i}} \qquad (2)$$

where: c_i = cardinality of cordon i (when this location's AADT is the subject to estimation)
$\widetilde{F}_{\bar{i}}$ = median of the other cordon's ratios of base AADT to their cardinality

To calculate $\widetilde{F}_{\bar{i}}$, the authors used median of the other cordon's ratios of base AADT to their cardinality, therefore, the base AADT of each location was treated as unknown in the estimation process. When cardinality was zero. That cardinality cordon was removed from the calculation of median value, but AADT was still estimated to be zero. The remainder of this section explains the datasets used for validation.

Data Preparation

In this research, the authors estimated AADTs based on point data provided by Safe2Save, LLC and compared them to AADT provided by the Texas Department of Transportation (TxDOT).

Base AADT

As a base AADT, estimated AADTs of 2017 was provided by the Texas Department of Transportation. The counts included short-term pneumatic tube counts and multi video coding. These data were used as a base AADT for calibration and comparison. The 2018 AADT was not publicly available at the time this research was conducted.

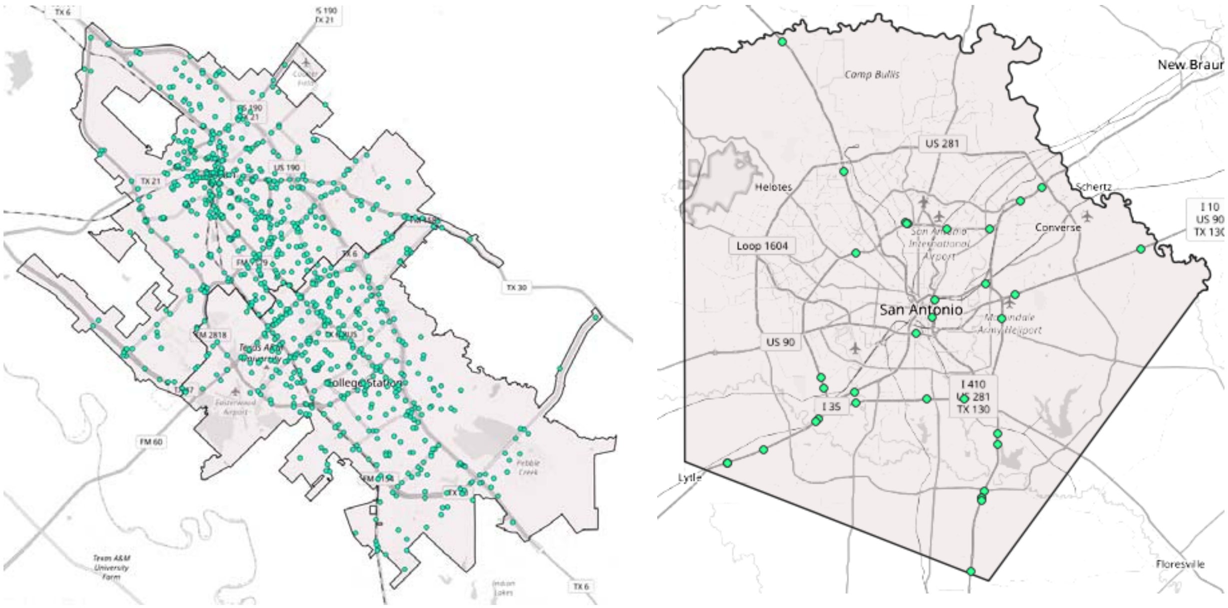
Driving Point Data

Anonymized point data were provided by Safe2Save, LLC, a Texas-based company that develops and operates a mobile phone app that rewards users for not touching the phone while driving by giving "SAFE 2 SAVE points" in the application (*15*). The application records user locations every a few minutes once it detected users were driving. The app had an upper-speed threshold at 100 mph so that it did not record user locations on flying airplanes. The dataset consisted of data points obtained by their iOS application. Each data point had a user's longitude, latitude, speed, and timestamp. Because the application's proprietary points can be redeemed as coupons at local businesses and most of the partnered businesses were located in Texas, the users of the application were mostly in Texas. As of July 2019, the app had more than 200,000 downloads.

