Depot-Based Vehicle Data for National Analysis of Medium- and Heavy-Duty Electric Vehicle Charging

Matthew Bruchon, Brennan Borlaug, Bo Liu,
Tim Jonas, Jiayun Sun, Nhat Le, and Eric Wood

National Renewable Energy Laboratory

Produced under direction of the U.S. Environmental Protection Agency by the National Renewable Energy Laboratory (NREL) under Interagency Agreement IAG-19-16388.

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.

Contract No. DE-AC36-08GO28308

Strategic Partnership Project Report
NREL/TP-5400-88241
February 2024

Depot-Based Vehicle Data for National Analysis of Medium- and Heavy-Duty Electric Vehicle Charging

Matthew Bruchon, Brennan Borlaug, Bo Liu,
Tim Jonas, Jiayun Sun, Nhat Le, and Eric Wood

National Renewable Energy Laboratory

Suggested Citation
Bruchon, Matthew, Brennan Borlaug, Bo Liu, Tim Jonas, Jiayun Sun, Nhat Le, and Eric Wood. 2024. *Depot-Based Vehicle Data for National Analysis of Medium- and Heavy-Duty Electric Vehicle Charging*. Golden, CO: National Renewable Energy Laboratory. NREL/TP-5400-88241. https://www.nrel.gov/docs/fy24osti/88241.pdf.

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.

Contract No. DE-AC36-08GO28308

Strategic Partnership Project Report
NREL/TP-5400-88241
February 2024

National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

NOTICE

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Support for the work was provided by the U.S. Environmental Protection Agency (EPA) under Agreement IAG-19-16388. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

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

U.S. Department of Energy (DOE) reports produced after 1991 and a growing number of pre-1991 documents are available free via www.OSTI.gov.

Cover Photos by Dennis Schroeder: (clockwise, left to right) NREL 51934, NREL 45897, NREL 42160, NREL 45891, NREL 48097, NREL 46526.

NREL prints on paper that contains recycled content.

Abstract

Medium- and heavy-duty vehicles (MHDVs) are a major source of greenhouse gases and local criteria air pollutants. Electrifying MHDVs may reduce these harmful emissions, which disproportionately impact disadvantaged communities. Due to their relatively high per-vehicle energy needs, consistent fleet operations, and frequent colocation of multiple vehicles at depots, MHDVs may have more spatially and temporally concentrated charging demands than light-duty passenger electric vehicles. That charging concentration means their electrification may require careful advance planning and coordination to manage potential impacts to the electrical grid via charge management or infrastructure upgrades. However, MHDV duty cycles and parking schedules are highly variable across vocations of operation, and there is a shortage of nationally representative, vocationally diverse public data describing typical MHDV operations. This report summarizes the methodology—designed with national representativeness in mind—used to create a new set of data describing typical daily driving distances, dwell durations, and normalized electric vehicle depot charging load curves for MHDVs. The dataset reflects the subset of MHDV operating patterns that may originate from a consistent depot each day and rely on the same depot for charging. In addition to trucks with depot-centric vocational patterns, the data describes operations of transit buses and school buses, each with a depot-centric focus. The dataset is available to the public and suitable for national analysis. It can inform research, infrastructure planning, and policymaking regarding the electrification of MHDVs.

Acknowledgments

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Support for the work was provided by the U.S. Environmental Protection Agency under Agreement IAG-19-16388. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

The authors would like to thank Susan Burke, Jessica Daniels, Mike Landgraf, James Sanchez, Eric Junga, and Meredith Cleveland from the U.S. Environmental Protection Agency for their guidance and support. We also thank Alicia Birky, Andrew Meintz, Matteo Muratori, and Jason Lustbader at the National Renewable Energy Laboratory for providing early guidance and input that helped shape this work. We especially thank Jessica Daniels from the U.S. Environmental Protection Agency, Aaron Sobel from the U.S. Environmental Protection Agency, and Alicia Birky from the National Renewable Energy Laboratory for providing helpful feedback as peer reviewers for this report.

Data Availability

The dataset summarized in this report is publicly available and can be accessed online via the NREL Data Catalog at https://data.nrel.gov/submissions/231. A combination of public and proprietary data was used in the creation of this report, and we link to public data sources in the body of this report or as citations.

List of Acronyms

API	application programming interface
BEV	battery electric vehicle
dVMT	daily vehicle-miles traveled
GTFS	General Transit Feed Specification
GVWR	gross vehicle weight rating
MHDV	medium- and heavy-duty vehicles
MiSA	Micropolitan Statistical Area
MSA	Metropolitan Statistical Area
NTD	National Transit Database
TAF	Traveler Analysis Framework
VIUS	Vehicle Inventory and Use Survey
VMT	vehicle-miles traveled

Table of Contents

List of Figures

List of Tables

1 Introduction

Medium- and heavy-duty trucks and buses are a major source of greenhouse gases and criteria air pollutants. The greenhouse gas emissions from these types of vehicles increased 76% from 1990 to 2021 in absolute terms and rose as a share within the transportation sector (itself the highest-emitting sector of the economy) [1]. Decarbonization of medium- and heavy-duty vehicles (MHDVs) has been described as an essential component of sector-wide deep decarbonization strategies [2], and one decarbonization pathway holding promise is transitioning from on-board combustion powertrains to battery electric vehicles (BEVs). A transition to BEVs could potentially occur for financial reasons alone, but substantial uncertainty remains. One recent study projects BEVs could reach 40% of total MHDV sales in the United States by 2030 and 83% of sales (66% of vehicle stock) in 2050 based solely on total cost of ownership considerations [3], but another projection suggests that all zero-emissions technologies combined (including non-BEV technologies) may be only 32% of global MHDV stock in 2050 without "strong additional measures" [4]. Public policy has begun to address these uncertainties, with California recently establishing a state-level requirement that zero-emissions technologies make up 100% of MHDV stock where feasible by 2045 [5].

Advance planning can ease potential challenges associated with a transition to electric MHDVs. Relative to light-duty passenger vehicles, MHDVs tend to consume more energy per distance traveled (due to their larger size and weight) and may be more highly utilized, increasing their daily charging demands. Also, MHDVs owned by fleets may operate from a shared business location (or depot) and may have consistent operations, resulting in battery charging demands of electric MHDVs that may be highly spatially and temporally concentrated. These concentrated charging demands may warrant additional care to manage the potential impacts to the electrical grid (whether by managing charge schedules or by upgrading infrastructure). Despite these challenges, MHDVs with relatively short-range daily operations, reliable depot charging, and reasonably long off-shift dwell periods could take advantage of overnight charging without the need for midday, en-route charging and may be relatively amenable to rapid electrification.

However, there is scarce public data to inform which types of MHDVs may be amenable to depot-based charging or to characterize operating patterns of depot-centric MHDVs. Within the depot-centric subset of MHDVs, driving and dwell patterns vary widely across operational vocations, which include garbage trucks, delivery vans, buses, and many more. The National Renewable Energy Laboratory's FleetREDI portal, which includes data from the Fleet DNA database, provides vehicle operating summaries collected from commercial vehicle fleets across a range of vocations, but it is difficult to know how representative of broad nationwide patterns these fleets' operations may be [6,7]. The U.S. Department of Transportation's Vehicle Inventory and Use Survey provides annual mileage and range of operation of the nation's MHDV fleet, but this information is provided in censored form with usage buckets that are not granular enough to support detailed electrification analysis [8]. The Federal Highway Administration's NextGen National Household Travel Survey National Origin-Destination Data provides geographically and temporally resolved MHDV usage information, but it aggregates data at the trip level rather than the vehicle level (complicating estimation of vehicle electrification needs) and aggregates across all body styles and vocations of MHDV [9].

This work addresses a gap in public MHDV data by providing a vocationally diverse dataset of typical driving and dwell patterns that is designed with national representativeness in mind and thus is suitable for national analyses. More specifically, it addresses the question: *what is the distribution of daily driving distances and dwell durations for depot-centric MHDV vocations, and what shape may their average daily depot charging load curve take?* In alignment with the data sources used in this analysis, the report is organized into three MHDV categories, all with depot-centric operation: transit buses, school buses, and trucks (the last of which includes all non-bus MHDV types). This work can inform research, planning, and policymaking to accelerate the electrification of MHDVs.

2 Methodology

We consider three disparate types of MHDV in this report: trucks (which themselves have a great deal of underlying vocational diversity), transit buses, and school buses. Our models for these three vehicle types leverage different underlying datasets and methodologies, but all three produced similarly structured outputs for the relevant vehicle types: percentiles describing the national distribution of daily driving distance, percentiles describing the daily dwell time at the vehicle's primary location of domicile, and normalized average daily depot load curves (assuming all charging occurs at that primary domicile location, spread across dwell durations). We describe the methods used to produce outputs for trucks in Section 2.1, transit buses in Section 2.2, and school buses in Section 2.3.

2.1 Trucks

To model the operating patterns of depot-centric MHDV trucks used for freight and other vocations, we estimated nationally representative behavior using data from a sample of seven regions. To identify which regions might be most nationally representative, all regions were clustered based on publicly available measures pertaining to MHDV operational characteristics with an exemplar region selected from each cluster (Section 2.1.1). Next, data from each exemplar region was analyzed using Geotab's Altitude application programming interface (API) to quantify typical operations (Section 2.1.2). Finally, cluster summaries were aggregated to form national average summaries using data regarding each cluster's population of active vehicles (Section 2.1.3).

2.1.1 Regional Operating Pattern Clustering

Substantial variation in MHDV operating patterns may exist across regions of the United States, but procuring and analyzing a telematics dataset covering the full nation is often cost-prohibitive. As an alternative to licensing a comprehensive nationwide dataset, we selected a set of publicly available regional variables that we anticipate affecting MHDV typical operating patterns, clustered regions of the country using these variables, and analyzed one representative subregion's data for each cluster.

2.1.1.1 Clustering Variable Selection

The objective of our clustering approach is to ensure that the analyzed subregions, taken together, capture as much variation as is feasible across measures that affect operating patterns and potential BEV charging patterns. Given this objective, we considered two broad categories of clustering variables: those that capture variation in travel distance of typical MHDV trips, and those that capture variation in MHDV travel concentration within a region. These two categories were chosen because we anticipate that they are closely related to variation in operating patterns and that they capture qualitatively different dimensions of regional variation. Within these two categories, we prioritized variables that have a substantial range from region to region and that also have good complementarity (i.e., relatively orthogonal rather than very strongly correlated variables).

To capture variation in driving distances, we estimated the average distance of short-haul trips (miles per trip) from the Federal Highway Administration's NextGen National Household Travel Survey National Truck Origin-Destination dataset [9]. This dataset joins trip data from multiple telematics providers and provides summary-level statistics for 583 Federal Highway Administration Travel Analysis Framework (TAF) zones: 447 zones based on Census-defined metropolitan statistical areas (MSAs) or micropolitan statistical areas (MiSAs)—with some MSAs and MiSAs further divided by state lines—and 136 additional zones based on the remaining rural areas of each state [10]. This dataset does not provide an average trip distance measure for a given region, but rather provides counts of trips per grouping of month, TAF zone origin-destination pair, and distance bin (e.g., 0–10 miles). To ensure each analyzed subregion reflected its full MHDV operations, Census-defined regions split into multiple TAF zones across state lines were recombined and treated as one unit, resulting in 526 total subregions. For each of those 526 regions, the dataset was filtered to origin-destination pairs starting or ending in the region and trips shorter than 150 miles to focus on local and regional operating characteristics rather than volume of interstate trips. The table of trip volumes by distance bin was reshaped into a list of repeated distance bins. For example, if 10% of all trips fell into the "0–10 miles" distance bin, then that distance bin would appear in 10% of the resulting list. Average short-haul trip distance was computed by fitting a distribution to this bin-censored data frame, using the SciPy scientific computing library in Python, and taking the average value of the fitted distribution [11]. A gamma distribution was chosen because it has support over positive values and is flexible to take on a wide range of shapes based on observed data.

To capture variation in concentration of MHDV movement, the log of employment-weighted employment density (employees per square mile) was estimated using U.S. Census data. In contrast to the standard unweighted measure of average employment density (employment divided by land area), employment-weighted employment density is the average weighted by the count of employees located within each Census tract in the area [12]. This can be thought of as the average density as perceived by the average employee in the region. Unlike unweighted density, it is not affected by large areas of empty off-road land and thus is more reflective of operations in the areas where more on-road operations occur. Density of employment was chosen over density of population due to this analysis's focus on MHDV fleets, whose domiciles may often be located at a place of business and whose operating patterns may often be business-to-business or business-to-residence rather than residence-to-residence. Three tables were combined to compute employment density for each TAF zone, weighted by Census tract-level employment: ZIP-level employment counts from the Census Bureau's County Business Patterns dataset (March 2020 release), Census tract-level land area from the Census Bureau's 2021 Planning Database, and a ZIP-to-Census tract crosswalk file produced by the U.S. Department of Housing and Urban Development preparation Office of Policy Development and Research [13–15]. A log transformation was chosen for this variable to reduce skew and ensure clustering algorithms could produce relatively similarly sized clusters. Before clustering was conducted, the two clustering variables were standardized to have an average of 0 and standard deviation of 1.

Because the final step of this approach aggregates operating and charging patterns using vehicle population information (Section 2.1.3), regional volume measures (e.g., total population or trip count) were not used for clustering.

2.1.1.2 Clustering Method

We partitioned the 526 modified TAF zones into clusters of similar regions along the dimensions of average short-haul trip distance and employment-weighted employment density. This step used the k-means clustering algorithm, which attempts to minimize within-cluster variance to create coherent partitions of data [16]. k-means was applied separately for two groups of modified TAF zones: those defined by a MSA or MiSA, and more-rural zones. This was done to account for drastic differences in the chosen clustering variables between these two regions and to guarantee representation of rural, suburban, and urban regions in our clusters.

