

Affordability of Household Transportation Fuel Costs by Region and Socioeconomic Factors

Energy Systems Division

About Argonne National Laboratory

Argonne is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC under contract DE-AC02-06CH11357. The Laboratory's main facility is outside Chicago, at 9700 South Cass Avenue, Lemont, Illinois 60439. For information about Argonne and its pioneering science and technology programs, see www.anl.gov.

DOCUMENT AVAILABILITY

Online Access: U.S. Department of Energy (DOE) reports produced after 1991 and a growing number of pre-1991 documents are available free at OSTI.GOV (<http://www.osti.gov/>), a service of the US Dept. of Energy's Office of Scientific and Technical Information.

Reports not in digital format may be purchased by the public from the National Technical Information Service (NTIS):

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Road
Alexandria, VA 22312
www.ntis.gov
Phone: (800) 553-NTIS (6847) or (703) 605-6000
Fax: (703) 605-6900
Email: orders@ntis.gov

Reports not in digital format are available to DOE and DOE contractors from the Office of Scientific and Technical Information (OSTI):

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
www.osti.gov
Phone: (865) 576-8401
Fax: (865) 576-5728
Email: reports@osti.gov

Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor UChicago Argonne, LLC, nor any of their employees or officers, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of document authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof, Argonne National Laboratory, or UChicago Argonne, LLC.

ANL/ESD-20/11

Affordability of Household Transportation Fuel Costs by Region and Socioeconomic Factors

by
Yan Zhou, Spencer Aeschliman, David Gohlke
Energy Systems Division, Argonne National Laboratory

December 2020

CONTENTS

ACKNOWLEDGEMENTS	vi
EXECUTIVE SUMMARY	1
1 INTRODUCTION AND LITERATURE REVIEW	1
2 OVERALL RESEARCH FRAMEWORK	4
3 DATA AND METHODOLOGY	7
3.1 Variables in VMT Modeling.....	8
3.2 Regression Analysis.....	10
3.3 VMT Modeling by Census Tract.....	14
3.4 VMT Model Validation	15
3.5 Stock-Weighted MPG.....	18
3.6 Fuel Price	20
3.7 Driving Cost Calculation	23
4 RESULTS AND DISCUSSION	24
4.1 Household VMT	24
4.2 Household Energy Burden.....	26
5 SENSITIVITY ANALYSIS	35
5.1 Stock Weighted MPG	35
5.2 Parametric Sensitivity Analysis	36
6 CONCLUSIONS.....	37
7 REFERENCES	39

FIGURES

ES-1 Transportation energy burden by county	3
1 Overall Research Steps	4
2 Feature importance of selected explanatory variables.....	8
3 NHTS mean household annual miles traveled for each of the 18 geographic groupings. Inset: United States map showing the 6 geographic regions.	9
4 Performance comparison (R^2) of LR, RF and GBM.	11
5 Training performance (R^2) of each regional model using the GBM	12
6 Distribution of errors: predicted $\log_{10}(\text{VMT})$ minus surveyed $\log_{10}(\text{VMT})$	13
7 Mean household annual miles traveled for each of the 18 geographic groupings: NHTS vs. Model Estimates. Inset: United States map showing the 6 geographic regions.....	16

FIGURES (CONT.)

8	Comparison of modeled VMT and Chicago regional travel survey	17
9	Stock-weighted vehicle MPG by county	18
10	Stock-weighted MPG and average vehicle age by zip code in Washington D.C.....	19
11	Alternative fuel vehicles adopted vs. stock-weighted MPG.....	20
12	ZIP3-aggregated gasoline prices in February, April, June, and September 2020. Local gasoline prices are highly correlated throughout the year, with consistently higher in the West and near major cities. February 2020 fuel prices were used in this study.....	22
13	Household annual VMT by county.....	24
14	Household annual VMT by average tract income	25
15	Transportation energy burden by county	26
16	Transportation energy burden by average tract income.....	28
17	Transportation energy burden by median income	29
18	Percentage of households nationwide and in four states spending above a given affordability threshold on transportation fuel	32
19	Key characteristics for transportation energy burden mapped by census tract in IL. Top left: Transportation energy burden; Top right: Median household income; Bottom left: stock-weighted fuel economy; Bottom right: household annual VMT.....	33
20	Key characteristics for transportation energy burden mapped by census tract in Chicago area. Top left: Transportation energy burden; Top right: Median household income; Bottom left: stock-weighted fuel economy; Bottom right: household annual VMT.	34
21	Stock-weighted MPG difference: 2018 vs. 2016.....	35
22	Parametric sensitivity analysis of transportation energy burden	36

TABLES

1	Percentages of household by number of vehicles, workers and income group for a representative census tract (Tract ID: 17031770602).....	6
2	Data used in this study	7
3	Comparison NHTS Mean Values and Regression Estimates	13

TABLES (CONT.)

4	Calculated VMT for a fixed number of vehicles, workers and income group for a representative census tract (Tract ID: 17031770602).....	15
5	Regression Estimates vs. California Household Travel Survey Estimates.....	17
6	Nationwide Household VMT and Transportation Energy Burden by Urban Type.....	27
7	Regional Household VMT and Transportation Energy Burden by Urban Type.....	27
8	Transportation Energy Burden by Tract Average Income.....	28
9	Comparison of Energy Burden by State.....	30

ACKNOWLEDGEMENTS

The efforts of all three authors of Argonne National Laboratory (under Contract DE-AC02-06CH11357) are supported by the Strategic Analysis (SA) Team, Energy Efficiency and Renewable Energy Office, U.S. Department of Energy. The authors would like to thank Ookie Ma and Kara Podkaminer of SA for the continued support of this project and useful and constructive comments. The authors also thank Dylan Hettinger of National Renewable Energy Laboratory for his support with IHS Markit data processing, and thank Zicheng Bi and Longwen Ou of Argonne National Laboratory for their assistance with data analysis and helpful discussion.

EXECUTIVE SUMMARY

Transportation energy is an important component of household budgets. Household vehicle fuel, one part of total transportation energy costs is over 3% of total household expenditures nationwide in the U.S. The average annual expenditure of over \$2,000 for vehicle fuel (mostly gasoline) is comparable to the total average household expenditures for electricity and natural gas combined. However, these average values vary geographically, and lower income households can face higher energy cost burdens. The highly resolved data developed in this study reveals the variation of household energy burdens across the country and provides localized data to support local decision making.

In this study, we enable the quantification of *household transportation energy affordability* in terms of the transportation energy burden, defined as the percentage of annual household income spent on the household vehicle fuel costs.

In addition to being a large household expenditure, fuel costs are also the most volatile cost component of total household transportation expenditures. Household transportation energy burden depends on the annual vehicle miles traveled (VMT), fuel price, and vehicle fuel efficiency. This study analyzes each of these quantities at high spatial resolution and then quantifies resulting burden of household transportation fuel costs at the census tract levels, shown in the equation below. This study focuses on fuel costs, but the finer resolution data and analysis framework developed and used in this study could also be used to assess additional affordability impacts of vehicle ownership costs and other household transportation costs.

$$\text{Transportation energy burden} = \frac{\text{Cost}}{\text{Income}} = \left[\frac{\text{mi}}{\text{Household} \cdot \text{year}} \times \frac{\text{gal}}{\text{mi}} \times \frac{\$}{\text{gal}} \right] / \frac{\$}{\text{Household} \cdot \text{year}}$$

There are four major research steps in this study to calculate the household transportation energy burden at the Census tract level:

- 1. Project annual household VMT based on demographic factors using machine-learning techniques.** We used the 2017 National Household Travel Survey data to develop 18 different regression models projecting annual household VMT for six census geographic regions and three levels of housing density. The geographies selected include the Northeast, Midwest, South Atlantic, South Central, Mountain, and Pacific divisions. We used the self-reported annual miles driven by all the vehicles from the same household in the NHTS as the dependent variable in the household VMT model. Eight independent variables were selected based on review of related published studies, data availability, and predictive capabilities of these variables for our VMT models: number of household workers, number of household vehicles, housing unit density, household income, and four variables for household lifecycle factors. Lifecycle factors represent the number of children and seniors in the household. For each census tract in the United States, we obtained household counts based on income group, number of vehicles, and number of workers. There are 11 income groups, 5 vehicle groups, and 4 worker groups for a total of 220 different classifications for each census tract. We made fine-grained VMT estimates within each tract by feeding in each combination of

income, vehicles, and workers to our model while holding other tract-level variables constant.

- 2. Estimate local differences in vehicle fuel efficiency based on household vehicle registrations.** Next, we estimated stock-weighted miles per gallon (MPG) by tract by combining vehicle registration data (IHS Markit 2018) with the EPA-reported fuel economy for each vehicle. The stock-weighted MPG for each fuel type was multiplied by the cost of fuel (in \$/GGE) to find the cost per mile of operating each vehicle.
- 3. Determine representative fuel price by census tract.** Then, we used high-resolution gasoline prices from GasBuddy.com and state-level alternative fuel prices from the U.S. Department of Energy to estimate typical household fuel costs for each vehicle at the tract level.
- 4. Quantify the resulting household transportation energy burden by census tract.** Finally, these three metrics (annual household VMT, stock-weighted MPG, and fuel price) were combined with tract-level income data to calculate the household transportation burden for every census tract in the U.S.

Analyzing the resulting household transportation energy burden results demonstrates variability across regions and shows sharp disparities in affordability by socioeconomic variables and by urban form (defined by housing density). The national average household transportation fuel cost has a population-weighted mean value of 3.3% of total income. However, with the increased geographical and socioeconomic resolution developed in this study, we found that the burden by census tract varies between 0.09% and 23.3% (for households with at least one vehicle). Figure ES-1 shows the household burden of transportation fuel cost. Despite increasing household VMT for higher household income groups, transportation energy burden decreases as the household income increases. States in the South Central region tend to have higher percentages of household income spent on vehicle fuel, with lower median household income. States in the Northeast region tend to have lower percentages of household income spent on vehicle fuel, with higher median household income.

This analysis identifies how the different factors influence overall burden in different areas or groups. Suburban and rural households spend more on vehicle fuel costs compared to urban households primarily due to less fuel-efficient vehicles and higher annual VMT. Lower-income groups have a wide distribution of transportation energy burden than the national average, with the most burdened quartile of census tracts in the lowest income group above 6.3%, but another quartile below 3.6%. In contrast, the highest income group, \$125,000/yr and above, has the narrowest distribution with half of the tracts falling between 1.8% and 2.6%. We can also illustrate the distribution of energy burden by assessing the portion of households that spend more than a given level. For example, almost half of households across the U.S. experience more than 2% transportation energy burden, while only 5% of households experience more than 7% transportation energy costs.

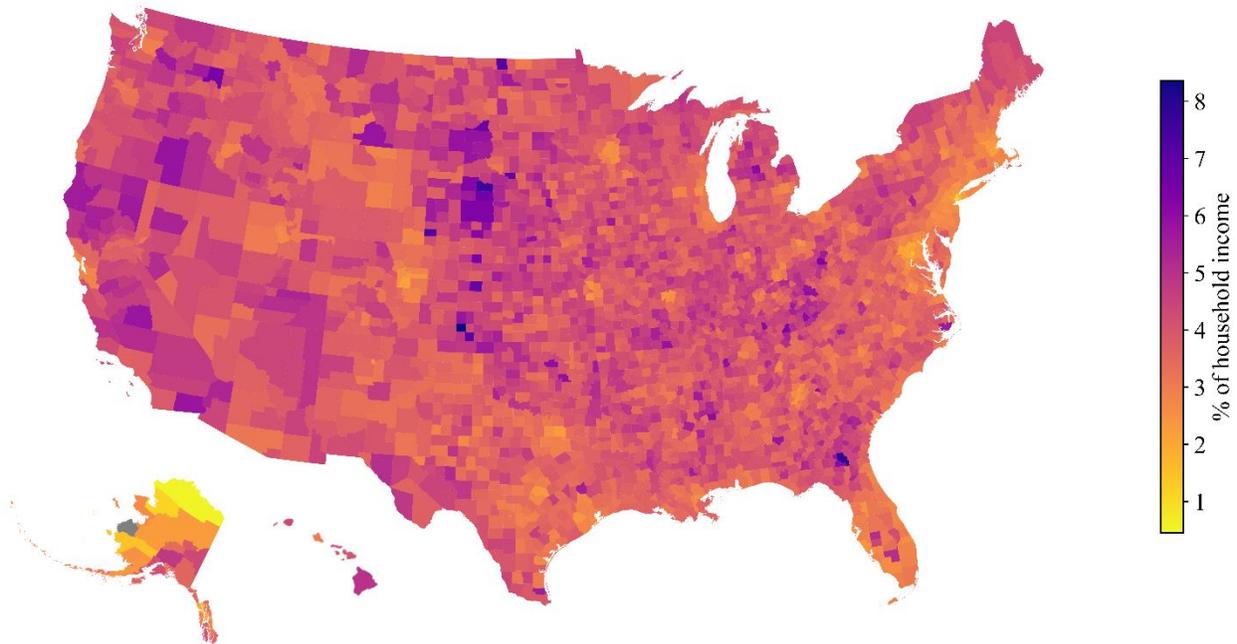


FIGURE ES-1 Transportation energy burden by county (grey area = no data)

This variation in transportation energy burden can be largely explained by vehicle fuel efficiency, measured as the stock-weighted MPG. Currently, wealthier census tracts have better fuel economy on average. Based on this analysis, the adoption of more fuel-efficient vehicles, especially among low-income households, could have the biggest impact on improving household transportation energy burden. This improvement in fuel efficiency of vehicles would result in lower transportation energy burdens, just as a previous 3% improvement in stock-weighted MPG from 2016 to 2018 saved American households \$8.2 billion in vehicle fuel costs.

This study provides a finer understanding of the spatial variation in household transportation energy burden by connecting VMT, vehicle fuel economy, fuel costs, and income data at the census tract level. The baseline data and framework developed here can be used to assess the spatially distributed impacts of additional transportation energy component costs or transportation policies on household transportation affordability. The full dataset is available to download and explore. Future research areas include identifying the households and communities that have potential to adopt alternative fuel vehicles or transportation modes based on their socioeconomic and other environmental factors, and quantifying the resulting affordability impacts.

This page left intentionally blank.

1 INTRODUCTION AND LITERATURE REVIEW

In the U.S., on average 21% of household transportation expenditures and 3.3% of total household expenditures are for vehicle fuel (1). The average annual expenditure for vehicle fuel was over \$2,000 in 2018 (mostly gasoline), which is comparable to the total average household expenditures for electricity and natural gas combined (1,2,3). Vehicle fuel costs are among the most elastic forms of energy consumption with respect to income (4). Further, vehicle fuel cost is also the most volatile cost component for household transportation expenses, being strongly dependent on oil prices, reaching 5.4% of household expenditures in 2012 (5). Nationwide, annual fuel expenditures of light-duty vehicles (LDVs) in 2018 were over \$300 billion (6). Annual fuel savings could reach \$94 billion to \$184 billion (2015\$) with the adoption of advanced vehicle technologies by 2050 (7). However, these benefits do not necessarily accrue evenly. Therefore, it is important to quantify the baseline annual household vehicle fuel cost and identify the variation by household characteristics and geography. Such a baseline will help decision-makers quantify the potential energy and economic impacts of promoting energy-efficient vehicle and mobility technologies at the household, regional, and national levels. Moreover, high levels of aggregation lose sight of variability; therefore, a finer resolution of data enables better insight into this spatial variability and can better inform decision-makers.

Household annual transportation energy costs depend on the vehicle miles traveled (VMT), fuel price, and vehicle fuel efficiency. Garceau found that U.S. states with higher automobile commute mode shares also have higher rates of VMT per capita, and a larger portion of household budgets devoted to transportation (8). For developing the baseline, the key research question is to estimate the household annual VMT based on household characteristics and geographic factors. The National Household Transportation Survey (NHTS) by the U.S. Department of Transportation's Federal Highway Administration (FHWA) gathers data on daily personal travel, including information on household and demographic characteristics, employment status, vehicle ownership, trips taken, modal choice, and other related transportation data pertinent to U.S. households (9). NHTS also provides a self-reported vehicle annual mileage. Oak Ridge National Laboratory (10, 11) estimates annual miles driven by each of the NHTS light-duty vehicles, referred as the BESTMILE, using a multi-variate linear regression model. The independent variables considered include odometer readings reported in NHTS, metropolitan statistical area (MSA) size, household income, vehicles per household, lifecycle factors, age, sex, worker status, and census region. However, despite being an excellent source of travel information for large geographic areas in the U.S., NHTS is a less suitable source for small geographic areas. Moreover, due to limited sample size, NHTS does not cover all census tracts. Henson and Goulias concluded that people in different geographic areas travel differently even if they share the same socio-demographic characteristics (12). Based on the NHTS, the Bureau of Transportation Statistics (BTS) developed a model to estimate average weekday household travel patterns including person miles traveled, person trips, vehicle miles traveled, and vehicle trips for six geographic areas (13). That finding serves as the basis of the LATCH model when separating the NHTS samples by region. LATCH developed a linear regression model to estimate daily household travel behavior at the census tract level as a function of household income, vehicles per household, workers in household, and lifecycle factors. Lifecycle factor is an indicator representing the number of children or senior people in the household. Hass et al. developed an

exponential function to predict the vehicle annual miles for Minneapolis–St. Paul area, based on four independent variables: household income, household vehicles, population density and household size (14).

The Center for Neighborhood Transportation developed an index to provide a comprehensive view of affordability that includes both the cost of housing and the cost of transportation at the neighborhood level (15). This study reports that transportation costs, which include both annual vehicle ownership costs and fuel cost, are considered affordable if they are 15% or less of household income. This study uses vehicle odometer readings from the Illinois Department of Natural Resources to derive a multi-variable linear regression model to estimate the vehicle annual miles as a function of 11 independent variables. These variables include: household income, vehicles per household, household size, household density, residential density, intersection density, employment access index, job diversity index, commute distance, per-capita income, and commuters per household. Several studies have noted that socioeconomic variables, such as the number of workers or children per household, influence travel behavior (16, 17). These variables are also closely associated with household income and size (14).

