



# The Development of Vocational Vehicle Drive Cycles and Segmentation

Adam Duran, Caleb Phillips, Arnaud Konan, and Kenneth Kelly

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

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

**Technical Report**  
NREL/TP-5400-65921  
September 2017

Contract No. DE-AC36-08GO28308



# The Development of Vocational Vehicle Drive Cycles and Segmentation

Adam Duran, Caleb Phillips, Arnaud Konan, and Kenneth Kelly

Prepared under the direction of the U.S. Environmental Protection Agency (EPA) by the National Renewable Energy Laboratory (NREL) as a product of Interagency Agreement IAG-14-1954 and Task No. WFGY.1000

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

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

National Renewable Energy Laboratory  
15013 Denver West Parkway  
Golden, CO 80401  
303-275-3000 • [www.nrel.gov](http://www.nrel.gov)

**Technical Report**  
NREL/TP-5400-65921  
September 2017

Contract No. DE-AC36-08GO28308

## NOTICE

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 any of their employees, 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 authors expressed herein do not necessarily state or reflect those of the United States government or any agency thereof.

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

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

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

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

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

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

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

NREL prints on paper that contains recycled content.

## List of Acronyms

AS	aerodynamic speed
CA	characteristic acceleration
CARB	California Air Resources Board
CH <sub>4</sub>	methane
CO <sub>2</sub>	carbon dioxide
DOE	U.S. Department of Energy
DRIVE	Drive Cycle Rapid Investigation, Visualization, and Evaluation tool
EISA	Energy Independence and Security Act
EPA	U.S. Environmental Protection Agency
ETC	European Test Cycle, also known as the FIGE Cycle
ft/s	feet per second
ft/s <sup>2</sup>	feet per second squared
GHG	greenhouse gas
GPS	global positioning system
HD	heavy-duty
HHDDT	Heavy Heavy-Duty Diesel Truck
Hz	hertz
kg	kilogram
MD	medium-duty
MOVES	MOtor Vehicle Emission Simulator
mph	miles per hour
N <sub>2</sub> O	nitrous oxide
NHTSA	National Highway Traffic Safety Administration
NREL	National Renewable Energy Laboratory
PCA	Principal Components Analysis
RMS	root mean squared
rpm	revolutions per minute
WHTC	World Harmonized Transient Cycle
WHVC	World Harmonized Vehicle Cycle

# Table of Contents

List of Acronyms .....	iv
List of Figures .....	vi
List of Tables .....	viii
Executive Summary .....	ix
<b>1 Introduction and Background .....</b>	<b>1</b>
1.1 Supporting EPA GHG Phase 2 Regulation Development.....	1
1.2 NREL’s Fleet DNA Database .....	1
1.3 Classifying Data in Fleet DNA .....	2
1.4 NREL’s DRIVE Tool.....	4
1.5 Segmenting the U.S. Vocational Vehicle Population and Developing Representative Drive Cycles and Weights.....	6
<b>2 Cluster Analysis.....</b>	<b>8</b>
2.1 Principal Components Analysis .....	9
2.2 Drive Cycle Clustering.....	12
2.3 Data Resampling .....	17
2.4 Vehicle Clustering.....	22
<b>3 Predictive Modeling.....</b>	<b>26</b>
3.1 Logistic Regression Model.....	29
3.1.1 Full Model.....	29
3.1.2 Univariate Diesel Model .....	31
3.2 Validation and Error Process.....	31
<b>4 Development of Representative Transient Drive Cycle Component.....</b>	<b>34</b>
4.1 Generation of Representative Low Speed Transient Component Using DRIVE™ .....	34
4.2 Comparison of DRIVE Generated Transient Cycles to Existing Standard Cycles .....	36
4.2.1 Standard Test Cycles Selected for Comparison .....	36
4.2.2 Drive Cycle Characterization and Comparison.....	40
4.2.3 Analysis of Mileage Accumulation by Vehicle Speed.....	42
4.2.4 Comparison of Acceleration/Deceleration Behavior .....	44
<b>5 Development of Representative Drive Cycle Weights .....</b>	<b>49</b>
5.1 Comparative Drive Cycle Characterization .....	49
5.2 Analysis of Mileage Accumulation.....	52
5.3 Weighting of Clusters by Mileage Accumulated in Broad Speed Categories.....	55
5.4 Characterization of Acceleration/Deceleration Behavior.....	56
5.5 Analysis of Idle Behavior.....	60
<b>6 Known Limitations and Potential for Future Work.....</b>	<b>63</b>
References .....	64
Appendix A: Standard Fleet DNA Data Channels .....	68
Appendix B: Determining Engine to Vehicle Speed Ratios via Kernel Density Estimation .....	70
Appendix C: Comparison of Fleet DNA and MOVES Resampled Distributions by Cluster .....	74
Appendix D: Detailed Mileage Accumulation Analysis for NREL-EPA Low Speed Transient Cycle .....	77
Appendix E: Generation of Representative High Speed Transient Component Using DRIVE™ .....	80

## List of Figures

Figure 1: Screen capture of DRIVE GUI processing data .....	5
Figure 2: Sample composition .....	8
Figure 3: A correlogram for the eight metrics .....	9
Figure 4: All drive cycles for all vehicles visualized in the first two principal component space.....	11
Figure 5: Scatter plot of all vocational drive cycles.....	12
Figure 6: Two clustering using k-medoids algorithm with cluster centers marked in black. ....	14
Figure 7: Three clustering using k-medoids algorithm with cluster centers marked in black. ....	15
Figure 8: Sub-cluster analysis for School buses .....	16
Figure 9: Resampling flow for vocational vehicles .....	17
Figure 10: Two cluster MOVES resampled and original Fleet DNA statistics .....	19
Figure 11: Two cluster MOVES subcategory resampled and original Fleet DNA statistics.....	20
Figure 12: Final cluster center convergence as a function of successive resamplings.....	21
Figure 13: Final cluster center convergence as a function of MOVES subcategory resamplings .....	22
Figure 14: Trace-mode clustering of Fleet DNA Vocational vehicles using two-cluster solution .....	23
Figure 15: Trace-mode clustering of Fleet DNA Vocational vehicles using three-cluster solution. ....	24
Figure 16: Additional results of trace-mode clustering Fleet DNA using two-cluster solution.....	25
Figure 17: Correlation between RPM at 55 mph and RPM at 65 mph for different vehicle fuel types.....	26
Figure 18: Engine RPM data available for different vocational categories .....	27
Figure 19: Engine RPM bimodality corresponding with two-cluster assignments.....	28
Figure 20: Distribution of Engine RPM characteristics for different types of vehicles.....	28
Figure 21: Logistic histogram of relationship between RPM at 65 mph and cluster 2 participation.....	30
Figure 22: Locations of 40 most often misclassified vehicles using the full logistic regression model. ....	32
Figure 23: Speed-time trace for representative low speed EPA transient cycle .....	35
Figure 24: Speed-time trace for CARB HHDDT transient cycle .....	37
Figure 25: Speed-time trace for WHVC cycle.....	38
Figure 26: Speed-time trace for ETC FIGE cycle.....	39
Figure 27: Speed-time trace for JP JE05 cycle .....	40
Figure 28: Comparative distribution of mileage accumulation binned by vehicle speed .....	44
Figure 29: Comparison of standard cycle acceleration frequency .....	45
Figure 30: Comparison of standard cycle deceleration frequency .....	46
Figure 31: Comparison of standard cycle peak acceleration rates.....	46
Figure 32: Comparison of standard cycle peak deceleration rates.....	47
Figure 33: Comparison of standard cycle average acceleration rates .....	47
Figure 34: Comparison of standard cycle average deceleration rates.....	48
Figure 35: Comparison of representative cycle zero speed time vs. mileage accumulated below 55 mph	50
Figure 36: Comparison of representative cycle average driving speed and kinetic intensity .....	51
Figure 37: Comparison of representative cycle standard deviation of speed and stops per mile.....	52
Figure 38: Comparison of representative cycle mileage accumulation .....	55
Figure 39: Comparison of representative cycle acceleration frequency .....	57
Figure 40: Comparison of representative cycle deceleration frequency .....	58
Figure 41: Comparison of representative cycle peak acceleration rates .....	58
Figure 42: Comparison of representative cycle peak deceleration rates .....	59
Figure 43: Comparison of representative cycle average acceleration rates .....	59
Figure 44: Comparison of representative cycle average deceleration rates .....	60
Figure 45: Sample potential dual mode test cycle .....	63
Figure 46: Sample heat plot of vehicle speed vs. engine operating speed.....	70
Figure 47: Single vehicle sample distribution of vehicle speed to engine speed ratio data .....	71
Figure 48: Single vehicle sample distribution of vehicle speed to engine speed ratio data with peaks.....	72
Figure 49: Identified gear ratios plotted atop sample vehicle speed vs. engine operating speed data.....	73

Figure 50: RMS Distance distributions for fleet DNA Dataset and MOVES resampled Cluster 1 ..... 74  
Figure 51: RMS distance distributions for Fleet DNA dataset and MOVES resampled Cluster 2..... 75  
Figure 52. RMS distance distributions for Fleet DNA dataset and MOVES resampled Cluster 3 ..... 76  
Figure 53. Comparison of representative cycle mileage accumulation using 2-mph speed bins..... 77  
Figure 54. Speed–time trace for representative high-speed EPA transient cycle ..... 80

## List of Tables

Table 1. Phase 1 and Proposed Vocational Drive Cycle Weightings .....	x
Table 2. Principal Component Loadings (All Vehicles).....	10
Table 3. Principal Component Loadings (Vocational Vehicles) .....	11
Table 4. Vocational Drive Cycles – 2-Clustering Centers.....	16
Table 5. Vocational Drive Cycles – 3-Clustering Centers.....	17
Table 6. Vocational Resampling by MOVES Categories.....	18
Table 7. Vocational Resampling by MOVES Sub-categories .....	18
Table 8. Fitted Coefficients for Full Logistic Regression (Fleet DNA Vocational Vehicles Only) .....	30
Table 9. Fitted Coefficients for Univariate Diesel-only Logistic Regression .....	31
Table 10. Comparison of Standard Cycle Drive Cycle Metrics.....	41
Table 11. Sub-45 mph Mileage Accumulation Comparison for Standard Cycles.....	42
Table 12. 45+ mph Mileage Accumulation Comparison for Standard Cycles .....	43
Table 13. Comparison of Cycle Acceleration Statistics.....	45
Table 14. Comparison of Representative Cycle and Cluster Drive Cycle Metrics.....	49
Table 15. Sub 45 mph Mileage Accumulation Comparison for Cluster and Representative Cycles.....	53
Table 16. 45+ mph Mileage Accumulation Comparison for Clusters and Representative Cycles .....	54
Table 17. Final Weights for Identified Clusters.....	56
Table 18. Comparison of Acceleration Behavior for Clusters and Representative Cycle .....	57
Table 19. Vocational Class 8 Breakdown.....	61
Table 20. Proposed In-Gear and Out-of-Gear Idle Weights .....	62
Table 21. Example GPS File.....	68
Table 22. List of Additional Logged Controller Area Network Channels.....	69
Table 23. Comparison of Representative Cycle Mileage Accumulation Using 2-mph Speed Bins .....	77
Table 24. Comparison of Standard Cycle Drive Cycle Metrics.....	81
Table 25. Sub-45 mph Mileage Accumulation Comparison for Standard Cycles .....	82
Table 26. 45+ mph Mileage Accumulation Comparison for Standard Cycles .....	82
Table 27. Comparison of Cycle Acceleration Statistics.....	83



## Executive Summary

Under a collaborative interagency agreement between the U.S. Environmental Protection Agency and the U.S. Department of Energy (DOE), the National Renewable Energy Laboratory (NREL) has performed a series of in-depth analyses to characterize the on-road driving behavior including distributions of vehicle speed, idle time, accelerations and decelerations, and other driving metrics of medium- and heavy-duty vocational vehicles operating within the United States. As part of this effort, NREL researchers segmented U.S. medium- and heavy-duty vocational vehicle driving characteristics into three distinct operating groups or clusters using real world drive cycle data collected at 1-Hz and stored in NREL's Fleet DNA database. The Fleet DNA database contains millions of miles of historical real world drive cycle data captured from medium- and heavy vehicles operating across the United States. The data encompasses data from existing DOE activities as well as contributions from valued industry stakeholder participants. For this project, data captured from 913 unique vehicles comprising 16,250 days of operation were drawn from the Fleet DNA database and examined. The Fleet DNA data used as a source for this analysis has been collected from a total of 30 unique fleets/data providers operating across 22 unique geographic locations spread across the United States. This includes locations with topology ranging from the foothills of Denver, Colorado to the flats of Miami, Florida. The range of fleets, geographic locations, and total number of vehicles analyzed ensures results which include the influence of these factors. While no analysis will be perfect without unlimited resources and data, it is the researchers understanding that the Fleet DNA database is the largest and most thorough publicly accessible vocational vehicle usage database currently in operation. Leveraging both drive cycle and vehicle chassis specification information also stored in Fleet DNA, NREL researchers then developed a statistically robust logistic model capable of predicting individual vehicle cluster membership based on vehicle chassis parameters for a simplified bicluster condition. Upon testing, the full logistic model demonstrated an accuracy level of 89% for bicluster prediction. For a subsample containing only diesel vehicles, the model demonstrated accuracy of 81%.

In addition to developing a predictive model, the results of the vocational segmentation analyses were also used by NREL researchers to develop a series of drive cycle weights along with a testable transient drive cycle representative of real world operating behavior of commercial U.S. vehicles. Based on the results of the vocational segmentation analyses, three unique operating clusters were identified and associated drive cycle weightings assigned as shown in Table 1. The three clusters of operation were reflective of urban, mixed urban, and highway driving conditions.

**Table 1. Phase 1 and Proposed Vocational Drive Cycle Weightings**

	<b>Transient Distance (% of Total Mileage)</b>	<b>55 mph Cruise Distance (% of Total Mileage)</b>	<b>65 mph Cruise Distance (% of Total Mileage)</b>	<b>Total Zero Speed Time (% of Total Cycle Duration)</b>
Phase 1 Vocational	42.0	21.0	37.0	NA
Phase 1 Vocational Hybrids	75.0	9.0	16.0	NA
Proposed Vocational Urban	96.5	3.5	0.0	47.3
Proposed Vocational Mixed Urban	64.7	32.2	3.1	52.4
Proposed Vocational Highway	26.5	62.5	11.0	47.9

Compared to Phase 1, the proposed weights developed by NREL for Phase 2 provide a greater level of detail by describing three modes of medium- and heavy-duty vocational vehicle driving behavior identified during the segmentation analysis with no differentiation between hybrid and non-hybrid vehicle weightings. In addition, after thorough evaluation, it was found by NREL researchers that there is an opportunity for improvements in the transient cycle used in the Phase 1 regulations, the CARB HHDDT Transient. Based on an analysis of mileage accumulation, it was found that the CARB HHDDT Transient does not accumulate significant mileage between 40 and 50 mph to match the characteristics of the urban segment identified as part of the Fleet DNA drive cycle cluster analysis. The representative cycle developed by NREL does accumulate mileage in these speed bins, and displays more aggressive driving behavior than most standard cycles examined, including the CARB HHDDT Transient.

This report includes an introduction to the Fleet DNA database and the data contained within, a presentation of the results of the statistical analysis performed by NREL, review of the logistic model developed to predict cluster membership, and a discussion and detailed summary of the development of the vocational drive cycle weights and representative transient drive cycles for testing and simulation. Additional discussion of known limitations and potential future work are also included in the report content.

# 1 Introduction and Background

## 1.1 Supporting EPA GHG Phase 2 Regulation Development

In August of 2011, the U.S. Environmental Protection Agency (EPA) and the National Highway Traffic Safety Administration (NHTSA) adopted a set of national standards to reduce greenhouse gas (GHG) emissions and improve fuel efficiency of medium-duty (MD) and heavy-duty (HD) trucks [1-4]. The standards were jointly developed under the Energy Independence and Security Act (EISA), and the Clean Air Act, the GHG standards included the establishment of limits for CO<sub>2</sub>, N<sub>2</sub>O, and CH<sub>4</sub> emissions. These limits would be enforced on model year 2014–2018 vehicles, and have come to be known as Phase 1 of the national GHG regulatory program. The development of the Phase 1 standards provided EPA, NHTSA, and the state of California with a set of fully aligned regulations, allowing manufacturers to build a single fleet of vehicles and engines for the U.S. market.

Soon after the implementation of the Phase 1 regulations, in response to the President's Climate Action Plan in February of 2014, President Obama announced efforts to update existing MD/HD vehicle regulations in Phase 2 of EPA's national GHG program [5-7]. As part of the EPA's proposed Phase 2 rulemaking, the U.S. Department of Energy (DOE) and EPA partnered to support a targeted project to refine and evaluate appropriate duty cycles for tractor-trailers and vocational vehicles to be used as part of MD/HD vehicle certification procedures for GHG emission standards. The National Renewable Energy Laboratory (NREL) has provided technical support utilizing DOE Vehicle Technologies Office-supported data, tools, and expertise to assist these efforts.

NREL's experience with large transportation database projects, including Fleet DNA [8] and the Transportation Secure Data Center (TSDC) [9], provide the prerequisite capabilities for tackling data-intensive problems. Additionally, NREL's data analysis tools including the Drive Cycle Rapid Investigation, Visualization, and Evaluation tool (DRIVE) [10-11] are used to distill large volumes of on-road vehicle data into statistically representative subsets suitable for testing and evaluation purposes. While the data contained in both Fleet DNA and the TSDC may not be statistically representative of the entire US population of commercial vehicles across all vocations, vehicle builds, and applications, NREL researchers believe this to be the most extensive database used to date to generate representative drive cycles. These skillsets combined with NREL's long-standing efforts in evaluating the on-road performance of conventional and advanced technology MD/HD vehicles for large commercial fleets are valuable tools to provide enhanced information for the EPA Phase 2 GHG rulemaking.

## 1.2 NREL's Fleet DNA Database

The Fleet DNA clearinghouse of commercial fleet transportation data was developed and is used to help vehicle manufacturers and developers optimize vehicle designs while assisting fleet managers with the selection of advanced technologies for their fleets, all with the goal of reducing petroleum consumption and improving energy efficiency. This online database and tool—available at <http://www.nrel.gov/fleetdna>—provides both static and interactive data

summaries and visualizations representative of the real-world “genetics” for medium- and heavy-duty commercial fleet vehicles.

The data for each visualization/report have been drawn from multiple fleets operating in a variety of vocations for each vehicle type, across multiple locations within the United States. This extensive breadth of data is necessary to capture the wide range of vehicle operation resulting from different geographies and fleet applications demonstrated by medium- and heavy-duty vehicles. In addition to providing public access to commercial fleet data summaries and visualizations, Fleet DNA serves as a portal for users to explore related DOE programs and projects such as NREL’s fleet testing activities and reports, the DOE Office of Energy Efficiency and Renewable Energy’s Alternative Fuels Data Center, and the Clean Cities’ National Clean Fleet Partnership program.

Designed by NREL in partnership with Oak Ridge National Laboratory, this online database and the accompanying reports help vehicle manufacturers and fleets understand the broad operational range of many of today’s commercial vehicle vocations. See Appendix A for examples of the type of 1-Hz data stored in the database [12]. In addition to the 1-Hz data stored in Fleet DNA, additional vehicle specifications information has been collected and stored to enable cross referencing. Examples of the types of vehicle specifications stored in Fleet DNA include classifiers for vehicle weight class, vocation, vehicle body type, drive train type, and fuel. Additional supplemental data used in this effort exist at the vehicle chassis and engine level for a subset of the overall Fleet DNA dataset.

For this project, a collection of over 16,250 days of data collected from 913 unique vehicles drawn from the Fleet DNA database were examined, analyzed, and segmented based on their corresponding driving behavior.

### 1.3 Classifying Data in Fleet DNA

As mentioned in the previous section, there are multiple vehicle, chassis, and engine classifiers which are used to sort and filter the data stored in Fleet DNA. For this study, vehicle classifiers of weight class, vocation, drivetrain, fuel, and vehicle type were all instrumental in helping to provide context for the drive cycle data analyzed. The vehicle classifiers are described as follows:

#### **Weight Class:**

- Class 2b (8,501 – 10,000 lbs.)
- Class 3 (10,001 – 14,000 lbs.)
- Class 4 (14,001 – 16,000 lbs.)
- Class 5 (16,001 – 19,500 lbs.)
- Class 6 (19,501 – 26,000 lbs.)
- Class 7 (26,001 – 33,000 lbs.)
- Class 8 (33,001+ lbs.)

**Vocation:**

- Beverage delivery
- Boom
- Delivery
- Food delivery
- Linen delivery
- Local delivery
- Long haul
- Mass transit
- Package delivery
- Parcel delivery
- Refuse pickup
- School bus
- Snow plow
- Telecom
- Towing
- Transit
- Utility
- Warehouse delivery.

**Vehicle Type:**

- Box
- Bucket truck
- City transit bus
- Dump truck
- Mini bus
- Refrigerated
- Refuse truck
- School bus
- Service van
- Step van
- Straight truck

- Tow truck
- Tractor
- Transit bus
- Walk in.