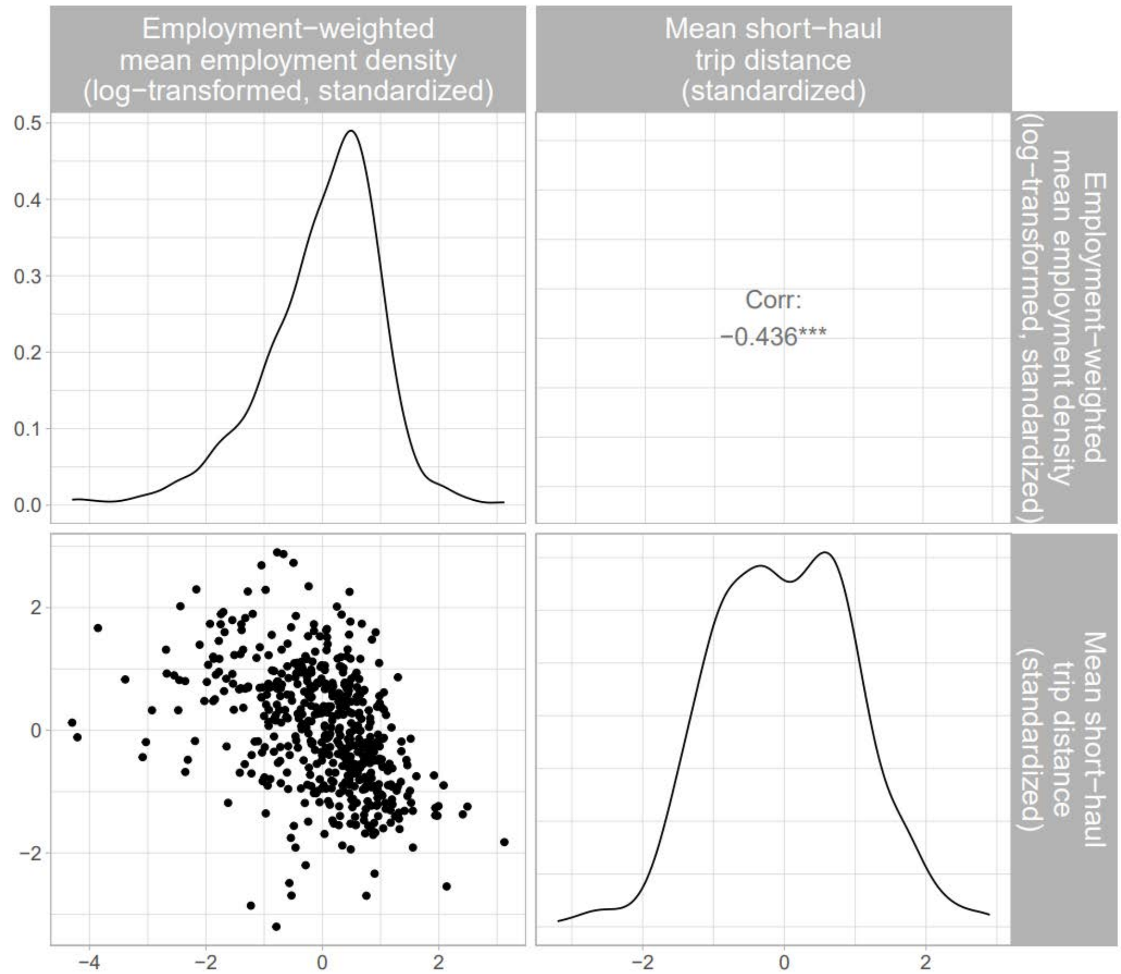

Figure 1. Scatterplot matrix showing univariate and bivariate distributions of clustering variables for each TAF region's truck operations

Several additional unsupervised clustering algorithms were considered to build segments of similar TAF zones but were found less satisfactory: hierarchical clustering using Ward's distance, k-medoids (similar to k-means but identifying specific representative medoid regions rather than average values) weighted by each zone's employment count, and a custom algorithm that enforces same-sizing of each cluster (i.e., similar employment count within each cluster). Hierarchical clustering was not preferred due to its tendency to produce clusters with very imbalanced sizes. k-medoids and the custom same-sizing algorithm were eliminated as candidate methods due to their overemphasis on dividing very high-population TAF zones (e.g., New York City and Los Angeles) into separate clusters.

A value of $k = 2$ was chosen for rural TAF zones and a value of $k = 5$ was chosen for the remainder of zones. These values were motivated primarily by budget considerations, because one representative from each cluster is used for data acquisition and analysis. There was not a pronounced point of inflection seen when the "elbow" method to setting k was considered, and silhouette scores did not differ drastically across values of k.

Table 1 summarizes the seven clusters that were developed, and Figure 2 shows the assignment of each modified TAF zone to a cluster. Descriptive analysis of each cluster's membership shows that the key difference between the two non-MSA/MiSA clusters was in their average short-haul trip distance. For zones associated with an MSA/MiSA, employment-weighted employment density serves as a primary differentiator between clusters; average short-haul trip distance divided the lowest-density ("Tier 3 density") zones further into three clusters.

Table 1. Descriptive Statistics for Clusters and Their Chosen Representative Zones

Cluster description	Approx. share of fleet-owned MHDV	# of TAF zones	Chosen zone	Weighted employment density (employees/sq mi)		Average short-haul trip distance (mi/trip)	
				Cluster average	Chosen zone	Cluster average	Chosen zone
Tier 1 density	20%	10	San Jose-Sunnyvale-Santa Clara, CA	5,058	3,565	21	21
Tier 2 density	30%	56	Pittsburgh, PA	1,649	1,811	24	23
Tier 3 density, shorter trips	10%	89	Evansville, IN-KY	715	796	23	24
Tier 3 density, medium trips	25%	146	Lafayette, LA	753	658	32	33
Tier 3 density, longer trips	5%	91	Janesville-Beloit, WI	608	712	42	42
Non-MSA, medium trips	5%	73	Southern ID non-MSA areas	213	340	35	33
Non-MSA, longer trips	5%	61	Eastern GA non-MSA areas	163	123	44	45

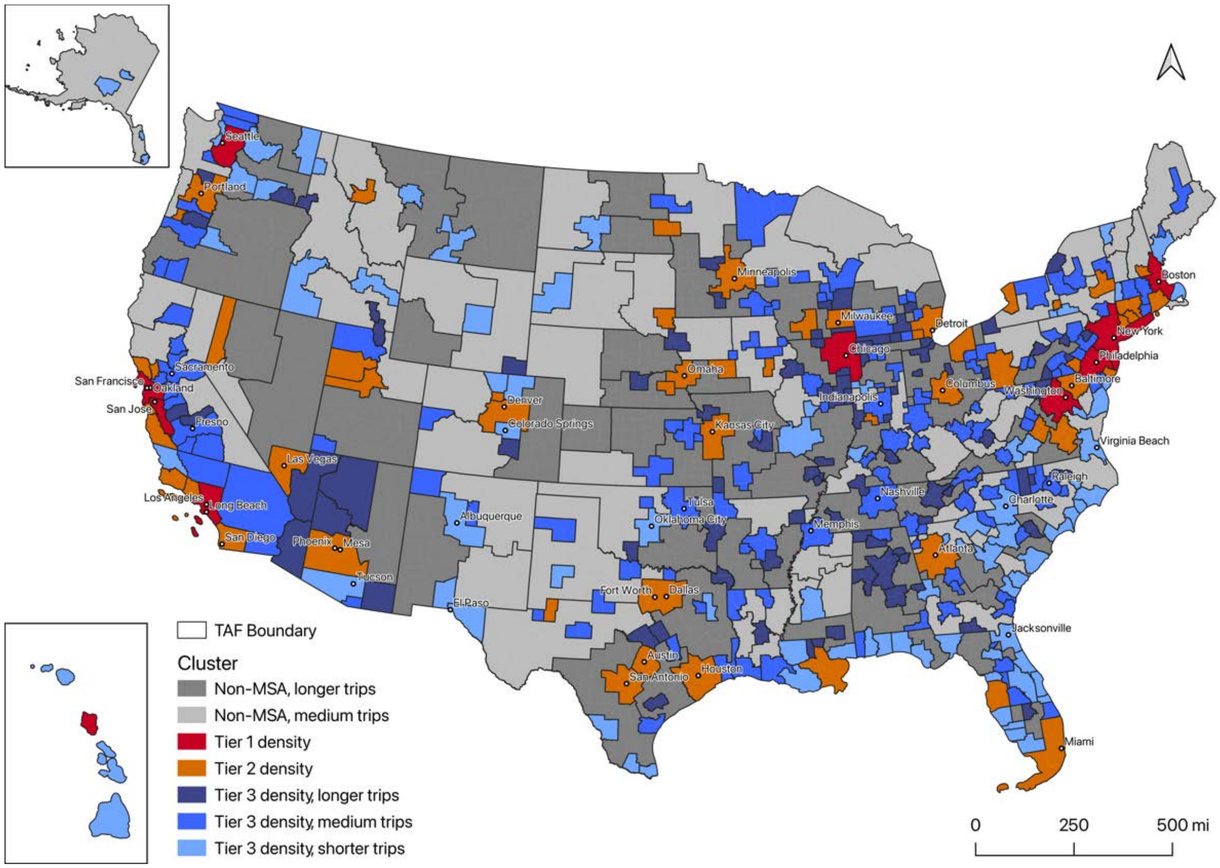

Figure 2. Clustering assignments for each modified Federal Highway Administration TAF zone

2.1.1.3 Representative Zone Selection

For each of the seven clusters, modified TAF zones were ranked by distance from the cluster's medoid zone within the Euclidean space of the chosen clustering variables. These medoids were determined using the R "cluster" library's implementation of partitioning around k-medoids function ($k = 1$) [17]. These similarity rankings informed the choice of representative zones for each cluster, along with each zone's anticipated cost of data acquisition (roughly proportional to the zone's population) and anticipated data coverage level. A final criterion used in zone selection was for the full set of selected zones to not skew too heavily in favor of one region of the country. The selected zones are shown in Table 1. Each cluster's selected representative is a relatively close match to the cluster medoid in terms of employment-weighted employment density and average short-haul trip distance.

2.1.2 Data Preparation

Access to Geotab's Altitude platform was acquired for each of the seven selected zones. This platform consolidates passively collected Geotab fleet telematics data into various summary measures of operating patterns and makes them accessible via several types of API query. For this dataset, the Altitude API's Regional Domicile Analytics query type was used to pull descriptive statistics for various categories of MHDV. The following filters were applied to API queries:

- Dates from September 7, 2022–September 30, 2022
- Minimum of two stops at the same location in that date range
- Body type of "Truck" (excludes buses, multi-purpose passenger vehicles, and "Other"/"Unknown" types).

September was chosen as a relatively typical month that avoids larger seasonal variations (e.g., mid-summer vacation periods and potential changes in freight patterns before and after the winter holidays); the first week of the month was excluded due to Labor Day holiday variation. Based on these filters, a total of 13,629 medium- and heavy-duty trucks were analyzed across the seven selected zones. For each vehicle category, descriptive statistics corresponding to percentile values were pulled from the API. Notation used to label these vehicle categories is summarized in Table 2, and notation used to label these descriptive statistics is summarized in Table 3. We use vocational driving style category definitions from Geotab's Altitude API, which are reproduced in Appendix Table A-1.

Table 2. Set Notation Used for Trucks

Description	Label	Values
GVWR class ranges	$c \in C$	$C = \{2\text{b} - 3, 4 - 5, 6 - 7, 8\}$
Vocational driving style categories	$v \in V$	$V = \left\{ \begin{array}{c} \text{Door to Door, Hub and Spoke,} \\ \text{Local, Regional, Long Distance} \end{array} \right\}$
Analysis regions	$r \in R$	$R = \left\{ \begin{array}{c} \text{San Jose} - \text{Sunnyvale} - \text{Santa Clara, CA,} \\ \text{Pittsburgh, PA,} \\ \text{Evansville, IN} - \text{KY,} \\ \text{Lafayette, LA,} \\ \text{Janesville} - \text{Beloit, WI,} \\ \text{Southern ID non} - \text{MSA areas,} \\ \text{Eastern GA non} - \text{MSA areas} \end{array} \right\}$
Days of week w within day-of-week groupings g	$w \in g \in G$	$G = \left\{ \begin{array}{c} \text{Monday} - \text{Friday,} \\ \text{Saturday} - \text{Sunday,} \\ \text{Monday} - \text{Sunday} \end{array} \right\}$
Percentiles	$p \in P$	$P = \left\{ \begin{array}{c} 5, 10, 20, 25, 30, 40, 50, \\ 60, 70, 75, 80, 90, 95 \end{array} \right\}$
Hour of day	$h \in H$	$H = \{0, 1, \dots, 23\}$

Table 3. Modeling Variables for Trucks

Description	Label	Source
Percentile range's daily driving distances (miles/operating day/vehicle)	$d_{c,v,r,w,p}$	Geotab
Percentile range's daily driving durations (hours/operating day/vehicle)	$t_{c,v,r,w,p}^{\mathrm{DRIVE}}$	Geotab
Domicile dwell duration (hours/operating day/grouping)	$t_{c,v,r,w}^{\mathrm{DOMICILE}}$	Geotab
Non-domicile dwell duration (hours/operating day/grouping)	$t_{c,v,r,w}^{\mathrm{NON\_DOMICILE}}$	Geotab
Domicile dwell duration (hours/grouping at time of day h across the full date range, operational and non-operational vehicle days)	$t_{c,v,r,w,h}^{\mathrm{DOMICILE\_TOTAL}}$	Geotab
Domicile dwell duration (hours/grouping across the full date range, operational vehicle days only)	$t_{c,v,r,w}^{\mathrm{DOMICILE\_TOTAL\_OP}}$	Geotab
Domicile dwell duration (hours/grouping at time of day h across the full date range, operational vehicle days only)	$t_{c,v,r,w,h}^{\mathrm{DOMICILE\_TOTAL}}$	Calculated
Proportion (0–1 range) of dwell time spent at domicile	$t_{c,v,r,w}^{\mathrm{DOMICILE\%}}$	Calculated
Daily domicile dwell duration percentile, assuming fixed share of dwell time at domicile (hours/day/vehicle)	$t_{c,v,r,w,p}^{\mathrm{DOMICILE\_FIXED}}$	Calculated
Percentile range's daily domicile dwell durations, assuming fixed share of dwell time at domicile (hours/day/vehicle)	$t_{c,v,r,w,p}^{\mathrm{DOMICILE\_ADJUSTED}}$	Calculated
Active vehicle-days within class bucket c, vocation v, region r, and day of week w	$n_{c,v,r,w}^{\mathrm{ACTIVE}}$	Geotab
Registered vehicles within class bucket c and region r	$n_{c,r}^{\mathrm{REGISTERED}}$	Experian
Scaling parameter assigned to results for grouping (c, v, r, d, p)	$s_{c,v,r,w,p}$	Calculated

For each specific combination (c, v, r, w) of gross vehicle weight rating (GVWR) class range group c, vocational driving style category grouping v, region grouping r, and day of week grouping w, percentile values were pulled for daily driving distances, $d_{c,v,r,w,p}$ (miles/day/vehicle), and driving time duration, $t_{c,v,r,w,p}^{\mathrm{DRIVE}}$ (hours/day/vehicle). For some groupings, too few vehicle fleets were active for the API to return percentiles (a privacy measure), or the number of operational vehicles was fewer than 25 (chosen for reasons of statistical robustness). In these cases, percentiles for the larger "parent" grouping (v, r, w), inclusive of all GVWR class ranges c, were used instead. When this "parent" grouping itself was smaller than 25 vehicles, percentiles were taken from the larger of that grouping and the second "parent" grouping: (c, r, w) inclusive of all vocational driving style categories.

For each vehicle in Geotab's data, the Regional Domicile Analytics query functionality determines the stop location having the longest cumulative dwell time over the analysis period and designates that location as the vehicle's domicile.

