Case Study

Developing a Practical Method to Compute State-Level Bus Occupancy Rate

Meixin Zhu[1]; Wenbo Zhu, Ph.D.[2]; Jerome M. Lutin, Ph.D., P.E.[3]; Zhiyong Cui[4]; and Yinhai Wang, Ph.D., P.E., F.ASCE[5]

Abstract: This work supported the Federal Highway Administration (FHWA) in presenting state and metropolitan area vehicle occupancy information in compliance with Title 23 of the US Code of Federal Regulations, Part 490 National Performance Measures. The basic goal is to provide and introduce a statistically effective and realistic approach to approximate bus occupancy rates for each US state and Washington, DC. Bus occupancies were calculated separately for transit buses, school buses, and motorcoaches. The average total bus occupancy was determined by aggregating the average occupancy of the vehicle for the three groups weighted by annual vehicle miles traveled. Regarding transit buses, the National Transit Database (NTD) of the Federal Transit Administration (FTA) was used to measure the occupancy of transit buses. For school buses, state transportation statistics from 2015–2016 were used to measure average school bus occupancy for each state. Data provided by the Port Authority of New York and New Jersey (PANYNJ) for the Port Authority Bus Terminal (PABT) in New York City, the largest US bus terminal, were used to measure the occupancy rates for motorcoaches. Results on state-level bus occupancy rates are summarized. The paper concludes with guidelines for future data collection, validation, and training. The date, code, and user guide of this study can be found at an online GitHub repository. **DOI: 10.1061/JTEPBS.0000493.** © *2021 American Society of Civil Engineers.*

Author keywords: Bus occupancy; Transit; School bus; Motorcoach.

Introduction

While automobile miles driven continue to rise in the US and significantly outpace the introduction of highway lane miles, several transport authorities are developing approaches to promote high occupancy vehicle (HOV) driving. To track the progress of transport demand management approaches and to better understand the movement of people and multimodal travel in general, collecting and estimating vehicle occupancy factors has become a research subject of interest.

Vehicle occupancy data are valuable for traffic engineers, planners, and decision-makers because they can (1) be used to calculate trip metrics related to passengers, such as person delays and passenger miles traveled; (2) help identify routes or areas with high demand needing service expansion; and (3) be useful for service optimization, such as determining the number of buses required for certain corridors. Moreover, in compliance with Title 23 of the US Code of Federal Regulations, Part 490 National Performance Management Measures, the Federal Highway Administration (FHWA) is mandated to provide the vehicle occupancy details to states and metropolitan areas. The calculation of average vehicle occupancy (AVO) is used by state DOTs for states and/or metropolitan areas as input to the Highway Performance Monitoring System (HPMS). The main use for occupancy of vehicles is measuring the peak hour excessive delay (PHED) metric.

A variety of methods for estimating the vehicle occupancy have been proposed, which can be divided into two main categories: manual/video observation techniques and analysis of external datasets. The former relies on the identification or counting of passengers in cars, manually (or theoretically automated in the case of video). The latter depends on the retrieval of occupancy details from external sources, such as law enforcement crash report data. Nonetheless, some of the following issues influence both types of methods: (1) they are limited in scope, only applicable to geographic regions that are no larger than a state; (2) manual counting approaches are time- and resource-consuming, rendering them unsuitable for calculating nationwide vehicle occupancy; and (3) external data sources such as crash data are prone to biases due to factors such as overrepresentation of certain groups in the crash data.

Another major limitation in past vehicle occupancy research is that the buses were not covered sufficiently: the aforementioned two major methods, although they work for passenger cars, do not work well in collecting information from buses. Physical counting, for example, is not an effective way to collect bus data because of the small percentage of heavy vehicles in the traffic flow. Precise occupancy reading from a loaded bus often takes a lot of time and energy, and can typically not be calculated without interrupting the bus operation. Also, crash report data sets typically have very few crashes involving buses, making it difficult

[1]Research Assistant, Dept. of Civil and Environmental Engineering, Univ. of Washington, More Hall 101, BOX 352700, Seattle, WA 98195. ORCID: https://orcid.org/0000-0003-3291-3616. Email: meixin92@uw.edu

[2]Applied Scientist, Zillow, 1301 2nd Ave. floor 31, Seattle, WA 98101. Email: wbzhu@uw.edu

[3]Senior Director of Statewide and Regional Planning, New Jersey Transit (Retired), 1 Penn Plaza E, Newark, NJ 07105. Email: jerome.lutin@verizon.net

[4]Research Assistant, Dept. of Civil and Environmental Engineering, Univ. of Washington, More Hall 101, BOX 352700, Seattle, WA 98195. Email: zhiyongc@uw.edu

[5]Professor, Dept. of Civil and Environmental Engineering, Univ. of Washington, More Hall 121F, BOX 352700, Seattle, WA 98195-2700 (corresponding author). Email: yinhai@uw.edu; yinhai@u.washington.edu

Note. This manuscript was submitted on March 17, 2020; approved on October 7, 2020; published online on March 27, 2021. Discussion period open until August 27, 2021; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Transportation Engineering, Part A: Systems***, © ASCE, ISSN 2473-2907.

to measure vehicle occupancy at the level of each state or urban area.

To address these limitations, this research developed a method for estimating AVO levels exclusively for buses, based on data that is easily available nationwide and regularly updated. The purpose of this study is to provide and implement a practical approach for estimating bus occupancy rates for each US state and Washington, DC.

A multisource data collection and aggregation process was used for nationwide bus occupancy estimate to overcome the challenges of overrepresentation (e.g., public transit buses), underrepresentation (e.g., school and tourist buses), and the scarcity of data samples in some comparatively less-populated areas. The technique framework allows the use of information from multiple data types while addressing data bias by applying different weights to different data sources. Specifically, bus occupancy rates were estimated separately for each of three categories: transit bus, school bus, and motorcoach. The average total bus occupancy was estimated by aggregating the AVOs for these three categories, weighted by their annual vehicle miles traveled (VMT).

The contributions of this research can be summarized as follows:

- Proposed a practical method for estimating AVO levels for buses. Since the data this method relies on are easily available nationwide and are regularly updated, this method can be used to regularly update bus AVO estimation.
- Although the method developed in this study focused on state-level AVO calculation, it can be easily extended to urbanized area level.
- Investigated the occupancy rates of school buses and motorcoaches, which have seldom been studied.
- Proposed several regression models to address the missing data issues when estimating vehicle occupancy rates.
- Provided a practical user guide for collecting data and measurement calculation, which can be of great help for traffic engineering practitioners.

Literature Review

AVO values are typically obtained by methods of roadside video recording. Through setting up a roadside monitoring team to count passengers in the passing vehicles, Heidtman et al. (1997) concluded that the approach was most efficient in gathering data on low functional corridors but less effective on multilane freeways. Hao et al. (2011) designed an imaging technique to make the vehicle's passengers more apparent, utilizing infrared rays while using a video recording device simultaneously. An analysis of vehicle occupancy performed in Arizona used the carousel approach as a supplement to roadside observations for AVO estimate and applied a carousel system that used multiple vehicles with several observers in the traffic flow to identify passengers in other vehicles (MAG 2013).