National Household Travel Survey respondents ranked “Price of Travel” as the most important of six transport issues considered, indicating that affordability is an important concern to transport system users (18). Agrawal et al. conducted in-depth interviews and examined how financial stresses such as reduced income and high fuel prices affect low-income families’ travel behavior and transport expenditures, as well as the costs and benefits of various travel modes (19). Similarly, this study finds that most low-income households are concerned about their transportation costs. The U.S. Department of Housing and Urban Development’s (HUD) location affordability index reports that vehicle ownership and vehicle fuel cost are 12.7% to 25.2% of household income in 20 different U.S. cities (20). Their results show that although large, compact cities tend to have relatively high housing costs, these are offset by higher incomes and lower transportation costs, such as in the Washington D.C., San Francisco, Boston, and New York areas. As a result, they tend to be more affordable overall, considering both housing and transport costs, than sprawling, automobile-dependent cities such as Phoenix, Miami, Atlanta, and Boston. The U.S. Bureau of Labor Statistics reported that vehicle fuel cost accounted for 3.5% to 13% of total household income in 2007 (21). Litman quantified the transportation affordability in Victoria, Canada by including all household expenditures on purchasing transportation services, varying from 10% to 20% (22). He found that (1) lower-income households tend to spend less in total but more as a portion of income than higher-income households, and (2) suburban and rural households spend more than urban households. Miller et al. (23), Mattingly and Morrissey (24) and Saberi et al. (25) found that suburban locations tend to spend more on transportation overall (both vehicle fuel and public transit) than city locations. Fitzroy et al. found that households in rural and mixed geographic areas tend to drive longer distances, but they also drive less fuel-efficient vehicles and do not have the same levels of market penetration by electric vehicles (26).

Moreover, there is extensive literature on the role of population density, built environment factors, land use characteristics, social networks, spatial dependency, socio-demographics, and macroeconomic conditions on household and personal travel behavior. Some studies have also analyzed the impact of the choice of residential location, neighborhood

characteristics, and vehicle choice on VMT (27–32). Singh et al. analyzed the relative contribution of all these factors on household VMT and found that while socio-demographic variables explain 12%, and self-selection effects account for 11% of the variability in household VMT, 44% of the household VMT remains unexplained, calling for further research in the domain (33). While these studies provide good references for the household characteristics and geographic factors to be considered when estimating household VMT, they only estimate either household daily mileage or annual vehicle mileage. Furthermore, some of the factors mentioned in the literature are not available from publicly available standard database.

All the studies reviewed have quantified transportation affordability of the U.S. at the city or state level but not at the tract level, or focus on other countries. Most of these studies combine vehicle fuel cost with vehicle ownership cost or housing cost, and only report the total portion of household income spent on these items. This study seeks to utilize and synthesis publicly available data sets, fill in data gaps, and to focus on the energy component of these costs. This study develops a baseline understanding of energy affordability by quantifying transportation energy burden as a function of household characteristics and geography at the census tract level with full coverage of the United States. Transportation energy burden refers to the percentage of annual household income spent on the household vehicle fuel costs. This study applied the machine learning technique of gradient boosting to estimate household annual VMT by census tract based on household characteristics and geography and used this household VMT to quantify the resulting household transportation energy burden.

2 OVERALL RESEARCH FRAMEWORK

The household transportation energy burden is defined as the percentage of the household annual income spent on vehicle fuel cost. Equation 1 shows that vehicle fuel cost for a household is calculated using the vehicle fuel consumption rate, expected local fuel price and annual modeled household VMT, and Figure 1 summarizes the overall research steps. In this study, we use tract-level averages for fuel economy and fuel price, while our VMT is modeled based on the income and demographic characteristics of each household.

$$\text{Transportation energy burden} = \frac{\text{Cost}}{\text{Income}} = \left[\frac{\text{mi}}{\text{Household}\cdot\text{year}} \times \frac{\text{gal}}{\text{mi}} \times \frac{\$}{\text{gal}} \right] / \frac{\$}{\text{Household}\cdot\text{year}} \quad (1)$$

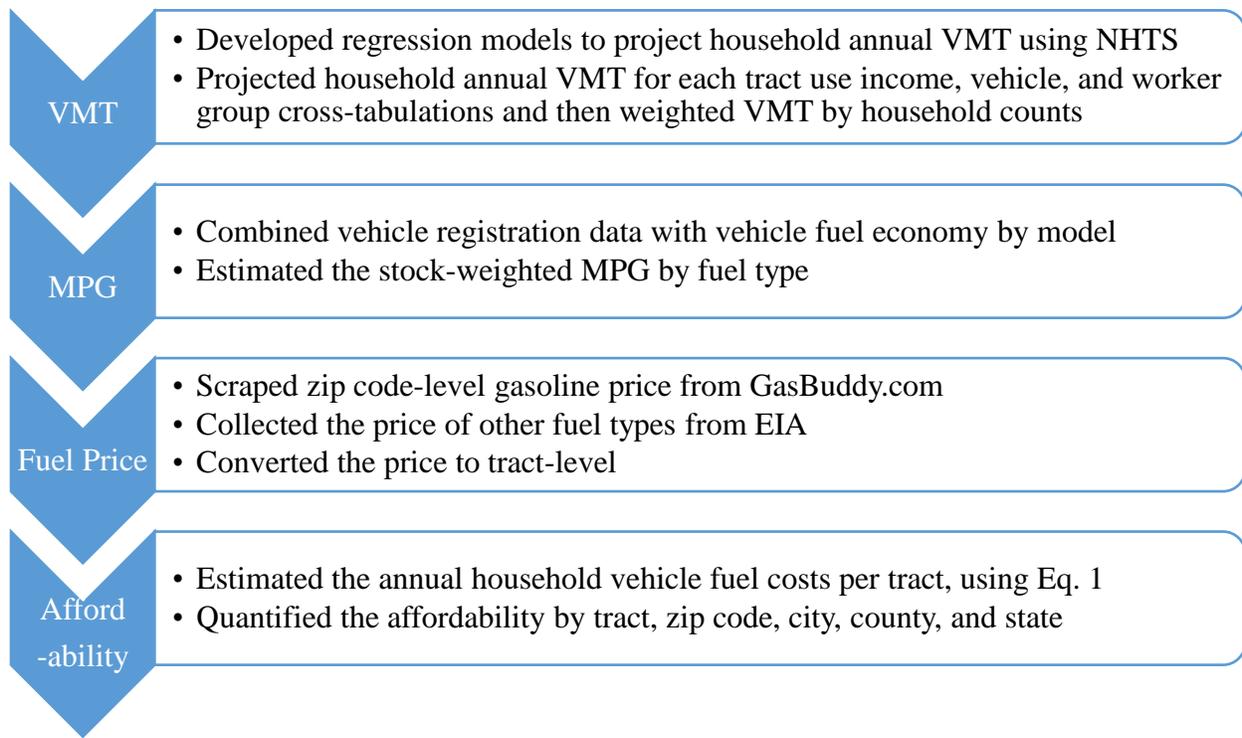


FIGURE 1 Overall Research Steps

In Eq. 1, the vehicle fuel economy, fuel price, and income information are compiled from exogenous data sources. However, household annual VMT at each census tract is not available and therefore is calculated in this work. There were two major considerations in the choice of data for the VMT prediction models. First, we selected a set of socioeconomic predictor variables supported by the literature as having an influence on household vehicle travel activity. Second, we identified data sets for the selected socioeconomic predictor variables that are available at a small geographic scale from standardized nationally available sources, such as the American Community Survey (ACS) from the U.S. Census (34). These data sources are also updated regularly so the results can be updated when new information becomes available.

These data sets are combined in Equation 2, which shows the functional dependence assumed for annual household VMT in this study.

$$\text{Annual miles per household} = f(\text{location, household workers, household vehicles, housing unit density, household income, household lifecycle factors}) \quad (2)$$

Choosing data variables that are available from standardized national databases enables us to estimate the household VMT for any census tract. Data from U.S. census and travel surveys are universally available at a fine resolution at the tract level, while other factors such as fuel price and vehicle registration are available at the zip code level. We convert between zip code and census tract as needed using the U.S. Department of HUD USPS Zip Code Crosswalk Files (35). The resulting model is used to provide a reliable prediction of annual VMT for a typical or average household by each census tract, but not to predict the travel behavior for each individual household. In particular, model prediction accuracy declines in regions with fewer households due to a higher chance of skewed average travel behavior. Household choices of vehicle model and trip distance could be based on preferences, rather than the energy efficiency and fuel cost. Moreover, certain exogenous factors, such as the ability to telecommute due to the COVID-19 pandemic, also influence household annual vehicle travel. However, these factors are not directly captured in this study.

After testing several variables and transformations of both dependent and independent variables, the final model for predicting the household annual VMT includes eight independent variables. The selected variables include the number of household workers, number of household vehicles, housing unit density, household income, and four variables for household lifecycle factors. These lifecycle factors represent the number of children and seniors in the household (see Table 2), which affect travel behavior. For example travel rates increase for people of working age and decrease for drivers over 65 years old (36).

For each census tract in the United States, we obtained cross-tabulated household counts based on income group, number of vehicles, and number of workers using the Low-Income Energy Affordability Data (LEAD) Tool methodology presented by the U.S. Department of Energy (37, 38). There are 11 income groups (\$0–5k, \$5–10k, \$10–15k, \$15–20k, \$20–25k, \$25–35k, \$35–50k, \$50–75k, \$75–100k, \$100–150k, and \$150k+), 5 vehicle groups (0, 1, 2, 3, and 4+), and 4 worker groups (0, 1, 2, and 3+) for a total of 220 different classifications for each census tract. For our model inputs, we use 3 and 4 to estimate 3+ and 4+ number of workers and number of vehicles, respectively. With these household counts in hand, we make fine-grained VMT estimates within each tract by feeding in each combination of income, number of vehicles, and number of workers to our model while holding other tract-level variables (described later in the report) constant. The household counts for each group then serve as weights when computing the average VMT in each tract. Using one census tract in suburban Chicago as the example, Table 1 demonstrates the resolution of income, number of workers and number of vehicles used in this study, showing the percentages of households that fall into each sub-category. Darker colors indicate that a larger fraction of households is within this classification, with all categories adding up to 100% for each census tract. In this census tract, the most common households are those with a single worker and a single vehicle, followed by households with two workers and two vehicles. The most common income group is between \$50,000 and \$75,000.

TABLE 1 Percentages of household by number of vehicles, workers and income group for a representative census tract (Tract ID: 17031770602). Data calculated for all census tracts.

Number of Workers	Number of Vehicles	Annual Household Income (thousand \$)										
		0-5	5-10	10-15	15-20	20-25	25-35	35-50	50-75	75-100	100-150	150+
0	0	1.99%	0.15%	1.13%	0.45%	0.66%	1.13%	0.26%	0.51%	0.09%	0.00%	0.06%
0	1	1.04%	0.08%	0.25%	0.02%	0.45%	2.09%	0.84%	0.88%	0.40%	0.26%	0.07%
0	2	0.20%	0.00%	0.23%	0.07%	0.32%	1.32%	0.60%	1.31%	1.41%	0.67%	0.48%
0	3	0.00%	0.02%	0.00%	0.00%	0.04%	0.12%	0.03%	0.01%	0.01%	0.02%	0.01%
0	4+	0.00%	0.00%	0.00%	0.00%	0.01%	0.01%	0.00%	1.14%	0.00%	0.00%	0.00%
1	0	0.17%	0.20%	0.43%	0.45%	0.11%	0.07%	0.01%	0.11%	0.00%	0.05%	0.00%
1	1	2.71%	0.37%	1.87%	0.87%	2.03%	4.06%	2.11%	5.36%	2.06%	0.81%	0.31%
1	2	0.59%	0.08%	0.38%	0.30%	0.76%	1.43%	0.89%	3.55%	1.61%	1.15%	0.57%
1	3	0.07%	0.00%	0.00%	0.00%	0.44%	0.29%	0.36%	0.40%	0.16%	0.11%	0.05%
1	4+	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.12%	0.03%	1.18%	0.96%	0.55%
2	0	0.00%	0.00%	0.08%	0.00%	0.00%	0.01%	0.02%	0.36%	0.01%	0.15%	0.01%
2	1	0.00%	0.01%	0.05%	0.03%	0.13%	0.90%	0.48%	1.11%	1.50%	1.16%	0.57%
2	2	0.17%	0.00%	0.03%	0.03%	0.12%	0.28%	0.47%	3.01%	2.67%	4.14%	3.21%
2	3	0.00%	0.00%	0.00%	0.02%	0.03%	0.44%	0.19%	0.88%	0.86%	1.27%	1.06%
2	4+	0.00%	0.00%	0.00%	0.00%	0.23%	0.08%	0.00%	0.00%	0.03%	0.03%	0.01%
3+	0	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
3+	1	0.00%	0.00%	0.00%	0.00%	0.00%	0.11%	0.00%	0.86%	0.33%	0.36%	0.14%
3+	2	0.00%	0.00%	0.00%	0.00%	0.00%	0.13%	0.10%	0.61%	0.33%	0.52%	0.50%
3+	3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.11%	1.21%	1.21%	1.64%	1.35%
3+	4+	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.02%	0.70%	1.53%	1.40%
All	All	6.9%	0.9%	4.5%	2.2%	5.3%	12.5%	6.6%	21.3%	14.6%	14.8%	10.4%

This study estimated the effective fuel price at a local level using gasoline prices from GasBuddy.com (39), and diesel prices and electricity prices from the Energy Information Administration (EIA) (40, 41). This study also estimated LDV stock-weighted vehicle efficiency by matching the vehicle registration from IHS Markit (formerly known as the R.L. Polk data) with real-world vehicle efficiency measured by the U.S. Environmental Protection Agency (EPA) by make, model, fuel type, and model year (42, 43). For each of these data sets, data is not directly available at the census tract level, so data was converted to the census tract level using the U.S. Department of HUD USPS Zip Code Crosswalk Files (36).

With this fuel price and vehicle efficiency data, we apply Equation 1 to each income, vehicle, and worker group in every census tract to estimate the percentage of annual household income that goes towards transportation energy. We also quantify the percentages of households in each tract that spend above or below certain affordability thresholds.

3 DATA AND METHODOLOGY

In addition to the 2017 NHTS data and LEAD tool-derived data, this study estimated household lifecycle factors and tract population from the ACS (34). Data sources used for VMT modeling and affordability analysis in this study are documented in Table 1. The newest vehicle registration data available to us when we conducted the study was the IHS Markit 2018 data, and we used 2018 ACS estimates to develop the baseline.

TABLE 2 Data used in this study

Data	Source	Description	Type
# of vehicles	NHTS (code: HHVEHCNT) (9) LEAD (38) / ACS 2018 (34)	Number of vehicles owned by the same household	Continuous
# of workers	NHTS (code: WRKCOUNT) (9) LEAD (38) / ACS 2018 (34)	Number of people need to go to work in the same household	Continuous
Household income	NHTS (code: HHGAMINC) (9) LEAD (38) / ACS 2018 (34)	Household annual income	Continuous
Housing units density	NHTS (code: HTRES DN) (9) LEAD (38) / ACS 2018 (34)	Housing units per square mile in the census tract of the household's home location	Continuous
Lifecycle factors	NHTS (9) ACS 2018 (34) BTS developed this variable from the age of the respondent in the NHTS data (13)	(5<=1+C<18): 1 or more children in household, less than 18 years old in the NHTS data but greater than or equal to 5 years old (1P hh<65): 1 person household, less than 65 years old (2+P hh, 0 65+): 2 or more person household, all less than 65 years old (2+P hh, 1+ 65+): 2 or more person household, at least one 65 or more years old	Categorical
Population	ACS 2018 (34)	Population per tract	Not used as an input to the VMT prediction model
Vehicle registration	IHS Polk Registration, 2018 (41)	Number of vehicle registered by powertrain type per zip code	
Fuel economy	FuelEconomy.gov (43)	Vehicle fuel efficiency by make and model in miles per gasoline gallon equivalent	
Gasoline price	GasBuddy (39), EIA (40)	\$ per gallon by zip code	
Diesel price	EIA (40)	\$ per gallon of diesel (regional), then convert to \$ per gasoline gallon equivalent	
Electricity price	EIA (41)	\$/kWh (state-level), then convert to \$ per gasoline gallon equivalent	
Compressed Natural Gas (CNG) price	Alternative Fuels Data Center Clean Cities Alternative Fuel Price Report, January 2020 (44)	\$ per gasoline gallon equivalent (regional)	

3.1 VARIABLES IN VMT MODELING

The dependent variable in the household VMT model is the self-reported annual miles driven by all the vehicles from the same household in the NHTS. We took a log transformation of the dependent variable. Eight independent variables are selected based on the literature, data availability, and feature importance as determined by our machine-learning models, as shown in Figure 2. Feature importance is a way of measuring how much a certain variable affects the prediction made by a machine-learning model. The variable with the most explanatory power is household income, followed by the number of workers and the number of vehicles in the household. Within most census tracts, there is a wide range of income, worker count, and vehicle count, and therefore these variables have been disaggregated for analysis, as described in Table 1, rather than use a single average value for the census tract.

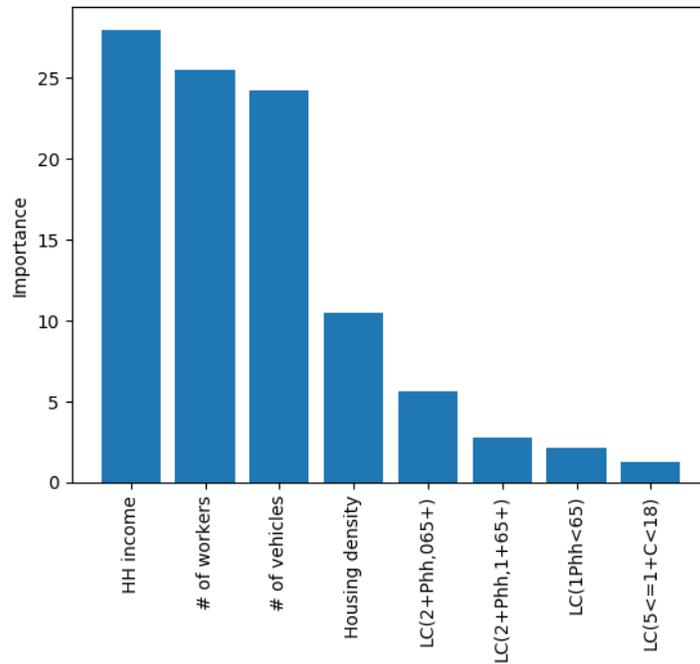


FIGURE 2 Feature importance of selected explanatory variables

The next most explanatory variable is housing density. We divided the 2017 NHTS sample of 129,696 households into six geographic regions (based on Census region/division boundaries) and then further disaggregated the regions into three levels of housing density, for a total of 18 separate groups. The VMT model produces estimates for each group separately. The geographic disaggregation enables more homogenous groupings of the households for the model. The geographies selected include the Northeast, Midwest, South Atlantic, South Central, Mountain, and Pacific Divisions. Some of these geographies are combined across multiple census divisions to have sufficient data for the machine-learning model. The division of the data into geographic areas follows from research by Henson and Goulias, which showed that people

in different geographic areas travel differently even if they share the same socio-demographic characteristics (12). Figure 3 shows the means for annual household vehicle travel as derived from the 2017 NHTS for households with at least one vehicle. The inset map shows the boundaries of the 6 geographic regions in this study.