2.1.3 Operations Profile Creation

For each grouping (c, v, r, w) of gross vehicle weight rating (GVWR) class range group c, vocational driving style category grouping v, region grouping r, and day of week grouping w, daily driving distances and driving time durations were derived using Geotab data (Section 2.1.2). Geotab does not currently provide domicile dwell duration percentiles at the vehicle level, so those metrics are computed as a post-processing step after queries are run. For each grouping in total, segment-wide total hours of domicile dwell, $t_{c,v,r,w}^{\text{DOMICILE}}$, and non-domicile dwells, $t_{c,v,r,w}^{\text{NON\_DOMICILE}}$, were pulled. From these two columns, each grouping's domicile dwell time share (ranging from 0 to 1) was computed:

$$t_{c,v,r,w}^{\text{DOMICILE\%}} = \frac{t_{c,v,r,w}^{\text{DOMICILE}}}{t_{c,v,r,w}^{\text{DOMICILE}} + t_{c,v,r,w}^{\text{NON\_DOMICILE}}} \tag{1.1}$$

Two alternative assumptions are used to compute two versions of dwell duration percentile: one assuming a fixed domicile dwell share, and one adjusting for potential variation within each grouping. In the former approach, for each grouping (e.g., all Class 3 local travel vehicles in Pittsburgh, PA on Wednesdays), it is assumed that domicile dwell time share is fixed across all vehicle operating days, such that for each percentile level p, the grouping's daily domicile dwell time can be computed using the percentile level $100 - p$ of the daily drive time distribution:

$$t_{c,v,r,w,p}^{\text{DOMICILE\_FIXED}} = \left(24 - t_{c,v,r,w,100-p}^{\text{DRIVE}}\right) * t_{c,v,r,w}^{\text{DOMICILE\%}} \tag{1.2}$$

This version of the measure is computed prior to national scaling.

In the latter approach, after the national scaling step, dwell duration percentiles are adjusted to incorporate observed variation within each vocational driving style category v. Because variation in dwell durations cannot be directly observed within each (c, v, r, w) grouping, within- v, across-$(c, r, w, \text{industry})$ variation was used as a proxy measure, with industry groupings defined as the primary two-digit North American Industry Classification System codes (e.g., Transportation and Warehousing) as determined by the fleet operating each vehicle and returned by Geotab's Altitude API. Across all queries run for each vocational group v, all values pulled for $t_{c,v,r,w}^{\text{DOMICILE\%}}$ across all weight classes c, regions r, days of week w—and, additionally, a set of queries for vocational group v that were further filtered to each available industry grouping—were combined into a single list. Across all those queries, the within- v interquartile range (IQR) of domicile dwell time shares, $\text{IQR}(t_v^{\text{DOMICILE\%}})$, is computed. This proxy measure for domicile dwell time share variation is used to adjust the first set of dwell duration percentiles by "extending" them out in each direction from the median, proportional to each percentile's size relative to the interquartile (25–75) range:

$$t_{c,v,r,w,p}^{\text{DOMICILE\_ADJUSTED}} = t_{c,v,r,w,p}^{\text{DOMICILE\_FIXED}} + \text{IQR}(t_v^{\text{DOMICILE\%}}) * \frac{p-50}{75-25} \tag{1.3}$$

For a given grouping, the median value is the same across both estimates. The first set of estimates, $t_{c,v,r,w,p}^{\text{DOMICILE}}$, can be viewed as underestimating the variation in domicile dwell durations within each grouping. The second approach's reliance on a proxy measure for variation requires an additional assumption be made, but it represents an attempt to model known variation within each group in a data-driven manner. It assumes that, all else equal, vehicle-days having longer driving durations also have shorter domicile dwell durations, and thus errs on the side of overestimating the occurrence of relatively "challenging" charge sessions (shorter domicile dwells with higher energy needs).

2.1.4 Charging Load Profile Creation

We produced nationally averaged charging load profiles for all groupings (c, v, r, w) of GVWR class range grouping c, vocational driving style category grouping v, and day of week grouping d. We excluded charging load profiles for the Long Distance vocational driving style category as defined by Geotab's Altitude API, because that category is less likely to be primarily served by domicile charging.

The Altitude API does not provide daily schedules including driving and dwell events at the vehicle level, but rather provides univariate distributions for summary metrics (e.g., vehicle-miles traveled [VMT] and dwell duration) across each category of vehicle. Rather than model individual trucks' daily drive cycles as random draws from univariate distributions assuming independence, we used a fleet-level set of summary metrics capturing the share of vehicles within each category parked at their primary domicile in each hour of the day.

For each grouping, the API provides a variable, $t_{c,v,r,w,h}^{\text{DOMICILE\_TOTAL}}$, describing the total domicile dwell hours (summed across all vehicles within the grouping and days in the analysis period) at hour of day h. However, this variable includes both operational and non-operational vehicle days, so non-operational vehicle days must be removed from these totals. The API also provides a variable, $t_{c,v,r,w}^{\text{DOMICILE\_TOTAL\_OP}}$, which measures each grouping's total domicile dwell hours, across all hours of the day and the full analysis period, but for operational vehicle days only. These two variables can be used together to subtract non-operational vehicle days from $t_{c,v,r,w,h}^{\text{DOMICILE\_TOTAL}}$, thus estimating the total domicile dwell hours at hour of day h on operational days only:

$$t_{c,v,r,w,h}^{\text{DOMICILE\_TOTAL\_OP}} = t_{c,v,r,w,h}^{\text{DOMICILE\_TOTAL}} - \frac{\sum_{h' \in H} t_{c,v,r,w,h}^{\text{DOMICILE\_TOTAL}} - t_{c,v,r,w}^{\text{DOMICILE\_TOTAL\_OP}}}{24} \; \forall \, h \in H \tag{1.4}$$

As a final step, load curves are normalized such that each grouping's sum across 24 hours equals one. This occurs after the national scaling step (Section 2.1.5).

2.1.5 National Scaling

To scale each percentile measure nationally, we considered its values for intervals of consecutive percentile values, $[\, p^{\text{LOW}}, p^{\text{HIGH}}]$. These intervals were taken for each grouping (c, v, r, g) of vehicle class bucket c, driving style vocation v, region r, and day of week w. For each grouping, the total scaling factor $s_{c,v,r,w,[\, p^{\text{LOW}}, p^{\text{HIGH}}]}$ was the product of several components:

$$s_{c,v,r,w,[\, p^{\text{LOW}}, p^{\text{HIGH}}]} = s'_{c,v,r,w} * s''_{c,r} * s'''_{[\, p^{\text{LOW}}, p^{\text{HIGH}}]} \tag{1.5}$$

Within groupings of vehicle class bucket c, "vocation" (in Geotab terminology) v, and region r, $s'_{c,v,r,d}$ represents the scaling of individual days of week d:

$$s'_{c,v,r,w} := \frac{n^{\text{ACTIVE}}_{c,v,r,w}}{\sum_{w' \in g_w} n^{\text{ACTIVE}}_{c,v,r,w'}} \quad (1.6)$$

Within each grouping of class bucket–vocation, vehicle registration counts per region type determine a scaling factor $s''_{c,r}$ for each region:

$$s''_{c,r} := \frac{n^{\text{REGISTERED}}_{c,r}}{\sum_{c' \in C_c} n^{\text{REGISTERED}}_{c',r}} \quad (1.7)$$

Finally, for each interval $[\,p^{\text{LOW}}, p^{\text{HIGH}}]$, a scaling factor $s'''_{[\,p^{\text{LOW}},p^{\text{HIGH}}]}$ is assigned based on the size of the interval. For example, the interval from the 25th percentile to the 30th percentile is assigned $s'''_{[25,30]} = \frac{30-25}{100} 0.05$:

$$s'''_{[\,p^{\text{LOW}},p^{\text{HIGH}}]} = \frac{p^{\text{HIGH}} - p^{\text{LOW}}}{100} \quad (1.7)$$

The combined scaling factors, $s_{c,v,r,d,[\,p^{\text{LOW}},p^{\text{HIGH}}]}$, are rescaled such that the smallest value equals 1,000 (determined from computational constraints), and a stacked list of observations each repeated $s_{c,v,r,d,[\,p^{\text{LOW}},p^{\text{HIGH}}]}$ times is made. For each observation in the stacked list, a value is randomly sampled from a uniform distribution between the upper and lower values associated with the percentile range $[\,p^{\text{LOW}}, p^{\text{HIGH}}]$. Nationally scaled percentiles are directly computed from the stacked list's sampled values.

Load curves are scaled in this manner prior to being normalized (using a scaling factor of $s'_{c,v,r,d} * s''_{c,r}$, because the percentile bin scaling factor $s'''_{[\,p^{\text{LOW}},p^{\text{HIGH}}]}$ does not apply).

2.2 Transit Buses

At time of publication, there were over 106,000 transit buses across 1,591 U.S. counties (Figure 3) [18]. The transit bus modeling primarily relies on two major data sources (Figure 4):

- The National Transit Database (NTD), which serves as a centralized hub for financial, operating, and asset information of transit agencies in the United States [18]. The 2021 Annual Database was used to calculate transit bus population, develop metrics for transit agency clustering, and support the generation of transit bus operating profiles.
- The General Transit Feed Specification (GTFS), an open standard that transit agencies use to publish their service schedules (referred to here as "GTFS Schedule") and real-time operations ("GTFS Realtime") for various software applications. GTFS Schedule and GTFS Realtime data were collected to support the generation of transit bus operating profiles.

We considered active vehicles with a bus body type (i.e., "bus", "articulated bus", "over-the-road bus", "double decker bus", and "cutaway") for developing the county-level inventory. According to the NTD 2021 Annual Database, 86% of cutaways (i.e., vehicles with a bus body and mounted on the chassis of a van or light-/medium-duty truck) are used for demand response, and GTFS data are only available for fixed-route services. For this reason, cutaways are only included in the fleet inventory and are not included in generating cluster bus shares for VMT/dwell time/load estimates.

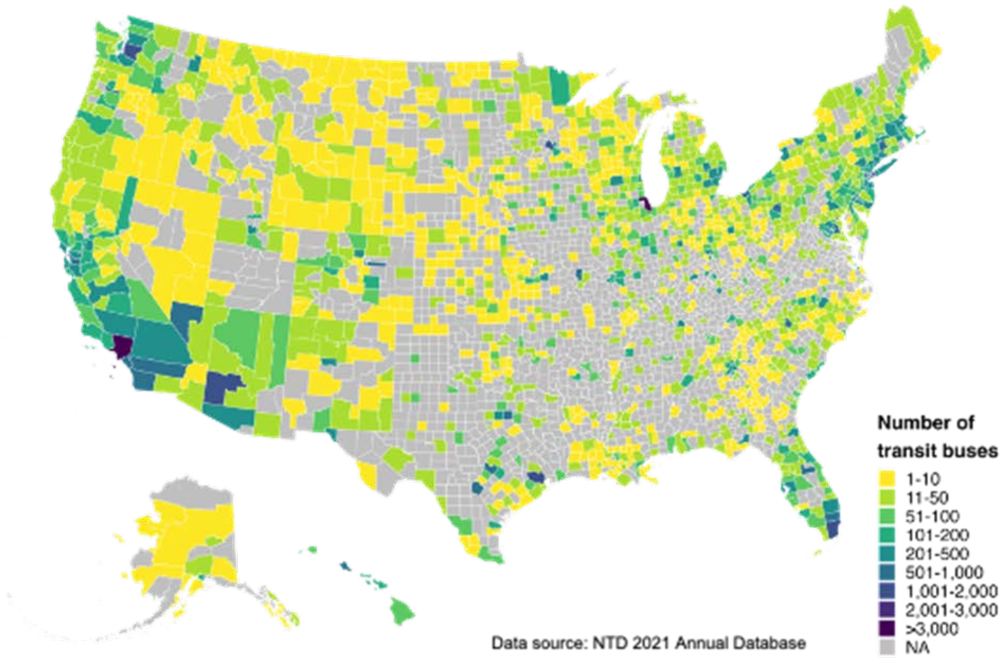

Figure 3. County-level transit bus inventory in the United States

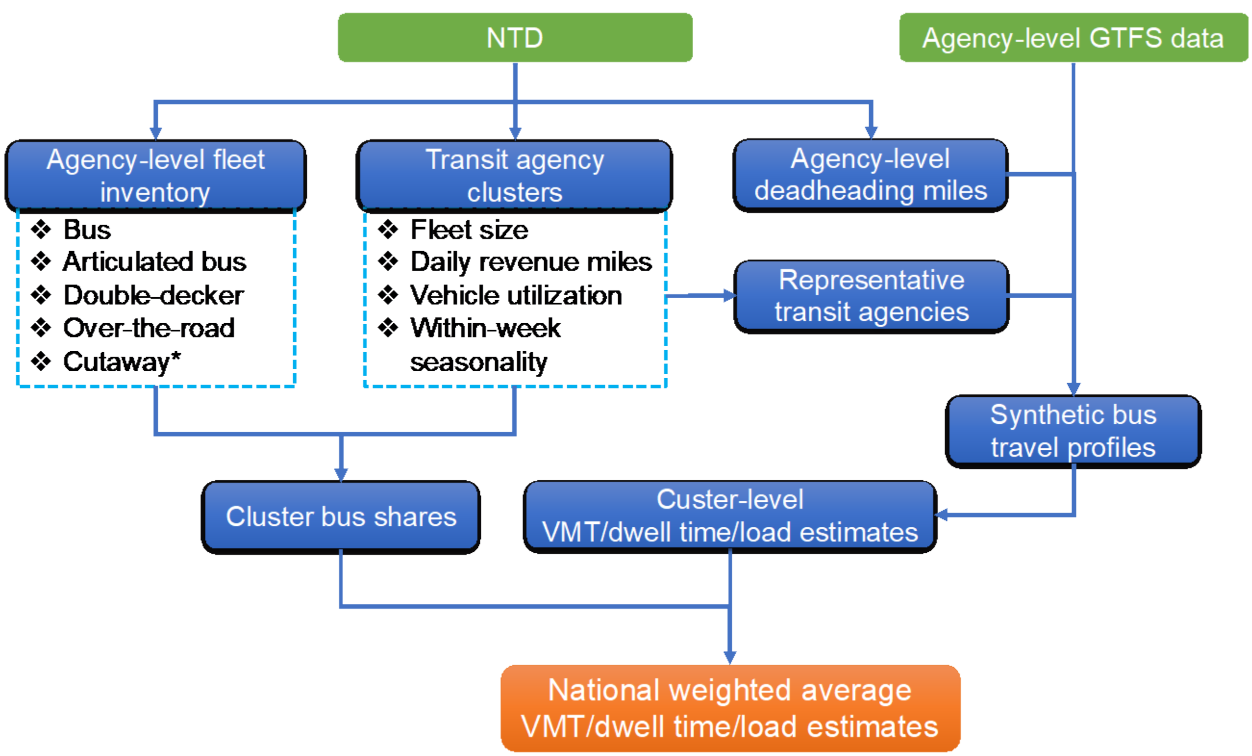

Figure 4. Overview of transit bus modeling approach

*Cutaways are only included in the fleet inventory and are not included in generating cluster bus shares for VMT/dwell time/load estimates.

2.2.1 NTD Transit Agency Clustering

Not every transit agency publishes their schedules or real-time operations in the GTFS format, and collecting every agency's real-time operations data would be highly labor-intensive and computationally expensive. Thus, we conducted a clustering analysis of transit agencies to extrapolate bus operations at a larger scale using the information from representative transit agencies with publicly available GTFS datasets.