**Drivetrain:**

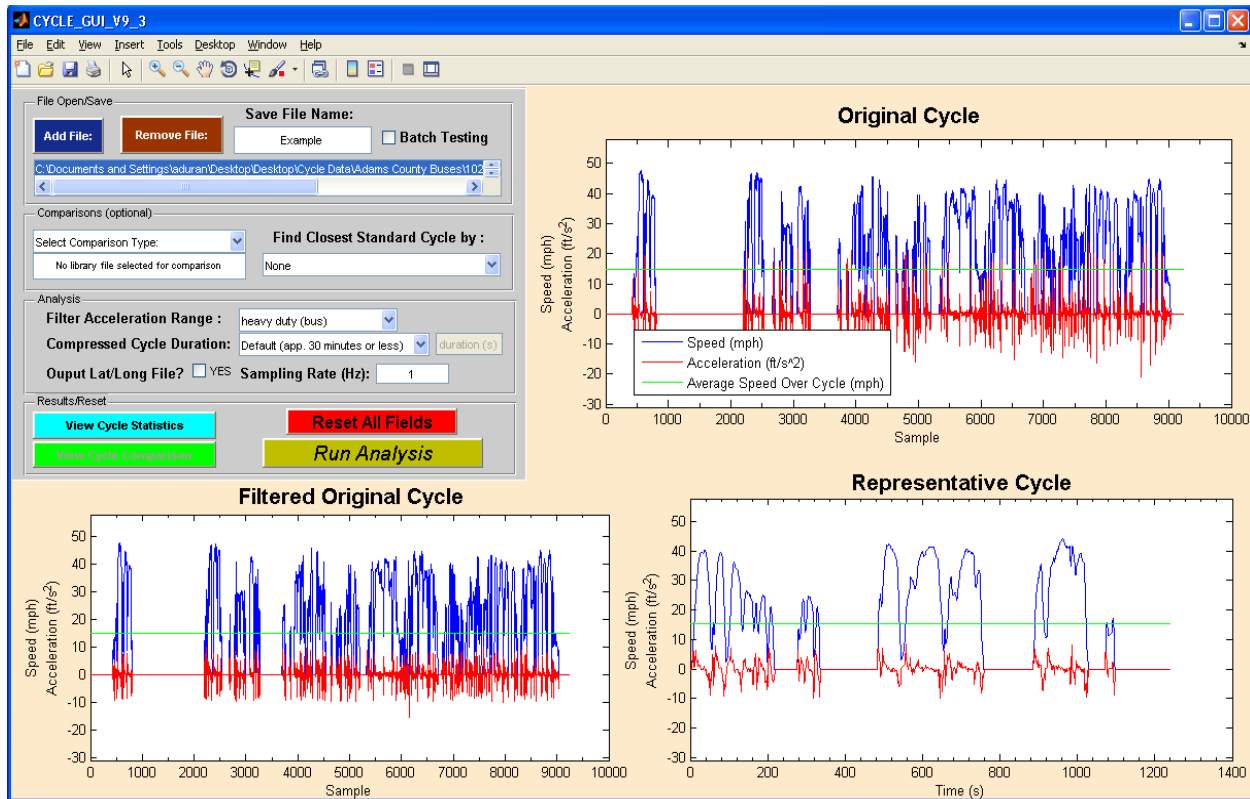
- Conventional
- Electric
- Parallel hybrid
- Series hybrid
- Hydraulic hybrid.

**Fuel:**

- BioDiesel
- CNG
- Diesel
- Electricity
- Gasoline
- Propane
- Renewable diesel.

## 1.4 NREL's DRIVE Tool

The DRIVE analysis tool shown in Figure 1 produces representative drive cycles from large amounts of vehicle data gathered via onboard logging devices. Developed by NREL, DRIVE uses global positioning system (GPS) and controller area network (CAN) data to characterize vehicle operation and produce custom vehicle drive cycles based on real-world activity, analyzing thousands of hours of data in a matter of minutes.



**Figure 1: Screen capture of DRIVE GUI processing data**

DRIVE filters large sets of raw data, removing erroneous data points and repairing missing data sections before performing analyses producing hundreds of unique vehicle drive cycle metrics [13]. Included in the list of metrics is Kinetic Intensity, a metric developed at NREL which describes the aggressiveness of a drive cycle as described by the ratio of acceleration to aerodynamic drag energy required over a drive cycle [14]. Upon filtration and analysis, the tool is used to generate custom representative drive cycles from “ideal” sections of filtered data using statistical clustering methods. The duration of these representative drive cycles, which aim to match the fuel economy and performance characteristics of the source data, can be specified via user-defined time limits, allowing for a wider range of testing and modeling applications.

In addition to DRIVE's ability to generate custom-duration representative drive cycles from raw data, the tool is capable of comparing vehicle operation to industry standard test cycles and can even select a representative standard cycle from its built-in library. Comparing the filtered data across the calculated drive cycle metrics allows users to select the standard cycle with the best fit to input data.

Algorithms mimic the logic and expertise of a human engineer, cutting testing and analysis time by days or weeks while establishing a repeatable process and making information accessible through a simple interface. Additional output results range from simple tabulated summary statistics to Google Earth route maps, providing information that has enough depth for scientific applications but is accessible to users without technical backgrounds.

## 1.5 Segmenting the U.S. Vocational Vehicle Population and Developing Representative Drive Cycles and Weights

In support of ongoing refinement of EPA's proposed GHG Phase 2 standards, NREL researchers were tasked with leveraging their broad collection of data, tools, and analysis expertise to successfully characterize and segment the U.S. Vocational Vehicle Population based on currently available chassis and drive cycle data. Multivariate statistical analysis approaches would be applied to the collection of drive cycle and chassis data stored in the Fleet DNA database to produce a robust set of statistically representative vehicle segments. Examples of some but not all of the key metrics that were examined include:

### Drive cycle metrics:

- Average driving speed
- Maximum driving speed
- Standard deviation of speed
- Kinetic intensity and its derivative components aerodynamic speed and characteristic acceleration
- Ratios of zero speed, low speed (0+-45 mph), and high speed (55+ mph) driving times and distances
- Stops per mile
- Percentage of operating time spent at zero speed/idle.

### Chassis/Engine Parameters:

- Final gear ratio
- Gross vehicle weight rating
- Vehicle weight class
- Number of axles
- Engine power rating
- Engine displacement
- Engine rpm in top gear at 65 mph.

Having successfully identified a set of distinct vocational vehicle segments, urban, mixed urban, and highway; NREL researchers were then tasked with developing a series of drive cycle weights which could be used in conjunction with existing proposed Phase 2 chassis dynamometer test cycles to describe each segment. In addition to developing drive cycle weights, NREL researchers also explored the development of representative transient test cycles using NREL's DRIVE tool. The results of the drive cycle development and weights determination was then compared to existing standard chassis test cycles such as the CARB Transient cycle to evaluate both the proposed and existing standard cycles ability to accurately describe the transient driving behavior of the national medium- and heavy duty vocational



vehicle population. This report documents the results of the completed work and conclusions reached.

## 2 Cluster Analysis

NREL researchers utilized a set of eight metrics describing each drive cycle to define representative segments through cluster analysis. Cluster analysis is an analytical method where groups of data within a larger dataset are identified through statistical analysis using a handful of target metrics to judge for similarity. The metrics in this study were chosen given their role in previous NREL research to characterize impacts of drive cycle on vehicle fuel economy and emissions production. The metrics used in the analysis were:

- Aerodynamic Speed (ft/s) - Describes the positive tractive energy required to overcome aerodynamic drag per unit distance over a given drive cycle. It is defined as:

$$AS = \frac{\sum_{i=1}^{N-1} \text{positive}(\frac{1}{2} \times (\text{velocity}_{i+1} - \text{velocity}_i) + g \times (h_{i+1} - h_i))}{D}$$

- Where: D = drive cycle cumulative distance (ft)  
g = gravitational constant (ft/s<sup>2</sup>)  
h = height of vehicle indexed from start of drive cycle in time by i, (ft)  
velocity = speed of vehicle indexed from start of drive cycle in time by i (ft/s)

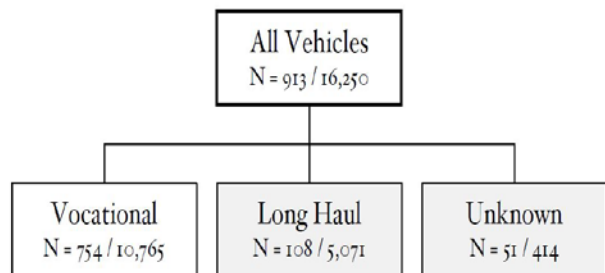
- Characteristic Acceleration (ft/s<sup>2</sup>) – Describes the positive tractive energy required to accelerate/raise a vehicle per unit distance over a given drive cycle. It is defined as:

$$CA = \frac{\sum_{i=1}^{N-1} \bar{v}_{i,i+1}^3 \times (t_{i+1} - t_i)}{D}$$

$$\bar{v}_{i,i+1}^3 = \frac{v_{i+1}^3 + v_{i+1}^2 + v_i + v_i^2 + v_i^3}{4}$$

- Where: D = drive cycle cumulative distance (ft),  
t = drive cycle time indexed from start of drive cycle in time by i (s)  
v = speed of vehicle indexed from start of drive cycle in time by i (ft/s)

- Percent (%) of total cycle distance accumulated at speeds below 55 mph
- Percent (%) of total cycle time duration accumulated at vehicle speeds of 0 mph
- Number of vehicle stops per mile
- Mean (nonzero) driving speed (mph)
- Maximum driving speed (mph)
- Standard deviation of (nonzero) driving speed (mph).

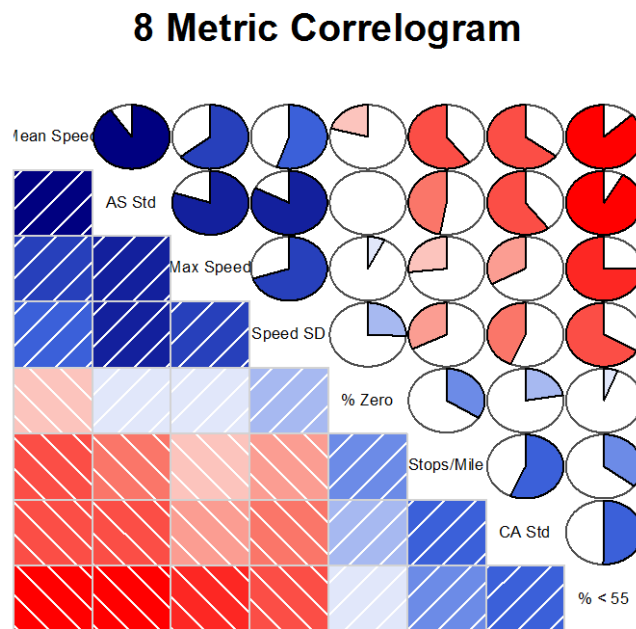


**Figure 2: Sample composition**

For this analysis, NREL utilized a collection 16,250 daily drive cycles from 913 vehicles. Among those vehicles, 108 (5,071 cycles) are long haul trucks with sleeper cabs and 754 (10,765) are vocational vehicles as shown in Figure 2. Additionally, contained in the database are 51 vehicles (414 cycles) classified as possessing an unknown vocation but possessing drive cycle data. Results for the entire sample population, as well as detailed results for the vocational segment will be presented in the following subsections. Additionally, for the vocational segment, resampling of the database based on reliable estimates of the true vehicle population in the U.S. was performed to ensure the sample dataset possessed a composition that is representative.

## 2.1 Principal Components Analysis

In order to explore the variability in the eight-metric dataset, NREL researchers performed a pairwise correlation (Figure 3) and a Principal Components Analysis (PCA) on the Fleet DNA dataset. The correlogram in Figure 3 shows the degree of linear dependence between each pair of variables. For instance, one can see from this plot that mean speed and percentage below 55 mph are strongly inversely correlated as shown by the dark red box and pie chart, while the mean speed and the Aerodynamic Speed (AS standard) are highly positively correlated as shown by the dark blue box and pie chart. The PCA is a dimensionality reduction process that allows us to describe a higher dimensional dataset with a smaller number of dimensions that can be easily visualized (in, e.g., two dimensions). In PCA, each observation can be described by a weighted sum of orthogonal loadings (Eigen vectors; principal components). As its focus is dimensionality reduction, the orthogonal loadings are always fewer than the number of starting dimensions.



**Figure 3: A correlogram for the eight metrics**

A correlogram gives a visual indication of the degree of linear correlation (dependence) between each combination of variables. The variables examined include average driving speed (mean speed), aerodynamic speed (AS Std), maximum driving speed (max speed), standard deviation of driving speed (Speed SD), percentage of total cycle time at zero speed (% Zero), number of stops (Stops/Mile), aerodynamic speed (CA Std), and percentage of time below 55 mph (% < 55).

per mile (Stop/Mile), characteristic acceleration (CA Std), and percentage of total drive cycle mileage accumulated below 55 mph (%<55).

A PCA on the eight metric drive cycle dataset suggests that two components are able to describe 75% of the variance in the complete dataset. The first four components are able to describe 91% of the variance and first six components describe 99%. As each component is a weighted sum of the 8 metrics, each component contains information from some combination of the metrics, and weights some more heavily than others. All 8 components together would reproduce the original 8 dimensional data exactly.

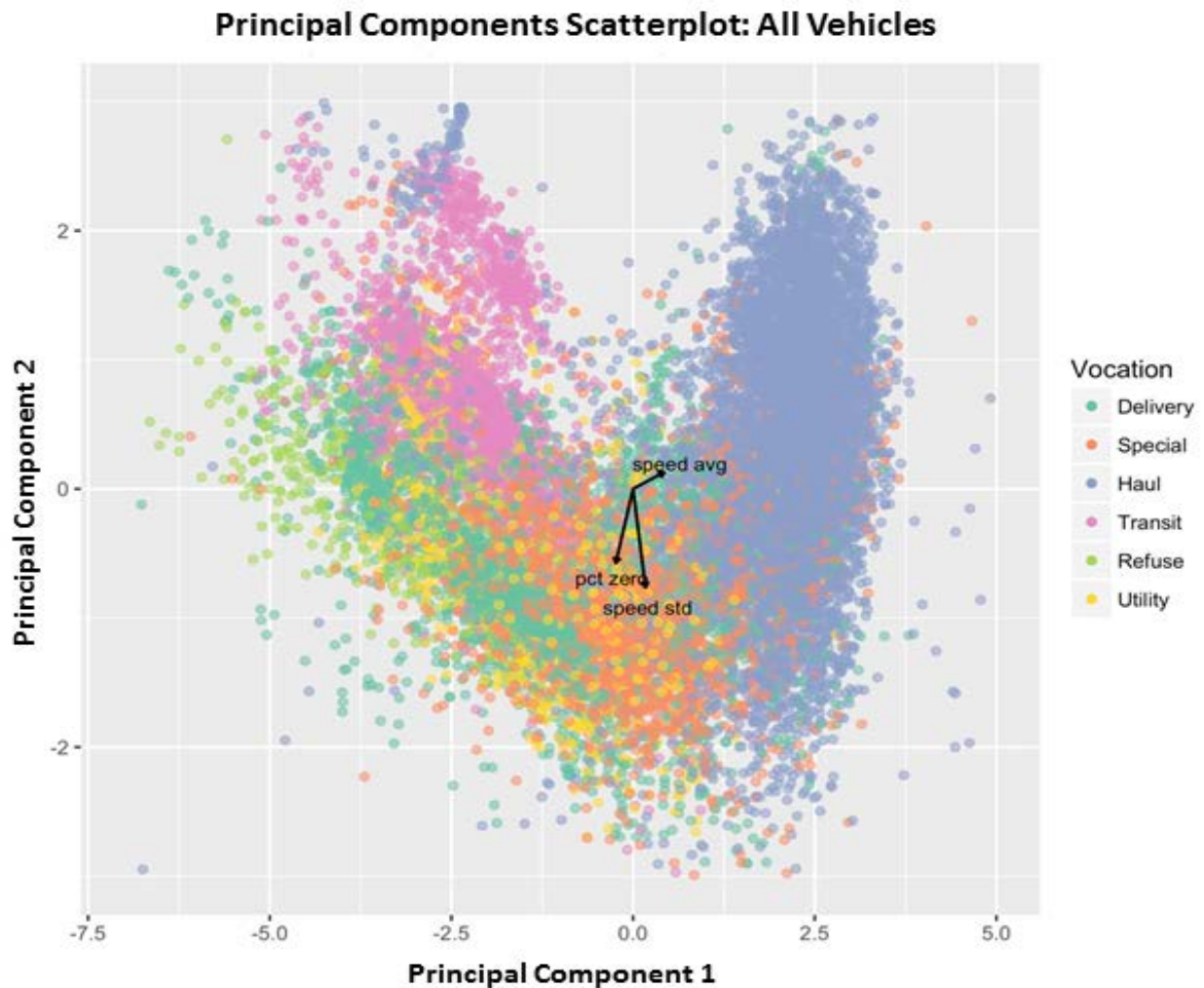
Tables 2 and 3 give the PCA loadings (rotation) for all vehicles and for vocational vehicles respectively. The first principal component (PCA 1) convolves up-weightings on driving speed and standard deviation. The second principal component (PCA 2) most heavily weights the percentage of time spent at zero. These metrics are heavily weighted in the first two components because they account for the greatest degree of variance among the samples. Figures 4 and 5 show how the population of drive cycle characteristics is distributed in a two-dimensional PCA space for all vehicles and for vocational vehicles respectively. Each point describes the position in the feature space, and hence describes the driving characteristics, for one day of driving for one vehicle. In this figure, some vocational categories are more distinct in their characteristics than others—Transit (which include school buses) and Haul vehicles are most similar within their categories (i.e., are most tightly and consistently clustered), although the variance within those groups is relatively high. Refuse and transit vocations stand out in their usage as compared to the rest of the sample. In both plots the rotation of a subset of underlying variables is given at the center of the plot. Note that with vocational vehicles, the optimal PCA loadings are slightly different and the second component (PC2) changes sign, which causes the resulting plots to appear flipped relative to one another. In both plots, there are two modalities visible in this dataset which appear to be largely differentiated by maximum speed and variability in speed, with a large cluster of mixed-mode driving somewhere in between.