Location Selection

The comparison of cardinality relies on the assumption that each AADT cordon has the same level of user penetration rates of a mobile application; hence, it is necessary to select areas that have homogeneous user penetration rates. Otherwise, the cardinality is largely affected by the difference in penetration rates, making the comparisons less meaningful. In addition, the higher the penetration rate, the smaller the sampling errors would be. For these reasons, this research selected two areas for the estimations of AADT: Bryan-College Station, Texas and Bexar County, Texas (**Figure 2**).

Bryan-College Station was chosen because the cities had the highest user penetration rates compared to other locations. The user penetration rate was assumed to be homogeneous because these two cities form a small college town community and are hours apart from large cities such as Dallas, Houston, and Austin. AADTs of the functional class 2 (Other Freeways and Expressways) through 6 (Minor Collectors) were estimated in these two cities. Because there were no Interstates in Bryan-College Station, Bexar County, Texas was chosen for AADT estimation for the functional class 1 (Interstates). The authors chose Bexar County because it had higher user penetration rate than other large cities. The user penetration rate was expected to be homogeneous on Interstates because users were spread out throughout the county. Basic information of the selected areas and is summarized in **Table 2**.



Bryan-College Station (*n* = 906) **Bexar County (*n* = 43)**

FIGURE 2 TxDOT's AADT Stations in Bryan-College Station and Bexar County

TABLE 2 Description of Dataset

Area	**Estimated population in 2018** (*16*)	**SAFE 2 SAVE downloads by July 2019**	**Number of all AADT stations**	**Period of SAFE 2 SAVE data**
Bryan-College Station, Texas	201,663	30,000	906	August, 28, 2018 12:12:48 – July 15, 2019 19:02:22
Bexar County, Texas	1,986,049	12,000	43	August 28, 2018 12:12:48 – July 29, 2019 23:59:59

To simplify the procedure while considering the isotropy of vehicular movements, circular cordons (**Figure 1**) were used in this paper. The cordon radii were 60 ft for functional classes 2 to 6 and 100 ft for functional class 1 (Interstates). The larger radii were created for Interstates so that cordons were sufficiently large to cover roadways. At this point, 7 "inactive" stations, 11 stations with 0 AADT, 2 cordons that overlapped other roadways, 3 cordons that had the exact same locations with other cordons were removed from the analyses. AADT stations with 0 volume were eliminated to avoid mixing up potentially closed segments. The final cordons were 889 in Bryan-College Station and 37 in Bexar County. The observed data points did not come from the same year of the base AADT, however, the authors considered it was still feasible to compare the SAFE 2 SAVE data points with 2017 AADT for initial comparison.

Accuracy Measures

The following measures were used as accuracy indicators (**Equations 3-6**). For each cordon, the base AADT of the same location was only referenced for this comparison purpose.

$$MSD = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x_i}) \tag{3}$$

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|x_i - \overline{x_i}| \tag{4}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|x_i - \overline{x_i}|}{\overline{x_i}} \tag{5}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x_i})^2}{n}} \tag{6}$$

where: n = number of comparisons
x_i = the *i*th estimated traffic volume based on cardinality
$\overline{x_i}$ = base traffic volume count for the *i*th comparison

FINDINGS

Table 2 and **Table 3** show accuracy measures by traffic volume level and highway functional class, respectively. As shown in **Table 2**, MSD, MAD, MAPE were better than what was reported in Turner and Koeneman (*13*) in most traffic volume levels. Overall, smaller MAPE was observed in locations that had large AADT. The largest MAPE in 1-to-299 (158.6 percent), indicating the difficulty of sampling small AADT. Because this AADT level was not even included or mixed with larger levels of traffic in existing reports, at least it might be valuable to know that AADT prediction accuracy can significantly drop at this small volume class.