We used a two-step approach for the clustering analysis. To better understand vehicle operations (especially at the largest transit agencies), we first separated out agencies with a fleet size of at least 750 transit buses. For the remaining NTD agencies (i.e., with a fleet size smaller than 750 transit buses), we conducted k-means clustering with four variables: (1) fleet size, defined as the number of active buses (NTD Revenue Vehicle Inventory table); (2) weighted daily vehicle-miles traveled (dVMT) per operating bus per day, including revenue miles only (NTD Service table); (3) maximum bus utilization rate, defined as vehicles operated in maximum service divided by the number of active buses (NTD Revenue Vehicle Inventory table); (4) within-week seasonality index, defined as the minimum value among an agency's weekday vehicles operated in maximum service, Saturday vehicles operated in maximum service, and Sunday vehicles operated in maximum service divided by the maximum value among the three metrics (NTD Service table). We applied log transformation to all four variables to reduce skewness and standardize the data to make variables comparable (Figure 5).

To determine an appropriate number of clusters, we used direct methods (i.e., elbow and average silhouette methods) and statistical testing methods (i.e., the gap statistic and 30 other indices) to compare results from two to six clusters. The results show that most of these methods suggest two and five as the optimal number of clusters. Given our goal of having a larger number of groups to represent various transit agency operating characteristics, we performed the final analysis and extracted the results using five clusters for agencies with a fleet size of less than 750 transit buses. Thus, we categorized the NTD agencies into six clusters (Table 4).

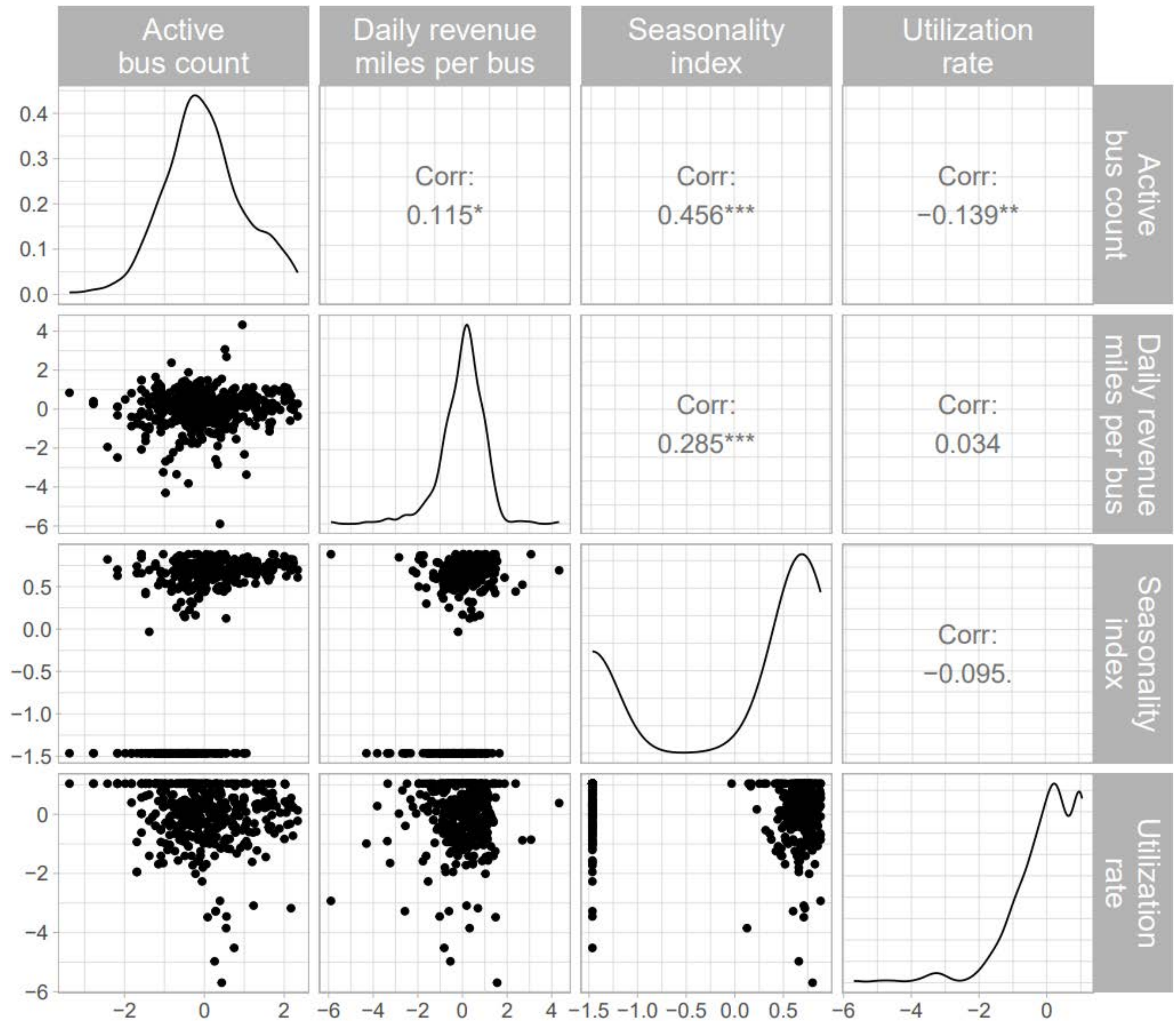

Figure 5. Scatterplot matrix showing univariate and bivariate distributions of clustering variables for transit bus operations. Variables are log-transformed and standardized.

Table 4. Overview of Transit Agency Clusters

Cluster	Cluster size	Cluster bus share	Average number of active buses	Average daily revenue miles/bus	Average vehicle utilization rate	Average within-week seasonality
Large agency (>= 750 buses)	13	37%	1719	116	71%	0.47
Mid-sized agency	84	39%	283	152	80%	0.47
Small agency, high dVMT, weekend service, high vehicle utilization	122	9%	43	152	84%	0.39
Small agency, high dVMT, weekend service, low vehicle utilization	60	8%	82	153	50%	0.42
Small agency, high dVMT, no Sunday service	110	5%	30	130	83%	0
Small agency, low dVMT	28	2%	46	59	73%	0.24

2.2.2 Data Preparation

In this section, we will describe the collection and processing of both GTFS Schedule and GTFS Realtime for transit bus modeling. GTFS Schedule data describes transit bus operations using several key concepts, including shapes, trips, stops, stop times, routes, and service blocks (Figure 6):

- A *trip* consists of one or many *shapes* that define vehicle movement along the trip. A trip may be the same length as an entire route or a portion of a route, depending on where the first and/or last stops are along the route.
- *Stops* and *stop times* define where and when buses make a stop along each trip.
- A service *block* consists of a sequence of trips assigned to a single bus. Service blocks may consist of services on the same route or different routes.

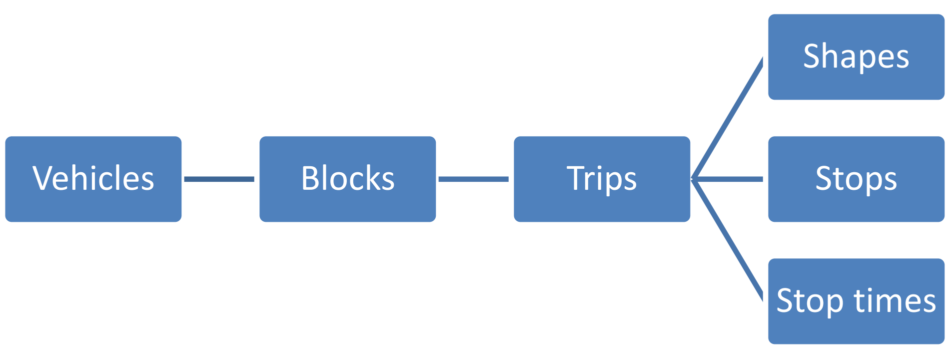

Figure 6. Key GTFS concepts

GTFS Realtime provides trip updates, vehicle positions, and service alerts. GTFS Realtime data contain mapping between vehicles and trips, and would be ideal sources for estimating vehicle-level travel distance. However, the availability of GTFS Realtime data is more limited, and the collection of real-time data requires additional processing time and resources. Liu (2020) developed a method to estimate daily transit bus VMT by extracting service block-level information from GTFS Schedule (Eq. 2.1–2.2), sequencing service blocks with a fixed time interval, and applying an agency-level deadheading-to-revenue-mile ratio to the vehicle level (where deadheading is defined as miles traveled out of revenue service, e.g., travel from the depot to the first bus stop of a route) [19]. We built on this method and improved it by collecting GTFS Realtime data to tune the time interval parameter for block sequencing and disaggregating agency-level deadheading miles to block sequence level (Eq. 2.3).

For the large agency cluster, we collected GTFS Schedule data and GTFS Realtime data (when available) for all agencies to tune between-block time intervals for sequencing service blocks and estimating vehicle-level operations. For the other five clusters, we ranked transit agencies by their distance to the cluster centroids to determine the most-representative agencies within each cluster. We used the top three agencies with both GTFS Schedule and Realtime data available to tune the between-block time interval at the cluster level, and then used the top three agencies with GTFS Schedule availability along with the cluster-level optimal between-block time interval to estimate vehicle-level operations at the agency-level within each cluster. In total, we analyzed GTFS data for 21,675 transit buses.

Larger transit agencies maintain their GTFS Realtime repositories and publish the repositories periodically. In most cases, three feeds are published and updated approximately every 10 seconds: trip updates, service alerts, and vehicle positions. Somes agencies host their own feeds and others use a third-party host such as Swiftly. The vehicle position feed of each agency was collected at the same frequency through agency and/or Swiftly API requests from August 11–14, 2023. These requests represented operations information of a typical workday (Monday/Friday), Saturday, and Sunday; we expect that agency service provisions change little enough from season to season that this date range should represent operations of other seasons well. Realtime data collection was programed to be queried every 5 minutes, a frequency believed to provide sufficient data for the analysis and keep API calls under the host-imposed request limit. After the initial data collection, duplicate records are removed to avoid redundancy in the data. The vehicle ID information, as revealed in GTFS Realtime, was essential for the purpose of this study. The total number of unique vehicle IDs was used for tuning the parameter for service block sequencing. Given the data availability during the period of data collection, we collected GTFS Realtime vehicle position information for 17 transit agencies across all 6 clusters.

To tune the between-block time interval parameter, we also collected GTFS Schedule data for the same 17 transit agencies during August 11–14, 2023, via the Mobility Database catalogs [20]. We used a previously developed algorithm for service block sequencing [19]. The sequencing consisted of two steps: (1) matching end stop of a block and start stop of another block and following time sequence, which generated the first round of block sequences; (2) the remaining blocks were sequenced by the following time sequence with a fixed time interval without matching the stop locations. This time interval can be perceived as the time for breaks in between blocks or traveling between the first/last stop of a block and the depot. Given that this time interval may vary by transit agency, we ran the sequencing algorithm for 18 parameter values: 10 minutes, 15 minutes, 30 minutes, 45 minutes, and 1 to 14 hours (with hourly steps). Thus, we generated 18 sets of block sequences (using each of the 18 time interval values) for each agency on a weekday, Saturday, and Sunday. We then compared the total number of block sequences against the total number of unique vehicle IDs from the Realtime datasets. For the large transit agency cluster, we calculated the percentage of differences for each agency. For the other five clusters, we calculated the average percentage of differences. The time interval value that yields the smallest average percentage of difference was considered the optimal parameter value. When the average percentage of difference stays the same for a few time interval values, the maximum time interval value was selected as the optimal to allow for as much time as possible in between service blocks. When the error stays the same for a few time interval values, the maximum value was selected as the optimal.

2.2.3 Operations Profile Creation

We relied on GTFS Schedule datasets, the estimated optimal time intervals, and the reported agency-level deadheading miles in the NTD to generate the operations profiles. We obtained GTFS Schedule data (October 14–16, 2023) for 37 transit agencies across all 6 clusters. The VMT estimates were calculated using Eq. 2.1-2.3, using the variables shown in Table 5. The deadheading time was estimated using the deadheading distance divided by the agency-level average deadheading speed, based on NTD reported total deadheading miles and hours at the agency level. The deadheading time was added to the beginning and the end of each block with associated deadheading trips. When the operating time of a bus was longer than 24 hours (around one percent or less of buses), we assumed that the bus would not be operated the next day and extend the dwell to the beginning of its next service day.

$$miles_{v,b} = \sum_{t \in b} \sum_{s \in t} length_{s,t} \tag{2.1}$$

$$miles_v^{\mathrm{REVENUE}} = \sum_{b \in B} miles_{v,b} \tag{2.2}$$

$$miles_v^{\mathrm{TOTAL}} = miles_v^{\mathrm{REVENUE}} + \left(\frac{count_v^{\mathrm{DEADHEAD}}}{\sum_{v \in V_a} count_v^{\mathrm{DEADHEAD}}} \right) * miles_a^{\mathrm{DEADHEAD}} \tag{2.3}$$

Table 5. Modeling Variables for Transit Buses

Description	Label	Source
Length in miles of shape s within trip t	$length_{s,t}$	GTFS Schedule
Reported daily deadheading miles at agency a	$miles_a^{\mathrm{DEADHEAD}}$	NTD
Set of vehicles v belonging to agency a	V_a	NTD
Miles traveled by vehicle v within block b	$miles_{v,b}$	Calculated
Revenue miles traveled by vehicle v at agency a	$miles_v^{\mathrm{REVENUE}}$	Calculated
Total dVMT by vehicle v	$miles_v^{\mathrm{TOTAL}}$	Calculated
Number of deadheading trips ran by vehicle v	$count_v^{\mathrm{DEADHEAD}}$	Calculated
Total daily depot dwell time (hours) for vehicle v	$time_v^{\mathrm{DEPOT}}$	Calculated
Charging power used for vehicle v	$power_v$	Calculated

2.2.4 Charging Load Profile Creation

As with other vocations, we assumed managed charging is spread evenly across each bus's depot dwell; however, unlike other vocations, we limited this charging to overnight dwell periods. As for dwell durations, when the operating time of a bus was longer than 24 hours (around one percent or less of buses), we assumed that the bus would not be operated the next day and that charging spanned from the end of the current service day to the beginning of its next service day. Each bus's charging power level was determined by its dVMT ($miles_v^{\text{TOTAL}}$) and overnight dwell time ($time_v^{\text{DEPOT}}$) (Eq. 2.5).

$$power_v = \frac{miles_v^{\text{TOTAL}} * \frac{2.28\ kwh/mile}{91.4\%}}{time_v^{\text{DEPOT}}} \tag{2.5}$$

The bus-level load profiles were aggregated to the cluster level for each hour during a 24-hour period. The cluster bus shares found in Table 4 were used for generating the nationally weighted load profiles. As a final step, the nationally weighted curves were normalized to get the per-vehicle hourly load profiles. The precise values assumed for energy consumption rate (2.28 kWh/mile [21]) and charging efficiency (91.4% [22]) did not affect the produced load profiles, because the assumptions were identical for all transit buses and the final load curves were normalized.

2.3 School Buses

The World Resources Institute estimates around 480,000 school buses are operating in the United States today, the vast majority of which are diesel-powered [23]. Diesel exhaust is a known carcinogen and exposes children to nitrogen oxides and fine particulates that impact lung function, exacerbate asthma, and lead to respiratory illness [24,25]. In a 2001 study, diesel exhaust levels inside school buses were found to be up to four times higher than inside passenger cars, and over eight times above the average outdoor concentration [21]. The health impacts on children are disproportionately distributed, with 60% of low-income students reportedly taking the bus, in contrast to just 45% for non-low-income students [26]. As a result, there is a strong push to decarbonize the U.S. school bus fleet through zero- and low-emission options. One recent example, the U.S. Environmental Protection Agency's Clean School Bus Program, allocates $5 billion from 2022–2026 to replace existing diesel buses with new cleaner alternatives [27].