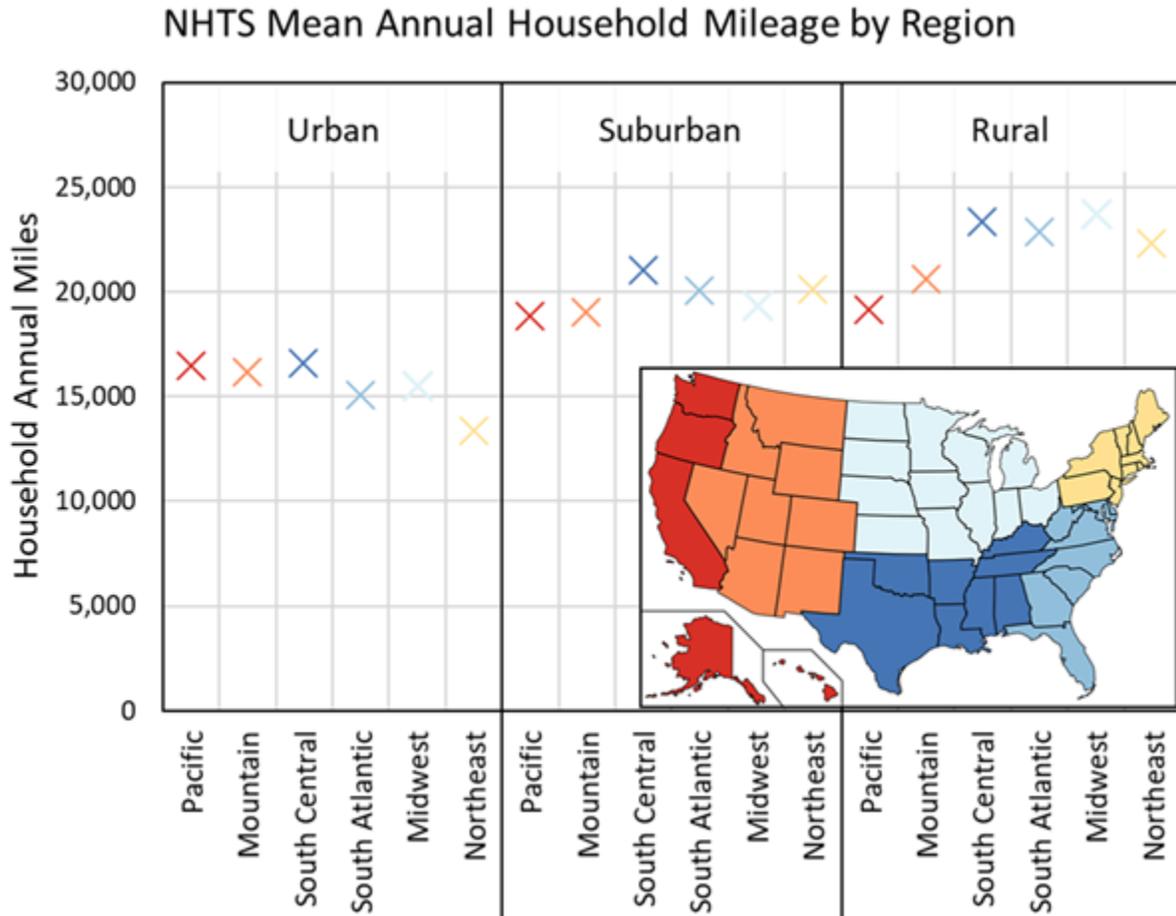


FIGURE 3 NHTS mean household annual miles traveled for each of the 18 geographic groupings. Inset: United States map showing the 6 geographic regions.

Urban households drive less than more suburban and rural households, as shown in Figure 3. To account for this difference in driving patterns, we classify each census tract as urban, suburban, or rural. Classifying urban form as a function of population density allows us to assign each census tract to a proper model for travel. BTS classifies census tracts as urban/suburban/rural using a method similar to the one used by Nielsen Claritas, Inc. to create the urban-rural continuum included in the 2009 NHTS. The classification uses information on the population density of a Census tract (converted to a centile score) and on whether the Census tract is in an urban area or urban region/division. Without access to the detailed methodology used by BTS and NHTS, we used the housing unit density (NHTS code: HTRES DN) to define urban, suburban and rural areas. With this method, we create separate regression models for each density group as well as use the actual value of HTRES DN as a numerical, independent variable

in the regression models. The NHTS data aggregates household density into eight bins, and so we rely on these delineations in our analysis. Areas with a density less than or equal to 99 housing units per square mile are considered as rural. Areas with a density between 100-1999 are considered as suburban, while areas with a density above 2000 housing units per square mile are considered as urban. These groupings are the closest to neighborhood descriptions recently classified by the U.S. Department of HUD (45).

Finally, the model accounts for lifecycle stage, with four descriptive variables representing the number of adults who are of working age or retirement age, the presence of children, and the number of occupants in the household. These factors do have correlations with travel behavior, but are less predictive than the aforementioned explanatory variables used in the model.

3.2 REGRESSION ANALYSIS

We applied three different regression analysis techniques, linear regression (LR), random forests (RF) and gradient boosting machine (GBM) on the NHTS samples to train the models. We implemented all models in Python, using the Scikit-learn package (46) for LR and RF, and using the Catboost package for gradient boosting (47). A random forest is a collection of decision trees whose outputs are all averaged for final prediction. A Gradient Boosting Machine (GBM) is an additive ensemble learning technique (somewhat similar to a random forest) where regression trees are constructed iteratively with each new tree an improvement on the last. Each “step” towards the next tree is based on the least-square errors of the previous tree, usually optimized via the steepest descent method. The final model is the sum of all the resulting regression trees (48).

To train all three of our machine learning models, our target variable was the base-10 logarithm of the reported household annual miles in the NHTS data set. The LR was implemented using Scikit-learn’s “LinearRegression” module, which implements ordinary least squares (OLS) regression. For the RF, we used Scikit-learn’s “RandomForestRegressor” module and set the number of regression trees to 300. On both the LR and RF models, we used 3-fold cross-validation to ensure the performances were robust and not over-fitting. The R^2 values of the best-trained LR and RF models are shown in Figure 4.

To train our GBM, we used the NHTS variables described in Table 2. We declared lifecycle factors as categorical and the rest as continuous. This is not technically the case for the NHTS dataset, because all of the selected variables are ordinal. However, treating them as numerical improves the model’s ability to generalize to the U.S. census data. There was no decrease in performance when treating the variables this way. Next, we split the entire data set into a training set (90% of the data) and test set (10%) using scikit-learn’s “train_test_split” function. On the training set, we perform hyperparameter optimization. This is a grid search on three different parameters: number of iterations, maximum decision tree depth, and learning rate, for a total of 120 different runs. Built into this parameter space sweep is 5-fold cross-validation, further splitting the training data into five permutations of training and validation sets. This is performed on each combination of parameters. The set of parameters that performs best in the

cross-validation (i.e., least amount of overfitting and lowest Root Mean Square Error) is returned. We chose the following hyperparameters: a learning rate of 0.1, maximum tree depth of 5, and maximum iterations of 1000.

After assessing the R^2 and Root Mean Square Error (RMSE) of LR, RF, and GBM models, we ultimately chose the GBM to determine annual household VMT in this study due to consistently better performance over the other two models across the 18 different geographic groupings. Figure 4 shows the test R^2 performance comparison of our LR, RF and GBM models when applied in the entire NHTS dataset, indicating that the GBM has the best predictive performance after about 200 training iterations.

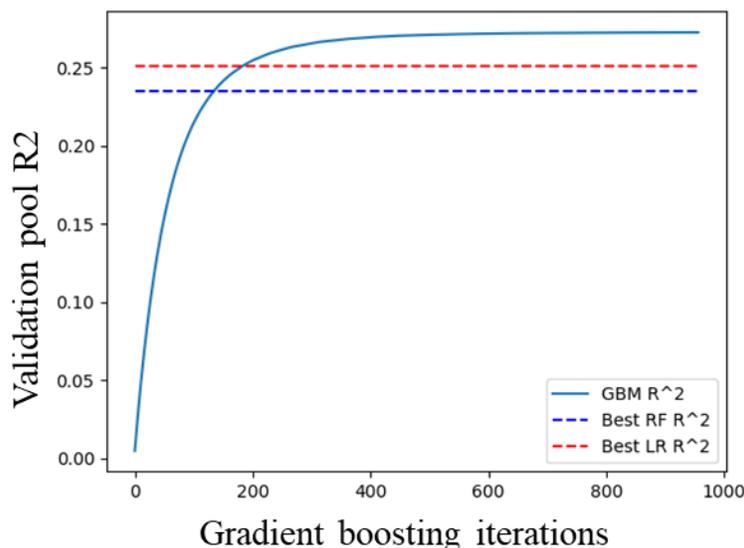


FIGURE 4 Performance comparison (R^2) of LR, RF and GBM.

After the GBM hyperparameter optimization, the entire data set is divided into six regions and three urban groups (18 total groups), since we find that many regions have better performance when trained individually than when trained as part of the entire data set. A new GBM is trained on each group using the selected hyperparameters so that each region and density level has its own model. In this step, a validation set is also derived from each region's training set to implement early stopping and to prevent overfitting. Each trained model is saved to be used later for tract-level VMT predictions on census data. The training performance of each regional model is shown in Figure 5 by comparing predicted annual household VMT with self-reported VMT by household. Taken at face value, the R^2 are not impressive, but our goal is not to accurately predict every single household. We aim to make a reasonable estimate of household VMT for any given tract. A difference between predicted $\log(\text{VMT})$ and actual $\log(\text{VMT})$ (from NHTS) appears to be lognormal distributed, and tapers pretty quickly, as shown in Figure 6. Literature shows high R^2 values for models predicting daily miles and daily trips. Our R^2 values are comparable with the values reported in (14), which also estimates annual miles. The H+T Index from the Center for Neighborhood Transportation has relatively higher R^2 when projecting

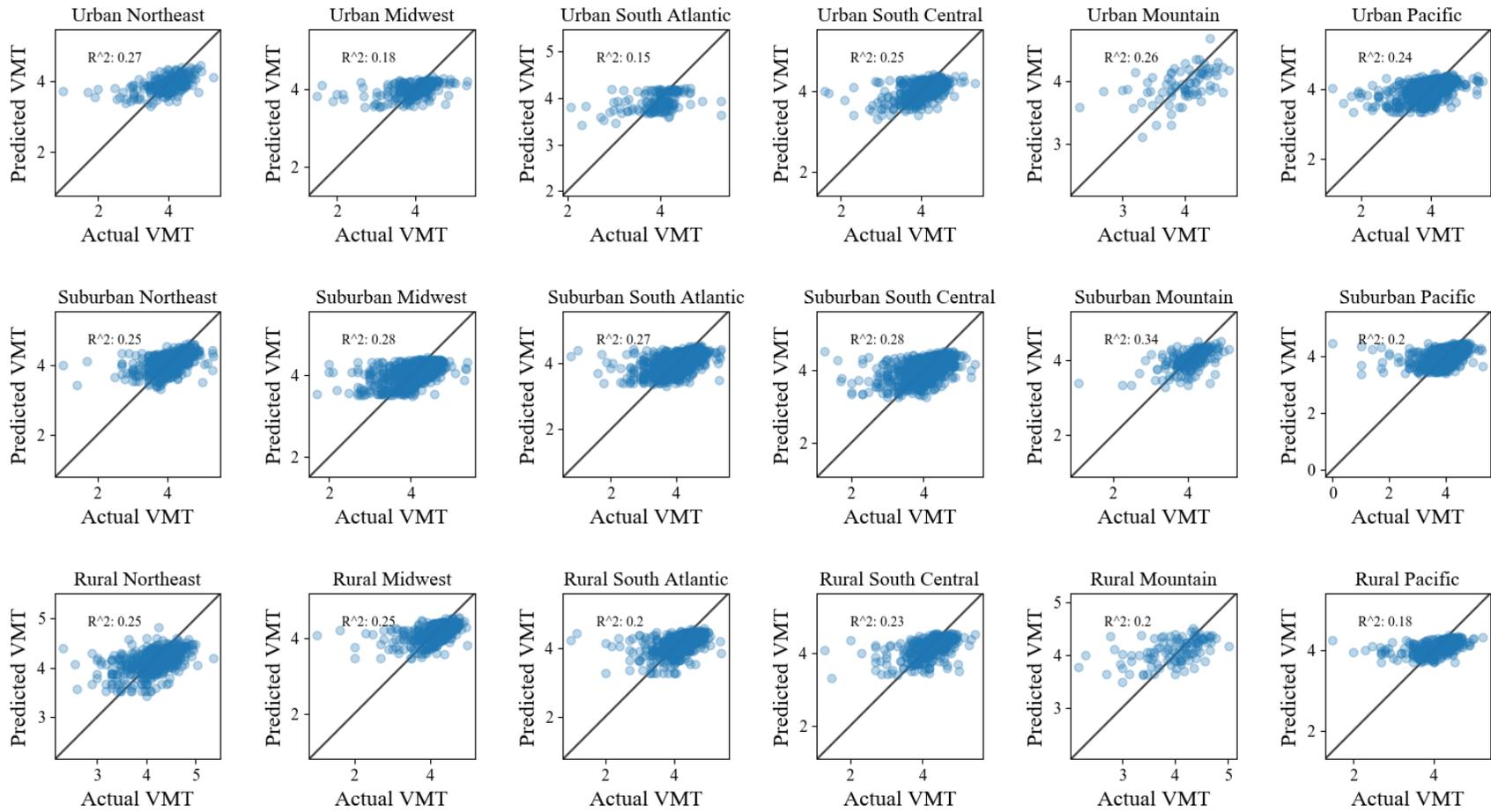


FIGURE 5 Training performance (R²) of each regional model using the GBM

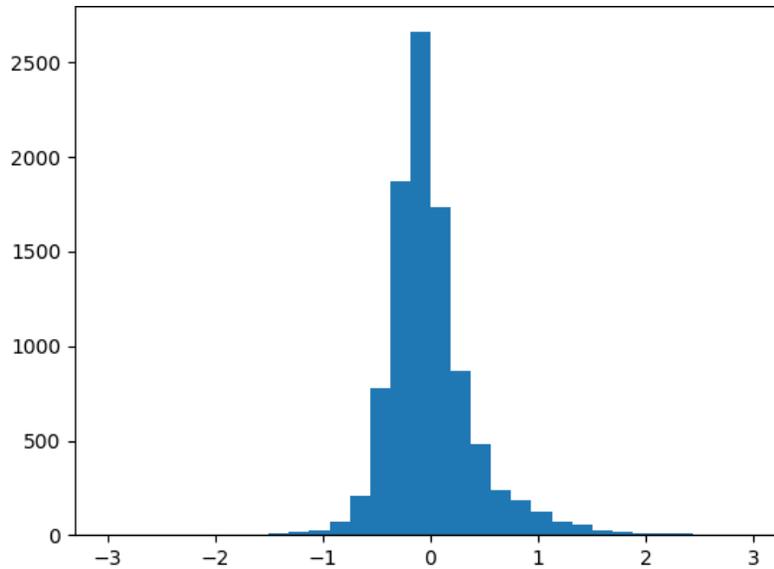


FIGURE 6 Distribution of errors: predicted $\log_{10}(\text{VMT})$ minus surveyed $\log_{10}(\text{VMT})$

vehicle annual miles by fitting additional parameters to improve their model performance (15). However, those additional variables are difficult to find at the zip code or census tract level from standardized databases. The literature also shows that estimating household VMT is a difficult problem because a significant portion of household VMT remain unexplained by the social-economic and demographic factors. Table 3 compares the regression estimates to the weighted values calculated from the NHTS sample in each Census region/division and urban group. The regression estimates are population-weighted means, while NHTS values are weighted using the household/vehicle weighting factor (WTHHFIN). All regression estimates are close to the NHTS mean values, and are within a 90% confidence interval of the mean values.

TABLE 3 Comparison NHTS Mean Values and Regression Estimates

Region	Density	NHTS 5 th percentile	NHTS Mean	NHTS 95 th percentile	Mean Regression Estimates (this study)
Northeast	Urban	1,000	13,365	35,293	12,076
	Suburb	2,000	20,085	44,000	23,431
	Rural	2,300	22,278	49,000	24,482
Midwest	Urban	1,200	15,530	37,875	16,248
	Suburb	1,500	19,364	44,000	20,668
	Rural	2,000	23,753	54,500	24,892
South Atlantic	Urban	1,200	15,070	33,620	15,464
	Suburb	1,200	20,057	47,356	20,051
	Rural	1,000	22,813	54,788	22,255

TABLE 3 (Cont.)

Region	Density	NHTS 5 th percentile	NHTS Mean	NHTS 95 th percentile	Mean Regression Estimates (this study)
South Central	Urban	1,211	16,580	40,000	17,877
	Suburb	1,301	21,019	49,671	19,768
	Rural	1,200	23,289	59,000	23,255
Mountain	Urban	1,200	16,155	36,350	17,140
	Suburb	1,200	19,006	41,337	18,736
	Rural	1,308	20,585	48,675	21,947
Pacific	Urban	1,200	16,414	37,000	16,817
	Suburb	1,460	18,803	40,000	22,146
	Rural	1,500	19,150	46,028	21,333

3.3 VMT MODELING BY CENSUS TRACT

In this analysis, VMT modeling is a function of eight variables shown in Figure 2: household income, number of workers, number of vehicles, housing density, and four variables representing the life stages of the household occupants. Every census tract has a housing density, similar to as described in Table 1, we disaggregate households with specific household income, number of workers, and number of vehicles into 264 different classifications. We cannot directly distinguish the demographics of the occupants of a household from knowledge of the number of workers, so we assumed that these variables are independent and that the distribution of households with different life stages is equivalent across all of these classes within each census tract.