**Table 2. Principal Component Loadings (All Vehicles)**

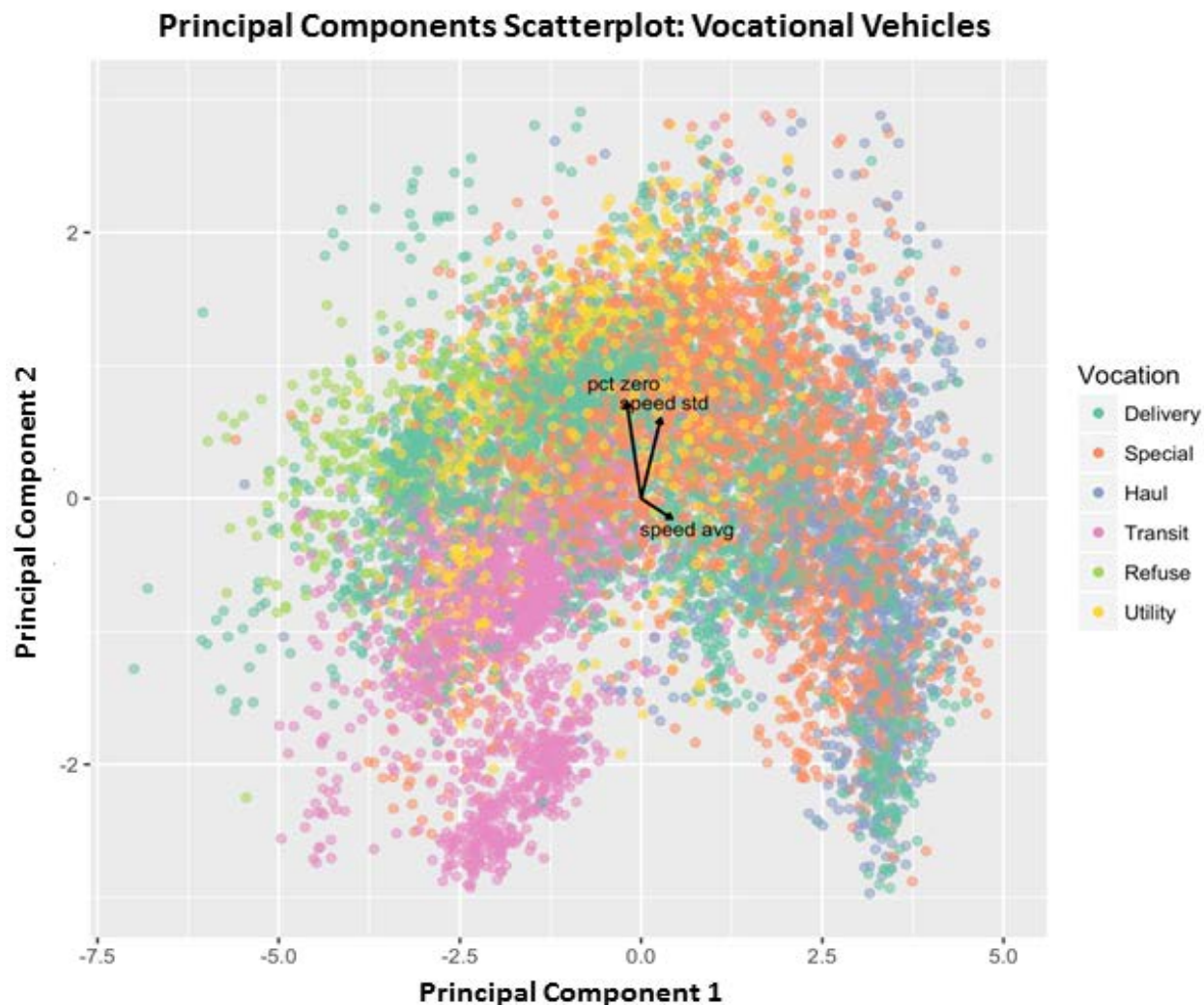
<b>Variable</b>	<b>PC1</b>	<b>PC2</b>	<b>PC3</b>	<b>PC4</b>	<b>PC5</b>	<b>PC6</b>
AS Standard	0.44	-0.1	0.13	-0.09	-0.02	-0.26
CA Standard	-0.34	-0.12	0.54	0.37	0.54	-0.4
Percent below 55 mph	-0.42	-0.04	-0.21	0.19	0.07	0.39
Percent at 0 mph	-0.24	-0.57	-0.11	-0.72	0.29	-0.04
Stops per mile	-0.34	0	0.58	-0.25	-0.68	-0.07
Mean driving speed (mph)	0.43	0.13	0.1	-0.2	0.11	-0.29
Max driving speed (mph)	0.36	-0.21	0.53	0	0.19	0.71
Driving speed SD (mph)	0.17	-0.76	-0.14	0.45	-0.34	-0.13

**Table 3. Principal Component Loadings (Vocational Vehicles)**

Variable	PC1	PC2	PC3	PC4	PC5	PC6
AS Standard	0.44	0.1	0.1	0.06	-0.02	0.24
CA Standard	-0.33	0.13	0.53	-0.55	0.36	0.41
Percent below 55 mph	-0.42	0.01	-0.26	-0.22	-0.02	-0.39
Percent at 0 mph	-0.2	0.72	-0.22	0.43	0.44	0.1
Stops per mile	-0.32	0.1	0.59	0.49	-0.51	0.04
Mean driving speed (mph)	0.43	-0.15	0.07	0.19	0.23	0.3
Max driving speed (mph)	0.35	0.23	0.48	-0.07	0.25	-0.72
Driving speed SD (mph)	0.27	0.6	-0.13	-0.42	-0.55	0.1



**Figure 4: All drive cycles for all vehicles visualized in the space defined by the first two principal components. Long-haul vehicles are present in a high speed grouping. Color is given by vocation.**



**Figure 5: Scatter plot of all vocational drive cycles. Refuse and transit vocations stand out in their usage as compared to the rest of the sample. Note that principal component 2 changes signs in this plot, which makes it appear inverted relative to Figure 4.**

## 2.2 Drive Cycle Clustering

NREL researchers aimed to group similar drive cycles and derive representative drive cycle characteristics based on the central tendency of each cluster using a variety of data clustering methods. After some work evaluating various clustering methods for this data (K-means, hierarchical, etc.), the K-medoids algorithm was selected as the best candidate. K-medoids functions well on large data sets by optionally clustering random subsamples. This algorithm works by first randomly selecting a number of cluster centers specified a priori. It then assigns all points to the closest cluster. New cluster centers are chosen and the operation repeats until a convergent set of optimal clusters are found. Unlike K-means which provides metric averages for each cluster, the K-medoids algorithm selects a most-representative data point which improves the interpretability of the results [15-16]. To determine an optimal number of clusters in the data, we utilize the silhouette method described in [17] where clustering is performed sequentially starting with two clusters and iterating with additional clusters. At each iteration, the silhouette

analysis describes the ratio between tightness (within cluster variance) and separation (between cluster variance). The optimal clustering is one that provides the greatest separation between clusters while being robust to small numbers of outlier points and keeping the number of clusters as small as possible. Based on this analysis, we select two clusters as the optimal clustering in the metric space. Figure 6 shows the resulting two clustering for vocational vehicles, which finds the underlying bimodal structure and splits the space logically. A small section of detached vehicles at the bottom of the left cluster are school buses with higher percentage of time at zero speed compared to the other vehicles in the cluster. Figure 8 focuses on school buses specifically with a sub-cluster analysis. Here it can be seen that there is additional structure present: school buses fall into two separate clusters within their own data. The first cluster is typified by its center point with 36.6% of time spent at zero speed while the center point of the lower cluster spends 3.3% of time at zero speed. As the vehicles in these two groups do not seem characteristically different, more investigation is needed to understand why school bus driving characteristics are partitioned this way. The balance of vocations appears relatively homogenous within the two dominant modalities.

For the sake of clarity, in the following sections we will refer to the left cluster, which contains slower speed cycles with more stops, as the ‘Slow’ cluster and the right cluster, which contains higher speed cycles with fewer stops, as the ‘Fast’ cluster. In this plot, as in the prior plots, each point represents one day of driving in the entire data set. Points are colored according to their optimized cluster placement. A single vehicle may have any number of drive cycles, which may over or under-represent individual vehicles in these plots.

### Drive Cycle Clustering - 8 Metrics (2 Clusters)

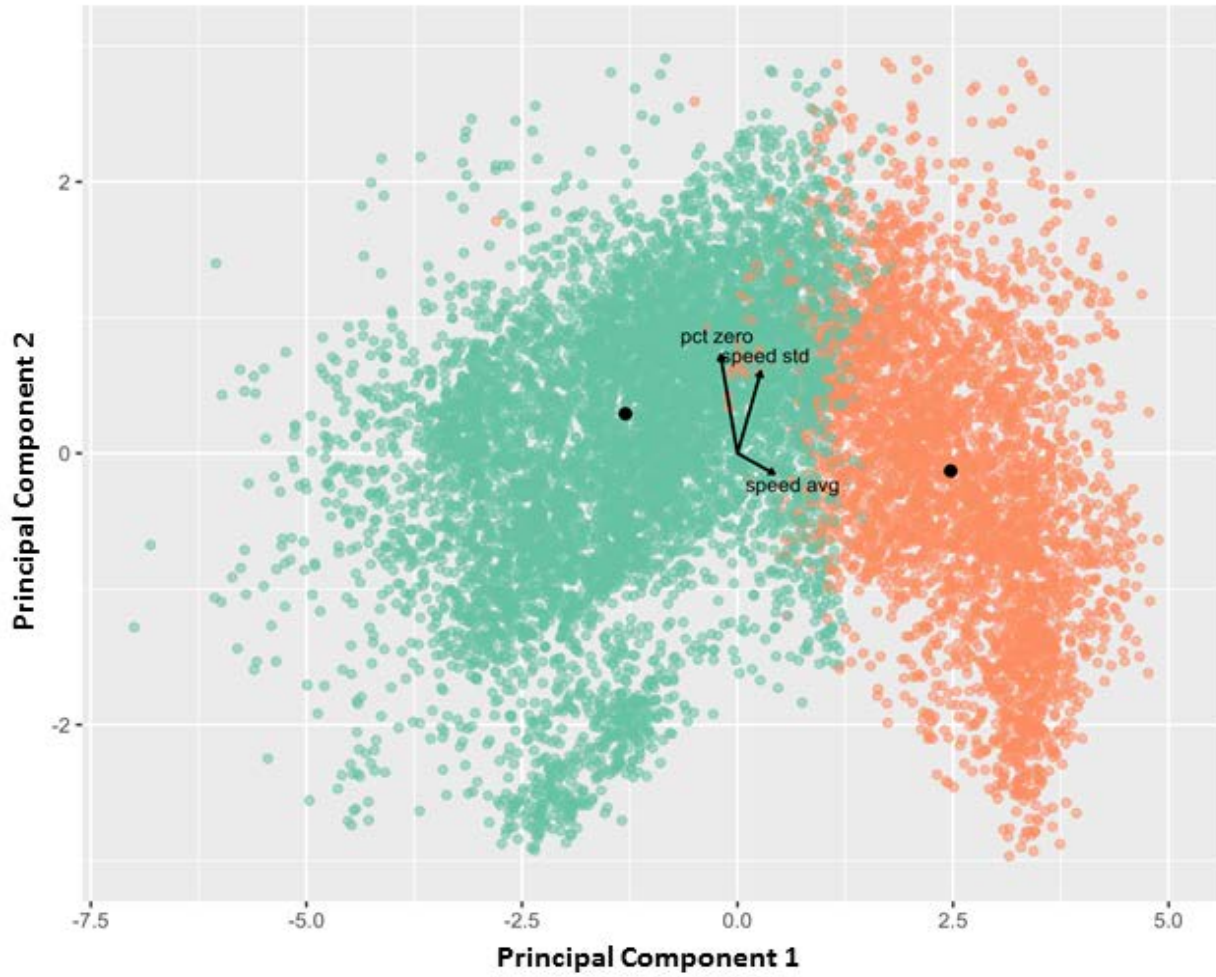
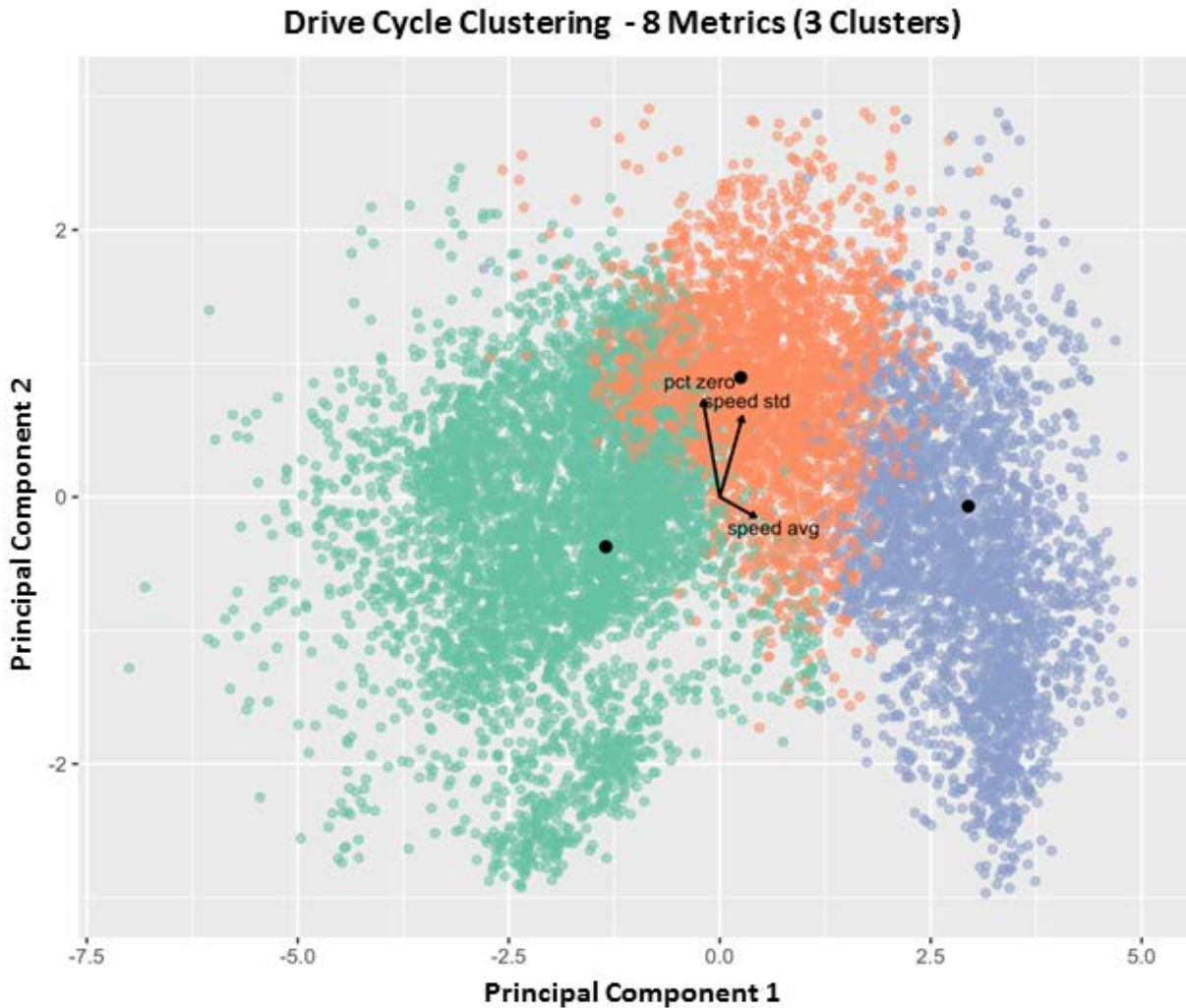


Figure 6: Two clustering using k-medoids algorithm with cluster centers marked in black. Only vocational vehicles are drawn in this plot.





**Figure 7: Three clustering using k-medoids algorithm with cluster centers marked in black. Only vocational vehicles are shown in this plot.**

For regulatory purposes where it may be useful to consider three classes of vehicles, an optimal solution with three clusters was calculated for the vocational vehicle drive cycle sample. Figure 7 shows this data. The first two clusters are joined in this plot by a middle cluster that contains those traces that do not clearly fall into either the left (slower) or right (faster) cluster. Tables 4 and 5 contain the drive cycle characteristics for the center point of each cluster along with the vocation of the vehicle located at the medoid center.

### Subcluster Analysis for School Buses

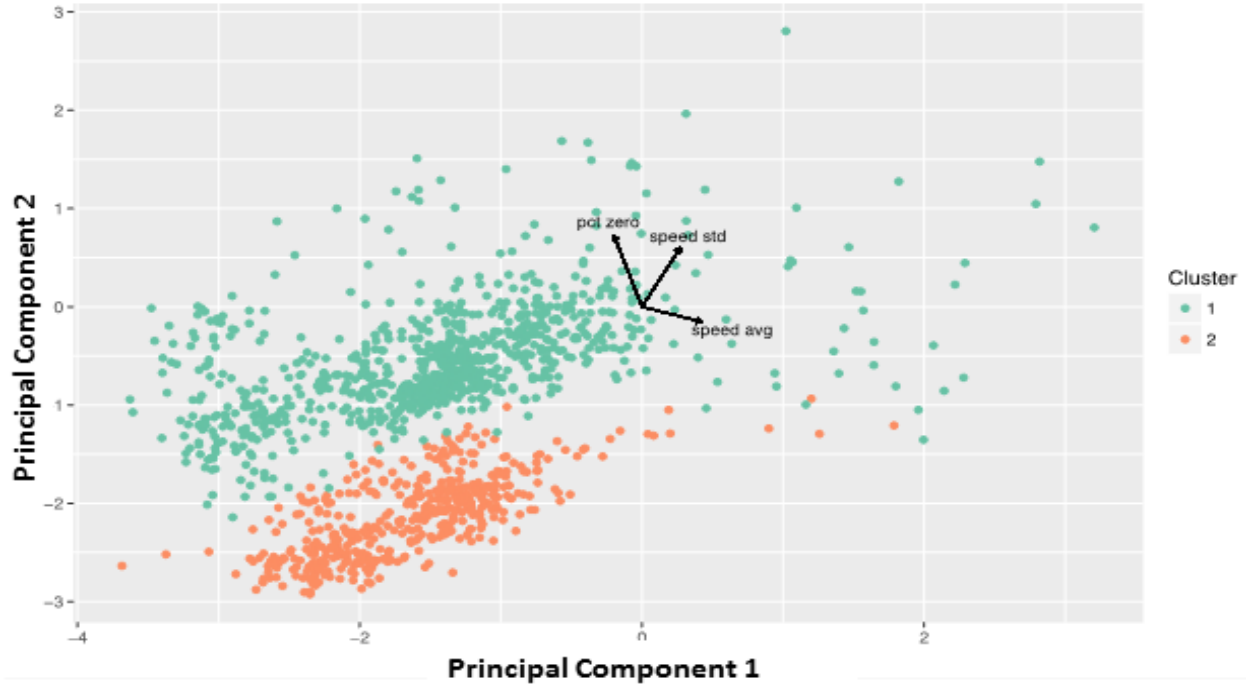


Figure 8: Sub-cluster analysis for School buses, which themselves fall into two clusters differentiated by slower and faster average speed.

Table 4. Vocational Drive Cycles – 2-Clustering Centers

	Medoid Cycle Vocation <sup>a</sup>	AS Standard (ft/s)	CA Standard (ft/s <sup>2</sup> )	Percent Mileage Below 55 mph	Percent Time at 0 mph	Mean stops/mile	Mean driving speed (mph)	Max driving speed (mph)	Driving speed Std Dev (mph)
Left (Slow) Cluster	Refuse Pickup	54.59	0.48	91.08	50.62	2.68	21.26	59.22	16.28
Right (Fast) Cluster	Drayage	81.21	0.28	26.62	22.11	0.24	43.76	65.01	21.06

<sup>a</sup> Note: the medoid vehicle may be representative for the overall cluster, while being abnormal for its own vocation. For instance, in the three-cluster solution the school bus that is selected as the medoid vehicle spends far less time at zero mph than other school buses. Nevertheless, this one vehicle is most representative of the entire class of vehicles irrespective of normality or abnormality for its own vocation.

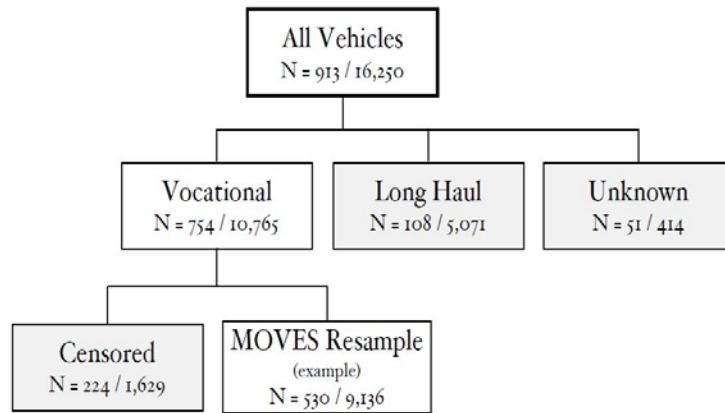
**Table 5. Vocational Drive Cycles – 3-Clustering Centers**

	Medoid Cycle Vocation	AS Standard (ft/s)	CA Standard (ft/s <sup>2</sup> )	Percent Mileage Below 55 mph	Percent Time at 0 mph	Mean stops/mile	Mean driving speed (mph)	Max driving speed (mph)	Driving speed Std Dev. (mph)
Left (Slow) Cluster	School Bus	48.88	0.50	97.85	47.87	1.22	24.12	62.61	13.03
Middle Cluster	Towing	69.27	0.68	64.46	52.68	0.98	34.50	67.86	18.75
Right (Fast) Cluster	Freight	84.85	0.20	25.38	25.43	0.26	47.70	70.79	20.48

In order to utilize cluster centers to define a representative drive cycle for testing/modeling purposes, the drive cycle characteristics from the top 50 days of data closest to these centers as ranked by multivariate least squares distance from the centroid were combined to determine the cluster averages for each metric.

### 2.3 Data Resampling

Although the Fleet DNA database provides detailed data for a large number of vehicles, in order to draw broad conclusions about how vehicles behave we must address sources of potential bias in this dataset. This bias may arise simply because those fleets most willing to contribute data to Fleet DNA may not be perfectly representative of the entire population of vehicles in the United States. As shown in Figure 9, in this section we perform resampling based on EPA’s MOTOtor Vehicle Emission Simulator (MOVES) study categorization, compute new clusters based on the new vehicle population, and then measure the difference in cluster center [18]. Using this analysis we can determine the degree to which the entire Fleet DNA sample population differs from the population of vehicles proportioned according to estimated US population statistics.



**Figure 9: Resampling flow for vocational vehicles**

It is important to note that due to differences in vehicle classification systems between the Fleet DNA and MOVES databases, that it was necessary for NREL researchers to aggregate Fleet DNA vehicle types into broader categorical groupings to match those of the MOVES and MOVES subcategory designations for resampling. In the case of a global MOVES resampling, the detailed list of vehicle types and vocations discussed in Section 1 were aggregated into four major vehicle categories as shown in Table 5.

Table 6 gives the MOVES-equivalent vehicle counts and proportions for resampling as compared to the entire Fleet DNA data. The four classes listed here are common to both the Fleet DNA data and the MOVES data. The Fleet DNA data appears to have a smaller fraction of Short Haul vocational vehicles while having relatively more school buses, refuse trucks, and transit buses. The resulting resampled population would contain 230 fewer vehicles if resampled according to these proportions. A MOVES-equivalent sub-category resampling, which includes vehicle weight class, is given in Table 7. In order to match proportions of each category within weight classes, the total population of vehicles must be decreased by 306 (nearly half of all vocational vehicles).