Unlike classifications by traffic volume, MAPE did not show an obvious trend by functional class, except for the largest MAPE (80.8 percent) was observed in the functional class 1. The reason that functional class 2 had the smallest MAPE (16.7 percent) was not clear. SAFE 2 SAVE's relatively lower user penetration rate in Bexar County might have affected MAPE on Interstates (37.3 percent).

Table 2 Accuracy Measures by Traffic Volume Level

TxDOT's AADT	Number of TxDOT Sites	*MSD*		*MAD*		*MAPE*		*RMSE*
1 to 299	87	176	N/A	257	N/A	158.6%	N/A	952
300 to 4,999	440	181	-520	1,139	-16	58.0%	-10.0%	1,591
5,000 to 4,999	131	806	-2,157	2,772	-1,251	39.9%	-4.1%	4,163
10,000 to 19,999	137	962	-4,081	2,832	-3,053	49.6%	5.6%	7,211
20,000 to 49,999	112	3,274	-3,270	7,573	-1,005	24.3%	-4.7%	9,543
50,000 or more	19	21,007	-11,135	29,959	-4,153	26.5%	-7.5%	46,254
All traffic levels more than 299	839	1,291	-1,765	3,573	-209	46.5%	-14.5%	2,355
All traffic levels	926	1,186	N/A	3,261	N/A	57.1%	N/A	2,095

Note: The highlighted area indicates the differences between Turner and Koeneman (*13*).

Table 3 Accuracy Measures by functional class

	Functional class	Number of TxDOT sites	*MSD*	*MAD*	*MAPE*	*RMSE*
1	**Interstates**	37	11,806	20,373	37.3%	34,786
2	**Other Freeways and Expressways**	11	-1,145	4,796	16.7%	7,377
3	**Other Principal Arterials**	131	1,602	6,070	29.9%	7,880
4	**Minor Arterials**	155	803	3,188	42.5%	4,258
5	**Major Collectors**	323	840	2,096	61.5%	4,368
6	**Minor Collectors**	19	-751	1,532	36.5%	4,321
7	**Local Roads**	250	328	872	80.8%	1,937
	All traffic levels	926	1,186	3,261	57.1%	2,095

Figure 3 is a scatterplot of the base AADT and the estimated AADT. The estimated AADT showed a strong positive correlation to the base AADT (r = 0.955). This correlation was not as strong as that of StreetLight Data (*12*) (r = 0.980), but the strong positive correlation still indicated that cardinality can be used as a morphism of real traffic.

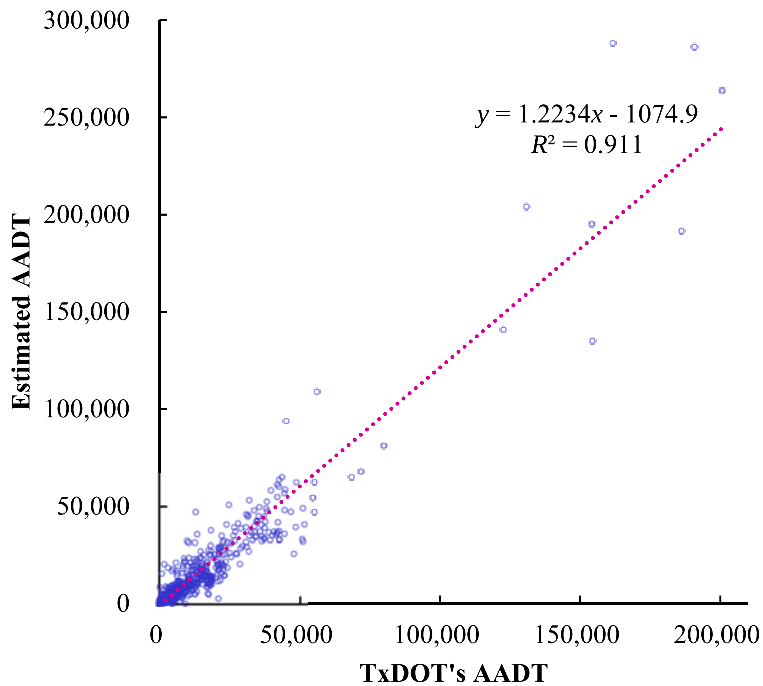


FIGURE 3 Scatterplot of TxDOT's AADT and the estimated AADT.