BEVs are an option for zero-emissions school buses that have seen growth recently due to their growing model availability, reduced costs, and their aptness for the school bus duty cycle, which is characterized by limited mileage and substantial downtime in normal operations (excluding non-routine operations such as field trips, sporting events, and other extracurricular activities). As of June 2023, there were 2,277 electric school buses that were either on order, delivered, or operating in the United States with a total of 5,982 committed (including buses awarded but not yet ordered), a greater than 100% increase from June 2022 [23]. While the market for BEV school buses is nascent, it will likely mature rapidly if state and local decarbonization targets are to be achieved. For example, New York State has committed to all new school bus purchases being zero-emission vehicles by 2027 and 100% of buses in operation being zero-emission vehicles by 2035 [28], and many other jurisdictions are targeting similar trajectories.

Still, there are large uncertainties regarding the driving and charging patterns for electric school buses that may dictate their future growth and infrastructure needs. This section describes the approach taken to clarify those uncertain patterns through analysis and modeling of real-world operating data. First, we introduce the Fleet DNA school bus data sample used to characterize and represent school bus operations in this study (Section 2.3.1). Next, we describe how school bus depot dwells are assigned for subsequent analysis (Section 2.3.2). Finally, we introduce the EVI-Pro model, used to simulate electric school (and transit) bus charging at depots (Section 2.3.3).

2.3.1 Data Preparation

The National Renewable Energy Laboratory maintains a database of real-world commercial vehicle operating data called Fleet DNA, hosted on the FleetREDI platform [6,7]. FleetREDI offers statistical summaries derived from 1-hertz (or greater) drive cycle data captured using on-board data loggers for commercial fleets across diverse vocations and geographical regions spanning all 50 U.S. states. To produce the datasets described in this report, we deduced from the 1-Hz speed and location data the daily driving (i.e., trip) patterns of school buses, including their driving distance and durations. We also determined when and for how long they parked at their depot, presenting potential charging opportunities for BEVs. Fleet DNA is updated regularly with data for additional fleets. Table 6 summarizes the sample of conventional school bus operating data that were included at the time of this study. In total, we analyzed 7 fleets, 279 buses, and over 1,700 operating days with over 106,000 miles driven from 2009 to 2017.

Table 6. Fleet DNA School Bus Operating Data Summary

Location	Year	Bus count	GVWR	Operating days	VMT
Austin, TX	2009	2	6	10	429
Thornton, CO	2010	99	6	428	29,371
Schenectady, NY	2010	3	6	22	565
Redmond, WA	2011	108	6	468	14,712
Torrance, CA	2015	33	2,3,8	231	11,454
Napa, CA	2015	8	8	88	4,830
Rialto, CA	2017	26	8	492	45,381
Total	**2009–2017**	**279**	**2,3,6,8**	**1,739**	**106,742**

The representation of GVWR classes in the sample aligned closely with the distribution of classes observed in the United States. According to vehicle registration data curated by Experian, school buses falling under Classes 2–5 represent less than 2% of the national total, while 25% are classified as Class 6, 41% as Class 7, and 32% as Class 8.

Two versions of the school bus operating dataset were developed for this work. The first is an "unfiltered" sample containing all school buses, operating days, and trips reported in Table 6. The second is a "filtered" sample excluding trips shorter than 2 miles, often attributed to repositioning at the depot and not transporting students, and only considering depot dwells over 1 hour as realistic charging opportunities. As a result, the filtered sample contains 1,541 operating days (11% reduction) and 93,865 miles driven (12% reduction). Operating distributions for both the filtered and unfiltered sample were included in our dataset. As for interpretation, the unfiltered distributions are reflective of all school bus driving days, whereas the filtered distributions are more representative of typical operations (i.e., transporting students).

For the remainder of the report, all results and discussion pertain to the filtered sample. Both versions are reflective only of operations during times of the year that school is in session.

2.3.2 Operations Profile Creation

Chronological schedules of daily driving and dwell patterns are derived directly from the filtered sample (Section 2.3.1), but a remaining challenge is to accurately identify which dwells occurred at each vehicle's domicile location (and represent charging opportunities). To achieve this, a systematic approach was developed to precisely define, recognize, and confirm depots for school bus fleets in the Fleet DNA school bus data. First, we identified dwells in the dataset that were characteristic of depot dwells (i.e., lasting more than 6 hours and occurring overnight). The count of overnight dwells was calculated for each prospective depot site, using the latitude and longitude coordinates of these "depot-like" dwells, rounded to three decimal points (accuracy approximately within 40 feet).

This process yielded a lookup table containing all potential depot sites and the number of depot-like dwells (across all buses in the fleet) occurring at each. In most cases, buses in the same fleet shared the same depot-like dwell locations, simplifying the identification process. However, in cases where multiple depot-like dwell locations were detected for a single fleet, the actual depot sites were verified using aerial imagery (Figure 7). With the depot locations identified, the start and end points of each school bus trip were recognized as originating or culminating at a fleet depot, provided they were within 500 meters of a designated depot location.

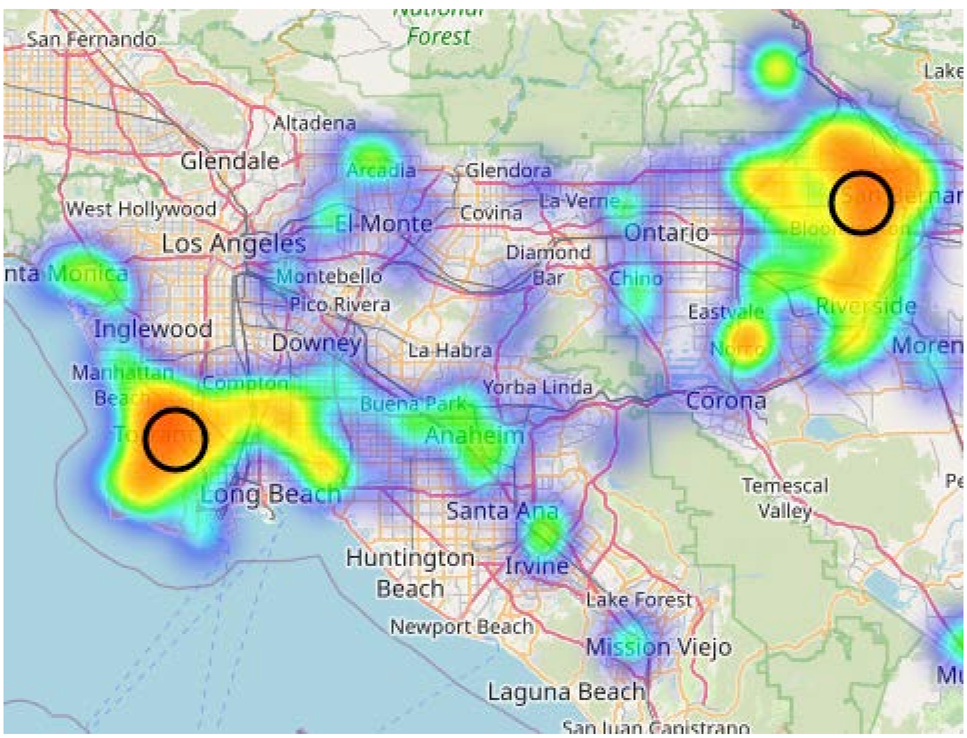

Figure 7. Heatmap of dwell locations and approximate depot locations for school bus fleets located in Torrance and Rialto, CA

2.3.3 Charging Load Profile Creation

EVI-Pro is a tool for projecting demand for plug-in electric vehicle (PEV) charging infrastructure under typical daily conditions [29]. EVI-Pro uses detailed data on vehicle operating patterns, vehicle attributes, and charging station characteristics in bottom-up simulations to model the charging behaviors, plug-in electric vehicle load profiles, and quantity and types of charging infrastructure necessary to support regional adoption of plug-in electric vehicles. A diagram of data flows within EVI-Pro is shown in Figure 8. EVI-Pro has been used in multiple detailed planning studies, including Wood et al. (2017, 2018, 2023), Moniot et al. (2019), and Alexander et al. (2021) [30–34].

For this study, EVI-Pro was used to simulate depot charging for battery electric school buses, mirroring the operations of diesel school buses detailed in the Fleet DNA data from Section 2.2.1. As with other vocations, we assumed that school bus charging is managed with the distribution of charging activities evenly spread across the hours when the buses are parked at the depot (and that all charging occurs at the depot), including the midday hours between morning and afternoon routes. This assumes BEV battery capacity and charging rates are adequately sized to support this managed charging scenario. As with transit buses, precise values for school bus energy consumption rate and charging efficiency did not affect the produced load profiles, because the assumptions were identical for all school buses and the final load curves were normalized.

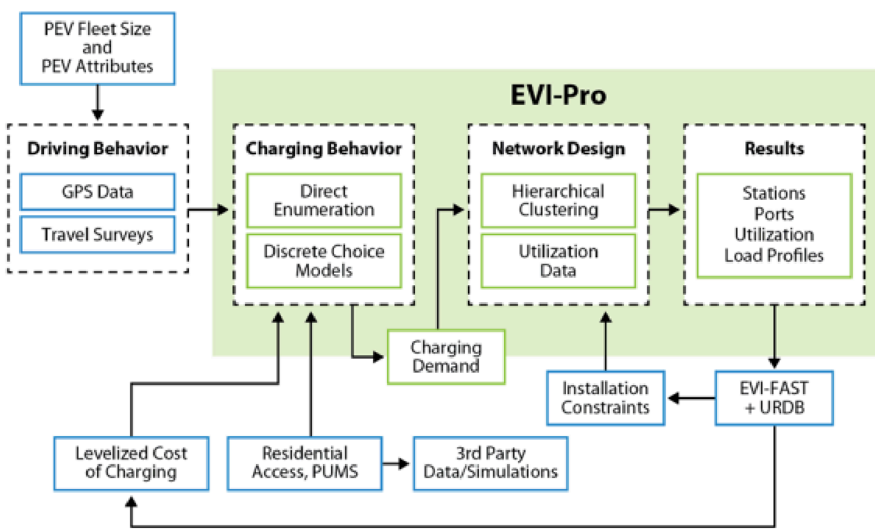

Figure 8. EVI-Pro block diagram for charging behavior simulations and network design

3 Summary of Produced Datasets

For each vocation, three sets of data were produced to characterize national average typical operations:

- Daily driving distance in miles per day per vehicle (excluding days the vehicle is not operated), expressed as vehicle-day percentiles ranging from 10 to 90.
- Daily dwell time spent per day at the vehicle's "domicile," or primary dwell location, in hours per day per vehicle (excluding days the vehicle is not operated), expressed as vehicle-day percentiles ranging from 10 to 90. For trucks and school buses, this measure includes all time spent at the vehicle's domicile location regardless of whether the hours of domicile dwell are in one continuous period or split across multiple shorter dwells. For transit buses, it includes the overnight domicile dwell only. Each vehicle's domicile is defined as the location where the vehicle spent the most time parked during the analysis period.
- A load curve for nationally aggregated daily depot-centric charging (excluding days the vehicle is not operated), normalized such that the sum of hourly values equals one for each category of vehicle. For trucks and school buses, these load curves assume that vehicle charging energy demand is spread evenly over the full duration of domicile dwell time. For transit buses, it is spread over the overnight domicile dwell only. They provide a comprehensive view of time spent at domicile and a stylized view of potential load curves (optimistic, but not a "best case" lower bound, in terms of avoiding peaks from coincident load across vehicles).

Each of these outputs is a nationally descriptive measure developed by aggregating across smaller subgroups of vehicles. Each one describes only operational days for a given category of vehicle, excluding days in which vehicles were not operated; for each category of vehicle, we include a separate variable for weekdays and weekends describing what share of the included vehicles are active on an average day ("Pct_Of_Included_Active"). Separate weekday and weekend estimates are provided for each variable's distribution and for load curves; for transit buses, due to substantial variation between Saturdays and Sundays, those days are provided separately. Load curves are not provided for the Long Distance vocational driving style category of trucks as defined by Geotab's Altitude API.

3.1 Trucks

Figure 9 shows a synthetic distribution of dVMT for each combination of GVWR class and vocational driving style category, as listed in Table 2 and defined by Geotab's Altitude API. These density plots and all subsequent ones are constructed using uniform random sample draws from each percentile bin (excluding the "tails" below the 10^{th} and above the 90^{th} percentile) and thus may look more "flat" within each percentile bin than the true distributions. Across weight classes, vehicle operating days from the Door to Door and Local vocations tend to have relatively similar distributions, with a median value (marked as a vertical line) between 60 and 80 miles. Hub and Spoke vehicles have a similar range of values, but within Classes 6–8, the median ranges from 95–115 miles per day, and there are relatively more vehicles toward the upper side of the distribution than Local or Door to Door.

As expected, Regional vehicles have substantially higher daily distances and Long Distance is higher still (except for Class 6–7; this may be related to Geotab's vocational definitions not relying entirely on distance, but rather a combination of typical radius of operation and whether the vehicle operates from a consistent domicile each day). It is worth noting that, while differences across vocations tend to be larger than across GVWR classes, larger vehicles tend to have longer daily distances, at least for Regional and Long Distance.

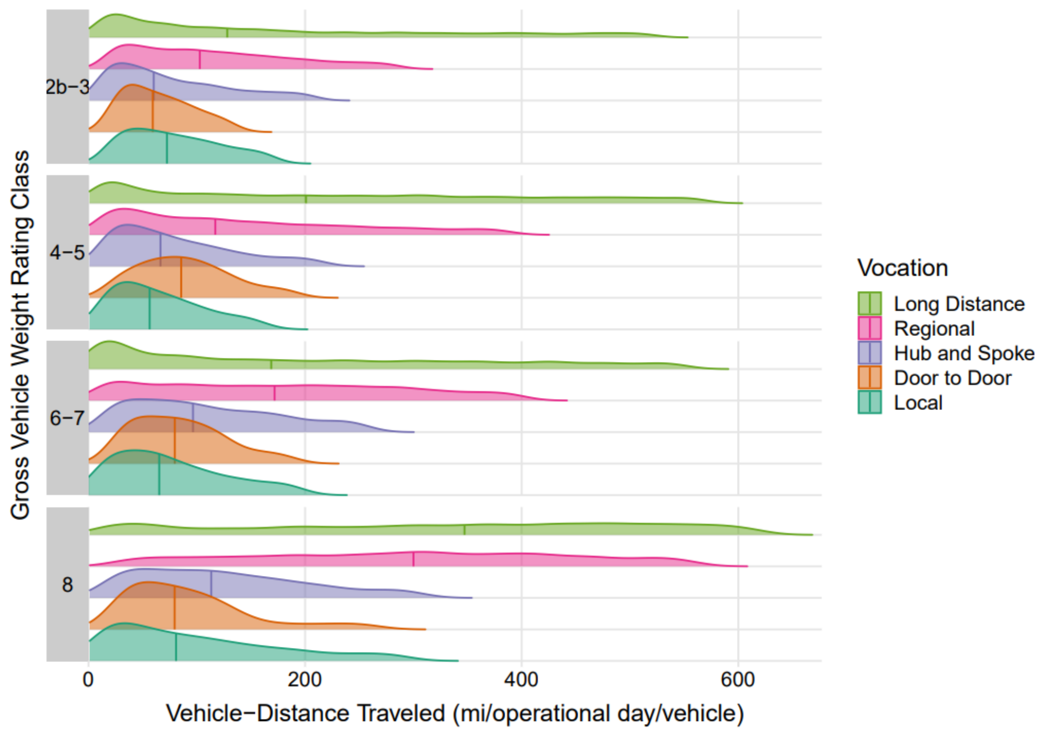

Figure 9. Distribution of dVMT for truck vocations. Vertical lines mark each group's median (weekdays, 10th–90th percentile range only).