In addition to technical methodologies, investigators have also used surveys and crash data sets to quantify vehicle occupancy. Gan et al. (2005) built a user-friendly software system that could be used to calculate occupancy rates in Florida from multiple years of crash data; the system also provides a dedicated GIS interface to enable the selection of regional features and present estimates of occupancy. Gan et al. (2008) also carried out an in-depth AVO prediction analysis utilizing existing traffic accident data comodeled with other variables such as area, city, hour, and week. However, they admit in their study that the results are highly vulnerable to potential bias arising from traffic crash reports. Jung and Gan (2011) provided a detailed procedure for calculating AVOs at the individual location, facility type, and county level, as well as a thorough sampling framework for choosing locations and dates for the data collection on different types of facilities.

As stated in the section "Introduction," although these methods function well in certain scenarios, they have limited scopes and can not be applied efficiently to nationwide bus occupancy estimation. To address their limitations, this study developed a method to calculate nationwide AVO rates specifically for buses, relying on data that are easily available nationwide and updated regularly.

Goal and Scope of the Research

The goal of this research is to provide and implement a practical method to estimate bus occupancy rate for each of the US states and Washington, DC. Buses are defined as Class 4 vehicles in FHWA's 13 Vehicle Category Classification (FHWA 2013): "all vehicles manufactured as traditional passenger-carrying buses with two axles and six tires or three or more axles. This category includes only traditional buses (including school buses) functioning as passenger-carrying vehicles. Modified buses should be considered to be a truck and should be appropriately classified."

This study further divided buses into three categories: transit bus (metro bus), school bus, and motorcoach, because their AVOs differ significantly. The first two categories are easy to understand. The third, motorcoach, is defined as a vehicle designed for long-distance transportation of passengers, characterized by integral construction with an elevated passenger deck located over a baggage compartment (American Bus Association 2017).

Methodology

Methodology Framework

As shown in Fig. 1, bus occupancies were estimated separately for each of three categories: transit bus, school bus, and motorcoach. Total average bus occupancy was estimated by aggregating the AVO for the three subgroups

$$\mathrm{AVO}_{Bus} = \frac{\mathrm{AVO}_{Transit} \times \mathrm{VMT}_{Transit} + \mathrm{AVO}_{School} \times \mathrm{VMT}_{School} + \mathrm{AVO}_{Motorcoach} \times \mathrm{VMT}_{Motorcoach}}{\mathrm{VMT}_{Transit} + \mathrm{VMT}_{School} + \mathrm{VMT}_{Motorcoach}} \quad (1)$$

where VMT = annual vehicle miles traveled.

The annual VMT for each bus category can be calculated based on their average vehicle VMT (national level) and the vehicle count data (state level) from the Polk dataset (R. L. Polk & Company, Southfield, Michigan) which contains detailed vehicle registration information (Polk City Directory 2018). The specific numbers for average vehicle VMT are as follows: $Average\,\mathrm{VMT}_{Transit} =$ 54,802 km (34,053 mi) per vehicle (DOE 2015); $Average$

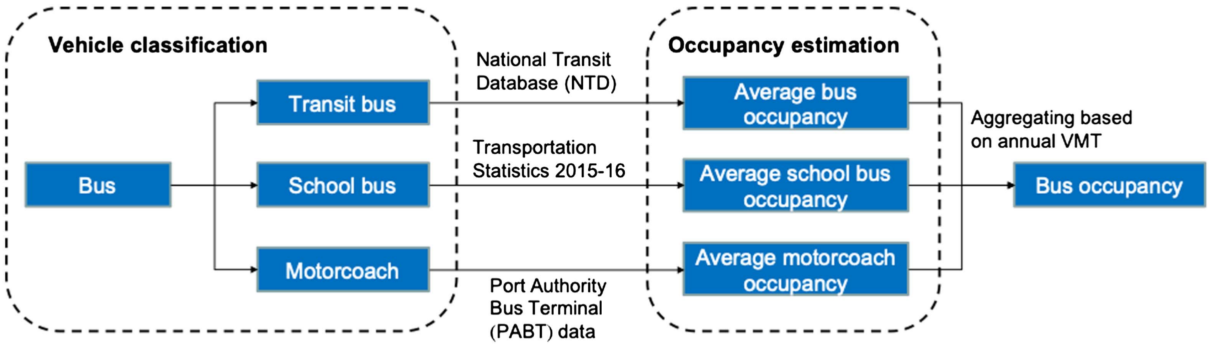

Fig. 1. Methodology framework of bus occupancy estimation.