**Table 6. Vocational Resampling by MOVES Categories**

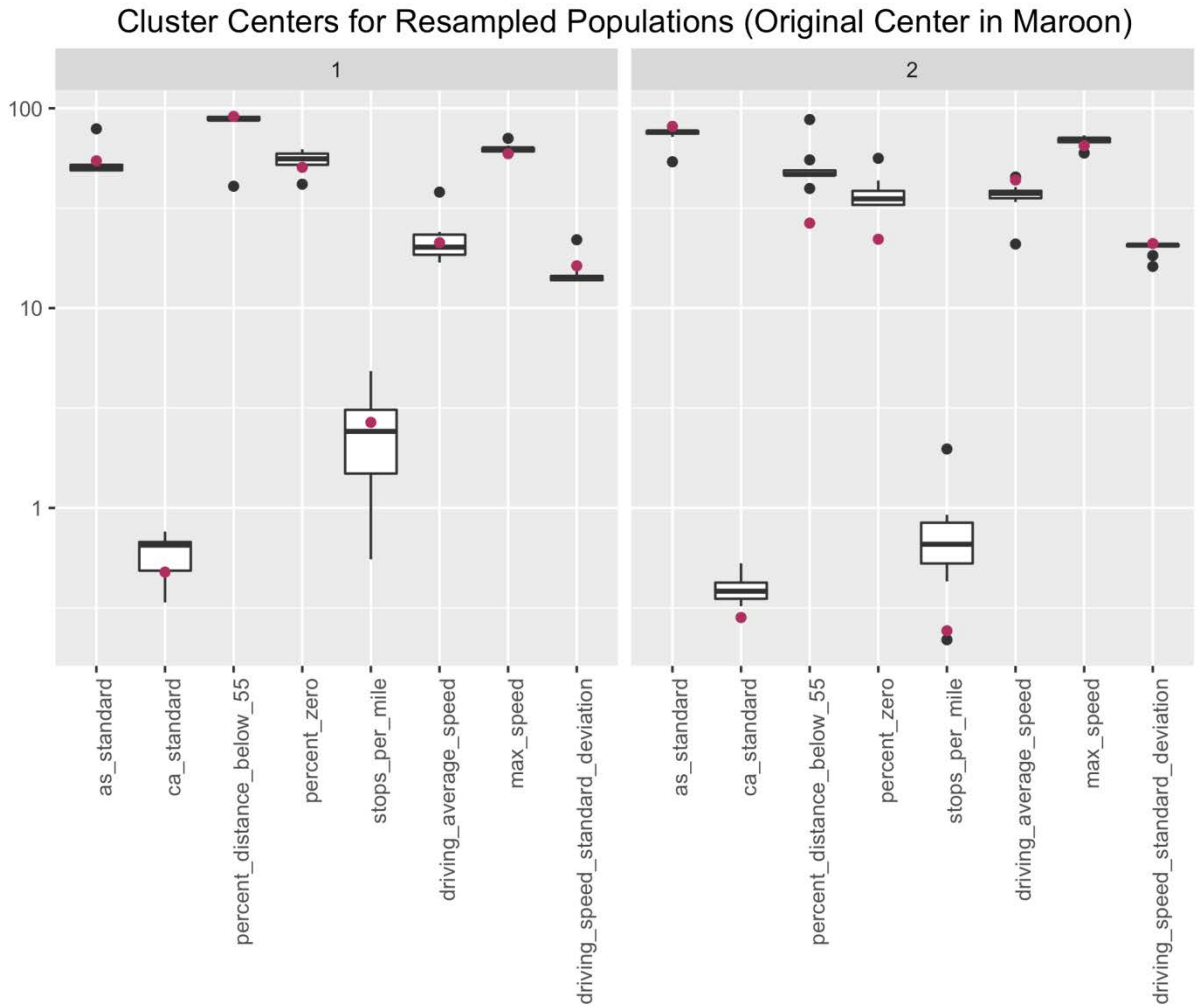
<b>Vocation</b>	<b>Entire Fleet DNA Sample Population</b>	<b>MOVES Resampled Fleet DNA Population</b>
Short Haul	441 (58.9%)	441 (84.1%)
School Bus	240 (23.6%)	57 (10.9%)
Refuse	50 (5.3%)	20 (3.9%)
Transit Bus	23 (2.1%)	6 (1.1%)

**Table 7. Vocational Resampling by MOVES Sub-categories**

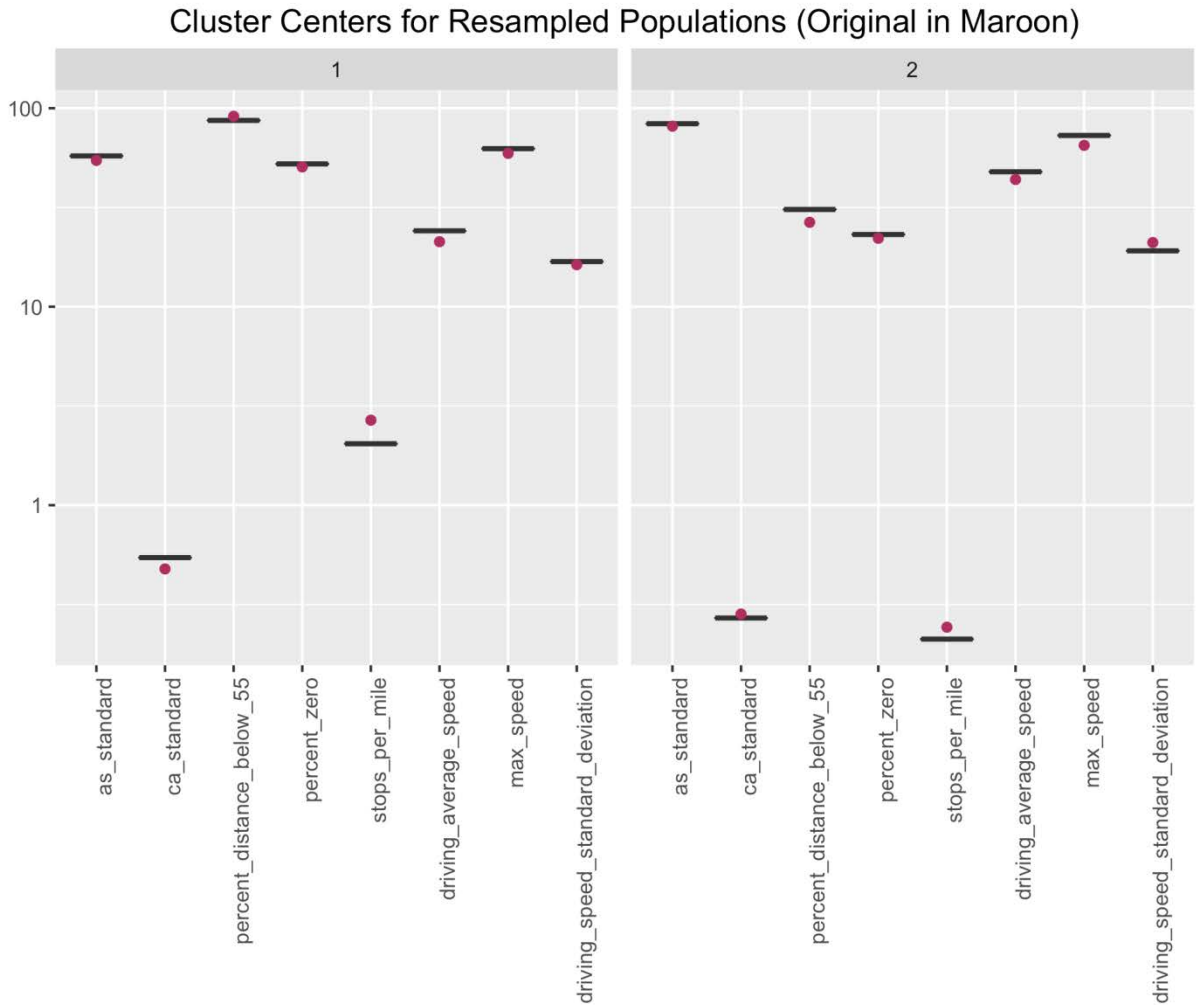
<b>Vocation</b>	<b>Entire Fleet DNA Sample Population</b>	<b>MOVES Fleet DNA Resampled Population</b>
Short Haul – Class 4/5	68 (6.3%)	53 (34.6%)
Short Haul – Class 6/7	155 (14.4%)	29 (19.0%)
Short Haul – Class 8	224 (20.9%)	21 (12.4%)
Refuse – Class 6/7	2 (0.2%)	0 (0%)
Refuse – Class 8	55 (5.1%)	4 (2.6%)
School Bus – Class 6/7	212 (19.8%)	12 (7.8%)
School Bus – Class 8	27 (25.2%)	1 (0.7%)
Transit Bus – Class 6/7	3 (0.3%)	0 (0%)
Transit Bus – Class 8	20 (1.9%)	1 (0.6%)
N/A	306 (28.5%)	

In order to evaluate the impact of resampling the underlying data on the extant drive cycle clusters, a 10-fold evaluation was performed, where 10 random subsamples of daily driving cycles that are consistent with the MOVES proportions were selected and used for cluster analysis. For instance, starting by selecting a random sample of vehicles from the Fleet DNA database that has the same categorical and fractional breakdown as given in Table 5, cluster analysis and statistical characterization were then performed. This process was repeated 10 times (folds) to obtain a notion of how much variability there was among random subsamples.

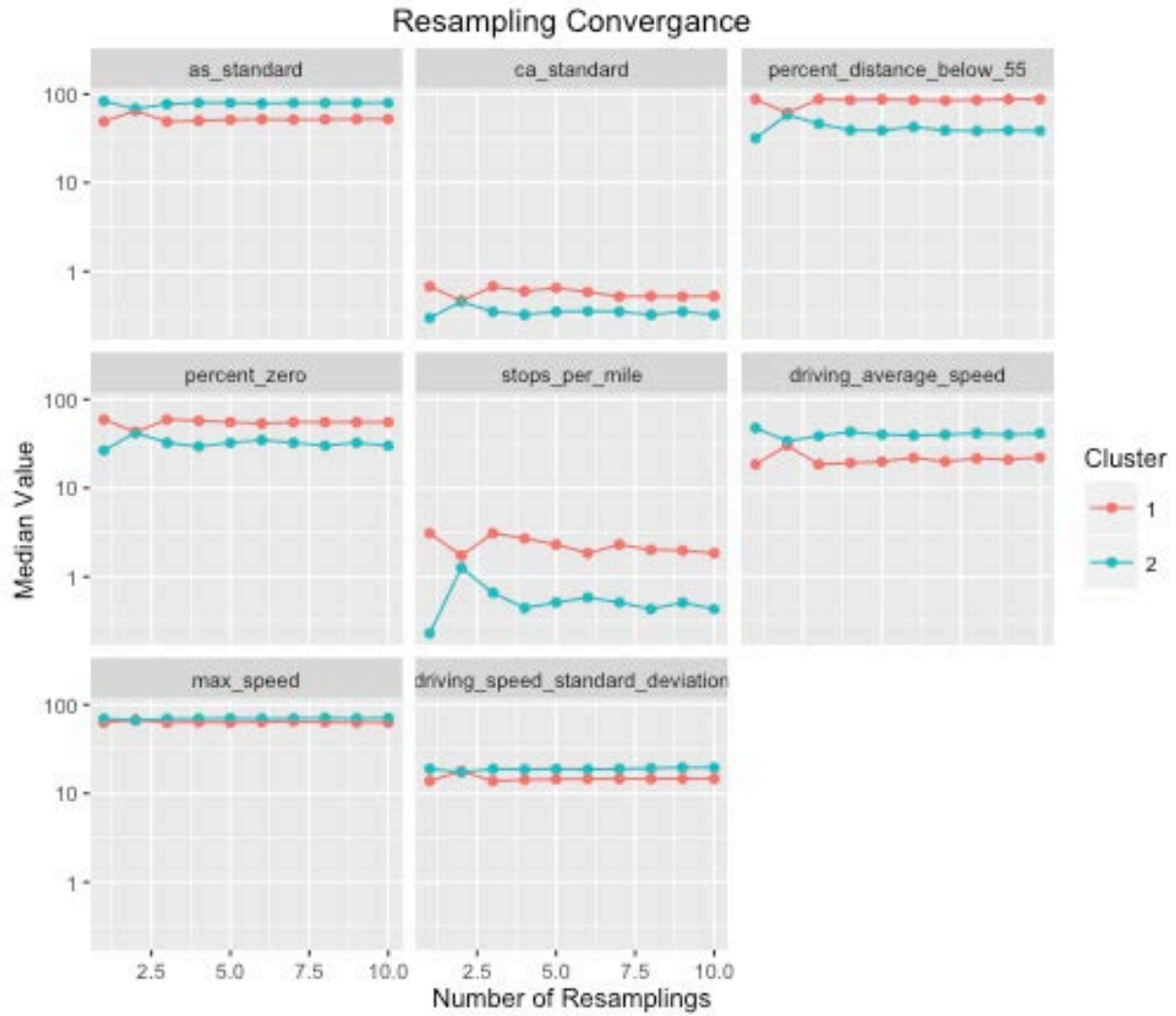
Figures 10 and 11 show the center value of each of the 8 metrics both with and without resampling for MOVES categories and subcategories. Figures 12 and 13 show how quickly the values stabilize when averaging results from iterative resamplings for both MOVES categories and subcategories. These plots show that after just 5 resamplings there is not a meaningful degree of additional variability.



**Figure 10: Difference between resampled cluster center statistics and original Fleet DNA (entire population) statistics for two clusters. The left pane shows the median value for the slow (cluster 1) population for each metric as a maroon dot. The right pane shows the same for the fast (cluster 2) population. In each pane, the boxplots give the distribution of median values with random MOVES-based resamplings. Black dots represent outliers.**



**Figure 11: Difference between resampled cluster center statistics and original Fleet DNA (entire population) statistics for two clusters. This plot uses MOVES subcategory resampling. The left pane shows the median value for the slow (cluster 1) population for each metric as a red dot. The right pane shows the same for the fast (cluster 2) population. In each pane, the boxplots give the distribution of median values with random MOVES-based resamplings. The subcategory resampling has smaller variance in the metrics because of a smaller sample size.**



**Figure 12: Convergence on final cluster center metrics as a function of successive resamplings**

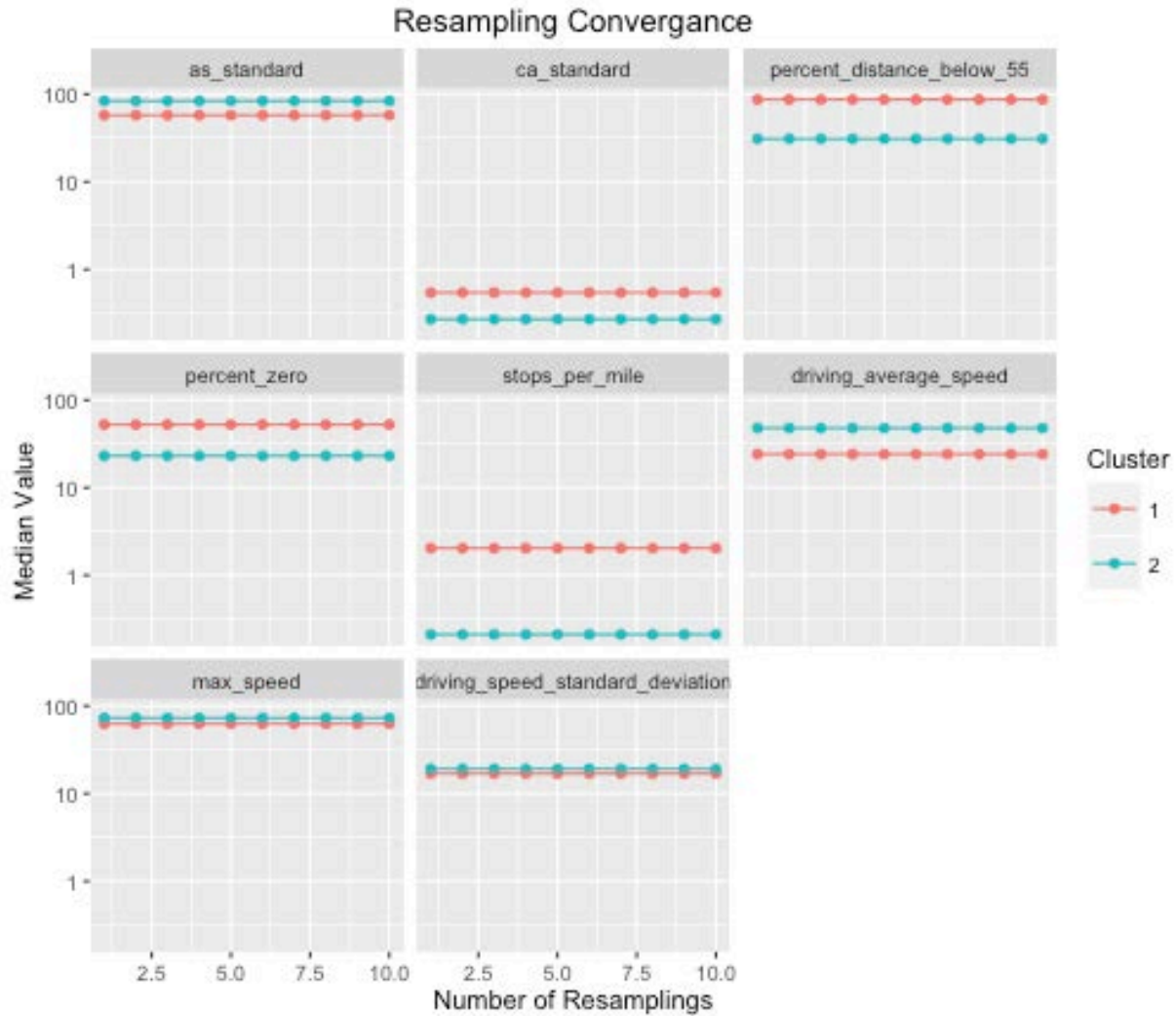


Figure 13: Convergence on final cluster center metrics as a function of successive MOVES subcategory resamplings

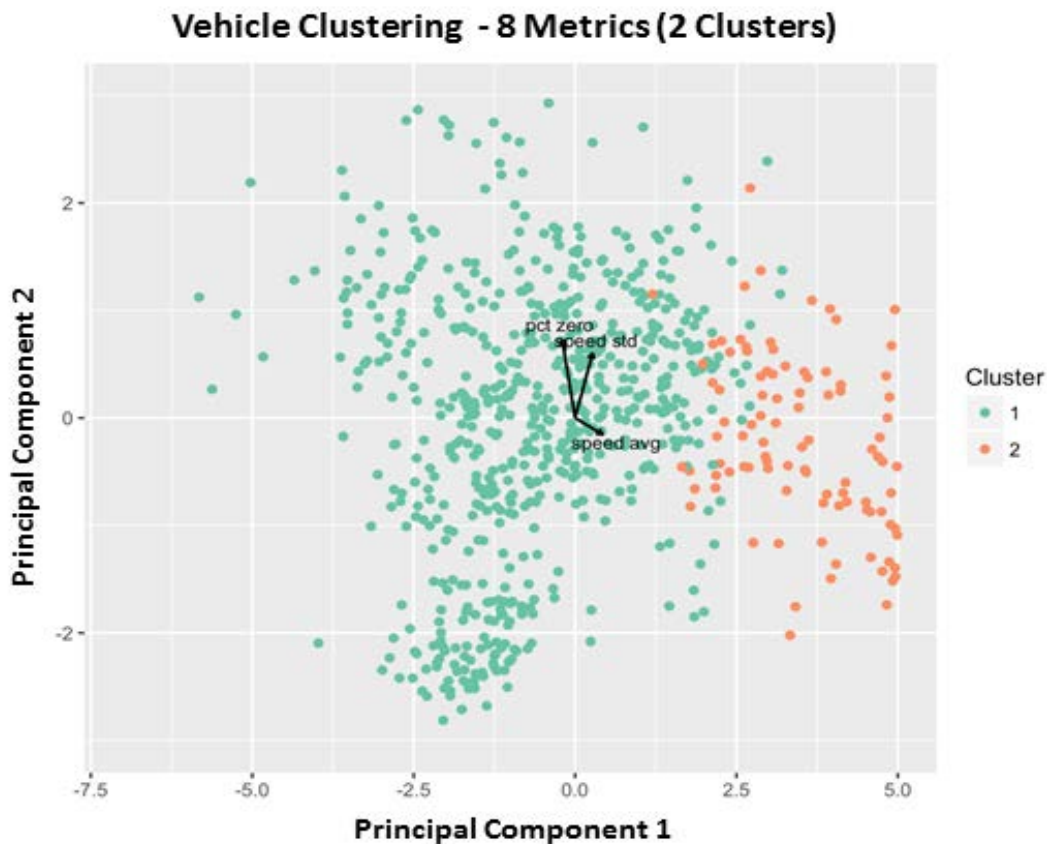
## 2.4 Vehicle Clustering

The prior sections have focused on daily drive cycle clustering. In this section we generalize the approach to the level of vehicles. Rather than segmenting vehicles based on a notion of their average use, which may hide multiple modalities or be skewed by outlier drive cycles, we instead assign each vehicle to the cluster that the majority of its drive cycles belong. This assignment can be computed trivially using the statistical mode function on each vehicle’s drive cycle cluster assignments. The mode function selects the value that appears most commonly in a collection of observations. For instance, a vehicle that spends 60% of its drive cycles within cluster 1 and 40% of its drive cycles in cluster 2 would be assigned to cluster 1. In the following discussion, we refer to this method as cycle-mode clustering. More complex definitions of vehicle cluster membership can also take into account the multi-modalities by, e.g., assigning vehicles with a mixture of modalities to a third (center) cluster. For instance, we might choose for vehicles that do not express a strong majority, i.e., greater than 2/3 cycles in a given cluster,

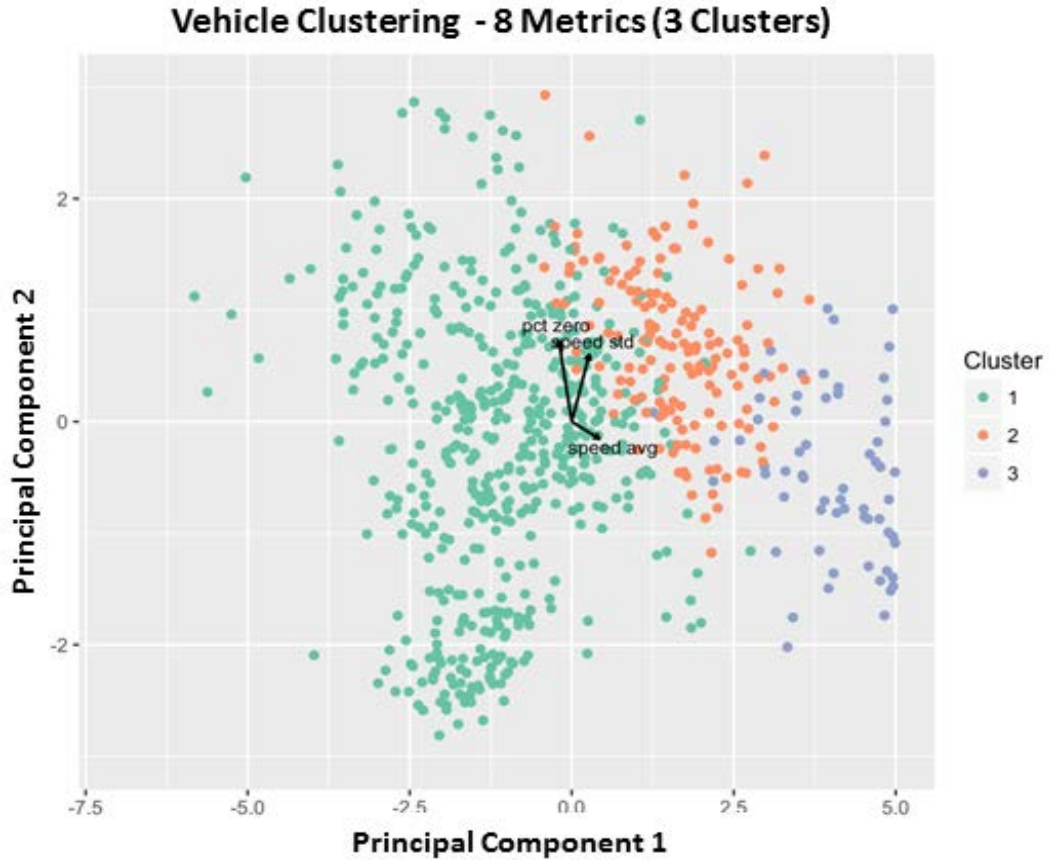


to be assigned to this third “center” cluster. In the work that follows we present results from cycle-mode clustering and leave these extensions for future work.