We estimated VMT using the machine-learning model for each classification. The modeled VMT and subsequent burden calculations used a single value for each income bin, using the midpoint of each bin except for the highest-income bin, which was assumed to be at \$175,000. Table 4 shows the calculated VMT for each of these classifications in a given census tract outside of Chicago, Illinois. Household VMT is assumed to be zero in households with no vehicles. As described qualitatively, VMT generally increases as household income increases for all other combinations of workers and vehicles. Household VMT also increases with higher vehicle ownership, though per-vehicle VMT tends to decrease, with secondary vehicles being driven less intensively than primary vehicles. Likewise, VMT increases as the number of workers in a household. However, none of these functions are strictly monotonic, indicative of the complexity of the data fed into the machine-learning model. When multiplied by the population distribution across the 220 classes presented in Table 1, the average annual travel distance per household within the census tract can be determined. Calculations for households with four or more vehicles (from Table 1) use VMT for 4 vehicles to match the available data.

TABLE 4 Calculated VMT for a fixed number of vehicles, workers and income group for a representative census tract (Tract ID: 17031770602). Data calculated for all census tracts.

Number of Workers	Number of Vehicles	Annual Household Income (thousand \$)										
		0-5	5-10	10-15	15-20	20-25	25-35	35-50	50-75	75-100	100-150	150+
0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	2,972	2,972	3,234	4,013	4,013	5,620	7,574	8,192	9,390	8,799	9,016
0	2	5,986	5,986	4,796	5,511	5,511	7,586	10,213	11,325	13,359	14,228	15,853
0	3	6,177	6,177	5,103	5,648	5,648	7,615	9,594	10,576	12,144	12,469	14,915
0	4	3,215	3,215	2,951	5,393	5,393	8,761	11,384	12,718	15,565	15,949	18,541
0	5+	4,200	4,200	3,805	6,476	6,476	10,360	12,977	13,443	15,948	17,075	20,237
1	0	0	0	0	0	0	0	0	0	0	0	0
1	1	6,614	6,614	7,097	7,419	7,419	8,267	9,695	10,168	11,978	12,824	13,733
1	2	11,305	11,305	8,830	9,361	9,361	10,836	12,628	13,491	16,107	18,703	20,642
1	3	10,348	10,348	8,356	8,802	8,802	10,200	12,174	13,181	15,976	19,249	22,099
1	4	4,959	4,959	3,548	7,065	7,065	9,586	11,342	12,282	15,801	19,258	22,345
1	5+	5,581	5,581	3,962	7,321	7,321	9,786	10,949	10,465	13,714	19,535	24,310
2	0	0	0	0	0	0	0	0	0	0	0	0
2	1	10,860	10,860	9,919	9,529	9,529	9,960	10,965	11,296	13,262	14,619	15,529
2	2	19,010	19,010	12,725	12,750	12,750	13,054	15,673	16,373	18,504	20,635	22,114
2	3	16,457	16,457	12,048	12,370	12,370	12,120	15,185	16,091	18,428	21,443	23,810
2	4	20,120	20,120	14,553	12,807	12,807	12,640	14,148	14,815	17,565	21,090	24,069
2	5+	21,518	21,518	16,081	14,681	14,681	14,888	15,802	13,880	16,961	23,737	28,890
3+	0	0	0	0	0	0	0	0	0	0	0	0
3+	1	7,572	7,572	10,028	12,251	12,251	13,084	13,639	13,045	14,291	14,830	15,389
3+	2	9,865	9,865	12,774	17,342	17,342	18,844	21,310	20,409	21,223	22,175	23,094
3+	3	8,371	8,371	12,390	17,287	17,287	18,629	21,356	20,548	21,514	22,973	24,948
3+	4	16,254	16,254	21,747	19,520	19,520	19,145	19,487	18,431	19,802	22,637	26,815
3+	5+	16,724	16,724	21,805	20,115	20,115	20,028	19,368	14,851	15,372	24,346	31,428

3.4 VMT MODEL VALIDATION

For validation of our VMT model, we compared our VMT regression estimates with both the national NHTS and with more local travel surveys. Figure 7 shows the calculated values for household travel across the same 18 geographic groupings, as compared with NHTS. Relative to the original NHTS data (as presented in Figure 3), there is good agreement by region between our modeled VMT and the average of the sampled VMT in NHTS. The largest differences are that our model outputs higher suburban VMT in the Pacific and Northeast regions. This is possibly because our modeled VMT is strongly dependent on household income (as can be seen for a single tract in Table 4), and these areas have large populations with relatively high incomes.

Surveyed and Modeled Annual Household Mileage

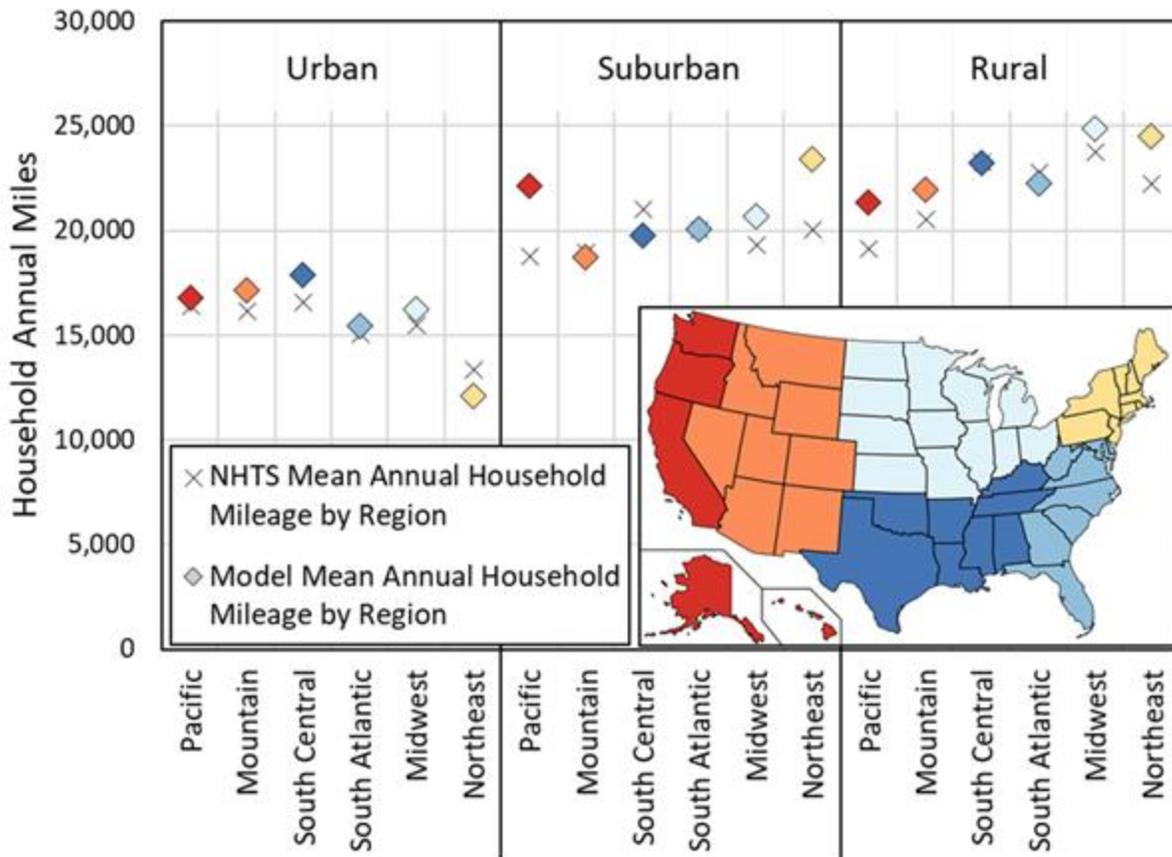


FIGURE 7 Mean household annual miles traveled for each of the 18 geographic groupings: NHTS vs. Model Estimates. Inset: United States map showing the 6 geographic regions.

For local validation, we considered regional household travel surveys conducted by Chicago Metropolitan Agency of Planning (CMAP) and The California Department of Transportation (Caltrans) (49, 50). These regional household travel surveys report daily vehicle travel activities in Chicago from 2007 to 2008, and in California from 2010-2013. To compare to the annual data in NHTS, we multiplied the daily vehicle mileages from the same household by 365 days. This method likely overestimates the annual household vehicle travel and may introduce a lot of noise due to the shorter sampling period. Figure 8 shows the comparison between regression estimates (Y-axis) and CMAP annual estimates (X-axis) per tract, by household income group. Each dot represents a tract from the survey, and the color of each dot represents the average household income within that tract. In general, our estimation is higher than the annual CMAP values. In addition to the possible overestimation from converting from daily to annual travel, our model likely overestimates travel for urban areas like Chicago in particular, because Chicago is the densest urban area in the Midwest. All urban areas in the Midwest share the same regression model, and higher density is correlated to lower household VMT. Moreover, the VMT model performs better for middle-income groups, while it tends to overestimate the VMT driven by higher income households and underestimate the VMT driven by the lower income households. In other words, the elasticity of the CMAP VMT data with

respect to income is lower than our VMT model. A similar trend was observed when we compared our modeled VMT with values derived from California Household Travel Survey, as shown in Table 5. Percentages represent the portion of our regression estimates that are higher than the values derived from the travel survey. A percentage of 50% reflects an unbiased estimate where half of the estimates are above and half of the estimates are below the survey data. For incomes near or above the median, our model largely agrees with the California survey data, but our model tends to underestimate travel of the lowest income groups.

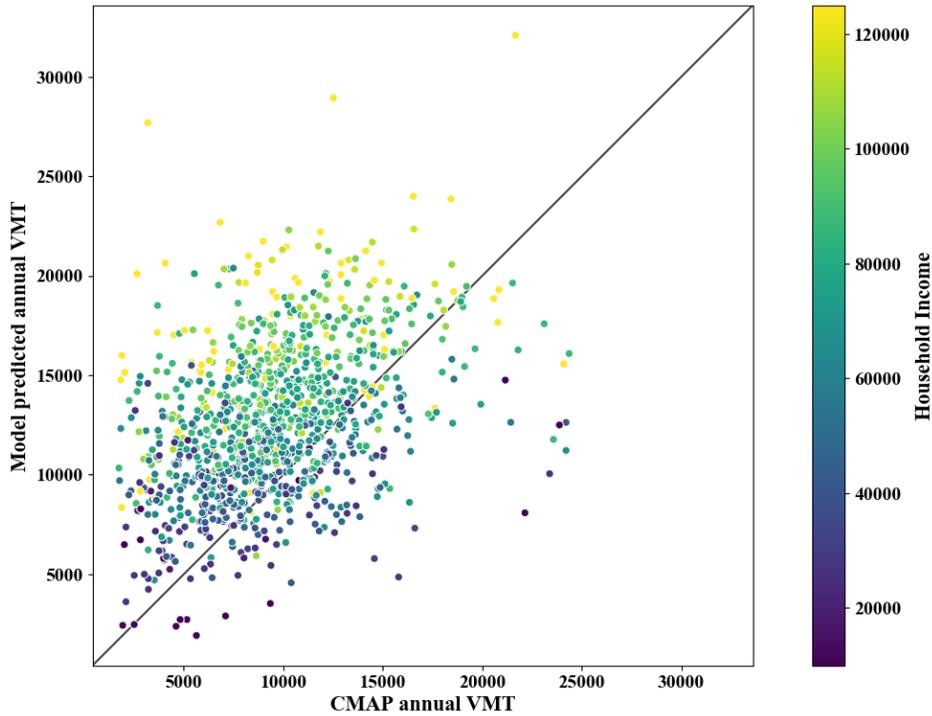


FIGURE 8 Comparison of modeled VMT and Chicago regional travel survey (Each dot represents a tract, CMAP annual VMT are extrapolated from daily miles)

TABLE 5 Regression Estimates vs. California Household Travel Survey Estimates

Income Group (\$/yr)	5000	17500	30000	42500	62500	87500	125000	175000	225000	275000
% > travel survey estimates	25%	39%	42%	47%	51%	54%	54%	56%	55%	54%

Stock-Weighted MPG

3.5 STOCK-WEIGHTED MPG

Light-duty vehicle registration data was sourced through IHS Markit, yielding the number of each vehicle make, model, and model year at the zip code level (42). This registration database was combined with the EPA-reported fuel economy to find the vehicle efficiency, measured in MPG for each vehicle (43). This process gives fuel economy for approximately 99% of vehicles in the IHS Markit registration data. The Fuel Economy Guide does not include vehicles older than 1984, and some vehicles in the registration database do not have sufficient information to ascertain their fuel economy. For each vehicle, we convert the fuel economy into miles per gallon of gasoline-equivalent (MPGGE) in order to average fuel economy across powertrains. The higher heat content of diesel fuel requires a factor of 0.885 to convert from a gallon of diesel to a gallon of gasoline. For electric vehicles, the conversion factor of 33.7 kWh/GGE was used (51). For plug-in hybrid vehicles, the Fuel Economy Guide uses the SAE J2841 multi-day individual utility factor to estimate the relative electricity and gasoline use for each vehicle, in order to estimate the electricity and gasoline fuel cost separately (52). Within each zip code, the stock-weighted MPG was calculated using the harmonic mean of MPGGE weighted by the number of registrations for each fuel type. The stock-weighted MPG for each fuel type was multiplied by the cost of fuel (in \$/GGE) to find the cost per mile of operating each vehicle. Figure 9 shows the stock-weighted MPG across the U.S., varying from 15.6 to 23.3 at the county level.

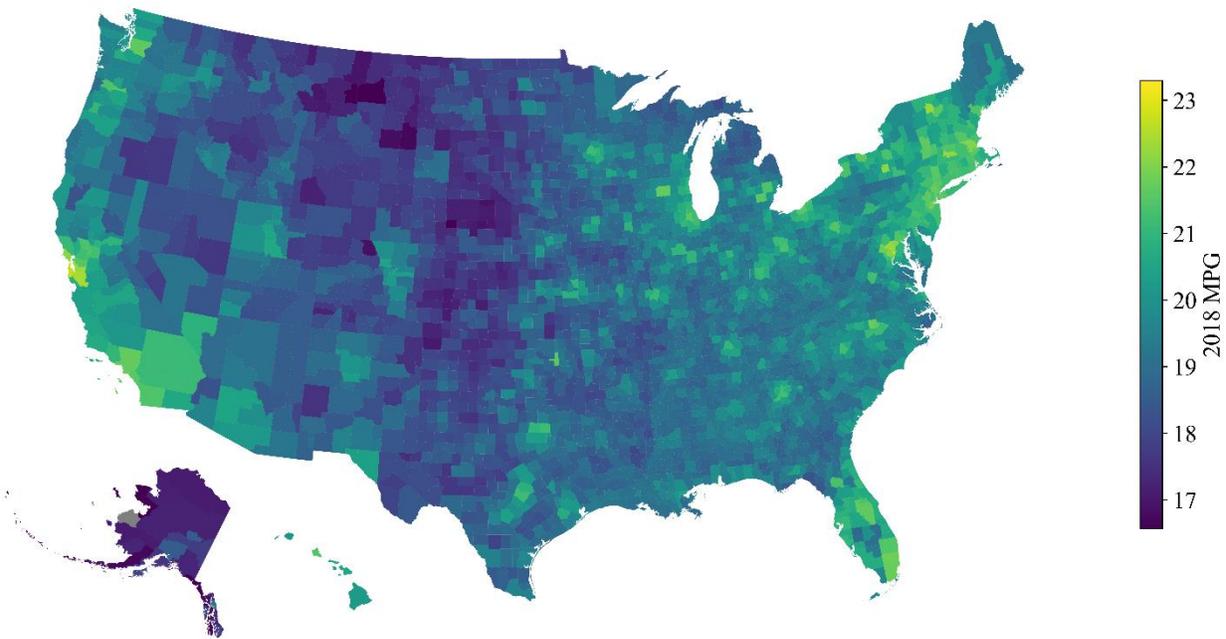


FIGURE 9 Stock-weighted vehicle MPG by county (2018 vehicle registration data)

Adoption of more fuel-efficient vehicles, either alternative fuel vehicles (AFVs) or newer gasoline vehicles with higher fuel economy, could lead to higher stock-weighted MPG. For example, Figure 10 shows that zip codes with newer vehicles have higher stock-weighted MPG in Washington D.C. Figure 11 also shows a positive correlation between number of alternative fuel vehicles adopted and stock-weighted MPG by zip code. AFVs include hybrid electric vehicles (HEVs), plug-in electric vehicles (PEV) and fuel-cell vehicles (FCEVs) which in general have higher fuel efficiency than their gasoline counterparts. Stock-weighted vehicle MPG at the zip code level were then crossed-walked with census tracts for the affordability calculation.

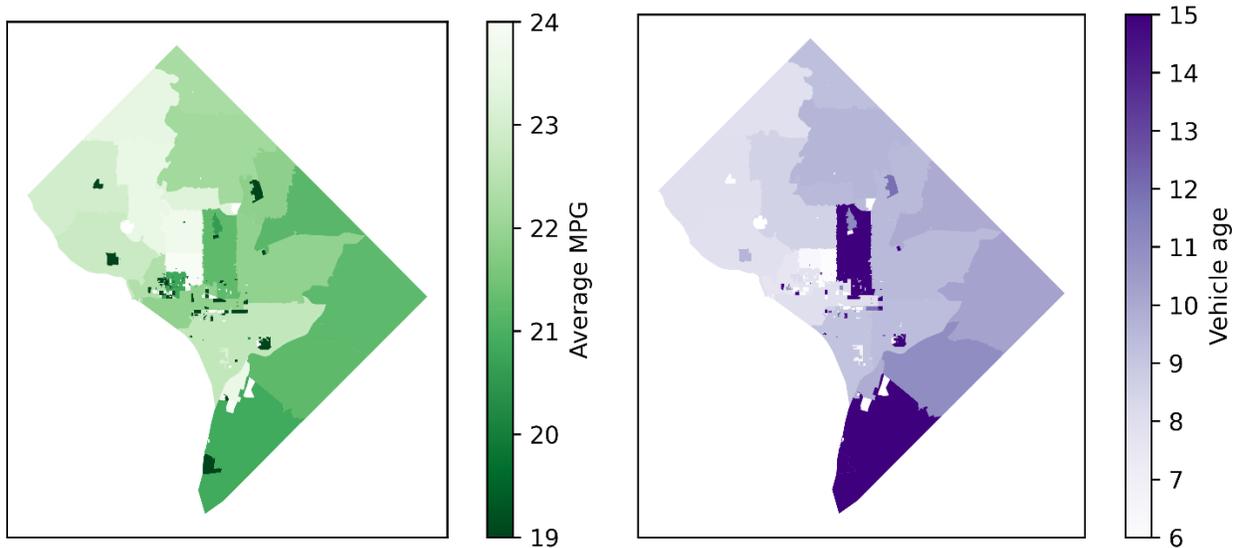


FIGURE 10 Stock-weighted MPG and average vehicle age by zip code in Washington D.C.