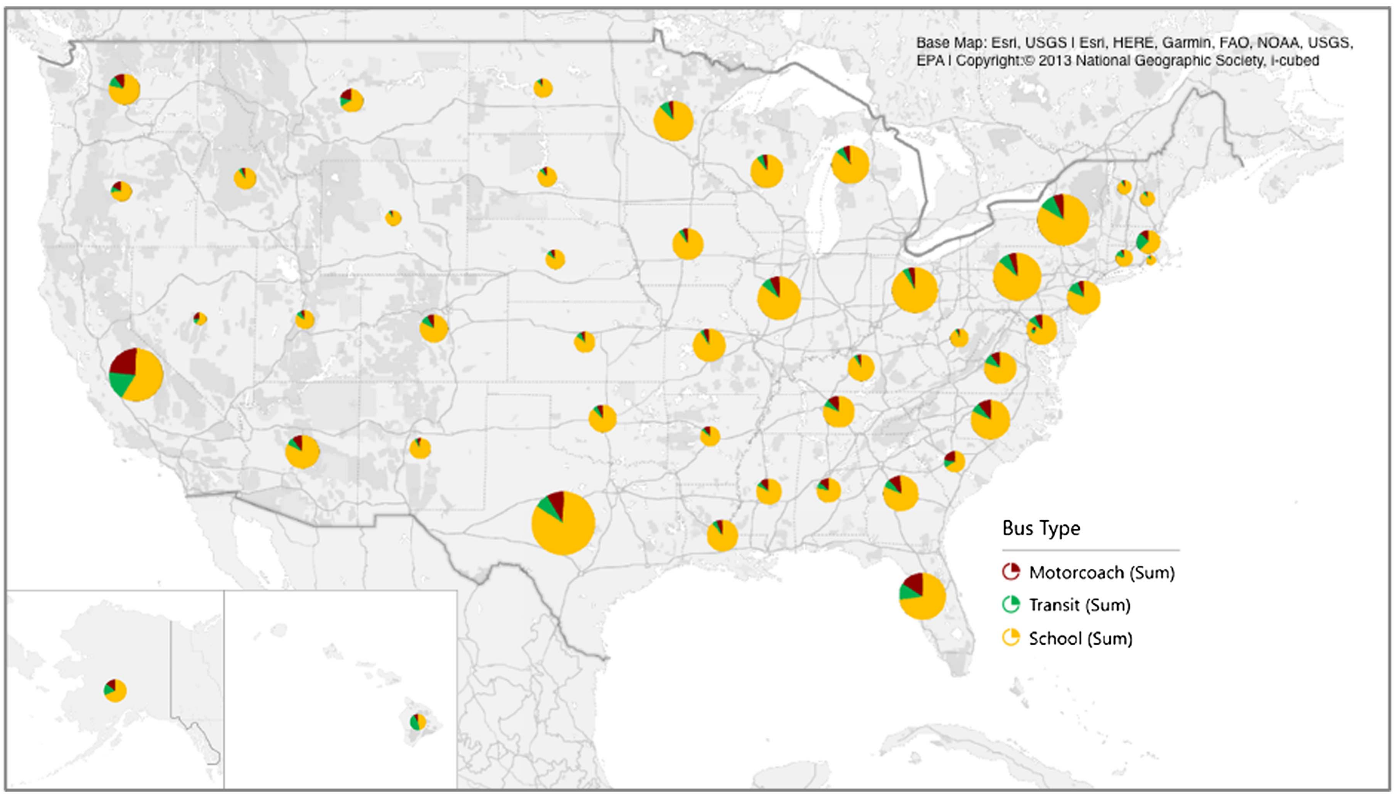

Fig. 2. State-level bus count by type. (Esri, USGS | Esri, HERE, Garmin, FAO, NOAA, USGS, EPA | Copyright: © 2013 National Geographic Society, i-cubed.)

VMT_{School} = 19,312 km (12,000 mi) per vehicle (American School Bus Council 2015); and $Average VMT_{Motorcoach}$ = 61,774 km (38,385 mi) per vehicle (American Bus Association 2017). The state-level bus count by type is shown in Fig. 2, in which the sizes of the circles represent total bus count.

Transit Bus

Data Source

Transit bus occupancy was calculated primarily using the Federal Transit Administration (FTA) National Transit Database (NTD) (FTA 2020). The latest dataset is for the year 2017. All public bus companies obtaining federal funding are required to report to the FTA annually on operating and financial data, including managed transit modes, the number of vehicles in operation, and hours of service. Also included in the NTD are passenger and vehicle miles traveled, which are the two major variables used in occupancy calculation.

Transit organizations are categorized into three categories of reporters: complete reporter, reduced reporter, and rural reporter. Only data from full reporters has been certified as accurate by the CEO of each agency and subject to audit as required by the FHWA (2017). After filtering out unrelated transit modes (e.g., train and ferry transport), the final data set included the following five transit types: commuter bus (CB), demand responsive (DR), motor bus (MB), rapid bus (RB), and trolley bus (TB). A total of 1,051 bus transit agencies have been classified as full reporters for those five modes as of 2016. The NTD data shows a very high reporting rate (i.e., about 99% across all transit modes), and the imputation of missing data is not needed.

Method

Passenger miles traveled (PMT) and vehicle revenue miles (VRM) are two important data elements to be used in the occupancy calculation of transit buses. VRM refers to miles that vehicles travel while in revenue service. It excludes miles during deadheading, operator training, and maintenance testing. The average occupancy of transit bus can be expressed as

$$\mathrm{AVO}_{transit} = average\_ridership + driver = \frac{\sum_i \mathrm{PMT}_i}{\sum_i \mathrm{VRM}_i} + 1 \quad (2)$$

where $\sum_i \mathrm{PMT}_i$ and $\sum_i \mathrm{VRM}_i$ = total PMT and VRM of all transit agencies in the investigated area. In this study, we include the driver as an occupant when calculating vehicle occupancy rates. Since FTA has different definitions (bus operator is not considered a passenger according to FTA's definition), we add one to the PMT/VRM results to include the driver.

All Wyoming transit agencies were classified as reduced reporter or rural reporter, so no PMT and VRM information was published. To estimate the transit occupancy of Wyoming, a multiple linear regression model has been developed. A previous study shows that the use of transit is closely linked to the local population and economy (Mittal et al. 2017). Therefore, the regression analysis used local gross domestic product (GDP) and population density data as two predictors. Data on population density is obtained from the US Census Bureau (2018). Annual state GDP information can be downloaded from the US Bureau of Economic Analysis (BEA 2018). A total of five years (2012–2016) of data were used to fit the regression model to expand the sample size and provide a more reliable estimation for model parameters. All candidate predictors had a right-dip distribution according to experimental results, so we used the log to logarithmically transform the data. The linear regression model that uses all states' data is expressed as

$$\mathrm{AVO}_{transit}(state) = 1.5122 + 1.0227 \times \log \mathrm{GDP}_{transit}(state) + 0.00226 \times pop\_density(state) \quad (3)$$

where $\mathrm{GDP}_{transit}$ = GDP for transit and ground transportation industry ($\$1 \times 10^6$); and $pop\_density$ = population density of the state [people per 2.59 km^2 (1 mi^2)]. The model fit information is given in Table 1. All the independent variables are statistically significant with p-values less than 0.01. As of 2016, the transit GDP for Wyoming is $\$33 \times 10^6$ and the population density is 5.980 people per 2.59 km^2 (1 mi^2). This gives a prediction of state average transit occupancy of 3.08.

The aforementioned illustrates the general procedure for estimating state-level transit bus occupancy when the data is missing. However, given that there is no urbanized area in Wyoming, it might be more appropriate to fit the regression model only using nearby states that are more similar to Wyoming in terms of transit GDP and population density. Thus, data from the nearby seven states (i.e., Idaho, Montana, North Dakota, South Dakota, Nebraska, Colorado, and Utah) that are relatively less populated were used to fit the model, which gives the following linear equation:

$$\mathrm{AVO}_{transit}(state) = 0.3353 + 0.9294 \times \log \mathrm{GDP}_{transit}(state) + 0.05524 \times pop\_density(state) \quad (4)$$

The model fit information is given in Table 2. The p-values for the independent variables are larger than 0.05, showing that they are not statistically significant for the linear model. This may be caused by the small sample size (only 7) because both variables are significant if we use data from all states as input. Since the main purpose of this linear regression is to predict missing occupancy data rather than investigating which factors affect the bus occupancy rates, a model with insignificant independent variables is still acceptable given its high R^2 metric (0.7188). Fig. 3 presents the ground truth and predicted transit occupancy rates. As can be seen, the model can generate good estimates for transit occupancy based on transit GDP and population density. This gives a prediction of state average transit occupancy of 3.92, which is slightly higher than the regression estimate using all states as model input.

Table 1. Linear regression information for Wyoming's transit occupancy estimation, using data from all states

Variable	Estimate	Standard error	t-value	p-value
(Intercept)	1.5122	0.6073	2.490	**0.01343**
$\log(\mathrm{GDP}_{transit})$	1.0227	0.1126	9.081	**<0.0001**
pop_density	0.00266	6.96×10^{-4}	3.241	**0.00136**
$R^2 = 0.3784$				

Note: Bold values denote statistical significance at the $p < 0.05$ level.