CONCLUSIONS

This research was significant from the following perspectives:

1. This paper proposed a method to estimate AADT based on point data without vehicle trajectory construction.
2. It showed predictability as high as other similar proprietary methods.

Overall, the results suggested that granular point data can be used to estimate traffic volumes as accurately as a method using reconstructed vehicle trajectories. The associated errors were not as small as StreetLight Data (*12*) but were as good as Turner and Koeneman (*13*). Considering there has been little existing research on this topic, it would be reasonable to say that the proposed method got an initial passing mark. Considering the authors only used the number of points and associated speed for predictions, the results were encouraging.

Including other similar methods, techniques to estimate AADT based on data obtained by mobile phone are still at their infancy. Therefore, there would still be room for accuracy improvements. In the first place, it is unknown what factors largely affect prediction accuracy. As higher accuracy achieved by StreetLight Data (*12*) implies, factors such as seasonal factors, demographics and weather could improve the prediction accuracy when they are included as prediction factors. At this point, it is also not clear what level of sampling rate is required to achieve what level of prediction accuracy. Although there is a study reporting that the sampling rate as low as 0.1 percent can be used to estimate hourly volume (*17*), the rate should also be confirmed for the new methods so that practitioners can make a proper decision.

This research used circular cordons for cardinality calculations. However, it may not always be desirable to use uniform-sized cordons when they largely overlap other roadways or non-roads (e.g., supermarket) (**Figure 4**), so the results could be improved if cordon shape and radius is optimized in some way. This can be a feature topic that needs to be addressed for greater applicability.

Overpasses/Underpasses **Overlapping to supermarket**

FIGURE 4 Undesirable cordons.

Furthermore, it may be possible to estimate hourly volume by the same or similar method as long as enough data points are available. While this paper focused on AADT estimation, it would be constructive for future work to address hourly volume estimations based on data derived from mobile devices. If such a technique is established, agencies may be able to filter traffic volume based on different conditions (*e.g.*, daytime and nighttime) to identify crash hotspots by detailed conditions.

Additionally, calculating cardinality without adjustment factor may not be suitable for time series comparisons (*e.g.*, February and August) when the number of app users changes drastically across the time series. If that is the case, an adjustment method should be developed to achieve higher prediction accuracy.

Limitations

There were some limitations to this method. First of all, the cardinality cannot be applied where and when the penetration rate of a mobile application is not homogeneous. For example, if Houston, Texas had 100,000 app users while Golden, Texas had 10 app users, cordons in Houston would have larger cardinality values even though that does not reflect the actual difference in traffic volumes in the two areas. Therefore, practitioners need to make sure that the users of mobile devices represent accrual road user populations in the areas of interest.

In addition, this research compared data points recorded from 2018 to 2019 with AADT in 2017. Since the base AADT in 2018 through 2019 were not available, the authors considered that it was feasible to use AADT from 2017 for an initial comparison purpose. However, from a technical standpoint, the accuracy should be confirmed using AADT in the same period. It is nonetheless anticipated that the difference in AADT for closely-related years will be minimum.

To quantify accuracy with greater a detail, future work should also include a true estimation of AADT by randomly selecting base AADT stations to calculate the ratio to estimate AADT for other cordons. In conjunction, accuracy assessment based on the base AADT availability (*e.g.*, how many base AADT counts are needed to achieve an acceptable estimation accuracy) should also be addressed in future work.