There is no directly comparable dataset in the public domain, but as a "back-of-envelope"-style comparison to existing estimates, we considered similar segments in two public datasets: The 2021 Vehicle Inventory and Use Survey (VIUS) dataset, published through the U.S. Census Bureau, and the Fleet DNA database [6,35]. While definitions for GVWR class and for vocation are not perfectly aligned across datasets, several groups have relatively similar comparison points. In general, VIUS seems mutually compatible with the statistics produced for this dataset; a more detailed analysis is presented in the Appendix.

Figures 10 and 11 show distributions for domicile dwell durations with and without an adjustment made to consider variations in dwell locations, respectively. The adjustment increases the spread of each regional cluster's distribution prior to the national aggregation step, so it alters both the spread and shape of the aggregated distributions.

Figure 10. Distribution of domicile dwell duration for truck vocations, adjusted to account for variability in dwell locations. Vertical lines mark each group's median (weekdays, 10th–90th percentile range only).

Figure 11. Distribution of domicile dwell duration for truck vocations, not adjusted to account for variability in dwell locations. Vertical lines mark each group's median (weekdays, 10th–90th percentile range only).

The adjusted version (Figure 10) shows that Long Distance and Regional vehicles spend less time each operational day at their primary domicile; this can be explained both by longer drive distances and by a lower propensity to park at the same place each day. Aggregated nationally, Hub and Spoke vehicle days tend to have the longest median domicile dwell; this could be due to having an extremely consistent "hub" home base of operation or could be because these vehicles are more likely to only operate during the 9-to-5 business day. Door to Door and Local vehicles tend to have very similar dwell patterns (with the exception of Class 2b-3), and very rarely do they have a full 16-hour (e.g., 5 p.m. to 9 a.m.) overnight domicile dwell between business days. One potential explanation for this trend is that some vehicles, such as package delivery vans, either have daily operating shifts much longer than 8 hours or could operate for multiple daily shifts. Differences across GVWR class are relatively minor for dwell durations.

Figure 12 shows average normalized curves for daily charging load at each vehicle's domicile (excluding trucks of the long-distance vocation, for which we did not produce load curves). Because these load curves were developed from an assumption that load is spread evenly across a vehicle's dwell time at its domicile, the load curves derived directly from the hours at which more vehicles are, on average, at their domicile. Across all GVWR classes and vocations, there tends to be a "trough" in the middle of the day when the average vehicle is less likely to be domiciled and higher levels of load in the evening. This trough is most pronounced on weekdays, when operational schedules are more consistent and thus the percentage of vehicles that could charge at their domicile in midday is lower. On weekends, operating schedules across vehicles are less similar and more variable, which has an effect of smoothing the average load curve. Local and Door to Door vehicles show more consistent schedules with shorter away-from-domicile periods. Door to Door vehicles, particularly, have a very consistent drop-off in availability to charge in the mid-morning (whereas Local, more of a catch-all category, shows more variation in morning departure time).

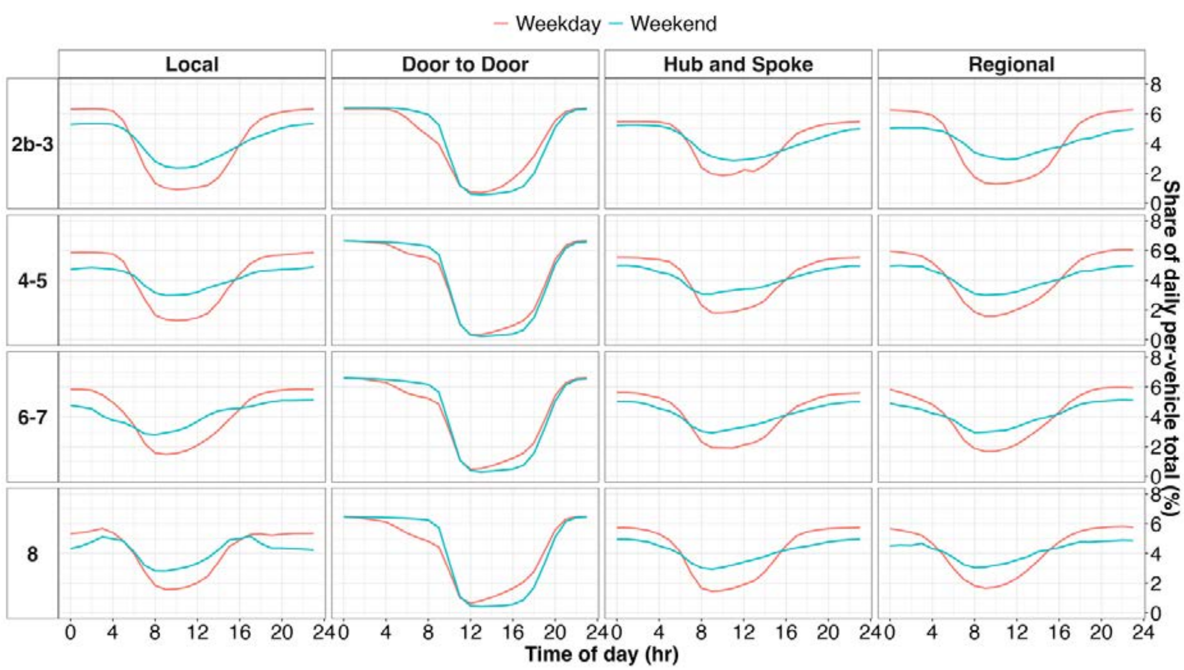

Figure 12. Normalized depot charging load curves for the national average truck operational day

3.2 Transit Buses

Figure 13 illustrates the distribution of nationally scaled weighted average dVMT for transit buses on a typical weekday. On average, the dVMT of transit buses ranges from 57 (10^{th} percentile) to 230 (90^{th} percentile) miles, with the median weekday operating day traveling 141 (50^{th} percentile) miles. Figure 14 illustrates the distribution of weighted average daily domicile dwell duration in hours per vehicle on a typical weekday. Half of bus operating days have a dwell time of at least 10.5 hours per day at the domicile, which exhibits the potential for nighttime depot charging with lower charger power requirements. For transit buses with shorter domicile duration, chargers with higher power levels can potentially support service provision without changing current schedules.

Consistent with the relatively long domicile dwell duration during nighttime, transit bus charging load typically occurs between 6 p.m. and 5 a.m. of the next day and peaks between 8 p.m. and 2 a.m. of the next day (Figures 15 and 16). As with other vocations, these curves assume that charging begins immediately when buses arrive at their depot and is spread evenly across their dwell. The variation in when peak load occurs across clusters is due to the average duration of service provision. For instance, large agencies would observe a substantial increase in charging load from 8 p.m. to 1 a.m. because there are many buses returning to the depot during this period. The variation between weekdays and weekends is similarly due to the varying operational needs and how vehicles are assigned to fulfill needs.

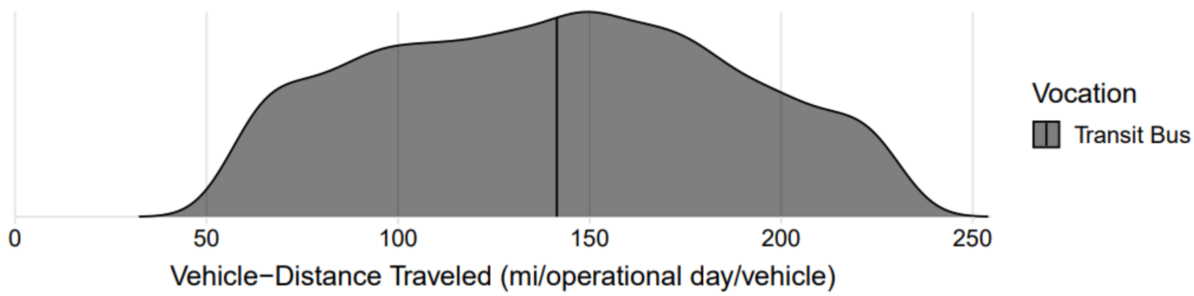

Figure 13. Distribution of dVMT for transit bus data. Vertical line marks the median (weekdays, 10th–90th percentile range).

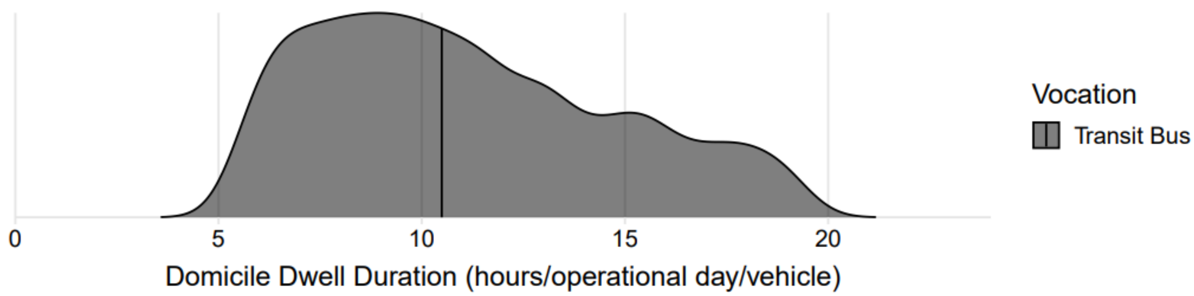

Figure 14. Distribution of domicile dwell duration for transit bus data. Vertical line marks the median (weekdays, 10th–90th percentile range).

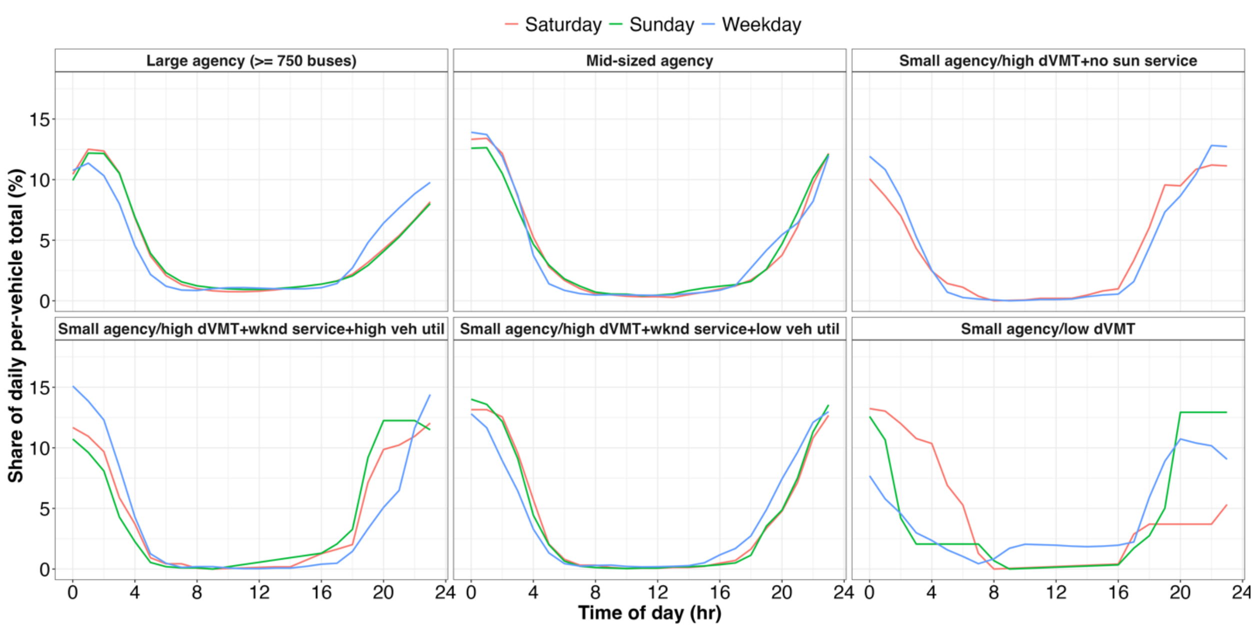

Figure 15. Normalized depot charging load curves for the national average transit bus operational day, by transit agency cluster

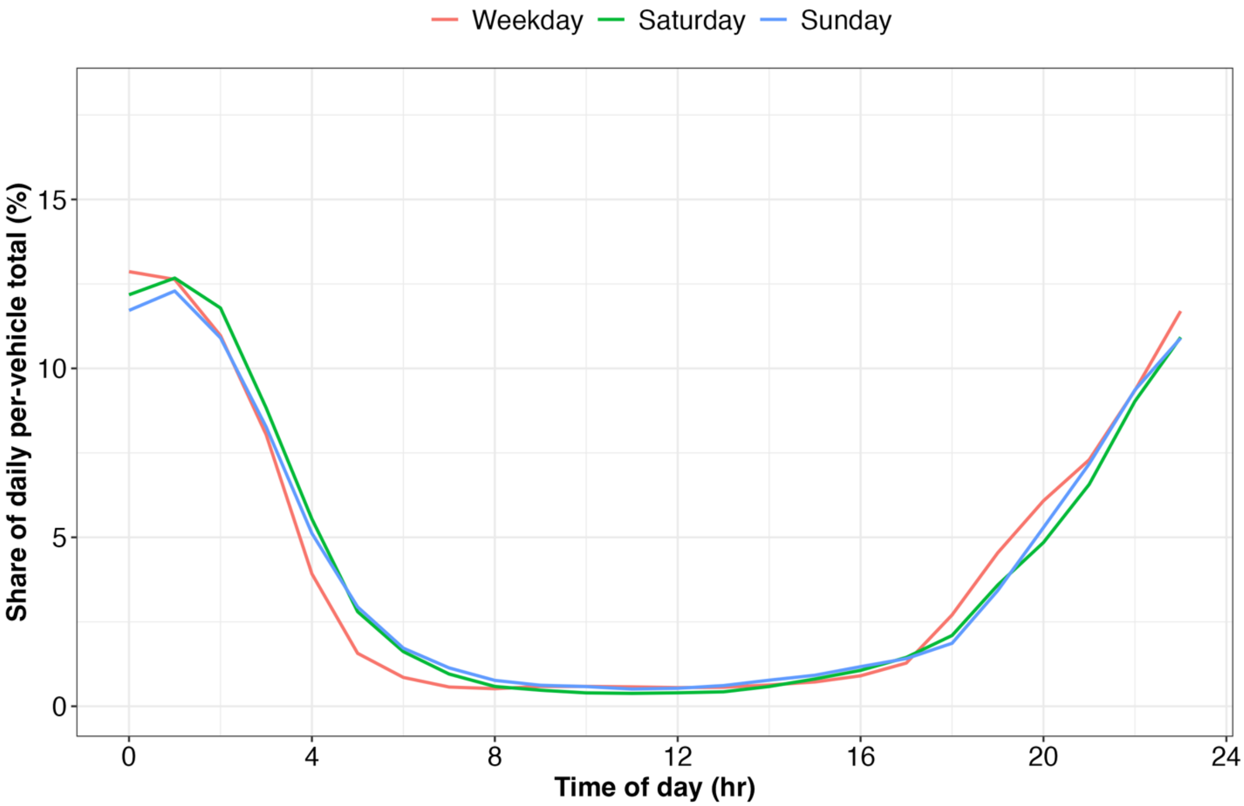

Figure 16. Normalized depot charging load curves for the national average transit bus operational day

3.3 School Buses

Figure 17 illustrates dVMT for vehicle operating days—including only times of year that school is in session—comparing unfiltered and filtered school bus data. Both datasets show similar distributions, with the median vehicle covering around 50 miles per operating day and few buses exceeding 100 miles driven in a single day (though we do not model or visualize the upper and lower 10th percentile tails of the distributions).