Table 2. Linear regression information for Wyoming's transit occupancy estimation, using data from seven nearby states

Variable	Estimate	Standard error	t-value	p-value
(Intercept)	0.33532	3.11915	0.108	0.915
$\log(\mathrm{GDP}_{transit})$	0.92943	0.85886	1.082	0.287
pop_density	0.05524	0.03856	1.435	0.162
$R^2 = 0.7188$				

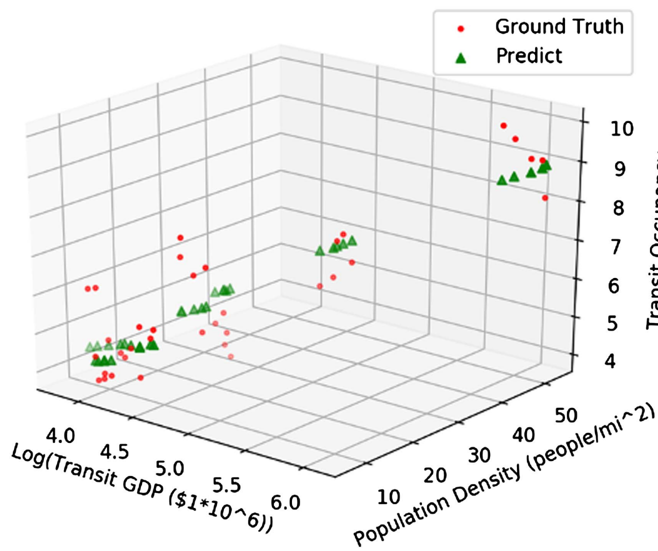

Fig. 3. Scatter plot of transit occupancy versus transit GDP and population density.

School Bus

Data Source

"US State by State Transportation Statistics 2015–16," (School Bus Fleet 2016) reported by School Bus Fleet were used to calculate average school bus occupancy for each state. The report provides a breakdown for each state of the number of K–12 public and private school students transported daily, the number of school buses, and the total annual route mileage. It should be noted that the number of school buses tells us how many school buses each state has

J. Transp. Eng., Part A: Systems, 2021, 147(6): 05021001

but does not tell us how the buses are scheduled (e.g., one bus may have multiple runs for a single day). When we directly divide the number of students transferred daily by the total buses count, the resulting occupancy rates tend to be extremely high (15–160), exceeding the maximum capacity of a standard school bus (72 passengers).

Therefore, average school bus occupancy for each state was estimated by dividing the number of students transported annually by the total annual route mileage. Data reported by the American School Bus Council (ASBC) (American School Bus Council 2015) were also utilized to derive the following important information: (1) the average distance from home to school for bus riders (ASBC estimate, miles) = 8.05 km (5 mi); and (2) the length of the average school year (days) = 180 days.

Education statistics from the website Governing (Governing 2015) were used to estimate the average school bus occupancy for states with missing annual route mileage data. The Governing website provides school district data including total districts, total schools, total public school enrollment, and average district enrollment for each state.

Method

The average school bus occupancy for each state can be estimated based on the following equation:

$$\mathrm{AVO}_{school} = \frac{\sum_i \mathrm{PMT}_i}{\sum_i \mathrm{VMT}_i} + 1 = \frac{Number\,of\,Student\,Transported\,Daily \times 180 \times (5 \times 2)}{Total\,Annual\,Route\,Mileage} + 1 \tag{5}$$

where (5 × 2) = average round trip distance from home to school. We acknowledge that using a national average bus route length will affect the accuracy of the estimation. However, considering that students generally live around their schools no matter in urbanized or unurbanized areas, the actual school bus route lengths will not deviate much from the national average. Therefore, the estimation errors should be within a reasonable bound.

The total annual route mileage data was missing for 14 states. To address the missing data issue, a local factor based weighted model was developed by incorporating local factors (e.g., total school enrollment, average district enrollment, total districts, total schools, total students transported daily, and total yellow school buses) as measurements of similarities among states. The data for these local factors can be found on the websites for School Bus Fleet and Governing. The estimation equation is

$$\mathrm{AVO}_{school}(State\ i) = \sum_{j=1}^{N} w(i,j) \times \mathrm{AVO}_{school}(State\ j) \tag{6}$$

where the weight $w(i,j)$ is defined as an indicator to describe the similarity between State i and State j. Let $F_l(\mathrm{i})$ be a local factor of State i, then the weight $w(i,j)$ can be defined as

$$w(i,j) = \frac{\sum_{l=1}^{L}\left(1 - \frac{|F_l(\mathrm{i}) - F_l(\mathrm{j})|}{max\{F_l(\mathrm{i}), F_l(\mathrm{j})\}}\right)}{\sum_{j=1}^{N}\sum_{l=1}^{L}\left(1 - \frac{|F_l(\mathrm{i}) - F_l(\mathrm{j})|}{max\{F_l(\mathrm{i}), F_l(\mathrm{j})\}}\right)} \tag{7}$$

where the design of the item $[|F_l(\mathrm{i}) - F_l(\mathrm{j})|]/[max\{F_l(i), F_l(\mathrm{j})\}]$ can guarantee that the value ranges between 0 and 1.

In this way, if the local factors of State i are close to those of State j, a high value of weight $w(i,j)$ will be generated. States with similar local factors to those of the target state will have more impact on the estimation. Based on the developed local factors–based weighted model, the average school bus occupancies for all states can be estimated.

Motorcoach

Data Source

Data provided by the Port Authority of New York and New Jersey (PANYNJ) for the Port Authority Bus Terminal (PABT) in New York City (PANYNJ 2015), the largest bus terminal in the US, were used to calculate motorcoach occupancy rates. The data include detailed bus and passenger hourly arrivals and departures for 256 routes identified by origin and destination. Motorcoach occupancy for each route can be estimated by dividing the total passenger departures and arrivals by total bus departures and arrivals. Since interstate buses usually cross several states, the average motorcoach occupancy for a state can be estimated by aggregating across all routes that pass through the state.

The Motorcoach Census Report 2015 (American Bus Association 2017) published by the American Bus Association Foundation and John Dunham & Associates was also used as a data source for collecting details on the occupancy of motor buses at the national level. In addition, some local reports, such as Motor Coach Tourism in Savannah, Georgia, created for the City of Savannah by Armstrong Atlantic State University (2013), were also referred to for verification of the results.

Method

Data from the Port Authority Bus Terminal provided thorough total bus and passenger hourly arrivals and departures for 256 routes specified by destination and origin. The occupancy of motorcoaches on a route can be calculated using the following equation:

$$\mathrm{AVO}_{motorcoach}(route) = \frac{Daily\,Passenger\,Departures + Daily\,Passenger\,Arrivals}{Daily\,Bus\,Departures + Daily\,Bus\,Arrivals} + 1 \tag{8}$$

Both the arrivals and departures data are used here to calculate the average occupancy of the motorcoach. As the interstate bus typically passes through many states, the average occupancy of the motorcoach for a state can be computed using the following equation

$$\mathrm{AVO}_{motorcoach}(state) = \frac{\sum_{route \ni state} \mathrm{AVO}_{motorcoach}(route) \times bus\_count(route)}{\sum_{route \ni state} bus\_count(route)} \tag{9}$$

Table 3. Weighted linear regression model for estimating Nevada's motorcoach occupancy

Variable	Estimate	Standard error	t-value	p-value
(Intercept)	58.81	4.82	12.19	$\mathbf{5.45 \times 10^{-11}}$
$\log(\text{GDP}_{transit})$	−2.82	0.71	−3.97	**0.0007**
pop_density	−0.00073	0.0050	−0.147	0.88
$R^2 = 0.4419$				

Note: Bold values denote statistical significance at the $p < 0.05$ level.