Figures 14 through 16 show the result of this approach to per-vehicle classification. Each point in the scatterplot is placed according to the vehicle’s average statistical behavior and shows the aggregate behavior of the 754 known vocational vehicles stored in the Fleet DNA database. The cluster assignment for each vehicle is the most common cluster among that vehicle’s multiple daily drive cycles (i.e., cycle mode cluster as discussed above). In Figure 16, the relationship between cluster membership, operational characteristics, and vocation is shown. Transit vehicles (city buses and school buses) are entirely contained in the left (Slow) cluster, while delivery, special purpose (tanker, concrete mixer, drayage, dump trucks, etc.), and utility vehicles are present in both clusters. In the next section we will explore possibilities for predicting vehicle cluster membership (and hence, drive characteristics) using information available at the time of regulatory testing.

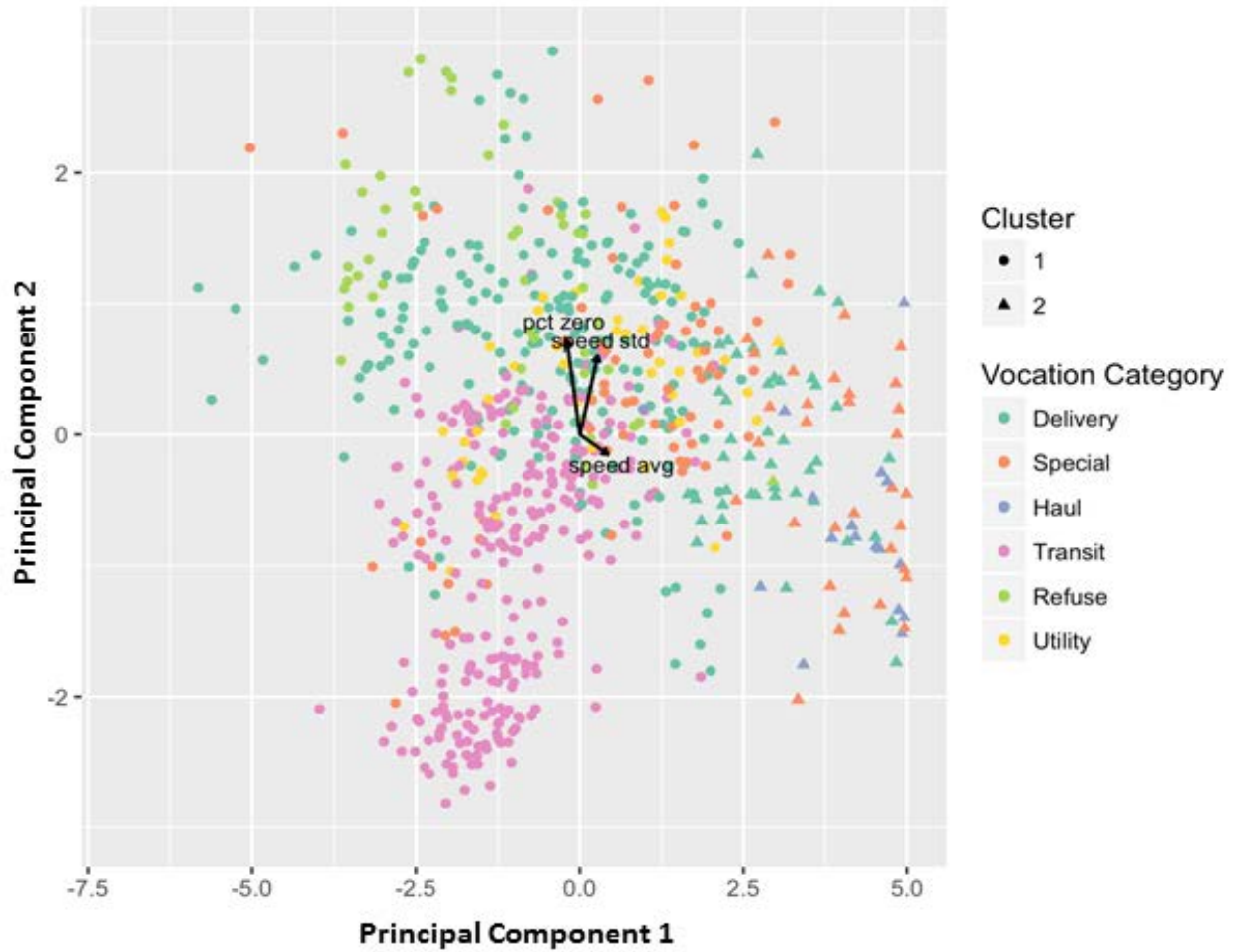


**Figure 14: Cycle-mode clustering of Fleet DNA Vocational vehicles using two-cluster solution. This plot provides a simplified (condensed) version of the entire drive cycle data. Each vehicle’s location in the scatterplot is given by average statistics. Cluster assignment is by the mode of trace cluster membership (most common cluster).**



**Figure 15: Cycle-mode clustering of Fleet DNA Vocational vehicles using three-cluster solution. This plot provides a simplified (condensed) version of the entire drive cycle data. Each vehicle's location in the scatterplot is given by average statistics. Cluster assignment is by the mode of trace cluster membership (most common cluster).**

### Vehicle Clustering - 8 Metrics (3 Clusters)



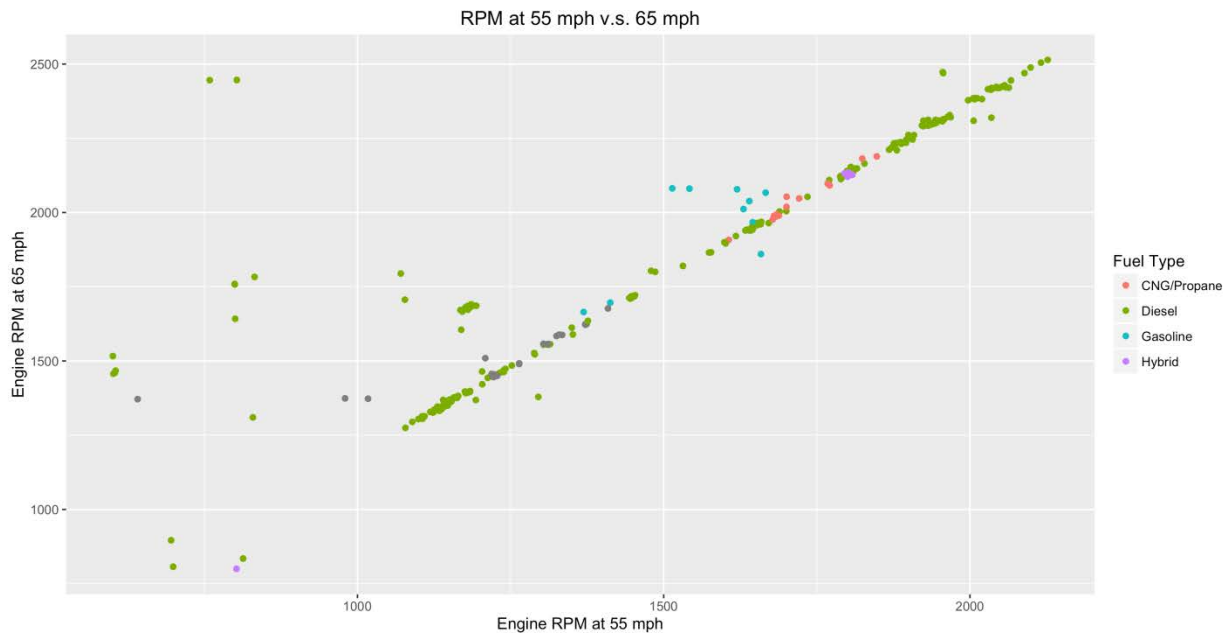
**Figure 16: Cycle-mode clustering of Fleet DNA Vocational vehicles using two-cluster solution. This plot provides a simplified (condensed) version of the entire drive cycle data. Each vehicle's location in the scatterplot is given by average statistics. Cluster assignment is by the mode of trace cluster membership (most common cluster). Vocations are provided to show vocational segmentation in the PCA space.**

### 3 Predictive Modeling

In this section NREL researchers explored whether it is possible to predict drive cycle characteristics using limited information available about vehicles at the time of regulatory testing. Additional information, quantifying the efficacy and general purpose accuracy of this method is also included. For a given vehicle, the information that a predictive model could use is limited to those variables that are in both the Fleet DNA database, and are present at the time of testing:

- Transmission Type
- Fuel Type
- Drivetrain Type
- Engine RPM at 55 mph
- Engine RPM at 65 mph.

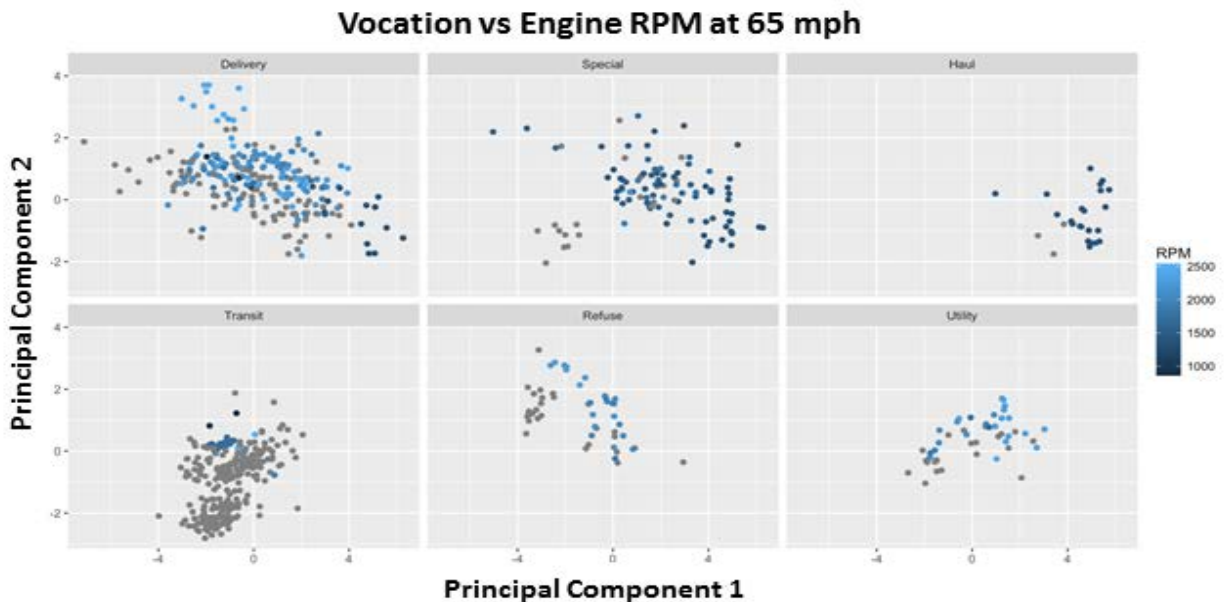
Engine RPM for vehicles is calculated using a method described in Appendix B. For the vast majority of vehicles, RPM at 65 and 55 mph are trivially correlated, as can be seen in Figure 17. Based on this finding, RPM at 65 mph was used, as it was the metric most likely to correspond to conditions in which the source vehicle would be operating in the final transmission gear for analysis. Figure 18 shows the data coverage for RPM data in the Fleet DNA data according to vocational category. All vocations are well represented aside from the Transit category (city buses and school buses) which appear under-represented in terms of availability of engine RPM data. This plot also makes clear that there is not an obvious or simple relationship between gearing and vocational category.



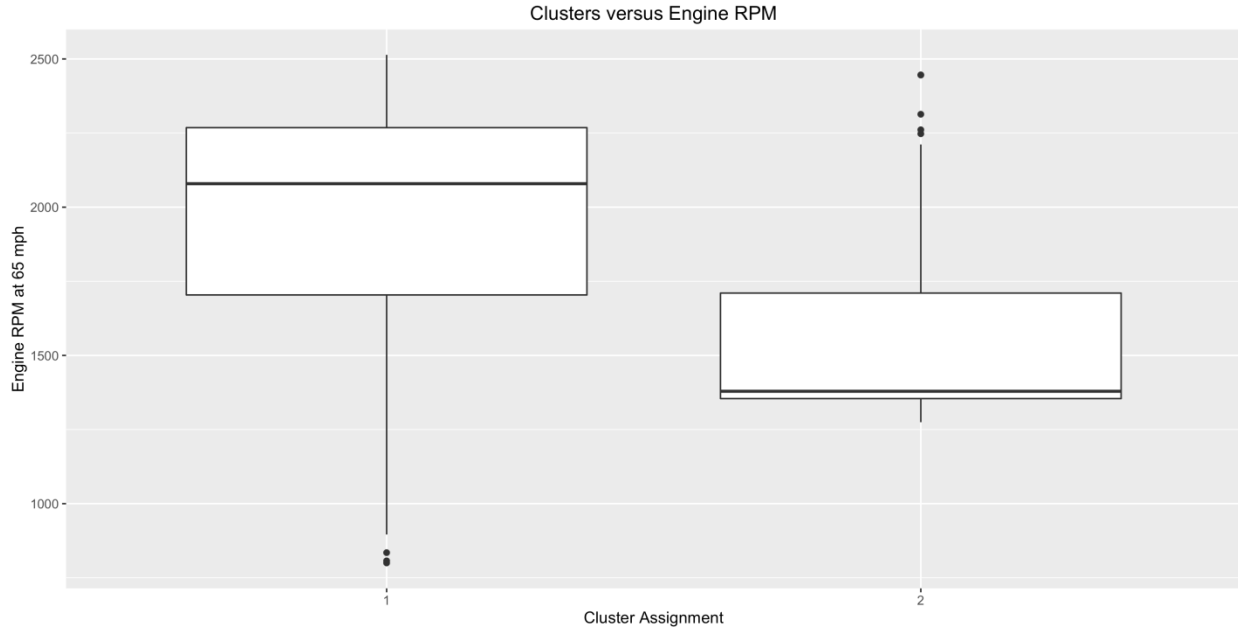
**Figure 17: Scatter plot showing correlation between RPM at 55 mph and RPM at 65 mph for different vehicle fuel types.**

Figures 19 and 20 illustrate the relationship between engine RPM and vehicle cluster categorization. The approach to modeling in this section stems from the observation that gearing (i.e., as indicated by RPM at high speeds) is indicative of typical driving characteristics and hence cluster membership. Mean engine RPM for vehicles with the majority of their operation in cluster 1 (slower cluster) is 1975 rpm while it is 1563 rpm for cluster 2 (faster cluster). The engine RPM inter-quartile range for cluster 1 (slower) is 1704 to 2268 rpm, while it is 1354 to 1710 rpm for cluster 2 (faster). Figure 20 draws out additional relationships between typical gearing and fuel and transmission type. One can see that CNG, Gasoline, and Hybrid vehicles are exclusively in cluster 1 in the Fleet DNA data. Vehicles with manual transmissions are exclusively in cluster 2, while vehicles with automatic transmissions are present in both clusters.

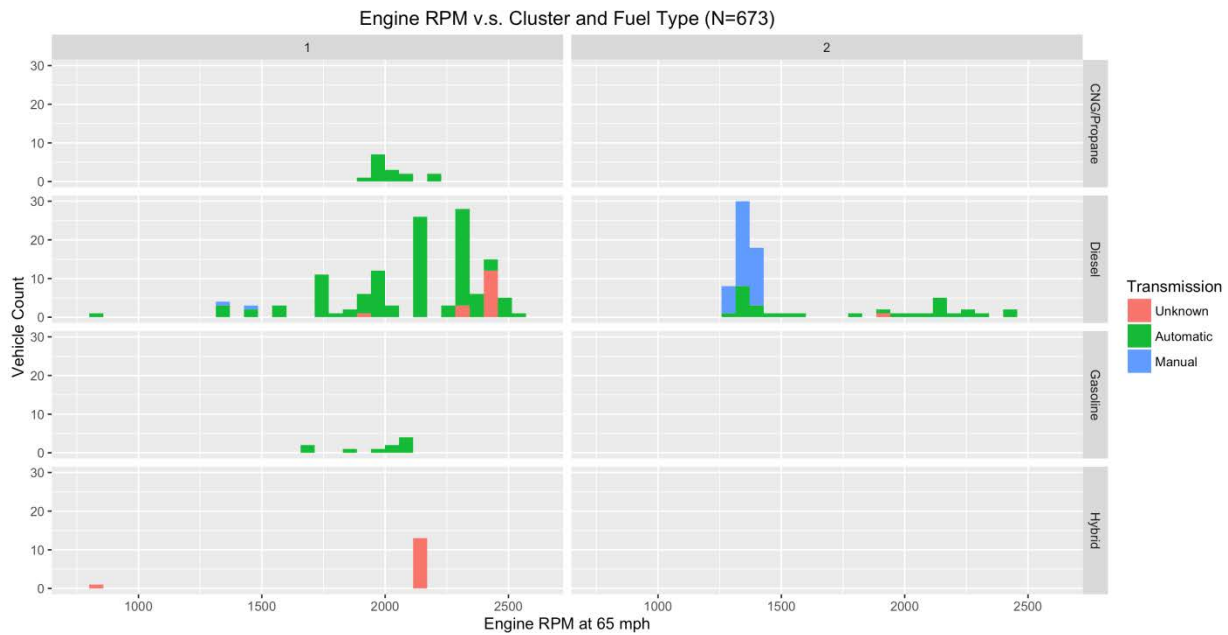
The remainder of this section explores using these features as parameters in a logistic regression to predict whether a vehicle will likely belong to cluster 1 (slower cluster) or cluster 2 (faster cluster). The three category case is not considered in this analysis, as vehicles whose operation lies in between these two clusters are not readily differentiable based on RPM data alone.