Findings regarding route distances were reported in Duran et al. (2013), which analyzed over 1,700 route shifts from the school buses available in Fleet DNA that were available at its time of publication [36]. This study found those route shifts averaged 32 miles and had a large spread, from 1 to 128 miles; this result was similar to a 2011 study considering 861 route shifts in Fleet DNA finding an average length of 35 miles [37]. Assuming a two-shift day, Duran et al.'s reported average VMT aligns relatively closely with our own findings, suggesting that more recent additions to Fleet DNA have not drastically altered operating patterns. However, other regions not currently represented in Fleet DNA may have different operating patterns.

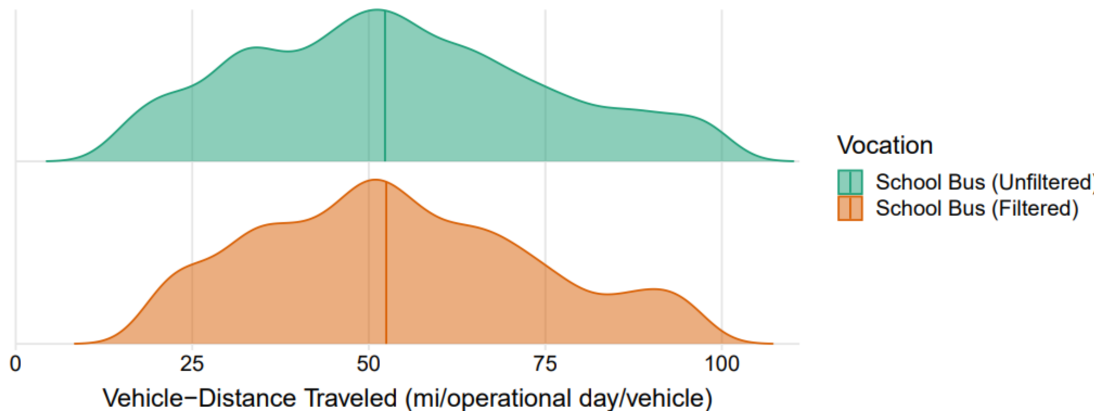

Figure 17. Distribution of dVMT for unfiltered and filtered school bus data. Vertical lines mark each group's median (weekdays, 10th–90th percentile range).

Figure 18 illustrates the distribution of domicile dwell durations across all school buses and operating days for both the filtered and unfiltered datasets. Within the Fleet DNA sample, school buses typically spend more than half of their operating day parked at the depot (median value of 16 hours per day). Domicile dwell times below 10 hours per day are rare within the sample and could represent non-routine travel such as field trips or sporting events.

Figure 19 presents the school bus load profiles derived from the filtered data for weekdays and weekends, as obtained from the EVI-Pro simulation results. As anticipated, the dominant charging opportunity occurs during nighttime hours, spanning from 8 p.m. to 4 a.m. Reflecting a standard bimodal operating schedule involving routes in the morning and afternoon, the charging opportunity is decreased in the morning during typical pickup hours and reaches its minimum around 8 a.m. This decline is more pronounced on weekdays, suggesting a less-predictable schedule when buses are operated on weekends. After the morning shift, there is an increase in depot charging opportunity, peaking at noon while most school buses are parked between shifts, before receding again in the afternoon.

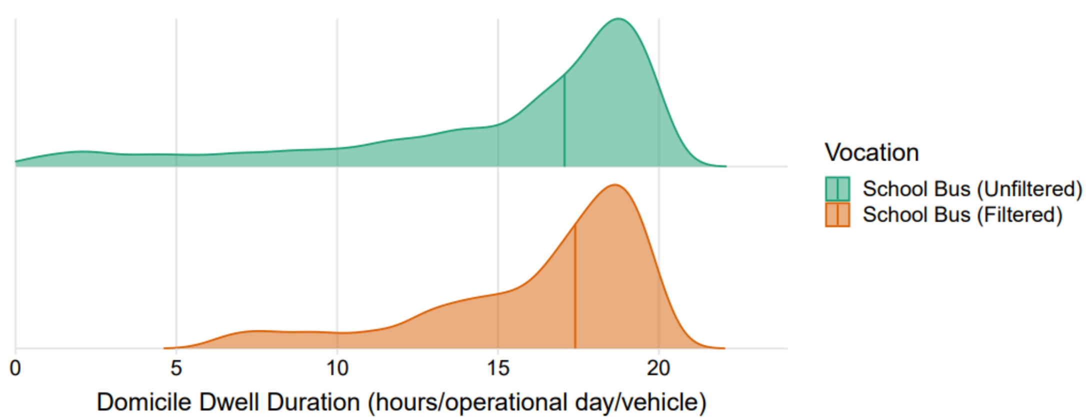

Figure 18. Distribution of domicile dwell duration for unfiltered and filtered school bus data. Vertical lines mark each group's median (weekdays, 10th–90th percentile range).

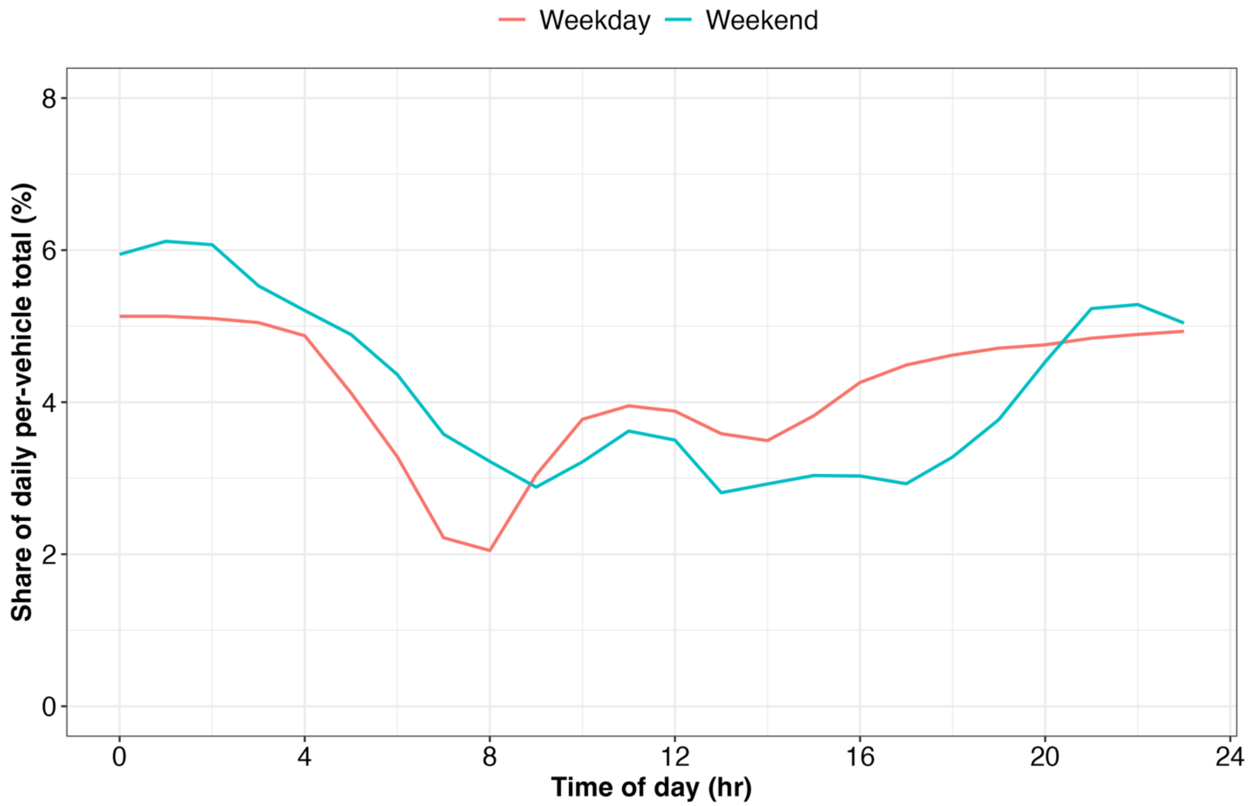

Figure 19. Normalized depot charging load curves for the national average school bus operational day (using filtered school bus data)

4 Conclusion

The datasets for daily distance, dwell time, and depot charging load summarized in this report are publicly available online. They provide insight into depot-based MHDV operational patterns, energy needs, and one potential shape the resulting depot-based charging load curves could take. These insights may be useful to researchers simulating MHDV operations, infrastructure planners preparing for the projected rapid electrification of MHDVs, and policymakers responding to the potential of BEVs to reduce greenhouse gas and criteria air pollutant emissions.

4.1 Context and Interpretation

The distributions and load curves are intended to describe operational days only. This means the distributions for each vocation must be taken in context: if the provided distribution suggests that the median driving distance is 10 miles on operational weekdays, but only half the fleet is operational on a given weekday, then the fleetwide median (including non-operational vehicles) is much lower. Similarly, the domicile dwell duration percentiles only reflect days on which each vehicle operated, and thus are considerably lower than if they were to include non-operational vehicles. For school buses, which have highly seasonal operations correlated with when school is in session, all operating distributions and EV load curves are reflective of the school year. For each category of vehicle, we provide as additional context a variable describing what share of the included vehicles are active on an average day ("Pct_Of_Included_Active").

The data is intended to describe the 10^{th}–90^{th} percentile range of operations, and we did not attempt to model the upper or lower tails of the distributions for driving distance or dwell duration (Vehicles with driving distances or dwell durations outside of those bounds are included in our normalized load curves, which represent the fleetwide average operational day across each vehicle category.).

Each vehicle type is summarized with a goal of present-day national representativeness, but their interpretations differ somewhat. For trucks, we incorporate as much variation as feasible from known sources of differences in operations to capture regional diversity and prioritize representativeness. Similarly, for transit buses, we cluster transit agencies using known sources of variation and merge cluster-level representatives with deadheading information from individual agencies. For school buses, the sample of available data is relatively limited and is randomly sampled to generate a national aggregation; representativeness is not easily optimized (though there is no *a priori* expectation of bias in a particular direction). Furthermore, for school buses, the operations data we analyze was collected from 2009 to 2017, and thus may be less representative of present-day operations than the data used for trucks (collected in 2022) or transit buses (collected in 2023).

We do not provide distributions for individual regions in this dataset. While the data represents nationwide operational trends, it is worth bearing in mind that the nationwide 90^{th} percentile driving distance for a given vocation may differ substantially from the 90^{th} percentile value observed within a specific region.

These load curves are intended to represent status quo vehicle movements. For each vocation, there likely are strategies that could improve operations of an all-electric fleet, but those strategies are not considered here. As one example, assigning different service blocks to each transit bus could help optimize load management across a transit system.

The load curves included in this data reflect one possible future scenario—one in which charge is managed to spread load out as evenly as possible when parked at the depot, thus reducing spikes in demand. This can be viewed not as a forecast, but rather as a relatively optimistic scenario in terms of reducing impacts to the electricity distribution system, relative to one in which each vehicle charges as quickly as possible upon arrival at the depot. However, it is not a lower bound on reducing facility-level load peaks, which would involve not only spreading out of individual vehicle load, but also coordinating across the fleet (and with other site-and system-level loads) such that vehicles parked during demand troughs charge more aggressively during those hours.

4.2 Limitations

For these datasets to be used as intended, their limitations and potential biases should be considered.

The analysis period of September 7, 2022 to September 30, 2022 for the Geotab data (trucks) was chosen to represent a relatively typical time period that occurred after the peak COVID-19 pandemic period and avoided extreme seasonal operating patterns (e.g., potentially atypical summer vacation periods or a pre-winter holiday uptick in local delivery traffic). However, even if it does capture typical operations in terms of seasonality, it cannot describe seasonal variation in operations.

Some categories of MHDV are excluded from these datasets. The truck analysis used an explicit filter for trucks, thus excluding multipurpose passenger vehicles and vehicles with an “unknown” or “other” body type (and deliberately excluding buses, as they were modeled separately). The Geotab Altitude API used to model trucks is likely to lack full vocational coverage in other regards, but it is impractical to ascertain fully what subcategories of body type and vocation may have robust or limited coverage beyond the GVWR class and vocational driving style fields included in these datasets.

These data are intended only to describe the subset of MHDV that are likely to rely primarily on depot-based charging at a relatively stable primary domicile location. This is why the Long Distance vocation was excluded from our depot charging load curves. While some Long Distance vehicles may charge at a primary domicile location from time to time, they are generally less likely to depend on it for the majority of charging needs, and are nearly certain to require en-route charging away from their depot.[38] Daily distance and domicile dwell duration data for the Long Distance vocation is provided alongside a variable capturing the percentage of the vocation that does not meet the chosen Geotab filters for domicile-centric operations; the excluded percentage is generally high. On the other hand, we assumed that each included vehicle’s primary domicile location could support BEV charging and that batteries were sized to last throughout the daily drive cycle. Actual battery sizes may depend on design factors including climate, vehicle age expectations, anticipated duty cycle, and degree of desired “oversizing” for robustness to variability, but we did not judge the feasibility of the implied battery sizes, battery energy density, or charging power levels. This may overestimate the potential of domicile-based charging to satisfy MHDV charging needs, but it is not evident that including these vehicles introduces bias into the estimated distributions or load curves in any specific manner.

More broadly, while our methodology for trucks selected a sample of regions capturing as much variation as feasible across two known sources of regional variation—and thus built national averages, including substantial diversity in operations—we did not assess whether the national averages were representative in a statistical sense. Doing so would have required a “ground truth” that was unavailable.

In scaling region- or agency-level results nationally, we averaged over all operating vehicles equally. This makes a simplifying assumption that each category of vehicle has a similar energy consumption rate per mile, whereas a true national average would weight higher-energy vehicle charging load curve shapes more heavily. One instance of this simplification is that we did not consider differences in subcategories across regions or agencies; for example, if Class 4 local operations-centric trucks make more stops per mile in one region than another, then the total energy consumption for that vehicle category may be different in reality across those two regions. Another instance of this simplification pertains to local factors such as weather and road grade: a true national average would not assume these factors affect all vehicles equally, but rather would estimate differences in total energy demand from region to region and weight a national average accordingly.