Table 4. Cities ranked by AllTransit performance score

Rank	Name	Score	Transit trips/week	% of commuters using transit
1	New York	9.6	13,771	59.00
2	San Francisco, California	9.6	11,896	36.50
3	Washington, DC	9.3	8,095	37.60
4	Boston, Massachusetts	9.3	8,241	34.80
5	Jersey City, New Jersey	9.3	4,741	50.10
6	Chicago, Illinois	9.1	6,190	29.50
7	Philadelphia, Pennsylvania	9	5,879	26.20
8	Portland, Oregon	8.9	8,189	13.40
9	Cleveland, Ohio	8.8	8,013	10.40
10	Newark, New Jersey	8.7	5,135	26.70
11	Miami, Florida	8.5	3,659	11.40
12	Seattle, Washington	8.5	6,152	23.00
13	Baltimore, Maryland	8.4	3,930	18.90
14	St. Louis, Missouri	8.4	4,495	9.90
15	Minneapolis, Minnesota	8.3	4,677	14.20

Source: Data from AllTransit (2020).

where $route \ni state$ reflects accumulation through all routes passing through the state. The average motorcoach occupancies for 25 states were determined based on Eq. (9).

A geographical distance-based, weighted linear regression model was developed to approximate average motorcoach occupancies for the remaining states. Population density and GDP in transit and ground passenger transportation were used as two regression predictors. The general equation of regression is formulated as

$$\text{AVO}_{motorcoach}(state) = \beta_0 + \beta_1 \times \log \text{GDP}_{transit}(state) + \beta_2 \times pop\_density(state) \tag{10}$$

Given that the motorcoach operation is spatially correlated, adjacent states appear to have comparable average occupancies of motorcoaches. The geographical distance between states was thus used as weights of regression. Weighted linear regression estimates the regression parameters by minimizing the weighted sum of squared residuals (WSSR), which is

$$\text{WSSR}_i = \sum_j w(i,j) \times (y_j - \hat{y}_j)^2 \tag{11}$$

where the weight of regression $w(i,j) = 1/dist(i,j)$ is defined as the inversed distance between two states. The central position for each state and the distances between states can be obtained from the DistanceFromTo website. The closer the two states are, the greater the weight of regression will be, indicating a strong association between the two states. Notice that this would lead to a particular

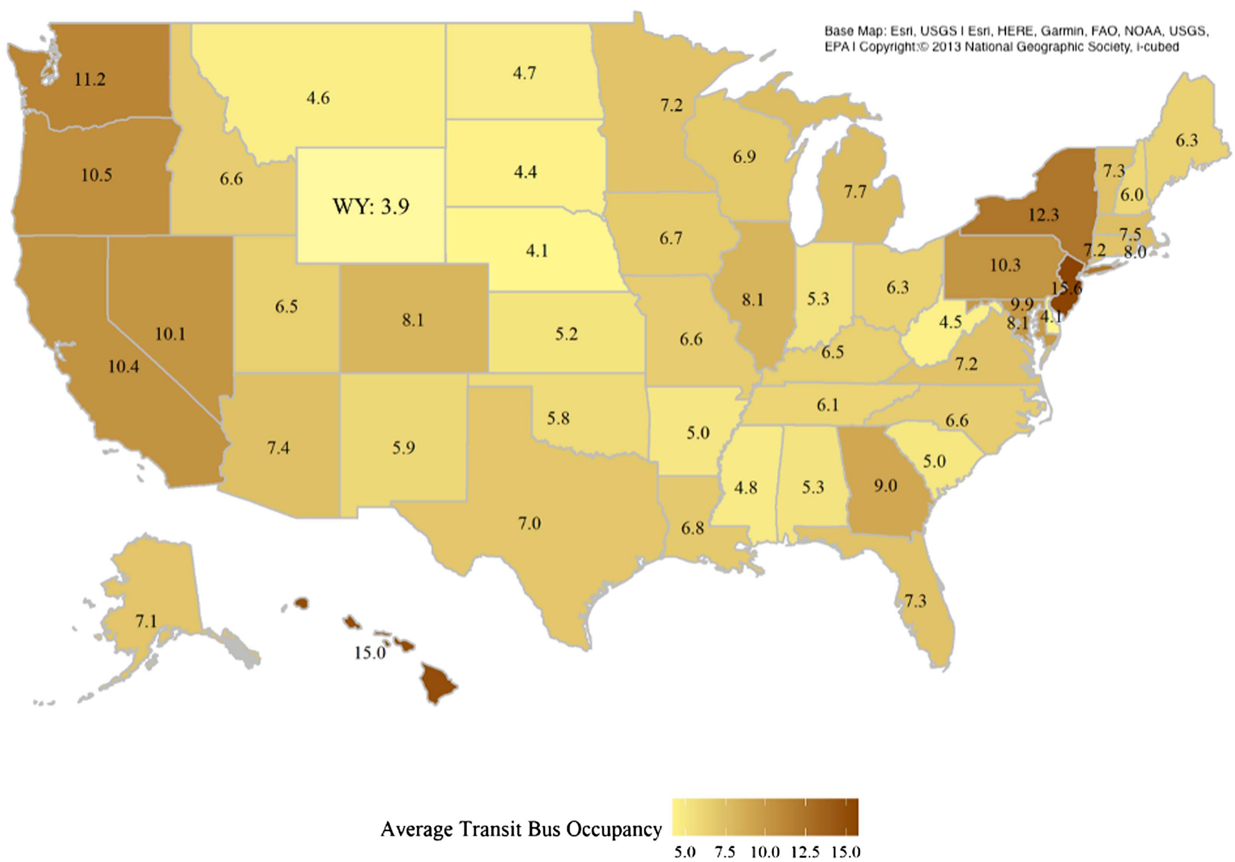

Fig. 4. Average transit bus occupancy by state. (Esri, USGS | Esri, HERE, Garmin, FAO, NOAA, USGS, EPA | Copyright: © 2013 National Geographic Society, i-cubed.)