**Figure 18: Engine RPM data available for different vocational categories. Transit vehicles are under-represented in the Engine RPM dataset. Grey points are vehicles that do not have RPM data available in the Fleet DNA database. The remainder of vehicles are colored according to their RPM at 65 mph.**



**Figure 19: Engine RPM has a clear bimodality which corresponds with two-cluster assignments. In this boxplot, the median is given as a horizontal line and interquartile range is described by the height of the boxes. From this plot, it is clear that the distribution (both central tendency and spread) in RPM between clusters 1 and 2 is different.**



**Figure 20: Distribution of Engine RPM characteristics for different types of vehicles. Vehicles with manual transmissions are exclusively in cluster 2 (Fast), with lower average gearing. Gasoline, CNG, and Hybrid vehicles are exclusively in cluster 1 (Slow) with similar central tendency to (mean/median) Diesel vehicles in the same cluster.**

### 3.1 Logistic Regression Model

Multiple logistic regression seeks to classify input data into two (binomial) outcomes by using a weighted combination of fitted logistic curves [19]. Each logistic curve describes the relationship between a single attribute and the probability of being in a particular category. In this application, information about a vehicle’s transmission, fuel type, and RPM at 65 mph was used to classify vehicles as being in one of two characteristic clusters.

Model evaluation was performed using standard k-fold cross-validation where 25% of data were withheld while 75% of data were used for fitting. The fitted model was used to predict the 25% of data that are withheld. This process was repeated 100 times and average model error rates were analyzed. Fitted parameters were obtained using a generalized linear least squares optimization using a binomial model.

#### 3.1.1 Full Model

Using all available features, the model was able to achieve 89% accuracy (N = 240) at predicting whether a vehicle is likely to be placed in cluster 1 or cluster 2. The accuracy per cluster is not quite symmetric since the training data is skewed towards cluster 1. Accuracy at correctly assigning vehicles to cluster 1 is 94%, while the accuracy at placing vehicles correctly in cluster 2 is 75%.

This model shows the strongest relationship between RPM at 65 mph and cluster mode, which has a statistically significant relationship to the outcome variable (p-value  $\ll 0.05$ , Wald chi-squared test [20]). Figure 21 shows the relationship between RPM at 65 mph and cluster assignment. Transmission type (i.e., Automatic vs. Manual) also appears to have a statistically significant relationship. The other variables do not show as strong a relationship with the cluster outcomes (i.e., a 95% Confidence Interval (CI) on their effect crosses zero), but still improve overall model accuracy, so we keep them here because they may be applicable to data with a different composition (the Vocational Fleet DNA database has very few Gasoline and Hybrid vehicles compared to Conventional Diesel vehicles). Table 8 gives fitted coefficients for the full model:

$$\Pr(\text{cluster 2} \mid \text{rpm65} \ \& \ \text{trauto} \ \& \ \text{dtconv} \ \& \ \text{fuelgas}) = \sigma(t) = \frac{1}{1 + e^{-t}}$$

$$t = 5.295 - 0.003415 * \text{rpm65} - 1.231 * \text{trauto} - 10.623 * \text{fuelgas} + 1.196 * \text{dtconv}$$

where:

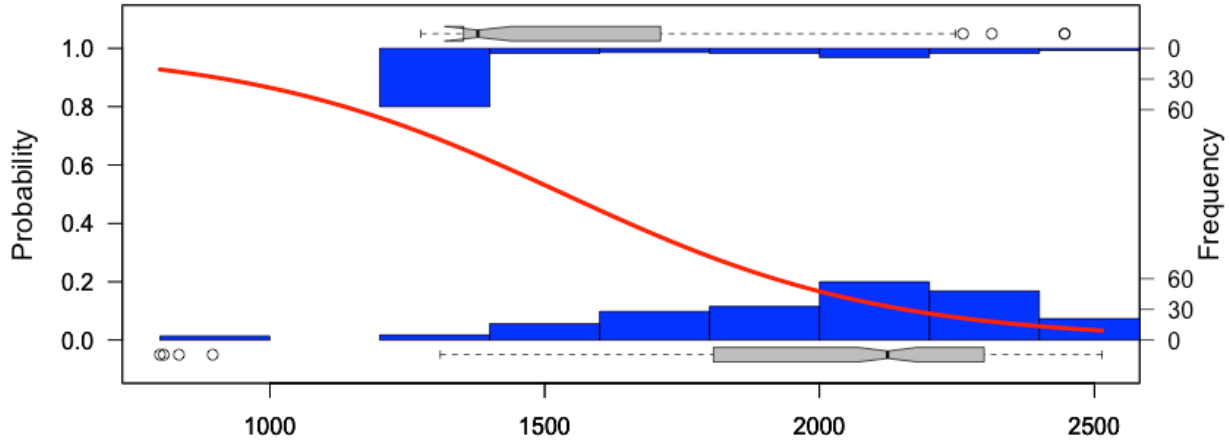
t is the logistic exponent

$\sigma(t)$  is the probability of being in cluster 2 (faster cluster) given the evidence

rpm65 is the RPM of the vehicle at 65 mph

trauto is True if the vehicle has an automatic transmission and is False otherwise

dtconv is True if the vehicle has a conventional drivetrain and is False otherwise  
 fuelgas is True if the vehicle has a Gasoline engine and is False otherwise



**Figure 21: Logistic histogram showing the relationship between RPM at 65 mph and the probability of being in cluster 2 (Fast). The blue bars show the actual distribution of RPM measurements in each cluster. The red curve shows the shape of the logistic function. This analysis is only for two clusters, hence there is no middle cluster.**

**Table 8. Fitted Coefficients for Full Logistic Regression (Fleet DNA Vocational Vehicles Only)  
 N=240**

	Estimated Coefficient (X)	Coefficient 95% CI	Wald Chi-Square p-value
(Intercept)	5.295	2.917 - 7.842	
RPM at 65 mph (rpm65)	-0.003415	-0.005 -0.002	< 0.01
Transmission = Automatic (trauto)	-1.231	-2.069 - 0.396	< 0.01
Fuel = Gasoline (fuelgas)	-10.623	NA - 76.432	0.99
Drivetrain = Conventional (dtconv)	1.196	-0.033 - 2.679	0.08



### 3.1.2 Univariate Diesel Model

To streamline the process of prediction, a simplified logistic regression for Diesel vehicles was built using only RPM at 65 mph to determine cluster membership. This model achieved 81% accuracy (N = 260). Accuracy is similarly asymmetric depending on cluster: accuracy for cluster 1 is 87% while cluster 2 is 70%. Table 9 provides fitted parameters for this model. The probability of membership in cluster 2 can be calculated as follows:

$$\Pr(\text{cluster 2} \mid \text{rpm65}) = \sigma(t) = \frac{1}{1 + e^{-t}}$$

$$t = 5.277 - 0.00328 * \text{rpm65}$$

**Table 9. Fitted Coefficients for Univariate Diesel-only Logistic Regression (Fleet DNA Vocational Vehicles Only) N=260**

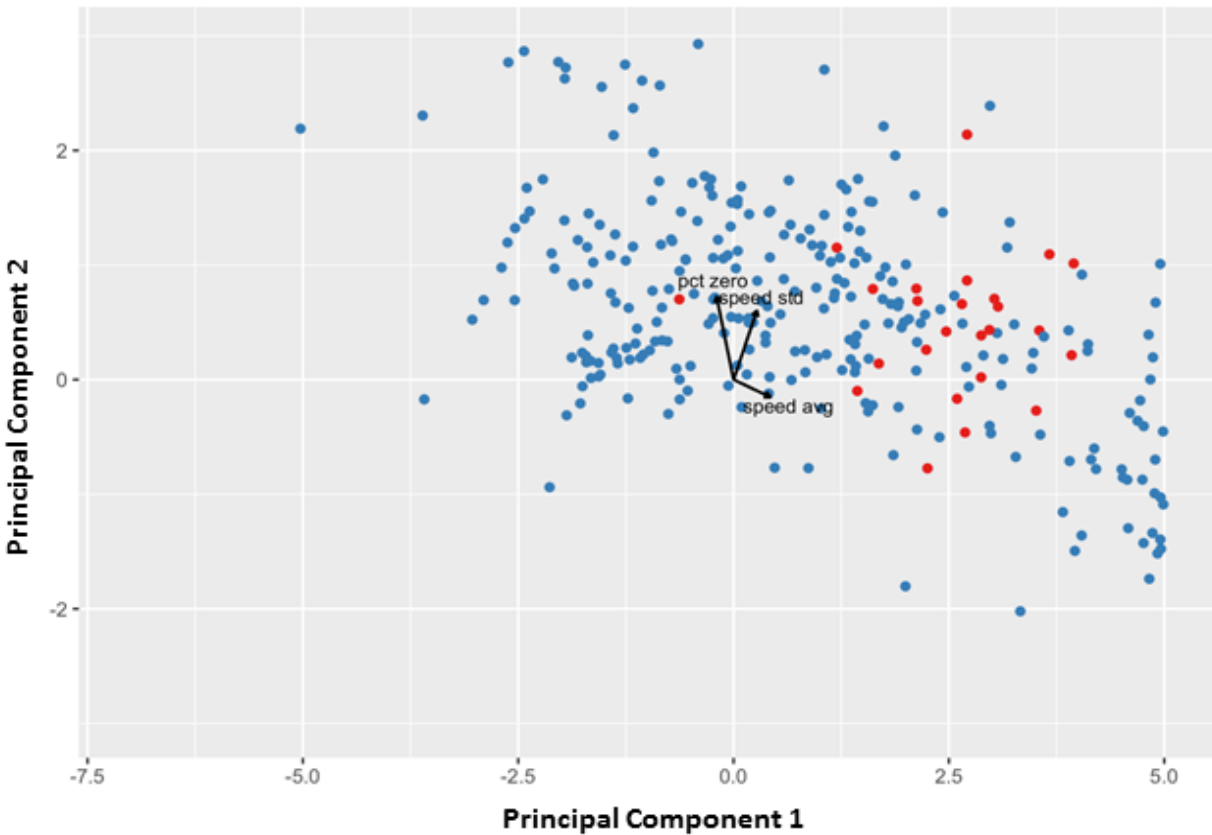
	Estimated Coefficient (X)	Coefficient 95% CI	Wald Chi-Square p-value
(Intercept)	5.177	3.733 - 6.731	
RPM at 65 mph (rpm65)	-0.003280	-0.004 -0.002	< 0.01

### 3.2 Validation and Error Process

When developing the predictive models, it was important to understand sources of error in the logistic regression classification. In particular, the process of characterizing errors as either systematic or random in nature is of particular value, as systematic errors can be strategically corrected. In this analysis, researchers made use of the full logistic regression model. Figure 22 shows the location of the 25 vehicles that account for 91% of misclassifications during the 100-fold cross validation. One can see that these vehicles' drive characteristics are roughly in the middle.

Cluster 2 (fast) vehicles that are consistently misclassified into cluster 1 are largely automatic transmission diesel delivery vehicles with conventional drivetrains and high RPM at 65mph (mean RPM 2138). It stands to reason that these vehicles would be misclassified since according to the Fleet DNA data, vehicles with a high mean RPM at 65mph are generally used for slow applications (cluster 1). Similarly, cluster 2 vehicles more often have manual transmissions than observed in this group of 25 vehicles.

## Vehicle Clustering - 8 Metrics (2 Clusters) – Classification Failures



**Figure 22: Locations (red points) of 40 most often misclassified vehicles (in 100-fold cross validations) using the full logistic regression model.**

Cluster 1 (slow) vehicles that are consistently misclassified into cluster 2 are largely conventional diesel vehicles with automatic transmissions used for special purposes (e.g., tanker, concrete, drayage, dump truck, etc.). These vehicles have a gearing that is consistent with faster moving (cluster 2) vehicles (mean RPM at 65mph = 1354).

Based on this analysis, it is observed that misclassification occurs in two scenarios:

1. Vehicles designed for one type of application (e.g., local deliveries) are used consistently outside of their intended use (e.g., highway transit). These vehicles are generally delivery vehicles (Beverage delivery, Food delivery, Concrete, or Linen delivery). Since these vehicles, whose design is inconsistent with their use, is a small fraction of overall vehicles, and the errors cannot be addressed systematically without additional knowledge (i.e., vocational categorization).
2. Vehicles have multi-modal usage, with a fraction of time spent in applications consistent with cluster 1 and a fraction of time spent in applications consistent with cluster 2. These vehicles are generally special purpose vehicles (e.g., drayage, utility, warehouse delivery) as well as parcel delivery vehicles. Misclassified drayage trucks, for instance, spend approximately 63% of their drive cycles operating in cluster 1 (presumably at port) and 37% of their drive cycles operating in cluster 2 (presumably doing highway transfers).

Since these vehicles are geared for slower use, the model misclassifies them into cluster 1 when they may actually spend a larger fraction of days operating in a way consistent with vehicles in cluster 2.

Both misclassifications could be addressed using additional information about the vehicle use at the time of testing (e.g., vocational information), since the limited available information (gearing and chassis) is insufficient to identify these special cases in practice. Nevertheless, they are a small percentage of the overall vehicles (approximately 10%), and from this work it is observed that vehicle chassis features (particularly engine RPM) are a generally reliable predictor of vehicle drive characteristics.

## 4 Development of Representative Transient Drive Cycle Component

Drive cycle has been shown to dramatically impact fuel consumption and emissions production for medium- and heavy-duty vehicles [21-26]. As such, controlled laboratory test procedures representative of real world operations conditions are necessary to accurately quantify these parameters. Numerous approaches have been developed to generate representative drive cycles from real world driving data [27-31], including NREL's DRIVE tool. Having already completed segmentation of the U.S. medium- and heavy-duty commercial vehicle data into a collection of three distinct clusters based on a multivariate drive cycle clustering analysis and development of a logistic model to predict cluster participation, NREL researchers then applied the results of the clustering analysis towards the development of both a representative low and high speed transient drive cycle representative of medium- and heavy-duty vocational vehicle use. Deploying NREL's DRIVE Tool within its High Performance Computing environment, researchers condensed thousands of hours of on road driving data down into representative speed-time drive cycles approximately 12 and 20 minutes in duration.

### 4.1 Generation of Representative Low Speed Transient Component Using DRIVE

To develop the representative low speed transient cycle, an iterative method was deployed where drive cycles from each of the 3 clusters described in section 3 were ranked by root mean squared (RMS) distance from medoid calculated using the 8 key metrics and then fed through DRIVE to generate candidate representative cycles which were then compared to the statistics for the low speed cluster to identify an ideal representative cycle. The procedure was as follows:

1. The top 50 cycles from each of the 3 clusters were fed into DRIVE Space to generate candidate representative cycles. The top 50 cycles were chosen as a means of limiting overall computation time while still capturing the overall representative behavior of the cluster. This decision also allowed researchers to explore and optimize the generated drive cycles by adjusting final drive cycle duration.
2. Having identified the top 150 total drive cycles most representative of their respective clusters, a variety of input parameters such as the desired and minimum cycle durations were then adjusted to generate over 100 unique cycles.
3. The representative cycles were then compared to the average values for the top 50 cycles in the low speed cluster using a non-weighted least squares approach using the eight drive cycle metrics from the clustering analysis described in section 3. The low speed cluster was chosen as the ideal target due to its inherent low speed transient behavior. This is especially true when compared to the mixed mode and high speed clusters.

Several additional constraints were applied when running the DRIVE tool including:

- Targeted cycle duration of 668 seconds to match the duration of the California Air Resources Board (CARB) Heavy Heavy-Duty Diesel Truck (HHDDT) Transient Cycle.
- A minimum allowable cycle duration of 300 seconds was established to ensure sufficient test duration.

- Exclusion of any drive cycles with a maximum speed in excess of 55 mph to ensure generated drive cycle is representative of low speed transient operation only

To examine the effects of population makeup on the resultant transient cycle, the drive cycle generation process was repeated a total of 3 times: one for each of the unique populations explored in the cluster analysis. The full Fleet DNA drive cycle population, the MOVES resampled drive cycle population, and the MOVES subcategory resampled drive cycle data were all examined – see section 3 for more details on each sample set. This was done to examine the sensitivity of representative transient cycle generation to source population and minimize the impact of any potential biasing as a result of source data composition (i.e. more school bus data than nationally representative). It was found that there were minimal differences between the weighting, and as such the MOVES resampled population was chosen given it produced the most representative drive cycle of the three populations. Additional detailed information regarding resampling and its impact on cycle generation can be found in the Appendix.

The transient cycle shown in Figure 23 was developed using 150 drive cycles drawn from the MOVES resampled Fleet DNA database following the procedure described previously.

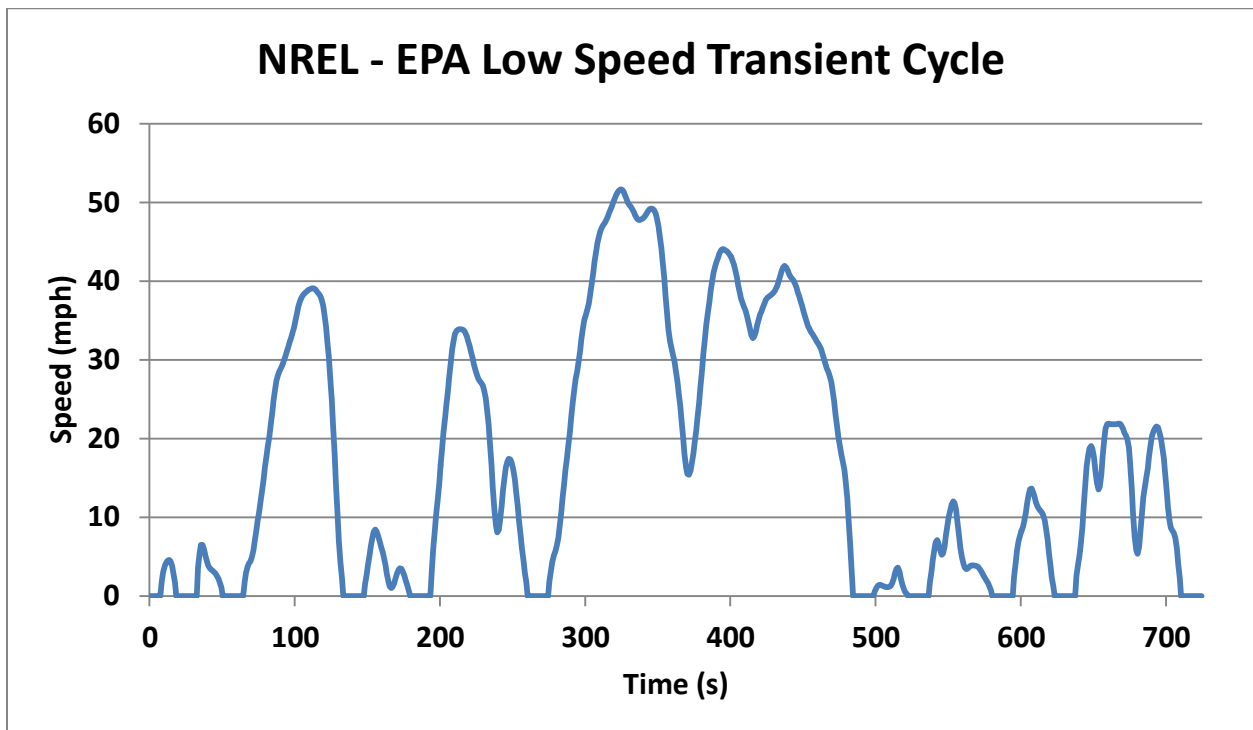


Figure 23: Speed-time trace for representative low speed EPA transient cycle

Key statistics for the final low speed transient cycle include:

- 724 seconds in duration
- Total of 10 microtrips
- Maximum speed of ~52 mph
- Average driving speed of ~21 mph
- ~22% of total cycle duration is at zero speed
  - The target zero speed duration for the representative transient cycle was less than 24.5%. This value was chosen as it represents the percentage of zero speed time observed from the high speed cluster identified in Section 3 (cluster 3-3). The high speed cluster possesses the lowest amount of zero speed time of all of the clusters, thus it was used as a lower bound when developing the low speed transient cycle because for any of the other driving conditions one can simply add an additional idle time segment to achieve an appropriate overall idle time weighting.