4.3 Future Work

This dataset provides a nationally aggregated view of daily distance, dwell duration, and potential load curves, but it does not assess regional variations in those distributions. Additional research is warranted to investigate what regional factors are strongly linked to variations in observed operating patterns and whether any of those linkages are causal in nature. Understanding these linkages better would assist in regional planning for MHDV electrification.

While this dataset characterizes a diverse set of depot-centric MHDV vocations, it does not fully capture the range of MHDVs that may electrify in the future or the full range of MHDV charging needs. Different modeling approaches are required to ascertain the operating patterns, charging demands, and infrastructure needs of MHDV that may lack a consistent depot, operate very far from it for multiple days, or simply have longer daily drive cycles than their battery size would permit.

The load curves described here provide one potential scenario for future load curves, but a more robust analysis of different load control strategies, charger configurations, and vehicle archetypes may reveal a diverse set of future load curves. These variations may impact future infrastructure needs and the nature of optimal charge management strategies.

References

(1) EPA (U.S. Environmental Protection Agency), *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2021*. https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks-1990-2021 (accessed 2023-11-29).

(2) Hoehne, C.; Muratori, M.; Jadun, P.; Bush, B.; Yip, A.; Ledna, C.; Vimmerstedt, L.; Podkaminer, K.; Ma, O. Exploring Decarbonization Pathways for USA Passenger and Freight Mobility. *Nat. Commun.* **2023**, *14* (1), 6913. https://doi.org/10.1038/s41467-023-42483-0.

(3) Ledna, C.; Muratori, M.; Yip, A.; Jadun, P.; Hoehne, C. *Decarbonizing Medium- & Heavy-Duty On-Road Vehicles: Zero-Emission Vehicles Cost Analysis*; United States, 2022. https://doi.org/10.2172/1854583.

(4) Colin McKerracher; Aleksandra O'Donomvan; Nikolas Soulopoulos; Andrew Grant; Jinghong Lyu; Siyi Mi; David Doherty; Ryan Fisher; Corey Cantor; Maynie Yang; Kwasi Ampofo; Yayoi Sekine; Andy Leach; Evelina Stoikou; Jiayan Shi; Peng Xu; Laura Malo Yague; Alex Haring; Philip Geurts; Charlotte Adriaenssens; Allen Tom Abraham; Komal Kareer. *BloombergNEF Electric Vehicle Outlook 2023 Executive Summary*. https://about.bnef.com/electric-vehicle-outlook/ (accessed 2023-12-26).

(5) *Zero-Emission On-Road Medium-and Heavy-Duty Strategies | California Air Resources Board*. https://ww2.arb.ca.gov/resources/documents/zero-emission-road-medium-and-heavy-duty-strategies (accessed 2023-11-29).

(6) *FleetREDI*. https://fleetredi.nrel.gov/#/ (accessed 2023-11-29).

(7) National Renewable Energy Laboratory. *Fleet DNA: Commercial Fleet Vehicle Operating Data*. https://www.nrel.gov/transportation/fleettest-fleet-dna.html (accessed 2023-11-22).

(8) *Vehicle Inventory and Use Survey (VIUS) | Bureau of Transportation Statistics*. https://www.bts.gov/vius (accessed 2023-12-22).

(9) Federal Highway Administration. *2020 NextGen NHTS National Truck OD Data*; 2020. https://nhts.ornl.gov/od/.

(10) Federal Highway Administration Office of Highway Policy Information. *Traveler Analysis Framework: Planned Passenger Travel Origin Destination Zone Information, Version 1*. Federal Highway Administration Policy & Governmental Affairs. https://www.fhwa.dot.gov/policyinformation/analysisframework/04.cfm.

(11) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2.

(12) Ottensmann, J. R. *On Population-Weighted Density*. Rochester, NY February 1, 2018. https://doi.org/10.2139/ssrn.3119965.

(13) U.S. Census Bureau. *County Business Patterns (CBP)*. Census.gov. https://www.census.gov/programs-surveys/cbp.html.

(14) U.S. Census Bureau. *2021 Planning Database*. Census.gov. https://www.census.gov/topics/research/guidance/planning-databases/2021.html.

(15) U.S. Department of Housing and Urban Development, Office of Policy & Research. *HUD USPS ZIP Code Crosswalk Files | HUD USER*. https://www.huduser.gov/portal/datasets/usps_crosswalk.html.

(16) Hartigan, J. A.; Wong, M. A. Algorithm AS 136: A K-Means Clustering Algorithm. *J. R. Stat. Soc. Ser. C Appl. Stat.* **1979**, *28* (1), 100–108. https://doi.org/10.2307/2346830.

(17) Maechler, M.; Rousseeuw, P.; Struyf, A.; Hubert, M.; Hornik, K. *Cluster: Cluster Analysis Basics and Extensions*; 2022.

(18) *2023 NTD Reporting Policy Manual | FTA*. https://www.transit.dot.gov/ntd/2023-ntd-reporting-policy-manual (accessed 2023-12-03).

(19) Liu, B. *Planning for Sustainable Transportation through the Integration of Technology, Public Policy, and Behavioral Change: A Data-Driven Approach,* UCLA, 2020. https://escholarship.org/uc/item/7595f70q (accessed 2023-12-03).

(20) *Mobility Database*. https://database.mobilitydata.org/ (accessed 2023-12-03).

(21) EPA. *Proposed Rule: Greenhouse Gas Emissions Standards for Heavy-Duty Vehicles – Phase 3*. https://www.epa.gov/regulations-emissions-vehicles-and-engines/proposed-rule-greenhouse-gas-emissions-standards-heavy (accessed 2023-12-01).

(22) Johnson, C.; Nobler, E.; Eudy, L.; Jeffers, M. *Financial Analysis of Battery Electric Transit Buses*; NREL/TP-5400-74832; National Renewable Energy Lab. (NREL), Golden, CO (United States), 2020. https://doi.org/10.2172/1659784.

(23) *Dataset of Electric School Bus Adoption in the United States - Data | World Resources Institute*. https://datasets.wri.org/dataset/electric_school_bus_adoption (accessed 2023-12-03).

(24) Weir, E. Diesel Exhaust, School Buses and Children's Health. *CMAJ Can. Med. Assoc. J.* **2002**, *167* (5), 505.

(25) Pandya, R. J.; Solomon, G.; Kinner, A.; Balmes, J. R. Diesel Exhaust and Asthma: Hypotheses and Molecular Mechanisms of Action. *Environ. Health Perspect.* **2002**, *110* (suppl 1), 103–112. https://doi.org/10.1289/ehp.02110s1103.

(26) Bureau of Transportation Statistics. *The Longer Route To School*. https://www.bts.gov/topics/passenger-travel/back-school-2019 (accessed 2023-11-22).

(27) *Electrification Coalition - Electric School Buses*. Electrification Coalition. https://electrificationcoalition.org/schoolbus/ (accessed 2023-12-04).

(28) NYSERDA (New York State Energy Research & Development Authority). *Electric School Bus (ESB) Roadmap*. https://www.nyserda.ny.gov/All-Programs/Electric-School-Buses/Electric-School-Bus-Roadmap (accessed 2023-11-22).

(29) *EVI-Pro: Electric Vehicle Infrastructure – Projection Tool*. https://www.nrel.gov/transportation/evi-pro.html (accessed 2023-12-28).

(30) Wood, E.; Rames, C.; Muratori, M.; Raghavan, S.; Melaina, M. *National Plug-In Electric Vehicle Infrastructure Analysis.* **2017**.

(31) Wood, E. W.; Rames, C. L.; Muratori, M.; Srinivasa Raghavan, S.; Young, S. E. *Charging Electric Vehicles in Smart Cities: An EVI-Pro Analysis of Columbus, Ohio*; National Renewable Energy Lab.(NREL), Golden, CO (United States), 2018.

(32) Wood, E.; Borlaug, B.; Moniot, M.; Lee, D.-Y.; Ge, Y.; Yang, F.; Liu, Z. *The 2030 National Charging Network: Estimating US Light-Duty Demand for Electric Vehicle Charging Infrastructure*; National Renewable Energy Laboratory (NREL), Golden, CO (United States), 2023.

(33) Moniot, M.; Rames, C. L.; Wood, E. W. *Meeting 2025 Zero Emission Vehicle Goals: An Assessment of Electric Vehicle Charging Infrastructure in Maryland*; National Renewable Energy Lab.(NREL), Golden, CO (United States), 2019.

(34) Alexander, M.; Crisostomo, N.; Krell, W.; Lu, J.; Ramesh, R. *Assembly Bill 2127 Electric Vehicle Charging Infrastructure Assessment: Analyzing Charging Needs to Support Zero-Emission Vehicles in 2030: Commission Report*; California Energy Commission, 2021.

(35) *VIUS213C: In-use Vehicles by Registration State and Vehicle Size for the U.S. (excluding New Hampshire) and States: 2021 - Census Bureau Table*. https://data.census.gov/table/VIUSC2021.VIUS213C?q=vius213c&g=010XX00US,$0400000&nkd=PRICHAR~18 (accessed 2023-12-03).

(36) Duran, A.; Walkowicz, K. A Statistical Characterization of School Bus Drive Cycles Collected via Onboard Logging Systems. *SAE Int. J. Commer. Veh.* **2013**, *6* (2), 400–406. https://doi.org/10.4271/2013-01-2400.

(37) Barnitt, R. A.; Gonder, J. *Drive Cycle Analysis, Measurement of Emissions and Fuel Consumption of a PHEV School Bus*; 2011; pp 2011-01–0863. https://doi.org/10.4271/2011-01-0863.

(38) Borlaug, B.; Moniot, M.; Birky, A.; Alexander, M.; Muratori, M. Charging Needs for Electric Semi-Trailer Trucks. *Renew. Sustain. Energy Transit.* **2022**, *2*, 100038. https://doi.org/10.1016/j.rset.2022.100038.

(39) Geotab ITS. *Altitude Data Dictionary: Term Definitions for the Altitude Platform.* 2023.

Appendix

Table A- 1 reproduces definitions from the Geotab Altitude Data Dictionary for each of the vocational driving style categories of truck included in the analysis.

Table A- 1. Geotab vocational driving style category definitions (Geotab Altitude Data Dictionary) [39]

Vocation	Definition
Local	The vehicle's range of activity is below 150-air-miles (regardless of miles traveled), thus qualifies for the short-haul exemption under Hours of Service Regulations. This is measured over a given 3 month period. In addition, the vehicle does not exhibit behavior in line with other vocations, such as hub-and-spoke and door-to-door.
Door to Door	The vehicle makes significantly more stops than most per workday, but also tends to spend very little time per stop.
Hub and Spoke	The vehicle spends many of its workdays making multiple round trips from a singular location (a centralized hub). Typically, the vehicle would average over one round trip per working day, with these round trips accounting for the majority of its total mileage.
Regional	The vehicle has a wide range of activity, over the 150-mile threshold for short-haul exemption, but tends to rest in the same location often. The vehicle is also neither hub-and-spoke nor door-to-door.
Long Distance	The vehicle has a very large range of activity and typically does not rest in the same location. The vehicle is also neither hub-and-spoke nor door-to-door.

Table A- 2 and Table A- 3 compare this dataset and 2021 VIUS data for Class 8 (the only grouping of classes aligned across datasets). VIUS Table 213C buckets vehicles by primary range of operation, which is one of several factors used to define Geotab Altitude vocations. The Local vocation included in Geotab's Altitude API explicitly has a typical range of operation radius below 150 miles, while Regional and Long Distance are above 150 miles; the Door to Door and Hub and Spoke vocations typically have daily driving distances most similar to the Local vocation. For this dataset's three locally oriented vocations, median VMT per day ranges from 75–110 miles per day. If scaled based on the percentage of vehicles operating each day (as a back-of-envelope measure of fleetwide central tendency), the range is from 52–75 miles per day. VIUS averages range from 59 miles per day for the "50 Miles or Less" operational range bucket to 162 miles per day for the "101 to 200 miles" bucket. Primary range of operation is correlated with average dVMT, but does not deterministically predict it, as, given an operational range, vehicles may still be used more or less heavily. Accounting for the tendency of the average to be skewed upward relative to the median, these sets of numbers appear to be mutually compatible. It is worth noting these numbers are still not directly comparable because VIUS results are the average from reported range of operation; include all vehicles, not only those with domicile-centric operations; and are survey responses rather than measured data and thus may have response biases.

Table A- 2. Class 8 Daily VMT From VIUS 2021 Dataset [35]

Primary Range of Operation	Share of Class 8 trucks	VMT per day (average across operational and non-operational days)
50 Miles or Less	41.3%	59
51 to 100 Miles	15.6%	106
101 to 200 Miles	8.0%	162
201 to 500 Miles	8.2%	211
501 Miles or More	11.6%	264
Not Reported	15.4%	149
Total	**100.0%**	

Table A- 3. Daily VMT of Domicile-Centric Class 8 Trucks By Vocation From This Dataset

Vocation	Share of domicile-centric class 8 trucks	% of vocation's vehicles excluded from domicile-centric dataset	% of domicile-centric vehicles that operate each day	Daily VMT (median across operational days only)
Local	30.6%	21.5%	62.0%	83
Door to Door	4.7%	1.5%	74.9%	75
Hub and Spoke	18.8%	3.6%	68.8%	110
Regional	26.6%	13.8%	67.5%	295
Long Distance	19.3%	47.1%	68.0%	334
Total	**100.0%**			

Table A- 4 compares this dataset's locally oriented vocations to Fleet DNA's "Freight–Local" vocation. For class groups 2 and 3, 4 and 5, and 6 and 7, this dataset's estimated median daily distance traveled is substantially higher than in Fleet DNA (58–70 miles per day versus 14 for Classes 2 and 3, 55–84 miles versus 43 miles for Classes 4 and 5, and 65–93 miles versus 46 miles for Classes 6 and 7, respectively.) Daily distance values for Class 8 are more similar (74–110 miles per day in this dataset's locally-oriented vocations versus 80 in Fleet DNA). These numbers are not directly comparable, as Fleet DNA is not intended to portray a national average but rather a diverse set of fleets. Furthermore, this dataset's vocational categories are not well-aligned with Fleet DNA (Fleet DNA has various other vocational definitions for local operations outside of Freight–Local, and this dataset's Local vocation is a catch-all for all local operations that do not typically follow a Door to Door or Hub and Spoke style of operations.).

Table A- 4. Comparison of Local Vocations in This Dataset and Fleet DNA [6] (approximate 25th-50th-75th percentile values)

Class Group	Vocation in this dataset			Fleet DNA
	Door to Door	Other Local	Hub and Spoke	Freight–Local
2 and 3*	33-59-94	34-70-119	24-58-124	7-14-25
4 and 5	47-84-123	25-55-102	28-64-127	16-43-70
6 and 7	41-77-116	27-65-125	39-93-172	19-46-77
8	38-74-124	27-83-178	44-110-196	35-80-194

*This dataset includes Class 2b and above, while Fleet DNA does not specify subclasses within Class 2.