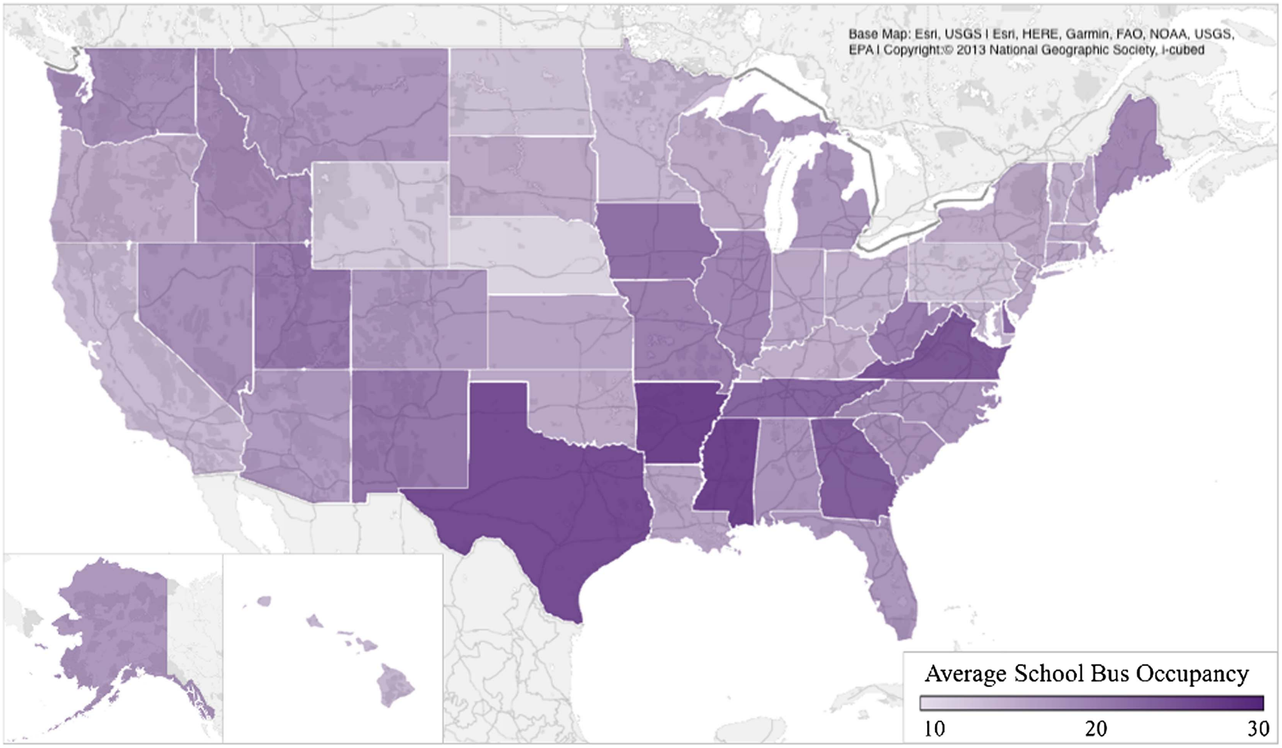

Fig. 5. Average school bus occupancy by state. (Esri, USGS | Esri, HERE, Garmin, FAO, NOAA, USGS, EPA | Copyright: © 2013 National Geographic Society, i-cubed.)

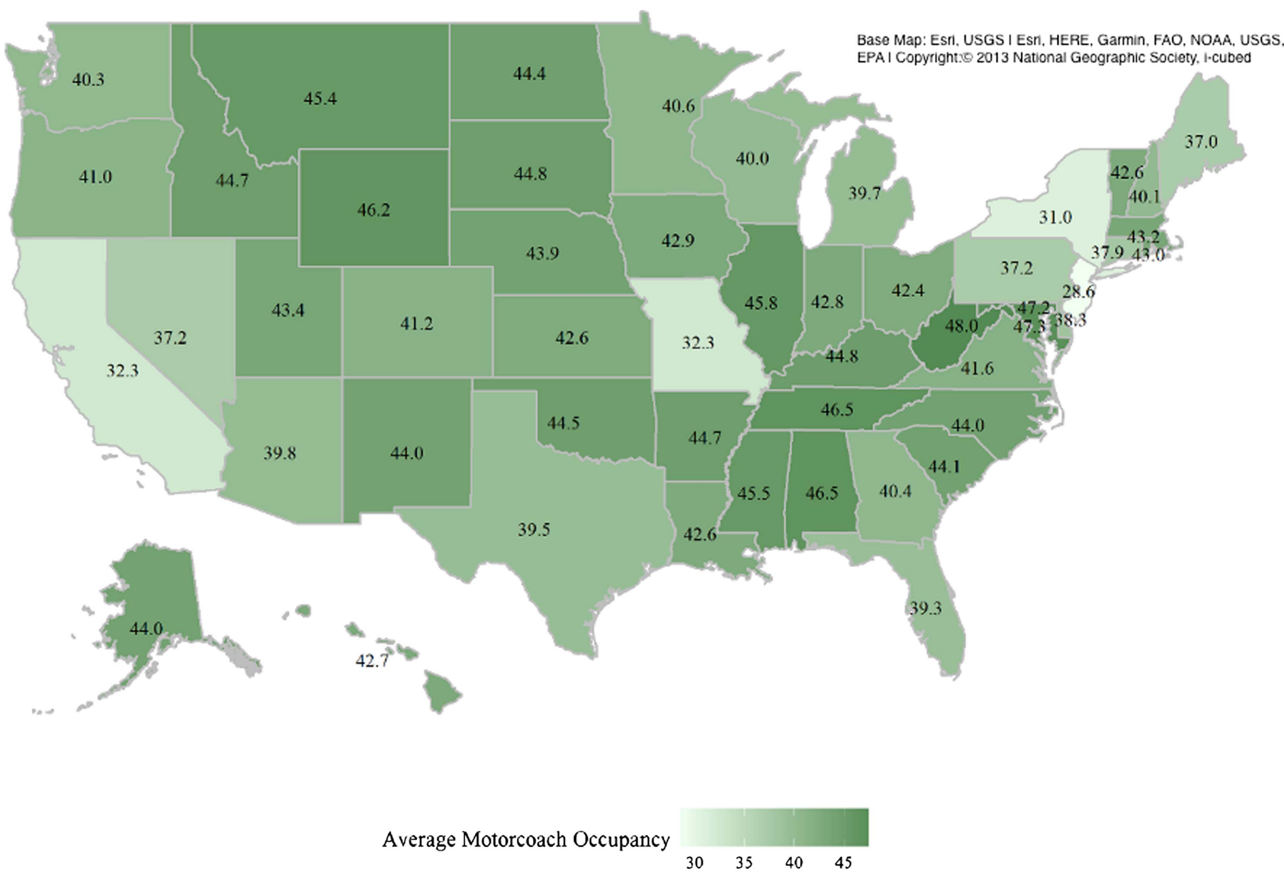

Fig. 6. Average motorcoach occupancy by state. (Esri, USGS | Esri, HERE, Garmin, FAO, NOAA, USGS, EPA | Copyright: © 2013 National Geographic Society, i-cubed.)

linear regression model for each state in which we need to approximate the average occupancy. One example model for estimating Nevada's motorcoach occupancy is provided in Table 3. The model has an R^2 metric 0.44, and transit GDP tends to be negatively correlated with motorcoach occupancy. Again, the high p-value for population density may be caused by the relatively small sample size (35).