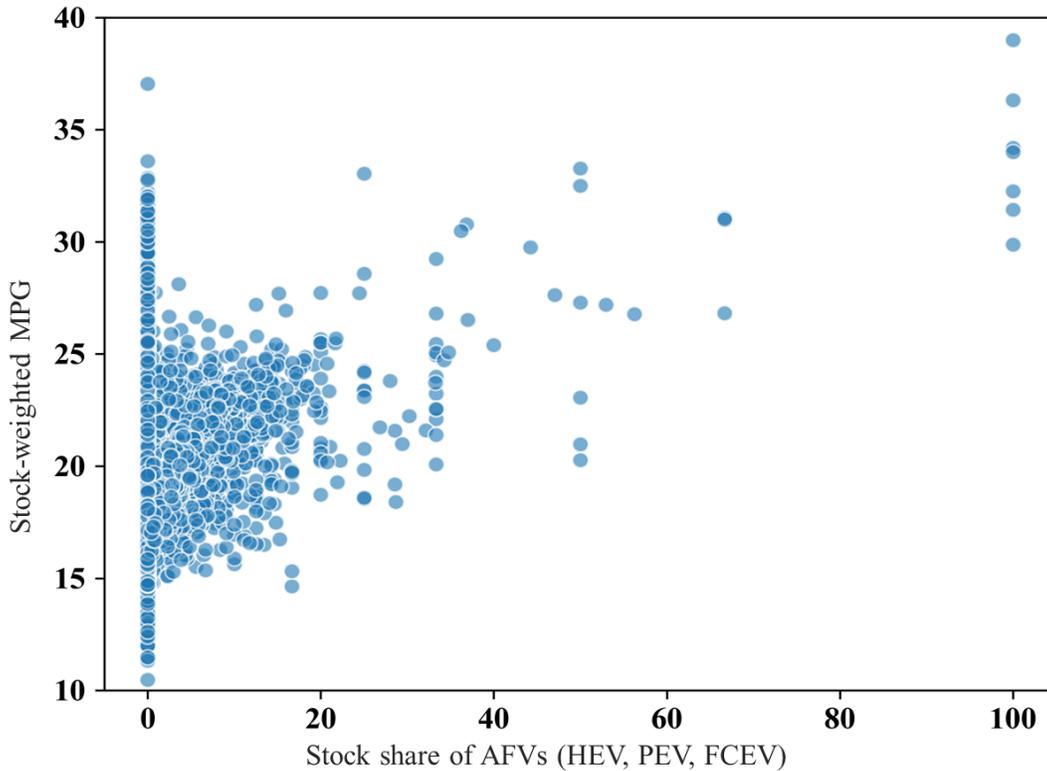


FIGURE 11 Alternative fuel vehicles adopted vs. stock-weighted MPG
(Each dot represents a zip code in the U.S.)

3.6 FUEL PRICE

To quantify the fuel price, we started with the finest resolution price data available for gasoline, diesel, electricity, and natural gas in this analysis. Since gasoline-fueled internal combustion engine vehicles (ICEVs) currently make up a vast majority of the private vehicles in the United States, and because there is a large degree of geographic variability, high-resolution gasoline price data are required. For this study, we scraped data from GasBuddy.com at the zip code level (39). GasBuddy data is populated by user-reported gasoline prices at individual stations, which are then aggregated to larger areas. To validate our use of GasBuddy data, we compared the data with historical fuel prices from the Energy Information Administration (EIA) (40). The EIA provides weekly gasoline prices for only nine states and ten metropolitan regions, which does not suffice for our study. At the national level, these two data sources have an average difference of \$0.014/gallon, or 0.6% of the total fuel cost, since January 2014. The two data sets agree with a Pearson correlation coefficient of 0.999.

Gas prices are neither uniform nationwide, nor are they constant in time. Gasoline prices have varied state-to-state by approximately 30 cents/gallon on average over the last six years. To further validate use of this data, we compare GasBuddy data with the nine states for which EIA presents data. Since January 2014, these data agree to within 2.2% for each state. At the state level, GasBuddy price data is biased higher than EIA, though typically only a few cents per gallon. At the metropolitan area, GasBuddy and EIA only compare similar geographies for

Cleveland, Denver, and Houston. In these three metropolitan areas, GasBuddy and EIA generally agree on gasoline prices to within 1%.

GasBuddy reports gasoline prices at the zip code level. These data were acquired for 30,635 zip codes nationwide. However, consumers may choose to fill up outside of their home zip code. Even within a city, fuel prices can vary based on location and marketing decisions, and consumers can select their fueling station accordingly. To account for this, we averaged gasoline prices by 3-digit zip code prefixes, which is a grouping by postal service sectional center facility, rather than by the entire 5-digit zip code, which is comparable in size to the census tract on average. The 30,635 zip codes with gasoline prices in the GasBuddy data are thus aggregated into 886 ZIP3 regions. This also serves to minimize statistical fluctuations from zip codes with outliers or missing gasoline prices. Figure 12 shows the gasoline price by ZIP3 in February, April, June, and September 2020. Zip codes trend from low on the east coast to high on the west coast, and a handful of specific states are labeled on the graph for clarity.

As demonstrated in Figure 12, gasoline prices have a high degree of temporal variability, but the relative fuel prices across the country exhibit similar trends. From January 2014 through July 2020, the average price for gasoline was \$2.552, with a standard deviation of \$0.441, or 17% of the total price. The gasoline prices in this study are based on a temporal snapshot of February 26, 2020. As described on the GasBuddy website: “The [Full Day Average] is published at 3:00AM eastern time and is an average of the last priced received at each station for the corresponding geography being viewed. The average price benchmarked for a particular day will be what was listed as the Full Day Average on that particular day.” (53) On this date, the national average for regular unleaded gasoline as reported by GasBuddy was \$2.46. This is in the 48th percentile of gasoline prices since January 2014, showing that this is a representative selection for gasoline prices at the national level. We further compared gasoline prices with high geographical precision. We compared the national gas prices at zip code level (or ZIP3) in February 2020, April 2020, June 2020, and September 2020. Over this timeframe, the crude oil price changed dramatically, as the COVID-19 pandemic greatly reduced travel demand (54) and disagreements between the Organization of the Petroleum Exporting Countries (OPEC) and partner countries on crude oil production cuts created volatility in the market (55). Average gasoline price dropped from \$2.46 to \$1.77/gallon from February 2020 to April 2020. Nonetheless, the fuel prices in every region were impacted, and localities with higher gasoline prices in February largely maintained higher prices in April, as indicated by a Pearson correlation coefficient of 0.821 between these two datasets. Over this timeframe, gasoline prices tended to drop the most in the Great Lakes region (e.g., MI, WI, IL); this local variability underscores the need for high geographic fidelity in gasoline prices. In Summer 2020, gasoline prices returned toward their values in February 2020, with a Pearson correlation coefficient of 0.918 between February and June data, indicating fewer low-price outlier zip-codes.

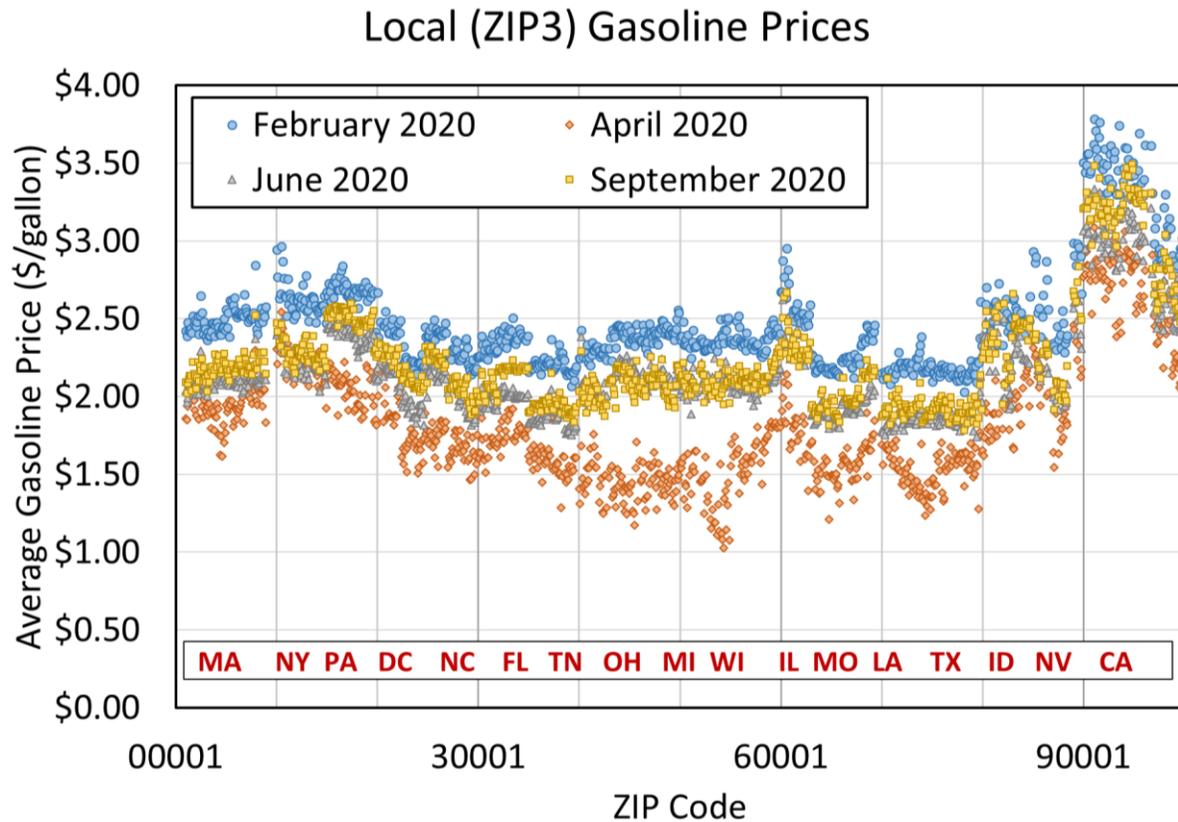


FIGURE 12 ZIP3-aggregated gasoline prices in February, April, June, and September 2020. Local gasoline prices are highly correlated throughout the year, with consistently higher in the West and near major cities (e.g., New York, Chicago). February 2020 fuel prices were used in this study.

As with MPGs, the fuel cost figures per zip code were cross-walked with census tracts for the driving cost calculation. In addition to gasoline-fueled ICEVs, we consider vehicles that use the following alternative fuels: diesel, electricity, and CNG. Diesel prices (\$/gallon of diesel) are from February 2020 EIA data, aggregated to the five Petroleum Administration for Defense Districts (40). Electricity prices (\$/kWh) come from EIA data on the state average residential price from February 2020 (41). CNG prices (\$/GGE) are from the January 2020 Alternative Fuels Data Center Clean Cities Alternative Fuel Price Report (44). Due to the data limitations at the zip code level, we did not consider the premium gasoline price used by some vehicle models.

3.7 DRIVING COST CALCULATION

We used the following equation to calculate cost per mile driven for each census tract:

$$C \left(\frac{\$}{\text{mile}} \right) = \frac{1}{\sum_i N_i} \times \sum_i \frac{P_i \times N_i}{MPGGE_i}, \quad (3)$$

where i is a specific type of fuel (gasoline, electricity, diesel, or CNG), P_i is the price per gasoline-gallon equivalent of fuel i , N_i is the number of vehicles that use fuel i in the tract, and $MPGGE_i$ is the stock-weighted MPGGE of vehicles that use fuel i in the tract. As with the fuel economy, we converted the fuel price for all fuels (electricity, diesel, and CNG) to \$/gasoline gallon equivalent within each tract, using lower heating values as appropriate. Because E85 fuel is generally more expensive than gasoline (44), in addition to the fact that flex-fuel vehicles (FFVs) run on E85 less than 2% of the total VMT nationally (6), we assumed that FFVs operate using gasoline rather than E85. The fuel economy and cost calculations presented here were then combined with census information on the number of vehicles per household and the above calculations on the number of miles per vehicle to find the total annual household cost for transportation fuels, per Eq. 1.

4 RESULTS AND DISCUSSION

We report here the household VMT and transportation energy burden by census tract. Given this large dataset, the discussion below highlights key trends. The full dataset is available to download and explore.

4.1 HOUSEHOLD VMT

There is wide variation in average household VMT by county across the U.S., from 2,507 to 40,985, as shown in Figure 13. All tract-level VMT and burden results are weighted by the number of households, then aggregated to a higher regional level (e.g., county, state, national) by population. The national average annual household VMT is 18,515. We used the VMT models developed for the Pacific West region to estimate the VMT for households in Alaska and Hawaii. However, households in these two states may have unique travel patterns, which are not captured by any of the 18 different VMT models we have developed for the lower 48 states. Future research is needed to improve models for these two states.

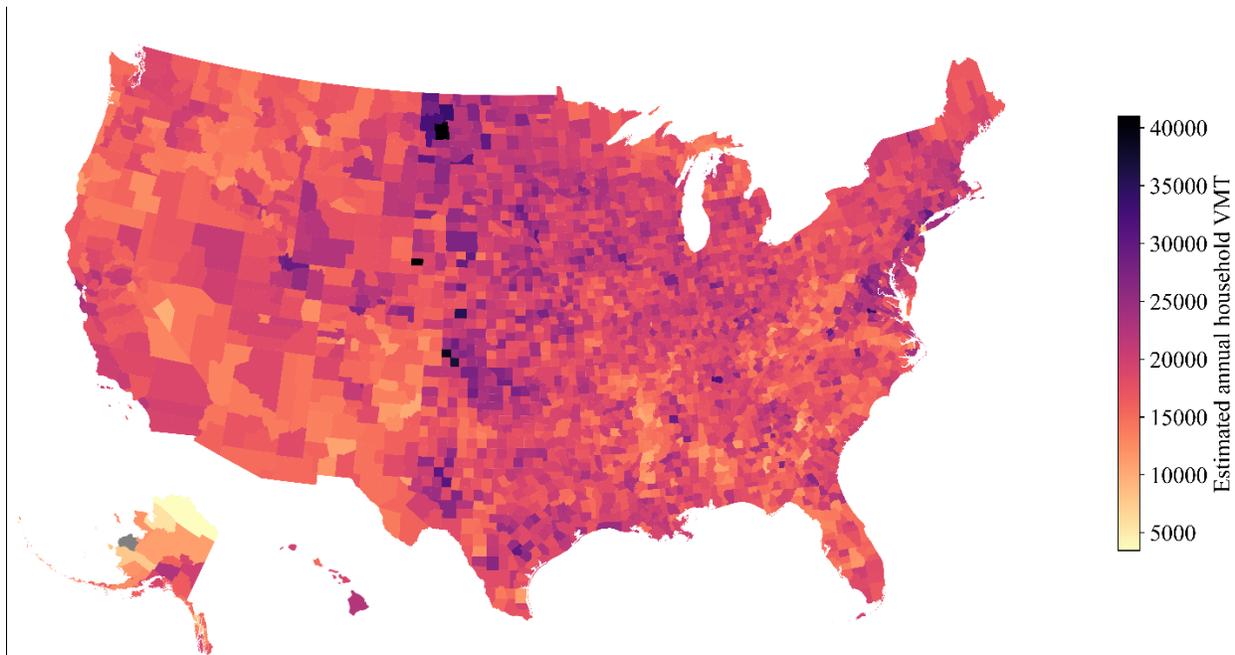


FIGURE 13 Household annual VMT by county (grey area = no data)

Household annual VMT varies by housing unit density and by region. Table 3 shows that suburban and rural households have higher annual VMT than the urban households across all regions. Midwest rural households have the highest annual VMT, followed by Northeast rural and South Central rural households, as shown in Figure 7. Households in the Pacific and Mountain regions tend to have lower annual VMT.

Moreover, household annual VMT varies by income group, as shown in Figure 14. In this figure, each dot represents a single census tract. The census tracts are binned by average income and presented in a raincloud plot, where individual data points are overlaid on a box-and-whisker plot, where the boxes represent the first and third quartiles, and the whiskers represent 1.5 times this interquartile range. This is plotted alongside a half-violin plot that represents the probability distribution function for census tracts within each income bin (56). Note that the largest probability distribution functions are in the income groups with annual incomes between \$40,000 and \$100,000. Because we use the regression model to determine household annual VMT, and there are outliers in the underlying explanatory variables in the model, some outliers in VMT are also generated in this process. For example, the highest estimated VMT in the country is a census tract centered on a naval shipyard in Kitsap County, WA, on the Puget Sound, where the average income is roughly \$154,000/year. The tract with the highest transportation energy burden is near a college campus, with income demographics skewed by college student housing, and housing density skewed by the presence of a golf course and football stadium. These types of census tracts are not particularly common, but the calculations of VMT and income at the tract-level are reasonable based on their demographic profiles. In order to make the figures easy to read, we truncated our VMT figure at 80,000 miles/yr, and the burden figure at 20%, without removing any samples from the subsequent energy burden calculation.

VMT increases as the average income increases. The average household annual VMT increases from 11,000 to 20,000 and then to 30,000 for the income groups \$0-19,999/yr, \$60,000-79,999/yr, and \$125,000/yr and above, respectively. Notwithstanding some outliers in other income bins, higher-income groups have the widest distributions of annual VMT, possibly due to the diversity of the numbers of vehicles owned by this group.