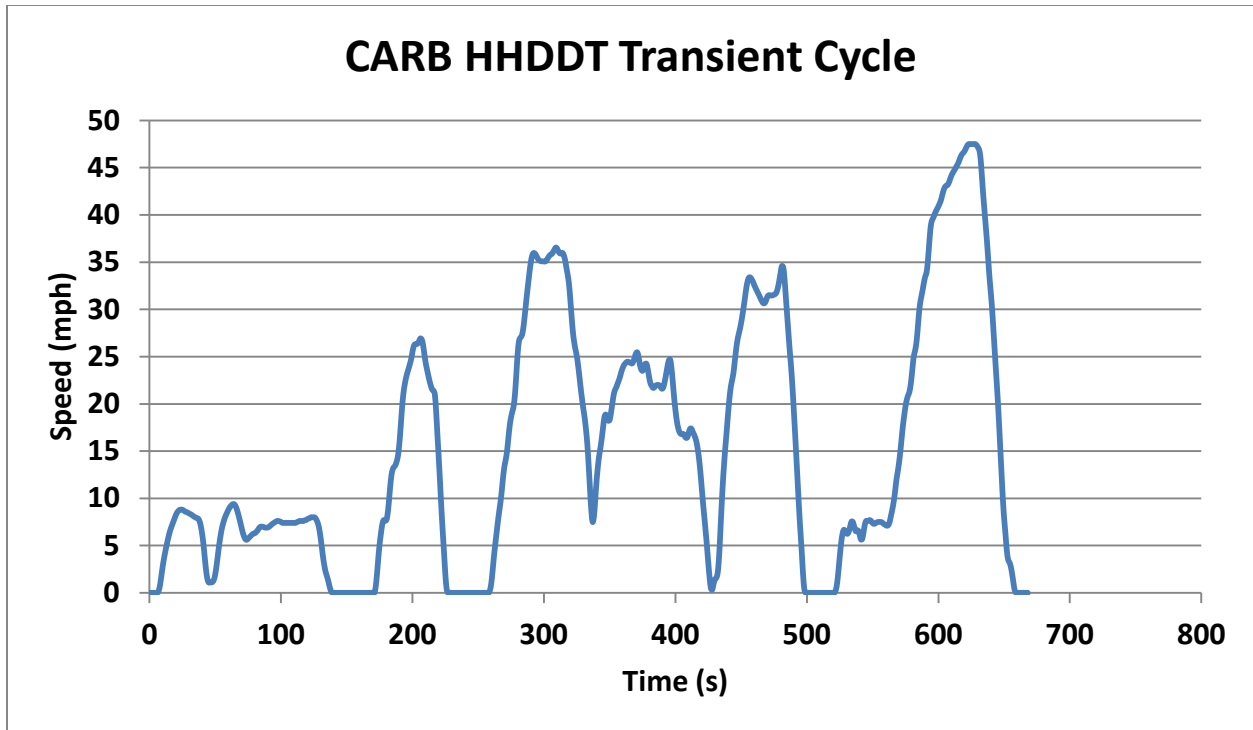
## **4.2 Comparison of DRIVE Generated Transient Cycles to Existing Standard Cycles**

Having developed a set of representative transient test cycles, the next step in the analysis was to compare the drive cycle characteristics of the generated drive cycles to establish standard test cycles from around the world. By comparing the representative drive cycles to existing standard cycles it was possible to evaluate the potential for existing standard cycles to characterize the driving behaviors demonstrated by each of the identified vehicle segment clusters. A series of four standard test cycles were chosen for comparison, the U.S. CARB Heavy Heavy-Duty Diesel Truck (CARB HHDDT) Transient, the International World Harmonized Vehicle Cycle (WHVC), the European Test Cycle (ETC) also known as the FIGE Cycle, and the Japanese JE05 cycle [32-35]. In sections 4.3.1 illustrations and brief descriptions of these standard cycles are provided, followed by more in-depth comparisons of the standard cycles to the representative cycles generated by DRIVE.

### **4.2.1 Standard Test Cycles Selected for Comparison**

#### **4.2.1.1 California Air Resources Board Heavy Heavy-Duty Diesel Truck Transient Cycle**

The CARB HHDDT Transient Cycle is a chassis dynamometer test cycle developed by CARB in collaboration with West Virginia University using real world data captured from trucks operating in the state of California. It is the current transient component defined in EPA's proposed GHG Phase 2 regulations, and is the transient cycle currently used in the Phase 1 regulations.



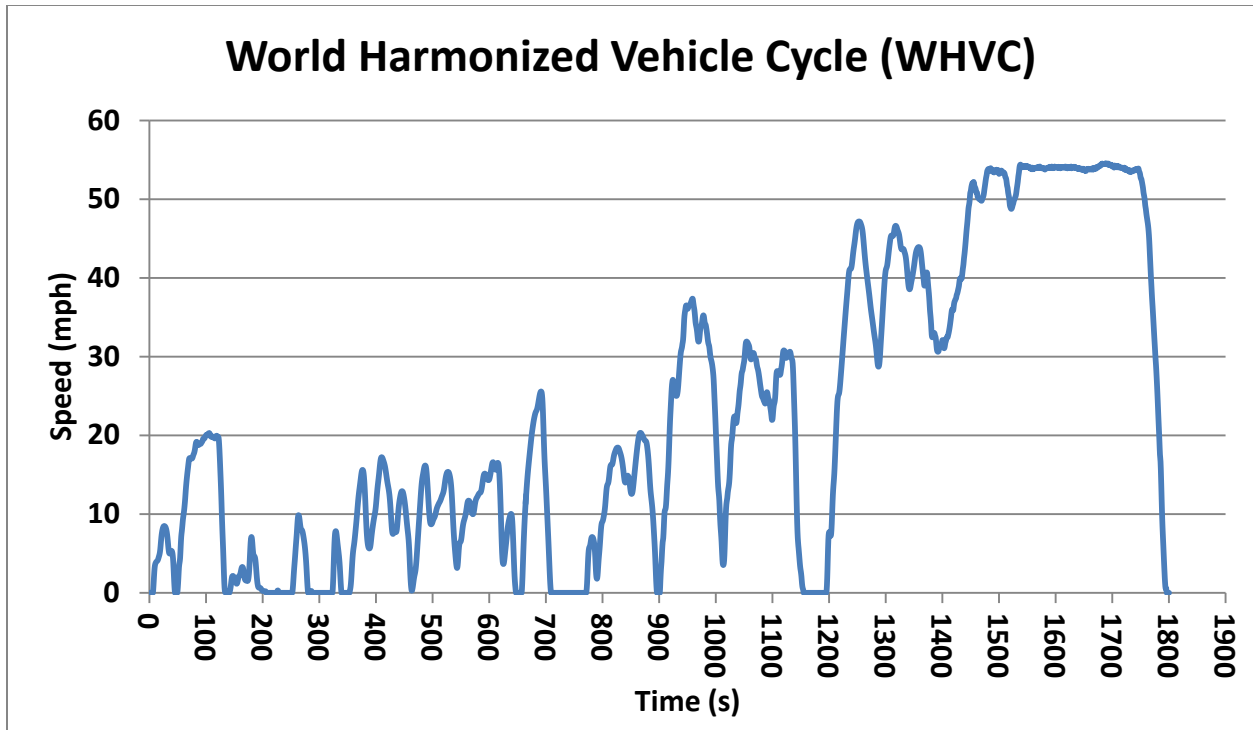
**Figure 24: Speed-time trace for CARB HHDDT transient cycle**

Some key drive cycle characteristics of the CARB HHDDT include:

- 668 seconds in duration
- Contains five microtrips
- Maximum driving speed of 47.5 mph
- 2.85 miles in length.

#### *4.2.1.2 World Harmonized Vehicle Cycle (WHVC)*

The WHVC cycle is a chassis dynamometer test cycle developed based on the same set of data used for the development of the World Harmonized Transient Cycle (WHTC). The first 900 seconds of the cycle are representative of urban driving, while the following 481- and 419-second sections represent rural and highway driving, respectively. Shown in Figure 25 is the speed-time trace for the WHVC cycle.



**Figure 25: Speed-time trace for WHVC cycle**

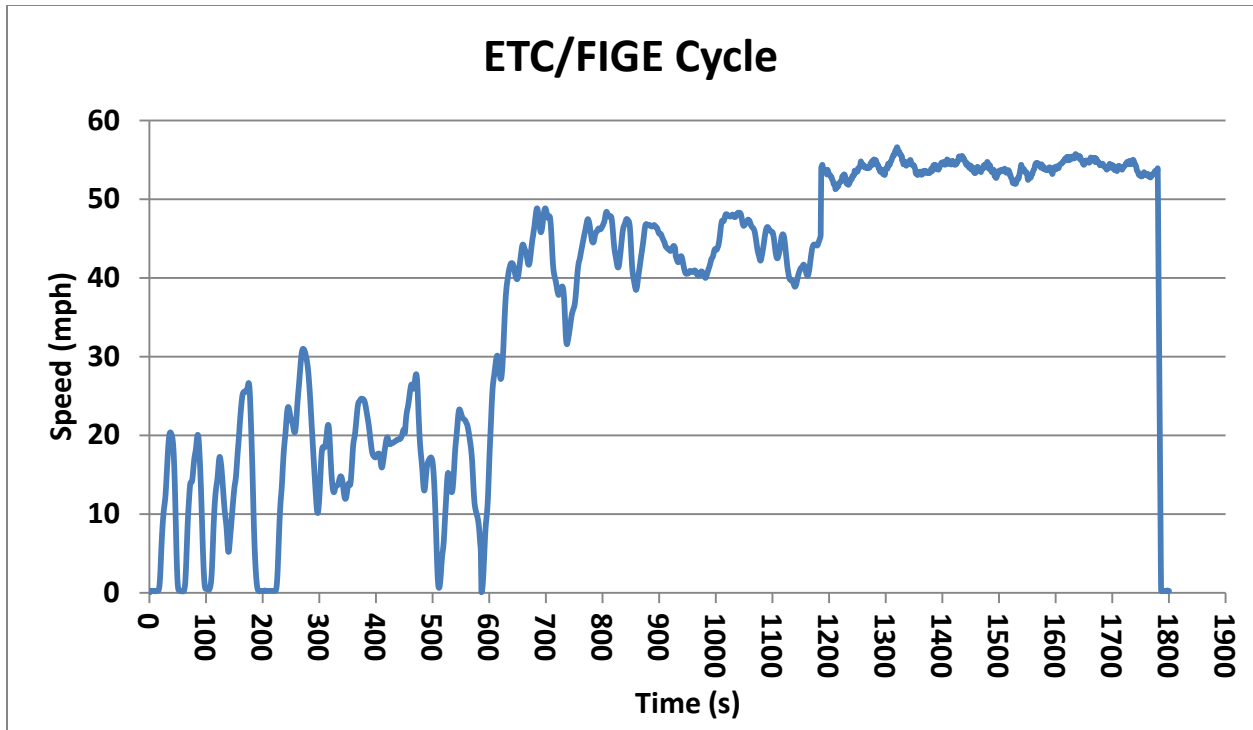
Some key drive cycle characteristics of the WHVC include:

- 1,800 seconds in duration
- Contains 11 microtrips
- Maximum driving speed of 54.56 mph
- 11.22 miles in length.

#### 4.2.1.3 European Transient Cycle (ETC)/FIGE Cycle

The ETC cycle was developed by the former FIGE Institute in Aachen, Germany based on real world drive cycle data captured from in-use heavy duty vehicles. The ETC cycle consists of three segments, representative of urban, rural, and highway driving. With a total duration of 1,800 seconds, each segment represents 1/3 of the total duration of the overall cycle. The first segment represents urban driving conditions, the second segment represents rural driving, and the final segment represents highway driving.





**Figure 26: Speed-time trace for ETC FIGE cycle**

Some key drive cycle characteristics of the ETC/FIGE include:

- 1,800 seconds in duration
- Contains 1 microtrip
  - Note: The speed trace specified in the report calls out no zero speed time, but rather series of extremely low speed creep in the range of 0.4 kph. This means that the ETC/FIGE cycle is 1 single microtrip as defined, however in reality the cycle is composed of three distinct driving sections each 600 seconds in duration. For more accurate chassis and simulation based testing, it is recommended that the low speed creep be replaced with zero speed driving conditions which would then produce a test cycle consisting of 6 distinct microtrips of which one represents high speed cruising, and the rest mix mode and transient driving operation.
- Maximum driving speed of 56.61 mph
- 18.33 miles in length.

#### 4.2.1.4 JE05 Cycle

In 2005, new Japanese emissions standards resulted in the development and introduction of the JE05 emission test cycle. The JE05 was developed using data captured from both diesel and gasoline commercial vehicles operating in Tokyo. The test cycle was developed for use by heavy vehicles with gross vehicle weights (GVW) above 3,500 kg or approximately 7700 lbs. just below the GVWR of 8,500 lbs. for a class 2b vehicle in the United States.

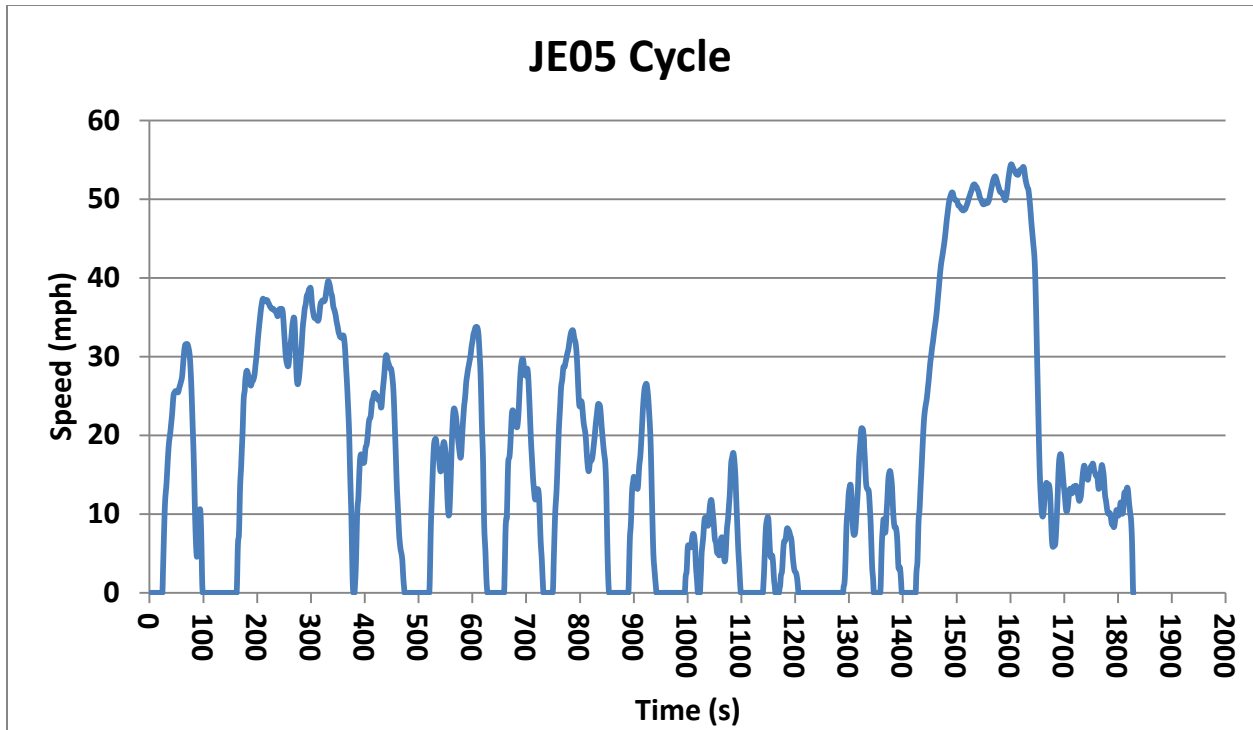


Figure 27: Speed-time trace for JP JE05 cycle

Some key drive cycle characteristics of the JE05 include:

- 1,829 seconds in duration
- Contains 14 microtrips
- Maximum driving speed of 54.43 mph
- 8.63 miles in length.

#### 4.2.2 Drive Cycle Characterization and Comparison

To benchmark the representative transient cycles developed from the Fleet DNA data, a detailed drive cycle characterization was performed using DRIVE. The eight metrics used in the drive cycle segmentation and clustering analysis were used as the basis to compare each of the different drive cycles. Shown in Table 10 are the results of the drive cycle characterization.

**Table 10. Comparison of Standard Cycle Drive Cycle Metrics**

Cluster	Aerodynamic Speed (ft/s)	Characteristic Acceleration (ft/s <sup>2</sup> )	Percent of mileage accumulated below 55 mph	Percentage of Cycle Time at Zero Speed	Number of Stops Per Mile	Average Driving Speed (mph)	Maximum Driving Speed (mph)	Standard Deviation of Driving Speed (mph)
CARB HHDDT Transient	43.36	0.49	100.00	15.57	1.40	18.20	47.50	13.38
WHVC	61.00	0.28	100.00	13.61	1.07	25.98	54.56	19.03
JE05	51.31	0.40	100.00	25.19	1.62	22.70	54.43	15.94
ETC FIGE	68.71	0.23	94.46	0.00	0.00	36.65	56.61	17.83
NREL - EPA Transient Low	50.99	0.59	100.00	21.69	2.98	21.31	51.67	16.16
FDNA 1 - Average	50.59	0.55	96.45	47.30	1.71	22.62	62.39	13.74

From Table 10, one can see that the NREL-EPA low speed transient generally matches the statistics of average for the top 50 drive cycles as ranked by RMS distance from the medoid selected from slow speed urban cluster number 1 (FDNA 1 – Average) and the CARB HHDDT Transient. The cycles are in very close agreement in both Aerodynamic Speed and Characteristic acceleration, with a notable difference in the percentage of total cycle time spent at zero speed, and the number of stops per mile. The higher values for stops per mile is expected when compared to the WHVC, JE05, and ETC cycles as they are composite cycles containing low and high speed driving operation while the NREL-EPA low speed transient is focused solely on low speed operation. The difference between the generated cycle and the CARB HHDDT illustrate just how aggressive and “stop and go” medium- and heavy-duty transient driving really is based on real world driving data from Fleet DNA. The representative transient stops over twice as frequently as the CARB transient and possesses a higher characteristic acceleration indicative of a greater proportion of tractive energy and thus fuel/emissions being associated with vehicle acceleration.

Beyond comparisons of driving aggressiveness, it is important to acknowledge that the representative low speed transient cycle developed by NREL using the Fleet DNA dataset possesses higher average and maximum driving speeds when compared to the CARB HHDDT transient. The CARB transient achieves a maximum driving speed of 47.5 mph where the NREL transient reaches 51.67 mph and the representative low speed driving cluster from the top 50 cycles average achieves a maximum speed of 62.39 mph. This is important as the final test procedure consisting of a transient, 55 mph cruise, and 65 mph cruise could be improved through the inclusion of a transient component possessing speeds in excess of 47 mph as indicated by the representative cluster average.

The difference in zero speed time as a percent of overall cycle time can be adjusted with post processing of test data, or the inclusion of additional idle time allowing the transient to be adjusted to match the idle characteristics of any of the clusters identified during the segmentation analysis.

### 4.2.3 Analysis of Mileage Accumulation by Vehicle Speed

As part of the drive cycle analysis, a detailed characterization of mileage accumulation by vehicle speed bin was performed to further aid in understanding the driving behavior represented by each of the cycles as well as the relative need for transient speed operation between 45 and 55 mph which is missing from the current EPA GHG test cycles. The mileage accumulation for each of the drive cycles examined can be found in Table 11 and Table 12, where mileage is accumulated in 5 mph speed bin increments as a percentage of total miles traveled over the cycle. By splitting the data into 5 mph speed bins, researchers were able to examine transient speed mileage accumulation (<45 mph), rural (45-55 mph), and highway type (55 mph) driving behavior at a finer level of detail than could be described by any single high level drive cycle metric.

**Table 11. Sub-45 mph Mileage Accumulation Comparison for Standard Cycles**

<b>Cycle</b>	<b>0+ - 5 mph distance (%)</b>	<b>5+ - 10 mph distance (%)</b>	<b>10+ - 15 mph distance (%)</b>	<b>15+ - 20 mph distance (%)</b>	<b>20+ - 25 mph distance (%)</b>	<b>25+ - 30 mph distance (%)</b>	<b>30+ - 35 mph distance (%)</b>	<b>35+ - 40 mph distance (%)</b>	<b>40+ - 45 mph distance (%)</b>
NREL - EPA Transient Low	2.66	4.40	5.42	7.08	8.37	8.30	15.15	18.69	12.55
FDNA 1 - Average	1.39	4.35	7.11	8.90	11.05	12.95	15.15	14.06	10.69
CARB HHDDT Transient	1.25	12.72	3.97	9.02	18.20	10.20	15.72	11.96	8.29
WHVC	1.07	3.87	6.11	7.47	4.34	6.34	9.30	7.42	8.09
JE05	0.97	4.68	8.98	9.53	9.32	13.53	12.14	12.10	2.06
ETC FIGE	0.19	0.60	2.40	4.24	3.81	2.36	1.31	3.50	17.49

Looking at the data in Table 11, a few trends become apparent. First is that the ETC FIGE cycle does not accumulate much mileage below 40 mph, while other cycles like the CARB HHDDT Transient, the JE05, and the NREL – EPA Transient accumulate a more equally distributed set of mileage at speeds below 45 mph. This makes sense intuitively as the ETC FIGE is a composite cycle with a high speed driving component that significantly weighs mileage accumulation results towards high-way type mileage accumulation. However, beyond the ETC FIGE displaying significant high speed driving, one can see through an examination of the data in the table that the CARB HHDDT accumulates the greatest number of low speed miles in the range between 20-25 mph and 30-35 mph while the NREL-EPA Representative Transient Cycle

mileage accumulation is greatest above 30 mph in agreement with the Fleet DNA cluster 1 average. While the HHDDT Transient appears to be the standard test cycle best matching the averages identified through the segmentation analysis, there are opportunities for improvement by increasing mileage accumulation in the 30-45 mph speed range.

**Table 12. 45+ mph Mileage Accumulation Comparison for Standard Cycles**

<b>Cycle</b>	<b>45+ - 50 mph distance (%)</b>	<b>50+ - 55 mph distance (%)</b>	<b>55+ - 60 mph distance (%)</b>	<b>60+ - 65 mph distance (%)</b>	<b>65+ - 70 mph distance (%)</b>	<b>70+ - mph distance (%)</b>
NREL - EPA Transient Low	12.75	4.64	0.00	0.00	0.00	0.00
FDNA 1 - Average	6.57	4.22	2.52	1.00	0.04	0.00
CARB HHDDT Transient	8.66	0.00	0.00	0.00	0.00	0.00
WHVC	7.28	38.71	0.00	0.00	0.00	0.00
JE05	9.46	17.23	0.00	0.00	0.00	0.00
ETC FIGE	15.41	43.14	5.54	0.00	0.00	0.00

Examining Table 12, one can see that the CARB HHDDT Transient does not accumulate any mileage at speeds in excess of 50 mph. Compare this to the roughly 8% of total mileage accumulated at speeds in excess of 50 mph for Fleet DNA slow speed cluster 1, and it can be seen that there is an opportunity for an improved transient including additional high speed (45-65 mph) transient driving. The NREL-EPA low speed transient cycle accumulates a more appropriate amount mileage in this speed range with a higher than expected mileage accumulation in the 45-50 mph speed bin.