Results

This section describes the bus occupancy estimation results for each subcategory, followed by the final aggregated bus occupancy rate for each state.

Transit Bus

Fig. 4 shows the average transit occupancy by state. In general, transit occupancy is higher on the east and west coasts. This is consistent with transit performance score ranking provided by the AllTransit platform (AllTransit 2020), as given in Table 4 in which cities like New York and San Francisco have high transit scores. The AllTransit platform gives an overall transit performance score for each local city based on the number of transit routes, transit trips, jobs accessible in a 30-min transit trip, and commuters who use transit.

Wyoming has the lowest transit bus occupancy (3.9), which is consistent with the AllTransit ranking, in which Wyoming's transit performance score is only 0.6 out of 10. According to AllTransit (2020), in Cheyenne, the capital of Wyoming, households on average have only three transit routes within 804 m (0.5 mi) and only 0.7% commuters using transit.

Being consistent with transit performance scores serves as a kind of validation for our results. Another guarantee of accuracy for transit occupancy estimation is that NTD has a unified data collection plan (National Transit Database Sampling Manual) for transit agencies (FTA 2009). Considering that the FTA has made specific data sampling plans, data collection guidelines, and requires that the estimates of annual consumed data meet 95% confidence and 10% precision levels, our proposed method should provide valid estimations.

School Bus

The estimated average school bus occupancy rates for the states are summarized in Fig. 5. School bus occupancy rates distributed in the range of 7–30, with Mississippi and Arkansas having the highest rates, and Nebraska having the lowest rate. According to the School Bus Fleet (2016) report, Mississippi and Arkansas have the highest number of students served by each school bus (around 155 students), while Nebraska has the lowest (around 15 students per school bus). This consistency serves as a validation for our estimation result for the school bus occupancy estimation.

Motorcoach

The complete state-level average motorcoach occupancy rates are presented in Fig. 6. Considering that the average capacity of motorcoaches is around 56 and most of the states have motorcoach occupancy rates around 40, the results seem to be reasonable. Different from the transit bus occupancy distribution, the average motorcoach occupancy rates are higher in less-populated areas and lower in more populated areas. This is easy to understand since people will rely more on motorcoaches if there are few transit buses in the less-populated areas.

For validation, according to the 2015 Motorcoach Census Report (American Bus Association 2017), the national average motorcoach occupancy is 36.4, which generally agrees with the results shown in Fig. 6. Also, according to Motor Coach Tourism in Savannah, Georgia, the City of Savannah has an estimated of motorcoach occupancy of 41.7, and our method's result (46.1) is close to their estimate (Armstrong Atlantic State University 2013).

Final Bus Occupancy

Table 5 provides the final bus occupancy rate for each state and Washington, DC. The results were obtained by aggregating the

Table 5. Bus occupancy results by state

State name	Transit occupancy	School bus occupancy	Motorcoach occupancy	Bus occupancy
Alabama	5.25	16.06	47.48	24.73
Alaska	7.10	14.92	45.06	23.80
Arizona	7.39	14.44	40.87	20.18
Arkansas	4.97	29.84	45.75	31.49
California	10.35	9.26	33.40	20.30
Colorado	8.11	15.26	42.30	18.90
Connecticut	7.20	12.52	38.94	15.32
Delaware	4.14	23.17	39.28	22.06
Florida	7.27	15.22	40.31	20.61
Georgia	9.00	25.21	41.53	27.94
Hawaii	14.98	9.67	43.79	18.01
Idaho	6.59	17.50	45.80	21.04
Illinois	8.13	18.21	47.13	23.32
Indiana	5.34	12.52	43.83	16.26
Iowa	6.72	21.83	43.94	24.78
Kansas	5.20	12.98	43.62	18.05
Kentucky	6.48	11.29	45.81	14.84
Louisiana	6.79	13.29	43.70	17.53
Maine	6.29	17.61	38.00	17.76
Maryland	9.91	10.13	48.19	15.93
Massachusetts	7.54	12.45	44.19	19.38
Michigan	7.65	15.17	40.69	19.44
Minnesota	7.16	9.78	41.70	12.66
Mississippi	4.80	30.07	46.57	32.59
Missouri	6.61	19.33	33.40	21.02
Montana	4.55	15.42	46.49	32.28
Nebraska	4.07	7.09	44.96	20.03
Nevada	10.14	16.28	38.28	24.93
New Hampshire	5.96	11.58	41.09	14.77
New Jersey	15.63	12.27	29.47	15.69
New Mexico	5.95	20.75	45.02	22.56
New York	12.32	12.26	31.86	15.44
North Carolina	6.61	16.47	45.03	21.68
North Dakota	4.72	8.43	45.49	11.58
Ohio	6.29	10.79	43.42	14.60
Oklahoma	5.78	12.13	45.53	17.34
Oregon	10.54	11.85	42.09	22.17
Pennsylvania	10.27	7.95	38.19	11.53
Rhode Island	8.03	12.25	44.03	16.55
South Carolina	5.00	17.01	45.23	26.50
South Dakota	4.40	12.09	45.79	17.49
Tennessee	6.09	23.77	47.48	28.69
Texas	6.97	28.20	40.58	26.98
Utah	6.50	20.55	44.46	23.12
Vermont	7.31	11.07	43.68	14.76
Virginia	7.24	26.19	42.71	27.08
Washington	11.24	16.87	41.39	21.90
Washington, DC	8.09	12.23	48.32	20.40
West Virginia	4.47	19.67	49.04	21.34
Wisconsin	6.90	12.29	41.06	14.90
Wyoming	3.92	7.23	47.26	12.63

Table 6. Inputs and outputs for the transit bus occupancy calculation program

Script name: 01_transit_bus.R	
File	Description
	Input
states.csv	US state and federal district names, codes, abbreviations, and coordinates
gdp_state.csv	2012–2016 state-level all industry GDP and GDP in transit and ground passenger transportation
pop_state.csv	2012–2016 state-level population and population density
2016_Service.csv	2016 National Transit Database service table
2016_Agency.csv	2016 National Transit Database agency information table
2015_Service.csv	2015 National Transit Database service table
2015_Agency.csv	2015 National Transit Database agency information table
2014_Service.csv	2014 National Transit Database service table
2014_Agency.csv	2014 National Transit Database agency information table
2013_Service.csv	2013 National Transit Database service table
2013_Agency.csv	2013 National Transit Database agency information table
2012_Service.csv	2012 National Transit Database service table
2012_Agency.csv	2012 National Transit Database agency information table
	Output
transit_state.csv	Average transit bus occupancy and VMT in each of the 50 US states and Washington, DC

average vehicle occupancies for transit bus, school bus, and motorcoach weighted by their annual vehicle miles traveled. The mean state-level bus occupancy rate is 20.29, with a standard deviation of 5.24.