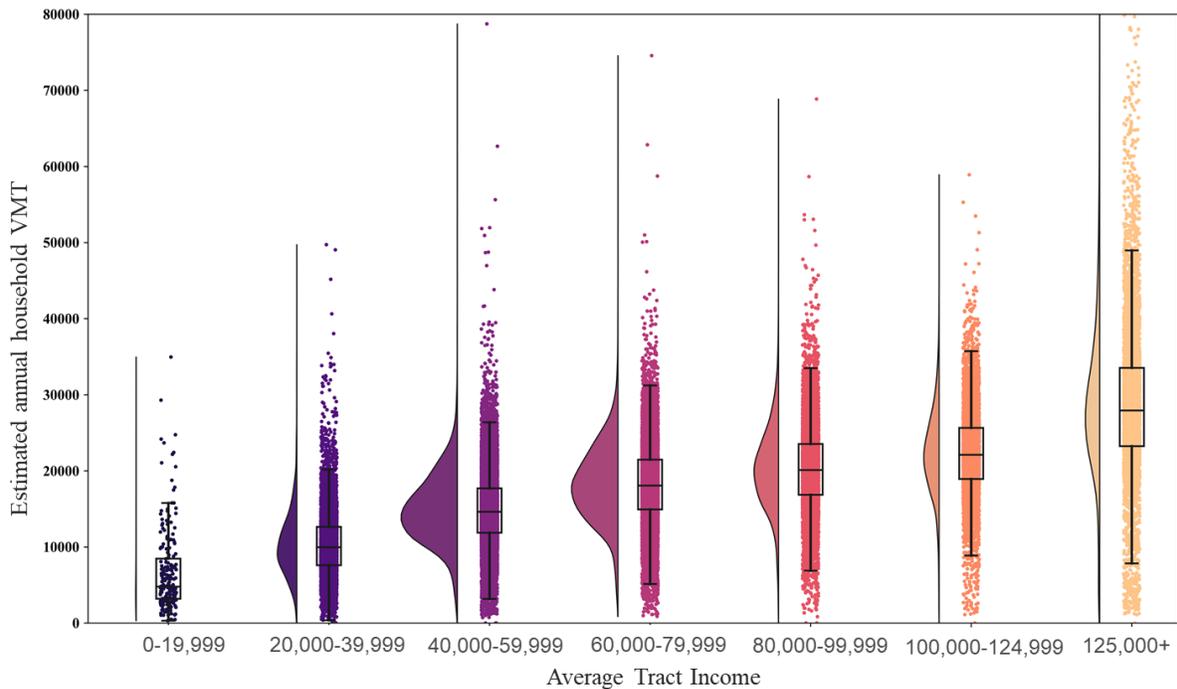


FIGURE 14 Household annual VMT by average tract income (each dot represents a tract)

4.2 HOUSEHOLD ENERGY BURDEN

The variations in household VMT, fuel price and stock-weighted MPG all contribute to the variation in the transportation energy burden (quantified from Equation 1) shown in Figure 15. At the national level, the distribution of household transportation energy burden has a mean value of 3.3%. However, county-level burden ranges from 0.5% to 8.4%, highlighting the importance of geospatial data. Clusters of counties with lower transportation energy burden align with the metropolitan areas. This is due to both the higher household income in metropolitan areas compared to rural areas and the reduced VMT in urban areas.

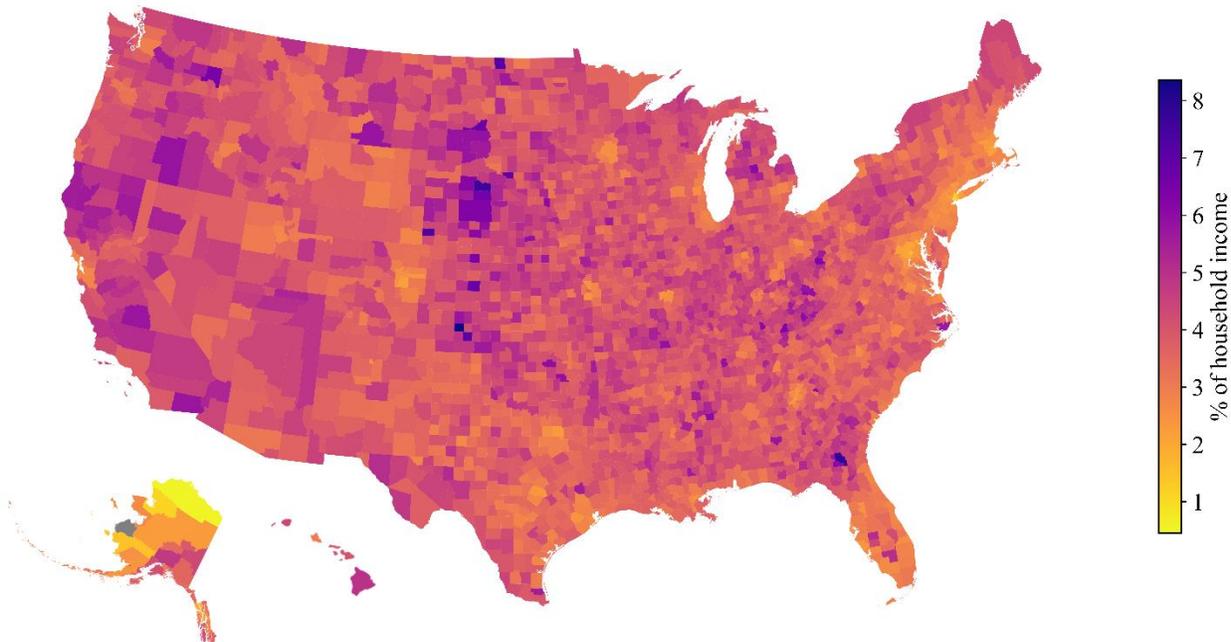


FIGURE 15 Transportation energy burden by county (gray area=no data)

Table 6 compares modeled VMT as a function of population density. Nationwide, rural households drive further and have a higher transportation energy burden than suburban and urban households. Rural households have higher burden for all regions, as shown in Table 7. Pacific rural households have the highest portion of their income spent on transportation fuel, followed by South Central rural and Midwest rural areas. Figure 16 and Table 8 show that despite increasing household VMT, transportation energy burden improves as the household income increases, with the average burden trending from 6.5% to 3.3%, then to 2.2% for census tracts where the average incomes are in the ranges from \$0-19,999/yr, \$60,000-79,999/yr, and \$125,000/yr and above, respectively. In contrast to the VMT trend, lower-income groups have the widest distribution of transportation energy burden, with the most burdened quarter of tracts in the lowest income group above 6.3%, but another quarter below 3.6%. The highest income group, \$125,000/yr and above, has the narrowest distribution with half of the tracts falling between 1.8% and 2.6%.

TABLE 6 Nationwide Household VMT and Transportation Energy Burden by Urban Type

Urban Type	Household Annual VMT			Burden		
	Median	5 th percentile	95 th percentile	Median	5 th percentile	95 th percentile
Rural	20,465	12,617	30,392	4.0%	2.8%	5.6%
Suburb	17,549	9,111	31,724	2.9%	1.9%	4.9%
Urban	13,418	4,593	24,023	2.8%	1.0%	5.1%

TABLE 7 Regional Household VMT and Transportation Energy Burden by Urban Type

Region	Urban Type	Household Annual VMT			Burden		
		Median	5 th percentile	95 th percentile	Median	5 th percentile	95 th percentile
Midwest	Rural	23,116	22,565	14,921	31,047	4.2%	3.1%
	Suburb	18,610	16,811	9,210	29,251	2.9%	2.0%
	Urban	14,786	13,217	5,720	24,246	3.0%	1.6%
Mountain	Rural	19,761	18,235	11,867	29,924	3.9%	2.7%
	Suburb	17,612	16,109	9,392	28,583	2.9%	1.9%
	Urban	15,228	14,667	8,350	22,239	3.3%	2.0%
Northeast	Rural	22,112	20,803	15,276	31,229	3.9%	2.6%
	Suburb	22,198	19,940	10,177	36,749	2.6%	1.8%
	Urban	11,435	10,493	2,752	22,607	2.1%	0.5%
Pacific	Rural	20,203	19,306	11,961	31,106	4.3%	2.9%
	Suburb	21,197	19,434	10,924	34,759	3.6%	2.2%
	Urban	15,481	14,919	7,849	24,041	3.4%	1.8%
South Atlantic	Rural	19,663	18,553	11,456	29,374	3.8%	2.7%
	Suburb	18,723	16,965	8,687	30,825	2.7%	1.8%
	Urban	14,967	13,450	4,862	25,377	2.4%	1.1%
South Central	Rural	21,014	19,843	12,108	29,685	4.0%	2.8%
	Suburb	18,771	16,698	8,211	29,660	2.9%	1.8%
	Urban	16,350	14,819	7,482	26,393	2.8%	1.7%

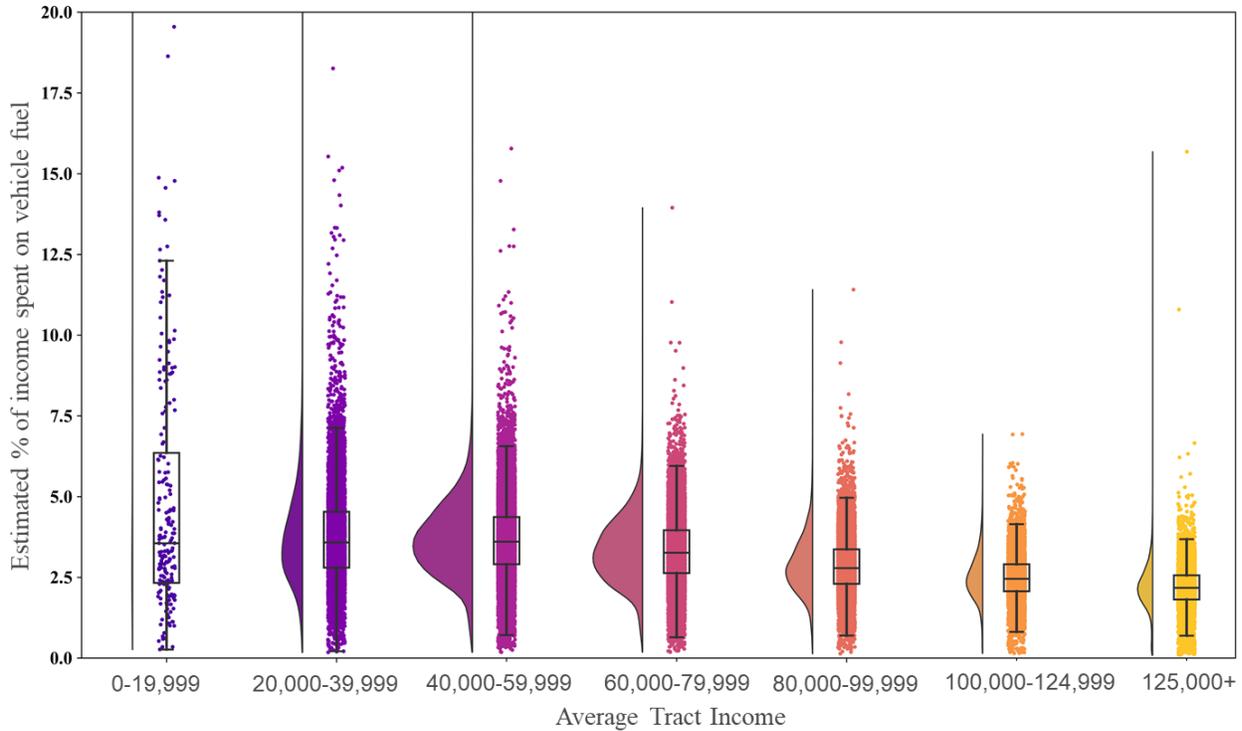


FIGURE 16 Transportation energy burden by average tract income (each dot represents a tract)

TABLE 8 Transportation Energy Burden by Tract Average Income

Income Range	Number of Tracts	Average Burden	Median Burden	Minimum Burden	Maximum Burden
\$0 – \$20,000	208	6.5%	3.6%	0.3%	21.4%
\$20,000 – \$40,000	7,876	3.9%	3.6%	0.2%	20.4%
\$40,000 – \$60,000	22,748	3.7%	3.6%	0.2%	23.4%
\$60,000 – \$80,000	18,782	3.3%	3.3%	0.2%	13.9%
\$80,000 – \$100,000	9,923	2.9%	3.3%	0.2%	11.4%
\$100,000 – \$125,000	6,213	2.6%	2.5%	0.1%	21.4%
\$125,000+	5,832	2.2%	2.2%	0.1%	15.7%

At the state level, the average transportation energy burden also declines as income increases, as shown in Figure 17. Bubble size stands for the population of the state, and the colors of the bubbles represent the geographic regions. Average transportation energy burden by state ranges from 1.4% to 4.0%. States in the South Central region tend to have higher percentages of household income spent on transportation energy, with lower median household income. States in the Northeast region tend to have lower percentages of household income spent on transportation energy, with higher median household income. South Dakota, Mississippi, and California have the highest household transportation energy burden, about 4.0%, while Washington, D.C. is the lowest, about 1.4%, followed by New York at 2.2%.

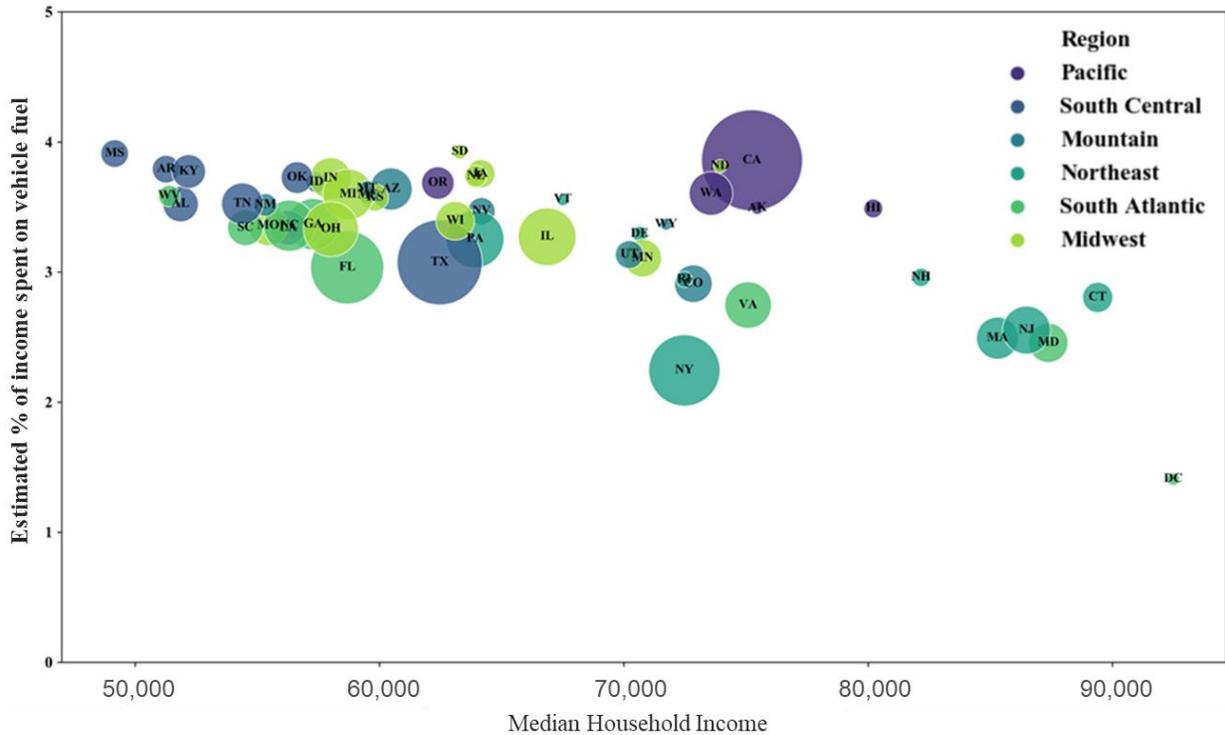


FIGURE 17 Transportation energy burden by median income

The variation in transportation energy burden at the state level is largely explained by the variation in on-road vehicle fuel efficiency. Table 9 shows a comparison of each of the components of the burden equation across each state. The components shown are stock-weighted fuel consumption rate in gasoline-gallon-equivalent per mile (GPM), median household income at the tract level, average household annual VMT, and fuel price (\$/gallon equivalent). The color-coding shows how each variable correlates to state average burden, with warmer (orange) colors representing states that are more affordable on that metric, and cooler (purple) colors representing higher burden. Darker colors are further from the national average, presented in the bottom row, with the deepest colors being three standard deviations from the mean. New York State has the lowest household transportation energy burden among states at 2.24%, though Washington, D.C. has a lower burden of 1.4%. South Dakota has the highest transportation energy burden, 3.95%, due to higher annual VMT and lower on-road vehicle fuel efficiency compared to most other states. Washington, D.C. is not included in Table 9 as it is an extreme outlier in VMT, GPM, and income, while Alaska and Hawaii are not included in the table due to lack of confidence in the VMT modeling.

Quantifying the similarity of these variables shows that the largest (anti-)correlation is between energy burden and income with a Pearson correlation coefficient of -0.67. In other words, as the average income in the state increases, the percentage of income spent on transportation fuel decreases (warmer color in Table 9). The second largest correlation is between energy burden and GPM, with a correlation coefficient of 0.60. This again indicates that

TABLE 9 Comparison of Energy Burden by State (continental U.S., purple = higher burden)

State	Average Burden	Average VMT	Fuel Consumption (gge/mi)	Fuel Price (\$/gge)	Median Income
NY	2.24%	14,800	0.0467	\$2.64	\$72,500
MD	2.46%	20,200	0.0468	\$2.45	\$87,400
MA	2.49%	19,900	0.0465	\$2.46	\$85,300
NJ	2.55%	20,800	0.0469	\$2.52	\$86,500
VA	2.75%	20,500	0.0479	\$2.27	\$75,100
CT	2.80%	22,800	0.0471	\$2.54	\$89,400
CO	2.91%	18,600	0.0503	\$2.38	\$72,800
RI	2.94%	18,900	0.0468	\$2.47	\$72,500
NH	2.96%	21,800	0.0470	\$2.39	\$82,200
FL	3.04%	16,900	0.0473	\$2.40	\$58,700
TX	3.08%	19,500	0.0498	\$2.14	\$62,500
MN	3.11%	20,400	0.0489	\$2.36	\$70,800
UT	3.13%	19,000	0.0491	\$2.47	\$70,300
PA	3.25%	17,700	0.0480	\$2.67	\$63,900
IL	3.27%	18,800	0.0478	\$2.68	\$66,900
DE	3.29%	21,200	0.0480	\$2.39	\$70,700
OH	3.33%	18,000	0.0480	\$2.39	\$58,000
SC	3.34%	17,900	0.0504	\$2.19	\$54,500
LA	3.34%	18,700	0.0508	\$2.14	\$56,300
MO	3.36%	18,500	0.0490	\$2.20	\$55,500
NC	3.36%	18,400	0.0488	\$2.28	\$56,300
WY	3.37%	18,600	0.0545	\$2.41	\$71,700
GA	3.38%	19,000	0.0492	\$2.29	\$57,000
WI	3.39%	19,400	0.0487	\$2.36	\$63,100
NV	3.47%	15,900	0.0486	\$2.93	\$64,200
NM	3.52%	16,800	0.0508	\$2.36	\$55,300
AL	3.52%	18,300	0.0507	\$2.20	\$51,900
TN	3.53%	19,000	0.0496	\$2.24	\$54,400
VT	3.56%	20,600	0.0465	\$2.58	\$67,500
KS	3.57%	20,500	0.0505	\$2.21	\$59,800
WV	3.58%	16,300	0.0499	\$2.41	\$51,400
ME	3.59%	18,800	0.0484	\$2.44	\$59,500
MI	3.60%	18,500	0.0494	\$2.44	\$58,700
WA	3.60%	18,500	0.0485	\$3.07	\$73,600
AZ	3.64%	16,500	0.0494	\$2.80	\$60,500
MT	3.64%	17,100	0.0538	\$2.42	\$59,500
OR	3.68%	16,600	0.0486	\$2.98	\$62,400
ID	3.69%	17,000	0.0517	\$2.52	\$57,400
IN	3.72%	19,100	0.0500	\$2.38	\$58,000
OK	3.73%	19,900	0.0506	\$2.19	\$56,600
NE	3.74%	20,200	0.0505	\$2.39	\$63,900
IA	3.76%	20,700	0.0505	\$2.35	\$64,200
KY	3.77%	18,500	0.0498	\$2.29	\$52,200
AR	3.79%	18,500	0.0518	\$2.19	\$51,300
ND	3.81%	23,300	0.0523	\$2.31	\$73,900
CA	3.86%	18,600	0.0467	\$3.45	\$75,300
MS	3.91%	18,700	0.0517	\$2.14	\$49,200
SD	3.95%	20,500	0.0523	\$2.33	\$63,000
USA	3.27%	18,500	0.0484	\$2.55	\$64,300

vehicle fuel economy is a major factor for the household transportation energy burden. Improved fleetwide fuel efficiency can occur for a variety of reasons, including a newer vehicle fleet, increased numbers of smaller, more efficient conventional vehicles, and a higher market share of alternative fuel vehicles. This implies that increased adoption of more fuel-efficient vehicles, especially among low-income households, could have the biggest impact on reducing household transportation energy burden. The anticorrelation between GPM and income indicates that wealthier households have more fuel-efficient vehicles. The correlation between fuel price and energy burden is weak, perhaps implying that people reduce their driving when fuel costs are higher. This concept is known as the rebound effect, and has been explored deeply in the literature (57, 58).