ACKNOWLEDGMENTS

The authors acknowledge the Texas Department of Transportation for providing detailed traffic count data. The authors also thank Safe2Save, LLC for providing the aggregated user footprint data. This work would not have existed without them.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: Iio, K.; data collection: Iio, K.; analysis and interpretation of results: Iio, K., Lord, D., Zhang, Y.; draft manuscript preparation: Iio, K., Lord, D., Zhang, Y. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. Federal Highway Administration. *Manual on Uniform Traffic Control Devices*, Federal Highway Administration, Washington D.C., 2009.

2. Transportation Research Board. *Highway Capacity Manual, Sixth Edition: A Guide for Multimodal Mobility Analysis*. Transportation Research Board, Washington D.C., 2016.

3. Schroeder, Bastian J., Christopher M. Cunningham, Daniel J. Findley, Joseph E. Hummer, and Robert S. Foyle. *Manual of Transportation Engineering Studies*, Institute of Transportation Engineers, Washington, D.C., 2010.

4. American Association of State Highway and Transportation Officials. *Highway Safety Manual*. American Association of State Highway and Transportation Officials, Washington D.C., 2010.

5. Krile, Robert, Fred Todt, Jeremy Schroeder, and Steven Jessberger. *Assessing Roadway Traffic Count Duration and Frequency Impacts on Annual Average Daily Traffic Estimation: Assessing Accuracy Issues Related to Annual Factoring*. No. FHWA-PL-16-012. United States. Federal Highway Administration, 2016.

6. Gene Hawkins Jr, H. “Traffic Flow Characteristics for Uninterrupted‐Flow Facilities.” *Traffic Engineering Handbook*, Institute of Transportation Engineers (2015): 203-234.

7. Texas Department of Transportation. *Traffic Data and Analysis Manual*, Texas Department of Transportation, (2001).

8. Obe, Regina O., and Leo S. Hsu. *PostGIS in Action*. Second Edition. Greenwich: Manning, 2015.

9. Zheng, Yu, Quannan Li, Yukun Chen, Xing Xie, and Wei-Ying Ma. "Understanding mobility based on GPS data." In *Proceedings of the 10th international conference on Ubiquitous computing*, pp. 312-321. ACM, 2008. doi: 10.1145/1409635.1409677

10. Marković, Nikola, Przemyław Sekuła, Zachary Vander Laan, Gennady Andrienko, and Natalia Andrienko. "Applications of trajectory data from the perspective of a road transportation agency: Literature review and maryland case study." *IEEE Transactions on Intelligent Transportation Systems* 20, no. 5 (2018): 1858-1869. doi: 10.1109/TITS.2018.2843298

11. StreetLight Data. *StreetLight Data’s AADT 2018 Methodology and Validation White Paper*, (2019).

12. StreetLight Data. *StreetLight AADT 2018 V2: Methodology and Validation White Paper United States + Canada*, (2019). Reirieved on July 12 from www.3yemud1nnmpw4b6m8m3py1iy-wpengine.netdna-ssl.com/wp-content/uploads/StreetLight-2018-V2-AADT-Canada-Methodology-and-Validation-White-Paper_190709.pdf

13. Turner, Shawn, and Pete Koeneman. *Using Mobile Device Samples to Estimate Traffic Volumes*. No. MN/RC 2017-49. Minnesota Department of Transportation, Research Services & Library, 2017

14. Federal Highway Administration. *Highway Functional Classification Concepts, Criteria and Procedures*. FHWA--PL--13--026, US Department of Transportation, Federal Highway Administration, 2013.

15. Corry, M. (2018). *Systems and methods for restricting mobile device use*. US20180374111A1.

16. United States Census Bureau. *Quick Facts*, (2019).

17. Ishida, Haruo, Hiroshi Miura, Naohisa Okamoto, and Hideki Furuya. "Sampling Rates for Travel Speed Survey with Advanced Information Systems." *Infrastructure Planning Review* 18 (2001): 81-88. doi: 10.2208/journalip.18.81