Implementation

To help practitioners in civil engineering to make better use of this study, we published the data, code, and user guide in Zhu (2020). The user guide has a detailed description of computation environment setup, data preparation, and bus occupancy calculation. Table 6 gives an example of the inputs and outputs for the transit bus occupancy calculation program. By directly running the code, users can reproduce the bus occupancy calculation results as described in this study. Users can also get updated bus occupancy estimation results if the source data are updated accordingly.

Summary and Discussion

This study implemented a statistically efficient and practical method for estimating the bus occupancy rate for each US state and Washington, DC. Bus occupancies were estimated separately for each of three categories: transit bus, school bus, and motorcoach. The average total bus occupancy was estimated by aggregating the average vehicle occupancies for the three categories weighted by annual VMT.

A multisource data collection and integration framework was used for nationwide bus occupancy estimation: transit bus occupancy was calculated by primarily using the FTA NTD; "US State by State Transportation Statistics 2015–16," reported by School Bus Fleet were used to calculate average school bus occupancy for each state; and data provided by the PANYNJ for the PABT in New York City, the largest bus terminal in the US, were used to calculate motorcoach occupancy rates. The proposed methodology framework utilizes information from all types of data while dealing with data bias by assigning different weights to different data sources. Results show that the mean state-level bus occupancy rate is 20.29, with a standard deviation of 5.24.

This study will continue assisting the FHWA in providing vehicle occupancy data to states following Title 23 of the US Code of Federal Regulations, Part 490 National Performance Management Measures. To facilitate future updates of the results, we also developed software code in R language and a user guide. The user guide explains how to reproduce the bus occupancy rate for each state and Washington, DC.

With data, code, and user guide published, this study will help traffic engineers, planners, and decision-makers in the following cases (1) calculating and updating state-level bus occupancy rates; (2) obtaining operation measurements like passenger miles for transit bus, school bus, and motorcoach; (3) calculating trip metrics related to passengers, such as person delays and passenger miles traveled; (4) identifying routes or areas with high demand needing service expansion; and (5) service optimization, such as determining the number of buses required for certain corridors or areas.

The following are some recommendations to improve the estimation methodologies in the future:

1. work with providers of the various relevant data sources to ensure access to regularly updated new data;
2. initiate a training program for the software code to ensure the results can be easily updated in the future; and
3. make use of alternative methods and datasets to validate results.

Data Availability Statement

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request. The data include state-level vehicle occupancy rates for transit bus, school bus, motorcoach, and the final aggregated bus occupancy; and state-level bus count by type.

Acknowledgments

The authors would like to thank the US Department of Transportation's Federal Highway Administration (FHWA) and the Pacific Northwest Transportation Consortium (PacTrans) at Regional University Transportation Center (UTC) for Federal Region 10, for funding this research. The authors would also like to thank Daniel Jenkins and Roger Mingo for providing constructive suggestions and Ruoxuan Wang for her helpful edits.

References

AllTransit. 2020. "AllTransitTM performance score ranking." Accessed March 8, 2021. https://alltransit.cnt.org/rankings/.

American Bus Association. 2017. "Motorcoach census report: A study of the size and activity of the motorcoach industry in the United States and Canada in 2015." Accessed March 8, 2020. https://www.buses.org/assets/images/uploads/pdf/Motorcoach_Census_2015.pdf.

American School Bus Council. 2015. "Environmental benefits—Fact: You can go green by riding yellow." Accessed March 8, 2021. http://www.americanschoolbuscouncil.org/issues/environmental-benefits.

Armstrong Atlantic State University. 2013. "Motorcoach tourism in Savannah." Accessed June 17, 2020. http://www.savannahga.gov/DocumentCenter/View/4364/FINAL-CoachStudy_AASU_031213?bidId=.

BEA (Bureau of Economic Analysis). 2018. "Regional data: GDP & personal income." Accessed October 15, 2018. https://apps.bea.gov/itable/iTable.cfm?ReqID=70&step=1#reqid=70&step=1&isuri=1.

DOE. 2015. "Maps and data—Average annual vehicle miles traveled of major vehicle categories." Accessed March 8, 2021. https://afdc.energy.gov/data/10309.

FHWA (Federal Highway Administration). 2013. "Traffic monitoring guide, appendix C., vehicle types." Accessed March 8, 2021. https://www.fhwa.dot.gov/policyinformation/tmguide/tmg_2013/vehicle-types.cfm.

FHWA (Federal Highway Administration). 2017. "Requirement to estimate vehicle occupancy for cars, buses, and trucks." Accessed March 8, 2021. https://www.gpo.gov/fdsys/pkg/FR-2017-01-18.

FTA (Federal Transit Administration). 2009. *National transit database sampling manual*. Rockville, MD: FTA.

FTA (Federal Transit Administration). 2020. "The national transit database." Accessed June 15, 2020. https://www.transit.dot.gov/ntd.

Gan, A., R. Jung, K. Liu, X. Li, and D. Sandoval. 2005. *Vehicle occupancy data collection methods*. Miami: Lehman Center for Transportation Research, Florida International Univ.

Gan, A., K. Liu, L. Shen, and R. Jung. 2008. "Prototype information system for estimating average vehicle occupancies from traffic accident records." *Transp. Res. Rec.* 2049 (1): 29–37. https://doi.org/10.3141/2049-04.

Governing. 2015. "Total school districts, student enrollment by state and metro area." Accessed March 8, 2021. https://www.governing.com/gov-data/education-data/school-district-totals-average-enrollment-statistics-for-states-metro-areas.html.

Hao, X. L., H. J. Chen, Y. Y. Yang, C. Yao, H. Yang, and N. Yang. 2011. "Occupant detection through near-infrared imaging." *Tamkang J. Sci. Eng.* 14 (3): 275–283.

Heidtman, K., B. Skarpness, and C. Tornow. 1997. *Improved vehicle occupancy data collection methods*. Rep. No. FHWA-PL-98-042. Washington, DC: Office of Highway Information Management, FHWA.

Jung, R., and A. Gan. 2011. "Sampling and analysis methods for estimation of average vehicle occupancies." *J. Transp. Eng.* 137 (8): 537–546. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000238.

MAG (Maricopa Association of Governments). 2013. *MAG 2012 vehicle occupancy study: Final report*. Phoenix: MAG.

Mittal, S., H. Dai, S. Fujimori, T. Hanaoka, and R. Zhang. 2017. "Key factors influencing the global passenger transport dynamics using the AIM/transport model." *Transp. Res. Part D* 55 (Aug): 373–388. https://doi.org/10.1016/j.trd.2016.10.006.

PANYNJ (Port Authority of New York and New Jersey). 2015. *Port authority bus terminal (PABT) data*. New York: PANYNJ.

Polk City Directory. 2018. *Automobile data*. Dallas: Data Axle.

School Bus Fleet. 2016. *U.S. state by state transportation statistics 2015–16*. Torrance, CA: Bobit Business Media.

US Census Bureau. 2018. *State population totals and components of change: 2010–2017*. Suitland, MD: US Census Bureau.

Zhu, M. 2020. "Bus occupancy." GitHub. Accessed June 17, 2020. https://github.com/MeixinZhu/bus_occupancy.