In a given locality, it is important to know the fraction of households spending above a given affordability level. At the national level, when the affordable threshold is set to be 2%, then almost 60% of households across the U.S. fall below this affordability level. If the level is set to be 7%, the portion of the household that lack affordable transportation energy decreases to about 6%. For a range of different thresholds from 2% to 7%, we calculated the overburdened fraction of households nationwide; this is shown in the dashed line in Figure 18. Figure 18 also shows the overburdened fraction for four specific states: California, Illinois, New Jersey, and Oregon. At the state level, California has higher percentages of households in the overburdened categories than the national average for all affordability levels. As shown in Figure 15 above, despite having lower transportation energy burden in some metropolitan areas such as the San Francisco Bay Area, Los Angeles and San Diego, households in most other counties in California have a higher transportation energy burden than most of the counties in the rest of the country, largely due to high fuel costs. Oregon also exhibits a greater share of overburdened households than the national average, while Illinois has a lower share. New Jersey has lower percentages of overburdened households than the national average, for all affordability levels, due to high income and stock-weighted MPG.

Transportation energy burden varies by census tract within each state due to the variation in socioeconomic and demographic factors. The transportation energy burden by tract in Illinois ranges from 0.6% to 12.7%, shown in Figure 19. Households in rural areas and suburban areas tend to spend more on transportation energy than urban areas, mainly due to higher annual VMT. However, they also drive older and less fuel-efficient vehicles and do not have the same levels of market penetration of fuel-efficient vehicle technologies such as hybrids and PEVs. In the Chicago area, shown in Figure 20, the downtown and near-north side have a lower percentage of income spent on transportation energy, with more efficient vehicles and higher incomes. This contrasts with the south and west sides of the city, which have higher percentages of income spent on transportation energy. For these census tracts, income and fuel economy are both generally lower than the state average, while total driving distance is comparable to the nearby suburbs.

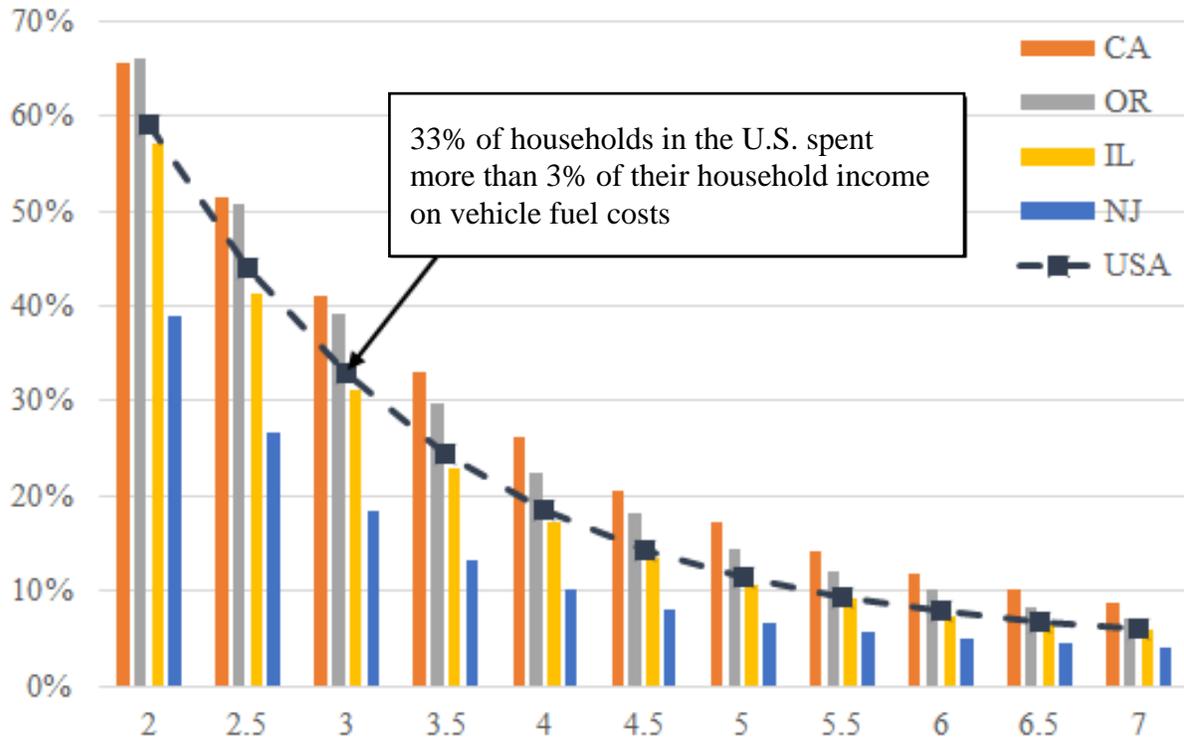


FIGURE 18 Percentage of households nationwide and in four states spending above a given affordability threshold on transportation fuel

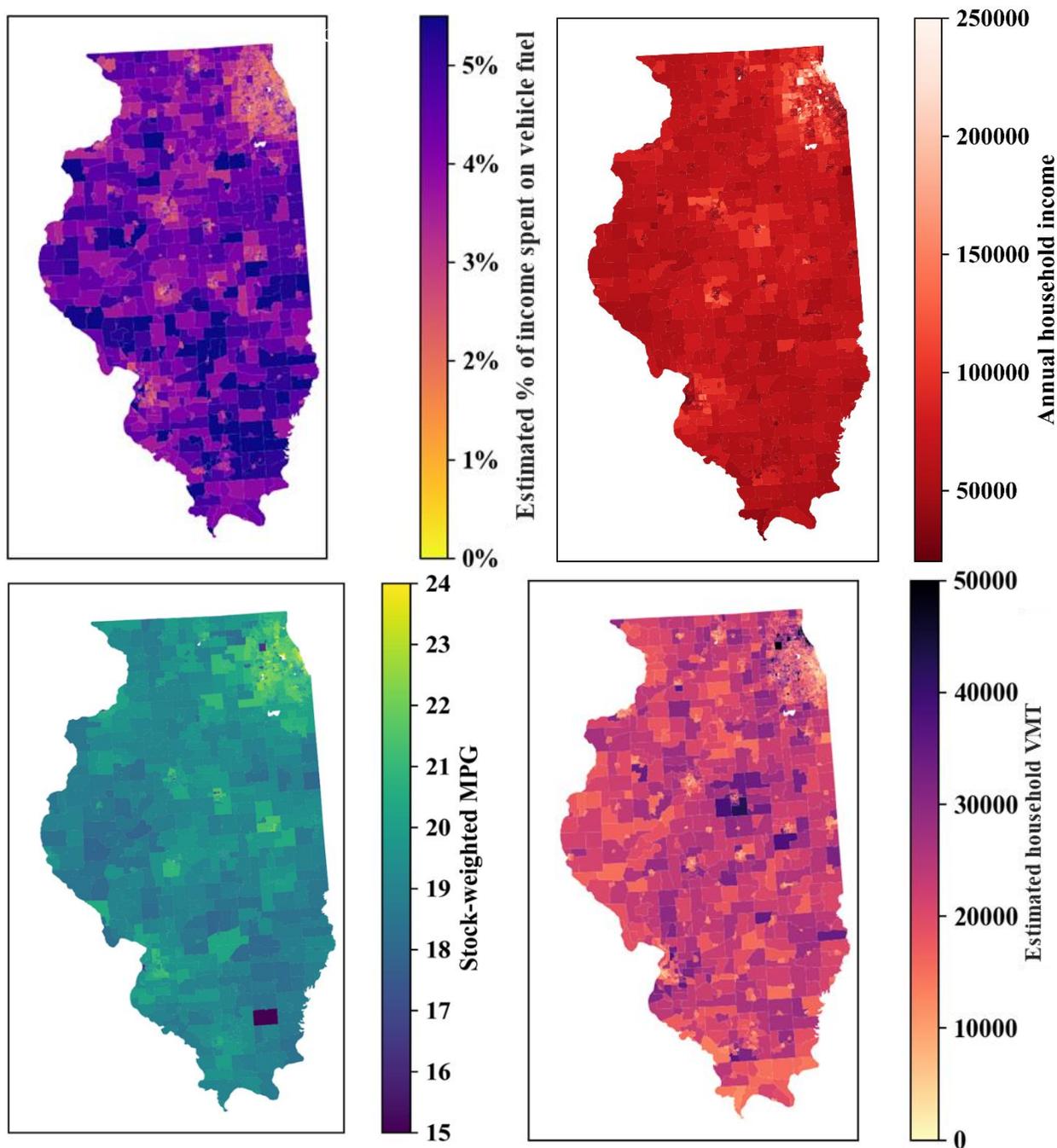


FIGURE 19 Key characteristics for transportation energy burden mapped by census tract in IL. Top left: Transportation energy burden; Top right: Median household income; Bottom left: stock-weighted fuel economy; Bottom right: household annual VMT.

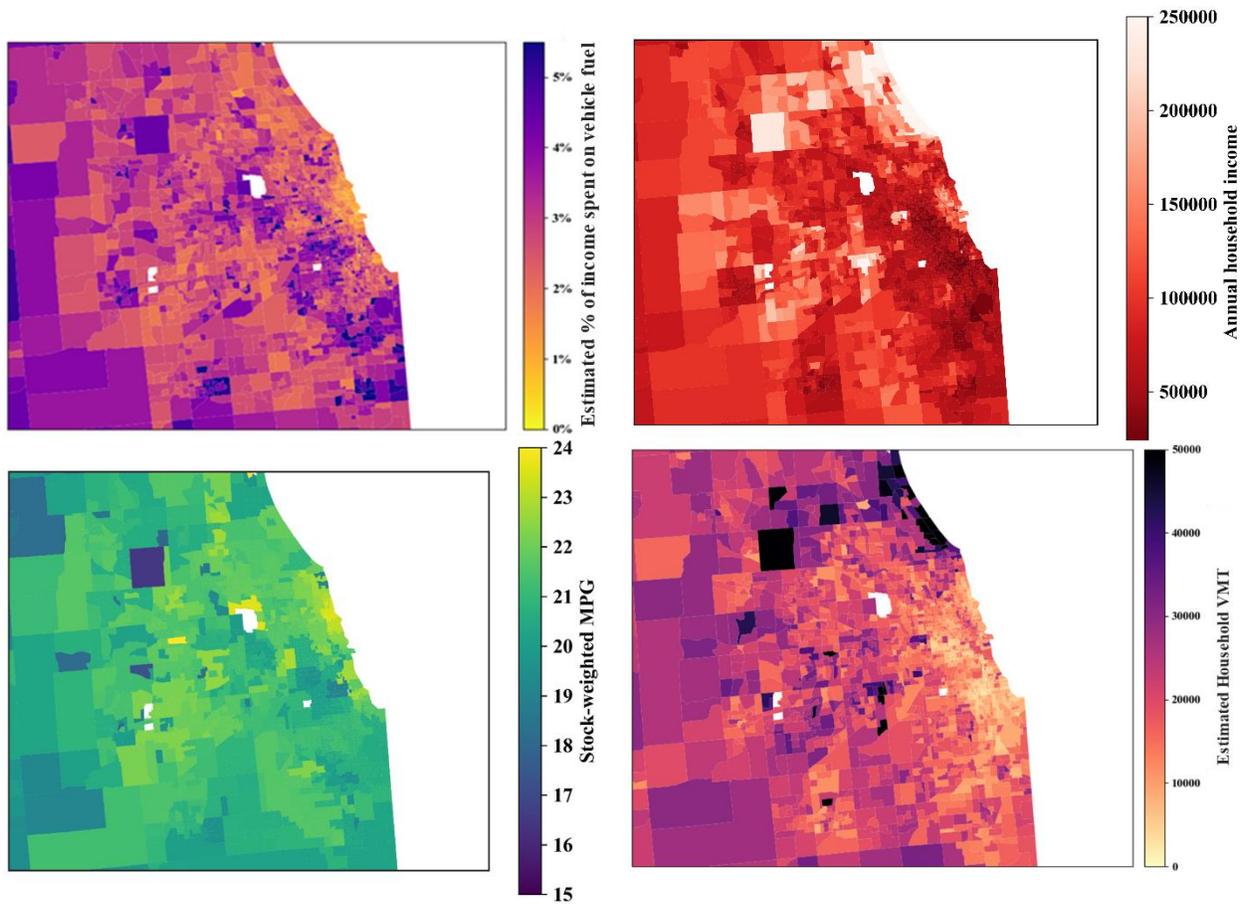


FIGURE 20 Key characteristics for transportation energy burden mapped by census tract in Chicago area. Top left: Transportation energy burden; Top right: Median household income; Bottom left: stock-weighted fuel economy; Bottom right: household annual VMT.

5 SENSITIVITY ANALYSIS

5.1 STOCK WEIGHTED MPG

Since fuel economy is one of the main causes identified for differences in transportation energy burden at the state level, we assess how real-world changes in fuel economy can improve affordability using historical data. Stock-weighted MPG from 2016 to 2018 vary from -4% to +8% by county, as shown in Figure 21. Higher numbers indicate more improvement in MPG, while negative numbers indicate a decrease in the MPG over that time period. Although most counties across the country show improvement, fuel economy of registered vehicles in Montana and Alaska decreased in 2018 compared to 2016. Counties in the San Francisco Bay Area and the Pacific Northwest show the most improvement in stock-weighted MPG, largely due to rapidly increasing PEV adoption. Nationally, the average stock-weighted MPG increased by about 3% from 2016 to 2018. Keeping the gasoline price and household VMT unchanged, the resulting overall reduction in transportation energy burden due to the 3% improvement in stock-weighted MPG is about \$8.2 billion. A retrospective analysis conducted by Greene and Welch also indicates the saving on fuel due to the increased light-duty vehicle fuel economy minus dollars spent to buy more fuel-efficient vehicles is substantial for every income quintile in the U.S. (59).

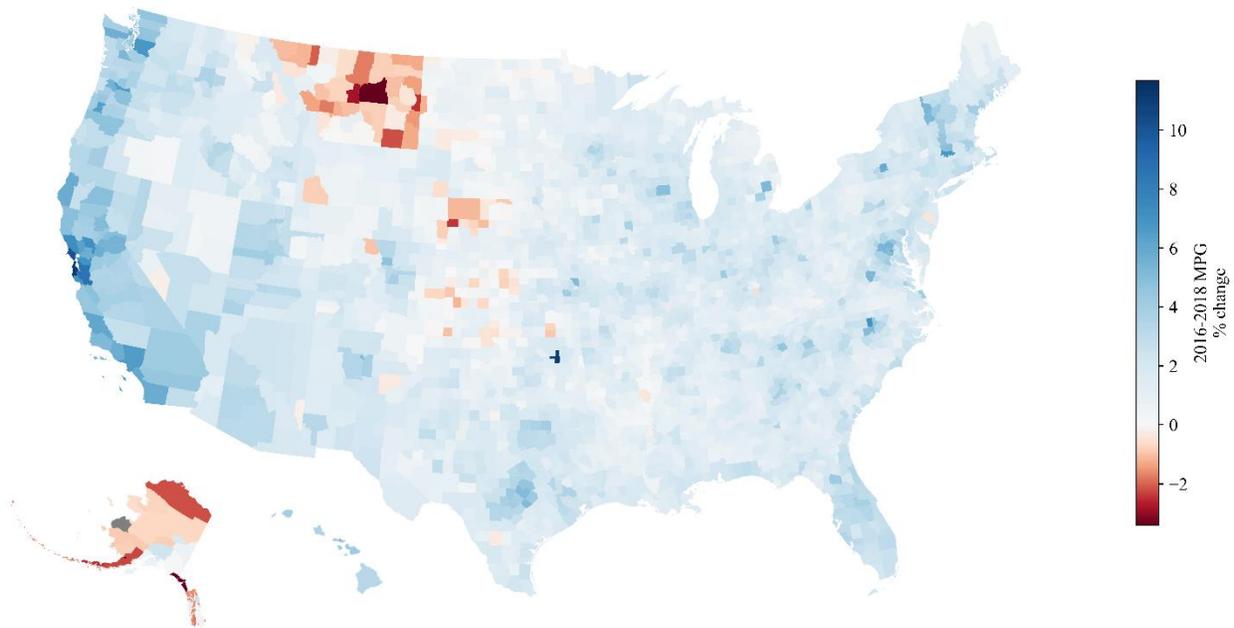


FIGURE 21 Stock-weighted MPG difference: 2018 vs. 2016

5.2 PARAMETRIC SENSITIVITY ANALYSIS

National average household transportation burden is about 3.3%, which is based on the national average household annual VMT of 18,515. We assessed the sensitivity of our results by varying the three key input variables at each tract by one standard deviation nationwide for each variable. These calculations were done at the tract-level, and then aggregated to find a national average. As shown in Figure 22, the energy burden results are presented using parametric sensitivity analysis highlighting the influence of three input variables: 1) household annual VMT, 2) stock-weighted MPG, and 3) gasoline price. In order to illustrate the sensitivities in a cross-comparable way, we vary each by one standard deviation from the national average. When household VMT for each tract is adjusted by one standard deviation from the national average at each tract ($\pm 7,402$ miles/year), the resulting energy burden at the national level varies from 1.8% to 4.7%. The energy burden results are most sensitive to the stock-weighted fuel economy change of ± 1.24 mpg at each tract; and the resulting energy burden varies from 3.1% to 3.5% nationally. The energy burden results are least sensitive to the existing variability in fuel price, varying from 2.8% to 3.8% when the price varies by $\pm \$0.39/\text{GGE}$.

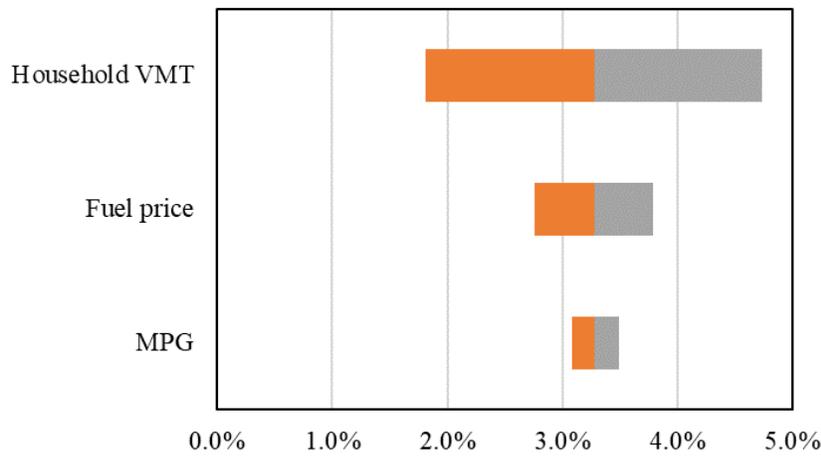


FIGURE 22 Parametric sensitivity analysis of transportation energy burden

6 CONCLUSIONS

This study quantified transportation energy affordability as a function of household characteristics and geography using standardized and localized transportation energy data at the census tract levels for the entire United States. Affordability in this study is quantified as the transportation energy burden, the percentage of annual household income spent on household vehicle fuel. This highly resolved data reveals the greater variation than national averages and provides data to support local decision making to reduce the transportation energy burden.

Household annual transportation energy cost depends on the vehicle miles traveled (VMT), fuel price, and vehicle fuel efficiency. Variability in household annual VMT, on-road vehicle efficiency and resulting burden is demonstrated geographically, socioeconomically and by urban form. Household annual VMT varies by housing unit density and income. Suburban and rural households have higher annual VMT than urban households across all regions, and VMT generally increases as the average income increases. Higher-income groups have a wider distribution of annual VMT, possibly due to the diversity of the number of vehicles owned by those groups.

The finer resolution data has identified how the different factors influence overall affordability in different areas or groups. National average household transportation energy burden has a mean value of 3.3%. With the increased geographical and socioeconomic resolution developed in this study, we found that the burden by census tract varies between 0.09% and 23.3%. Rural households have higher transportation energy burden than suburban and urban households for all regions. Despite increasing household VMT for higher household income groups, transportation energy burden improves as the household income increases. Suburban and rural households spend more on transportation energy compared to urban households due to the usage of less fuel-efficient vehicle technologies and higher annual VMT. Lower-income neighborhoods have a relatively higher transportation energy burden, with census tracts with annual income under \$20,000 averaging 6.5% burden. Lower-income groups have a wide distribution of transportation energy burden, with the most burdened quarter of tracts in the lowest income group above 6.3%, but another quarter below 3.6%. The highest income group, \$125,000/yr and above, has a narrow distribution with half of the tracts falling between 1.8% and 2.6%. By setting a threshold level on affordability, we can assess the portion of households that spent more than that level. For example, nearly 60% of households nationwide spend more than 2% of their income on transportation fuel. If the affordability level is set at 7%, the portion of households considered as overburdened decreases to only 6%. This higher resolution data can help support targeted efforts to reduce the transportation energy burden for areas that experience the highest energy burdens.

This analysis shows that the variation in transportation energy burden at the state level can be largely explained by the on-road vehicle fuel efficiency. The adoption of more fuel-efficient vehicles, especially among low-income households, could have the biggest impact on improving household transportation energy burden. Currently, wealthier census tracts have better fuel economy on average. Our analysis shows that, nationally, a 3% improvement in stock-

weighted MPG from 2016 to 2018 saved American households \$8.2 billion in transportation energy costs.

This study provides a finer understanding of the spatial variation in household transportation energy burden by connecting VMT, vehicle fuel economy, fuel costs, and income data at the census tract level. The baseline data and framework developed here can be used to understand additional components of transportation energy costs and assess spatially distributed impacts of transportation policies, adoptions of different vehicle technologies and fuel prices on household transportation affordability. Current work assumes that the mpg is uniform across a census tract. However, wealthier households purchase newer and more fuel efficient vehicles. Future research to expand this work could include identifying the households and communities that have potential to adopt more fuel-efficient vehicles or transportation modes, based on their social-economic and other environmental factors, and quantifying the resulting affordability.

7 REFERENCES

1. Davis, S.C. and R.G. Boundy, Transportation Energy Data Book, Edition 38, Oak Ridge National Laboratory, ORNL/TM-2019/1333, 2020, https://tedb.ornl.gov/wp-content/uploads/2020/02/TEDB_Ed_38.pdf
2. U.S. Energy Information Administration, Electric Sales, Revenue, and Average Price, October 6, 2020. https://www.eia.gov/electricity/sales_revenue_price/
3. American Gas Association, Average Residential Gas Bills by State, December 13, 2019. <https://www.aga.org/research/data/prices/>
4. Oswald, Y., A. Owen, and J.K. Steinberger. Large inequality in international and intranational energy footprints between income groups and across consumption categories. *Nature Energy*, 2020, 5:231–239. <https://www.nature.com/articles/s41560-020-0579-8>
5. Davis, S.C., & R.G. Boundy. Transportation Energy Data Book: Edition 33. Oak Ridge: Oak Ridge National Laboratory, ORNL-6990, 2014. <https://tedb.ornl.gov/archive/>
6. Argonne National Laboratory, VISION Model, <https://www.anl.gov/es/vision-model>
7. Stephens, T. S., Birky, A., and D. Gohlke, Vehicle Technologies and Fuel Cell Technologies Office Research and Development Programs: Prospective Benefits Assessment Report for Fiscal Year 2018. Argonne National Laboratory, ANL/ESD-17/22, United States, 2017. <https://doi.org/10.2172/1410412>.
8. Garceau, T., C. Atkinson-Palombo. N. Garrik, J. Outlaw, C. McCahill, H. Ahangari, Evaluating Selected Costs of Automobile-Oriented Transportation Systems from a Sustainability Perspective, *Research in Transportation Business & Management*, 2013, Vol. 7, pp. 43-53; <https://doi.org/10.1016/j.rtbm.2013.02.002>.
9. Federal Highway Administration, 2017 National Household Travel Survey, <https://nhts.ornl.gov/>.
10. Oak Ridge National Laboratory, Developing a Best Estimate of Annual Vehicle Mileage for 2009 NHTS Vehicles, 2011, <https://nhts.ornl.gov/2009/pub/BESTMILE.pdf>.
11. Oak Ridge National Laboratory, Developing a Best Estimate of Annual Vehicle Mileage for 2017 NHTS Vehicles, 2019, https://nhts.ornl.gov/assets/2017BESTMILE_Documentation.pdf
12. Henson, K. M. and K. G. Goulias. Travel Determinants and Multiscale Transferability of National Activity Patterns to Local Populations. *Journal of the Transportation Research Record*, 2001, 2231: 35-43. <https://journals.sagepub.com/doi/10.3141/2231-05>

13. Bureau of Transportation Statistics, 2017 Local Area Transportation Characteristics for Households Methodology, <https://www.bts.gov/latch/latch-methodology-2017>
14. Haas, P. M., C. Makarewicz, A. Benedict and S. Bernstein, Estimating Transportation Costs by Characteristics of Neighborhood and Household, *Journal of the Transportation Research Record*, 2008, 2077 (1): 62-70, <https://doi.org/10.3141/2077-09>.
15. Center for Neighborhood Technology, Housing + Transportation Affordability Index. 2012, https://htaindex.cnt.org/about/HTMethods_2016.pdf.
16. Ewing, R., and R. Cervero. Travel and the Built Environment: A Synthesis, *Journal of the Transportation Research Record*, 2001, 1780: 87–114. <https://journals.sagepub.com/doi/pdf/10.3141/1780-10>
17. Holtzclaw, J., R. Clear, H. Dittmar, D. Goldstein, and P. Haas. Location Efficiency: Neighborhood and Socioeconomic Characteristics Determine Auto Ownership and Use—Studies in Chicago, Los Angeles and San Francisco. *Transportation Planning and Technology*, 2002, 25 (1):1–27. <https://doi.org/10.1080/03081060290032033>
18. Chu X., An Assessment of Public Transportation Markets Using NHTS Data, National Center for Transit Research (NCTR), March 2012. <https://www.nctr.usf.edu/wp-content/uploads/2013/01/77920.pdf>
19. Agrawal, A.W., E.A. Blumenberg, S. Abel, G. Pierce, and C.N. Darrah. Getting Around When You're Just Getting By: The Travel Behavior and Transportation Expenditures of Low-Income Adults. Mineta Transportation Institute, CA-MTI-10-2806. January 2011. <https://transweb.sjsu.edu/research/getting-around-when-youre-just-getting-travel-behavior-and-transportation-expenditures-low>
20. U.S. Department of Housing and Urban Development, Location Affordability Index, accessed on September 30, 2020. <https://www.hudexchange.info/programs/location-affordability-index/>
21. Bureau of Labor Statistics, Consumer Expenditures (various years), www.bls.gov/cex/home.htm.
22. Litman, T. Transportation Affordability Evaluation and Improvement Strategies, 2020 <https://www.vtpi.org/affordability.pdf>
23. Miller, E., Travel and Housing Costs in the Greater Toronto Area: 1986-1996, Neptis Foundation (www.neptis.org), 2004.
24. K. Mattingly, J. Morrissey, Housing and transport expenditure: Socio-spatial indicators of affordability in Auckland, *Cities* 38 (2014) 69–83. <https://doi.org/10.1016/j.cities.2014.01.004>

25. Saberi, M., H. Wu, R. Amoh-Gyimah, J. Smith, D. Arunachalam, Measuring housing and transportation affordability: A case study of Melbourne, Australia, *Journal of Transport Geography*, 2017, 65: 134-146. <https://doi.org/10.1016/j.jtrangeo.2017.10.007>
26. Fitzroy, S.S., K. Schroeckenthaler, and J. Goldberg. Equity Effects of Revenue Neutral Mileage-Based Fees for Urban and Rural Households, *Journal of the Transportation Research Record*, 2018, <https://doi.org/10.1177/0361198118794714>
27. Cervero, R., and J. Murakami. Effects of Built Environments on Vehicle Miles Traveled: Evidence from 370 US Urbanized Areas. *Environment and Planning A: Economy and Space*, Vol. 42, No. 2, 19 2010, pp. 400–418. <https://doi.org/10.1068/a4236>.
28. Zhang, L., J. Hong, A. Nasri, and Q. Shen. How Built Environment Affects Travel Behavior: A Comparative Analysis of the Connections between Land Use and Vehicle Miles Traveled in US Cities. *Journal of Transport and Land Use*, 5(3). 2012. <https://doi.org/10.5198/jtlu.v5i3.266>
29. Hymel, K. M. Factors Influencing Vehicle Miles Traveled in California: Measurement and Analysis. 2014. https://sor.senate.ca.gov/sites/sor.senate.ca.gov/files/ctools/CCS_Report--Factors_Influencing_Vehicle_Miles_Traveled_in_California.pdf
30. Akar, G., and J.-M. Guldman. Another Look at Vehicle Miles Traveled. *Transportation Research Record: Journal of the Transportation Research Record*, Vol. 2322, No. 1, 2012, pp. 110–118. <https://doi.org/10.3141/2322-12>.
31. Cao, X., and Y. Fan. Exploring the Influences of Density on Travel Behavior Using Propensity Score Matching. *Environment and Planning B: Planning and Design*, Vol. 39, No. 3, 2012, pp. 459–470. <https://doi.org/10.1068/b36168>.
32. Brownstone, D. Key Relationships Between the Built Environment and VMT. *Transportation Research*, No. Special report 298: Driving and the Built Environment: The Effects of Compact Development on Motorized Travel, Energy Use and CO₂ emissions, 2008, p. 14. <https://doi.org/10.17226/12747>
33. Singh, A.C., S. Astroza, V.M. Garikapati, R.M. Pendyala, C.R. Bhat, and P.L. Mokhtarian, Quantifying the relative contribution of factors to household vehicle miles of travel, *Transportation Research Part D: Transport and Environment*, Volume 63, August 2018, pp. 23-36. <https://www.sciencedirect.com/science/article/pii/S1361920917307551>
34. U.S. Census Bureau, American Community Survey (ACS) 2018 estimates, <https://www.census.gov/programs-surveys/acs>
35. U.S. Department of Housing and Urban Development (HUD), USPS Zip Code Crosswalk Files, https://www.huduser.gov/portal/datasets/usps_crosswalk.html

36. U.S. Department of Transportation, Bureau of Transportation Statistics. Passenger Travel Facts and Figures 2016. <https://rosap.ntl.bts.gov/view/dot/35471>
37. U.S. Department of Energy, Office of Energy Efficiency & Renewable Energy. Low-Income Energy Affordability Data (LEAD) Tool. <https://www.energy.gov/eere/slsc/maps/lead-tool>
38. Ma, O., K. Laymon, M. Day, R. Oliveira, J. Weers, and A. Vimont, Low-Income Energy Affordability Data (LEAD) Tool Methodology, National Renewable Energy Laboratory, <https://lead.openei.org/assets/files/LEAD-Tool-Methodology.pdf>.
39. GasBuddy.com. Gas Price Map. <https://www.gasbuddy.com/gaspricemap>
40. U.S. Energy Information Administration, Gasoline and Diesel Fuel Update, <https://www.eia.gov/petroleum/gasdiesel/>
41. U.S. Energy Information Administration, Electric Power Monthly, Table 5.6.A. https://www.eia.gov/electricity/monthly/epm_table_grapher.php?t=epmt_5_6_a
42. IHS Markit, Vehicles in Operation (VIO) & Vehicle Registration Data Analysis, <https://ihsmarkit.com/products/automotive-market-data-analysis.html>
43. U.S. Department of Energy and U.S. Environmental Protection Agency. FuelEconomy.gov: the Official U.S. Governmental Source for Fuel Economy Information. Accessed October 20, 2020. <https://www.fueleconomy.gov/>
44. Bourbon, E. Alternative Fuel Price Report, Alternative Fuel Data Center, 2020, https://afdc.energy.gov/files/u/publication/alternative_fuel_price_report_jan_2020.pdf
45. Bucholtz, S., E. Molfino, and J. Kolko. The Urbanization Perceptions Small Area Index: An Application of Machine Learning and Small Area Estimation to Household Survey Data. Working paper, June 12, 2020. <https://www.huduser.gov/portal/AHS-neighborhood-description-study-2017.html>
46. Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay. Scikit-learn: Machine Learning in Python, JMLR, 2011, 12: 2825-2830. <https://www.jmlr.org/papers/v12/pedregosa11a.html>
47. Dorogush, A.V., V. Ershov, A. Gulin, CatBoost: gradient boosting with categorical features support, Workshop on ML Systems at NIPS 2017. http://learningsys.org/nips17/assets/papers/paper_11.pdf
48. Friedman, J. H., Greedy function approximation: a gradient boosting machine, The Annals of Statistics, 2001, 29:1189-1232. <http://dx.doi.org/10.1214/aos/1013203451>

49. Chicago Metropolitan Agency of Planning (CMAP), 2007-08 Travel Tracker Survey, <https://www.cmap.illinois.gov/data/transportation/travel-survey>.
50. The California Department of Transportation (Caltrans), 2012-2013 California Household Travel Survey, <https://dot.ca.gov/programs/transportation-planning/economics-data-management/transportation-economics/ca-household-travel-survey>.
51. U.S. Environmental Protection Agency, New Fuel Economy and Environment Labels for a New Generation of Vehicles. EPA-420-F-11-017, May 2011. <https://nepis.epa.gov/Exe/ZyPDF.cgi/P100BAV0.PDF?Dockey=P100BAV0.PDF>
52. SAE International, Utility Factor Definitions for Plug-In Hybrid Electric Vehicles Using Travel Survey Data J2841_201009, 2010, https://www.sae.org/standards/content/j2841_201009/.
53. GasBuddy, 2020. Fuel Insights Methodology. Accessed July 31, 2020. <http://fuelinsights.gasbuddy.com/FAQs>
54. Ou, S., X. He, W. Ji., W. Chen., L. Sui., Y. Gan., Z. Lu., Machine learning model to project the impact of COVID-19 on US motor gasoline demand, Nature Energy, 2020 <https://www.nature.com/articles/s41560-020-0662-1>.
55. U.S. Energy Information Administration, This Week in Petroleum, https://www.eia.gov/petroleum/weekly/archive/2020/200311/includes/analysis_print.php
56. Allen, M., D. Poggiali, K. Whitaker, T.R. Marshall, R.A. Kievit, Raincloud plots: a multi-platform tool for robust data visualization, Wellcome Open Res 2019, 4:63. <https://wellcomeopenresearch.org/articles/4-63>
57. Sorrell, S., J. Dimitropoulos, and M. Sommerville, Empirical estimates of the direct rebound effect: A review. Energy Policy, Volume 37, Issue 4, 1356–1371. April 2009. <https://www.sciencedirect.com/science/article/pii/S0301421508007131>
58. Greene, D.L., J. Kahn, and R. Gibson. Fuel Economy Rebound Effect for U.S. Household Vehicles. Energy Journal, vol. 20, no. 3 (1999), pp. 1–31. <http://www.iaee.org/en/publications/ejarticle.aspx?id=1314>
59. Greene, D.L., and J.G Welch. Impacts of fuel economy improvements on the distribution of income in the U.S. Energy Policy, vol. 122, 2018, pp. 528-541.



Energy Systems Division

Argonne National Laboratory
9700 South Cass Avenue, Bldg. 362
Lemont, IL 60439-4832

www.anl.gov