Visualizing the mileage accumulation for slow speed cluster 1, NREL-EPA low speed transient, and CARB HHDDT Transient distributions shown in Table 11 and Table 12, one can see that the cluster mileage accumulated is normally distributed with either a slight skew towards low speed (CARB HHDDT Transient) or high speed (EPA Transient Low) for the test cycles. These results can be seen in Figure 28.

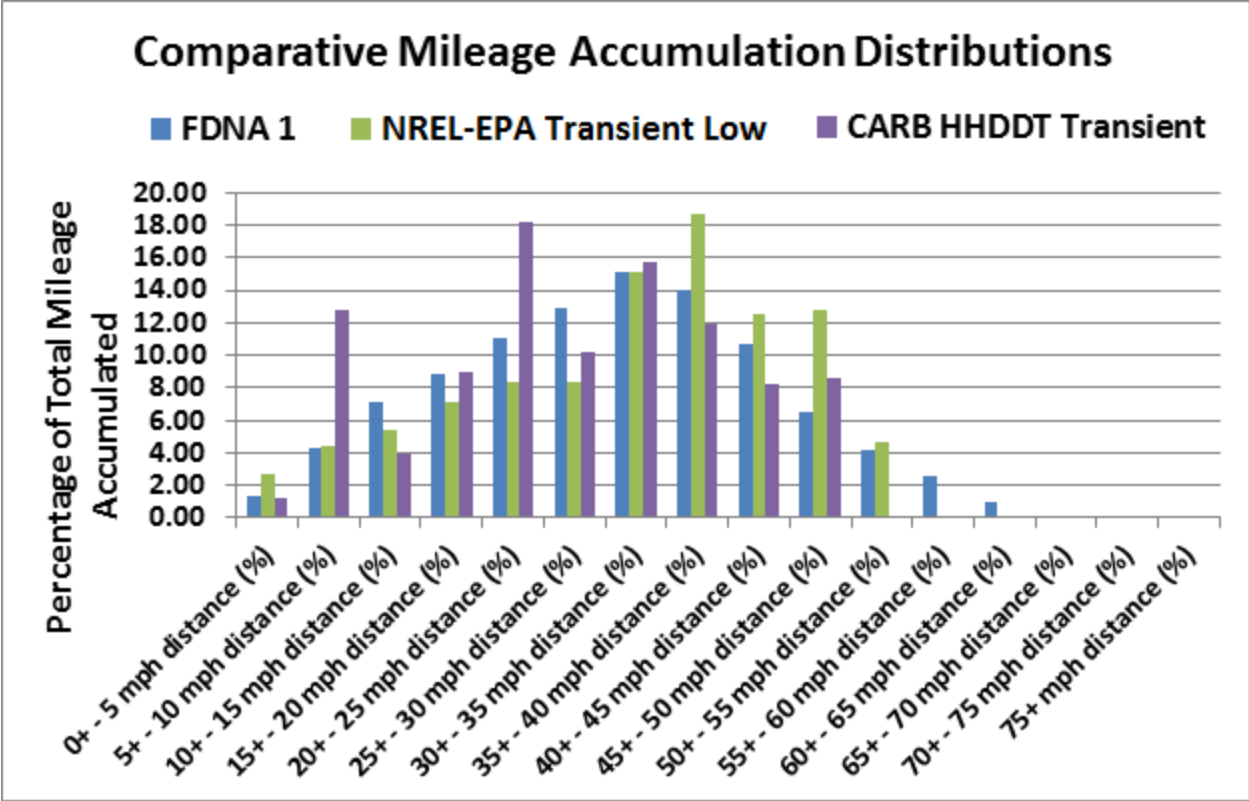


Figure 28: Comparative distribution of mileage accumulation binned by vehicle speed for low speed transient

Figure 28 illustrates one of the limitations of the CARB HHDDT Transient when compared to the Fleet DNA low speed cluster averages, which is no mileage accumulation at speeds in excess of 50 mph.

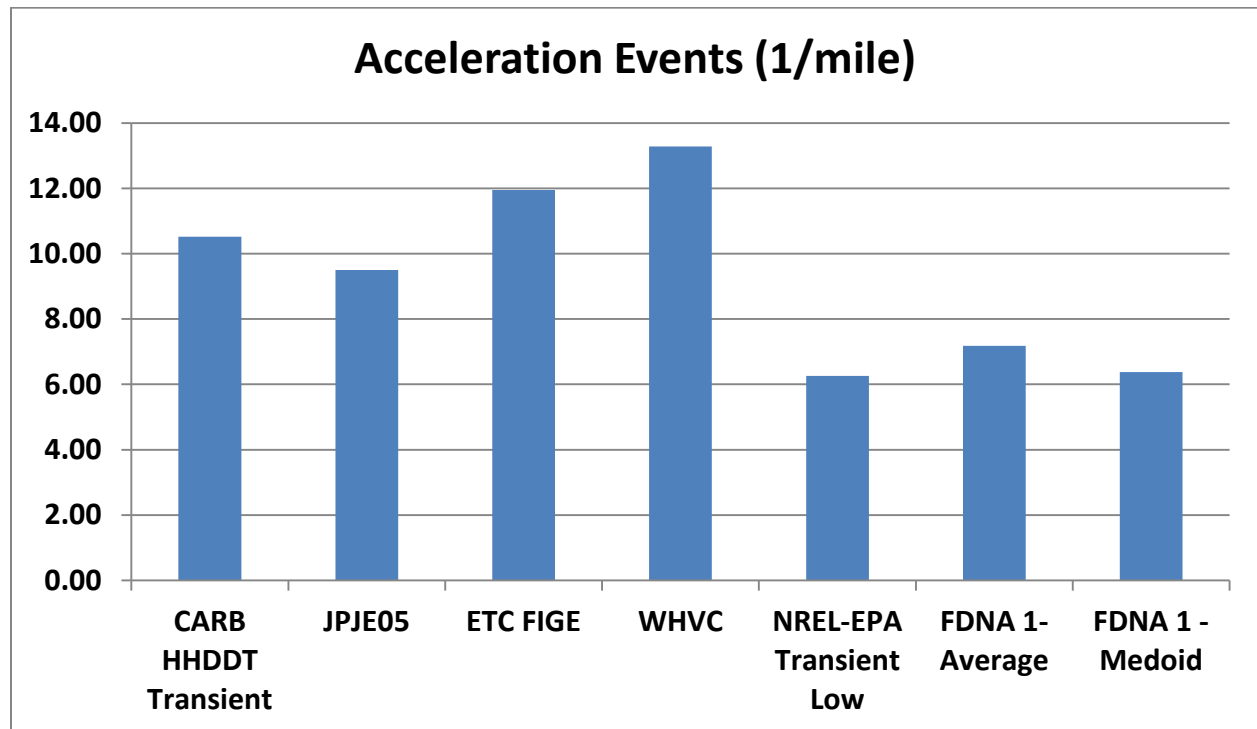
**4.2.4 Comparison of Acceleration/Deceleration Behavior**

A detailed analysis of the acceleration and deceleration behavior of both the representative transient cycles and the standard cycles was performed to benchmark both the NREL-EPA transient cycles and the CARB HHDDT as it relates to slow speed urban cluster 1 identified in the segmentation analysis. Metrics describing the frequency (per mile), and aggressiveness (average and max rates) were calculated and collected in Table 13.

**Table 13. Comparison of Cycle Acceleration Statistics**

Cluster	Acceleration events (1/mile)	Deceleration events (1/mile)	Max accel (ft/s <sup>2</sup> )	Max decel (ft/s <sup>2</sup> )	Avg accel (ft/s <sup>2</sup> )	Avg decel (ft/s <sup>2</sup> )
NREL - EPA Transient Low	6.26	6.26	5.34	-5.66	1.41	-1.34
FDNA 1 - Average	7.18	7.19	9.90	-9.98	1.25	-1.43
CARB HHDDT Transient	10.52	10.17	4.25	-4.06	0.94	-1.24
WHVC	13.28	13.10	3.84	-4.99	0.71	-0.81
JPJE05	9.50	9.38	5.22	-6.01	1.03	-1.08
ETC FIGE	11.95	11.73	12.58	-13.21	0.63	-0.72

Figures 29-34 illustrate the results shown in Table 13. Looking at the results, one can see that both the Fleet DNA low speed cluster data and NREL-EPA representative cycles displayed fewer acceleration/deceleration events per mile than any of the standard chassis test cycles examined. The low-speed transient cluster data (FDNA 1) demonstrated higher maximum accelerations than all cycles except for the ETC FIGE (Figures 31 and 32). The NREL-EPA representative cycle and the low speed cluster data show somewhat higher average acceleration and deceleration rates than all other cycles. Figures 29-34 also illustrate the close agreement over a wide range of metrics between the representative transient cycles and Fleet DNA dataset.



**Figure 29: Comparison of standard cycle acceleration frequency**

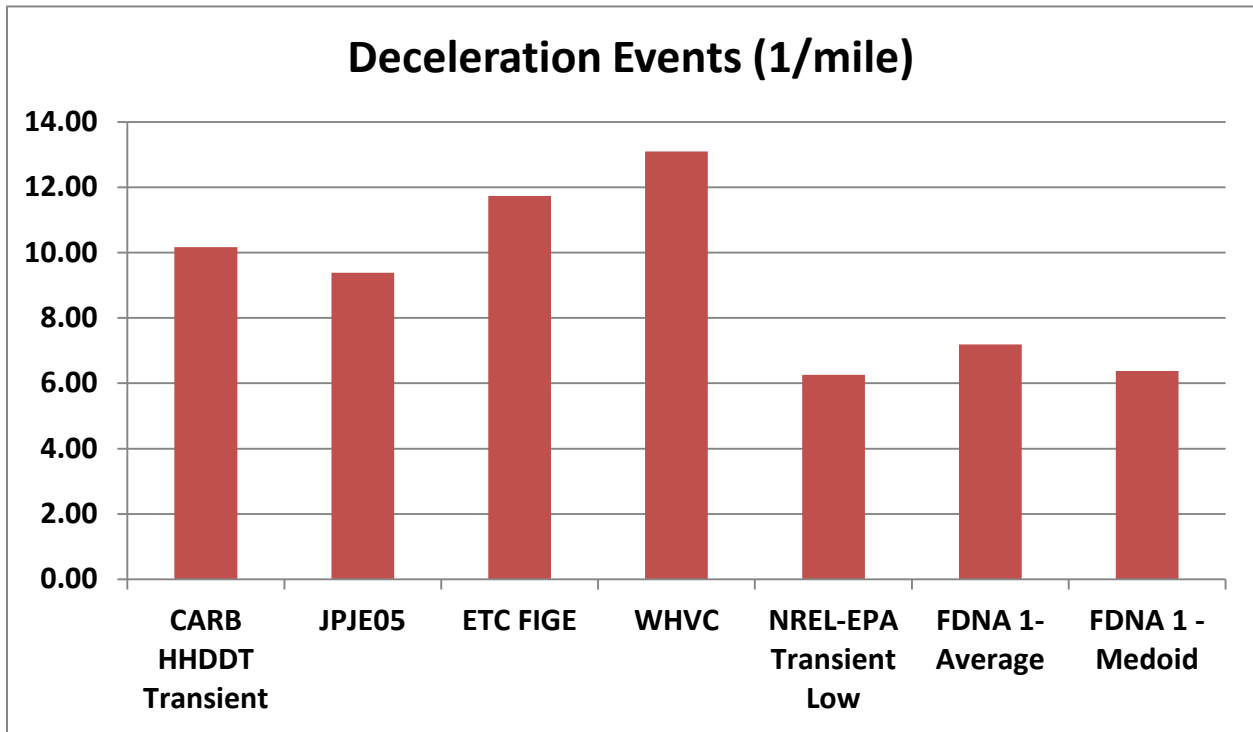


Figure 30: Comparison of standard cycle deceleration frequency

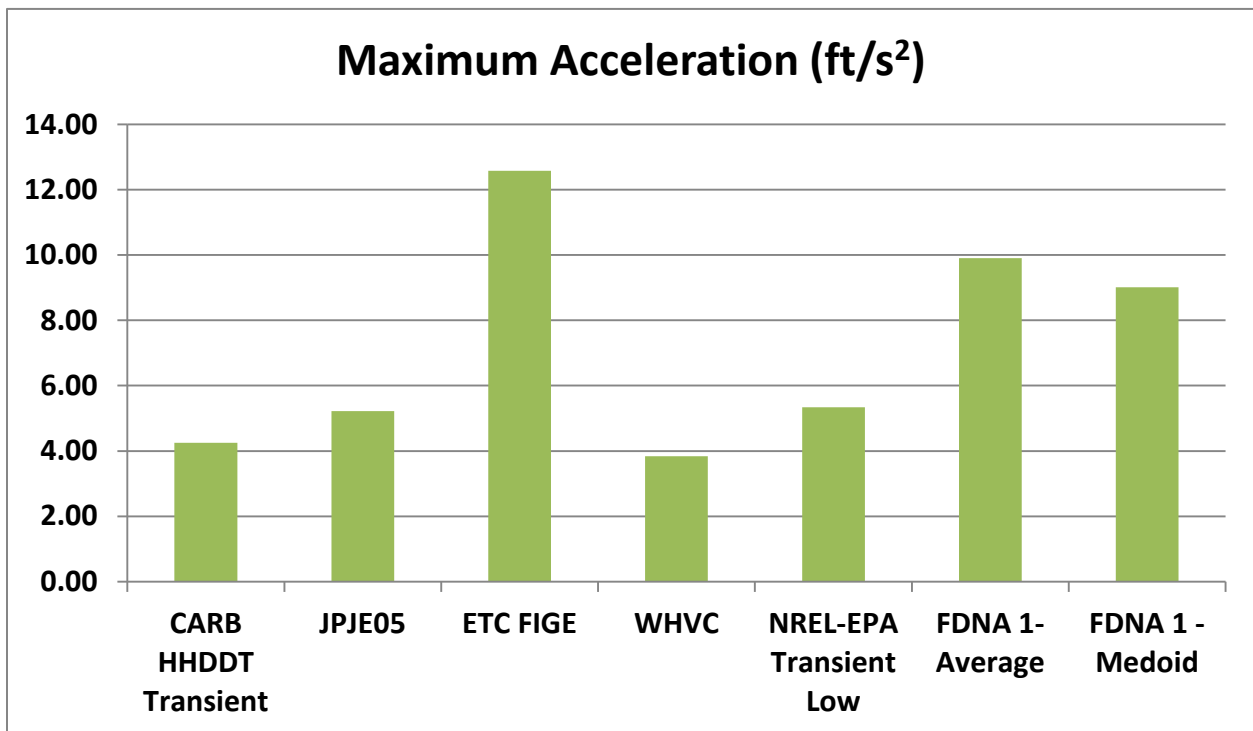


Figure 31: Comparison of standard cycle peak acceleration rates



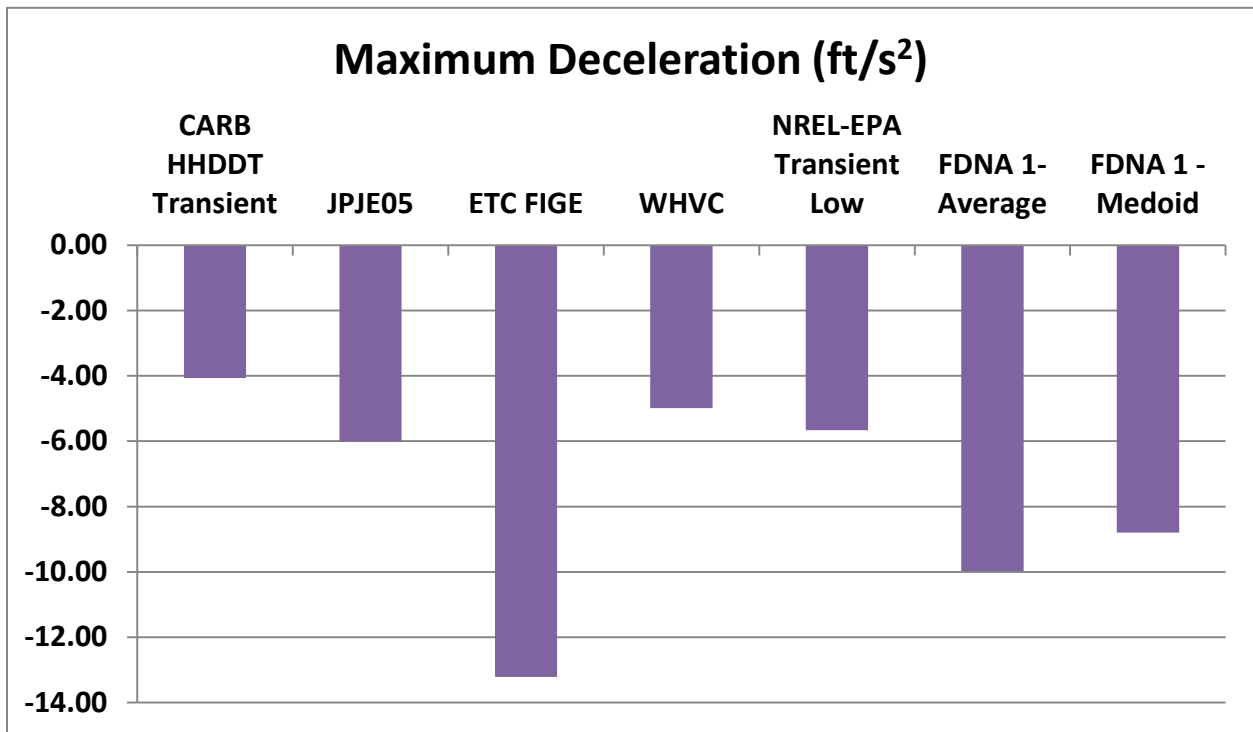


Figure 32: Comparison of standard cycle peak deceleration rates

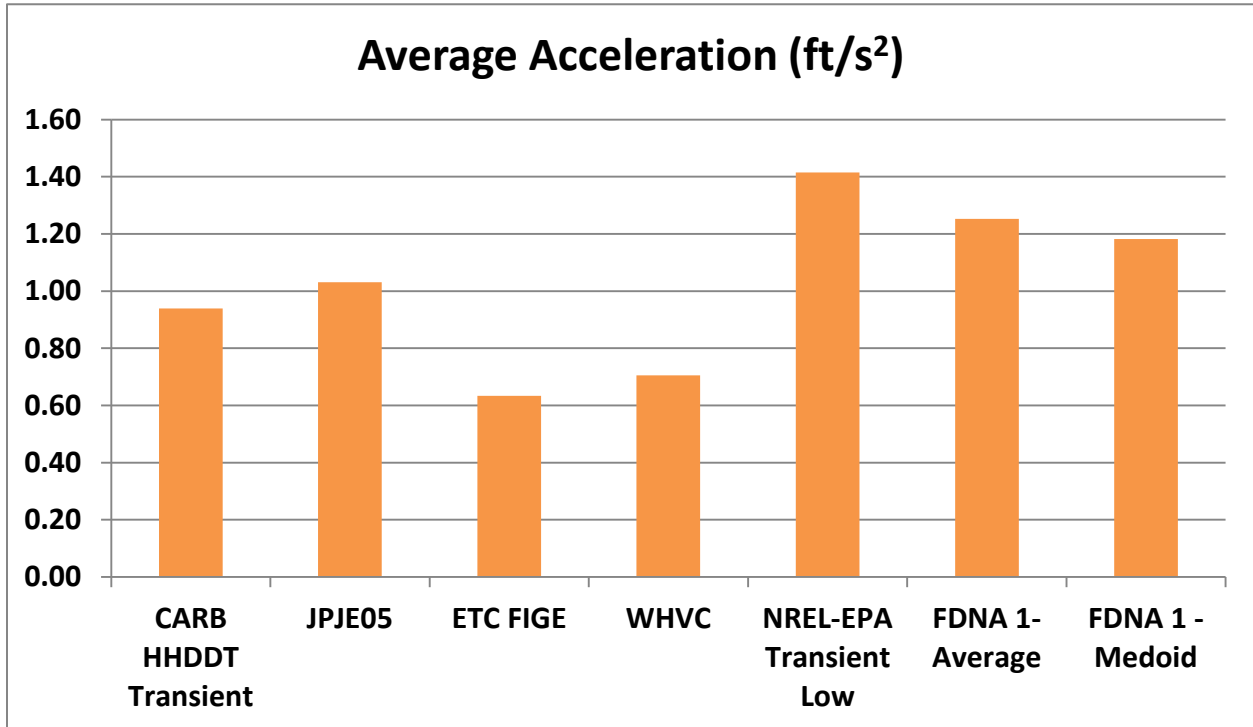


Figure 33: Comparison of standard cycle average acceleration rates

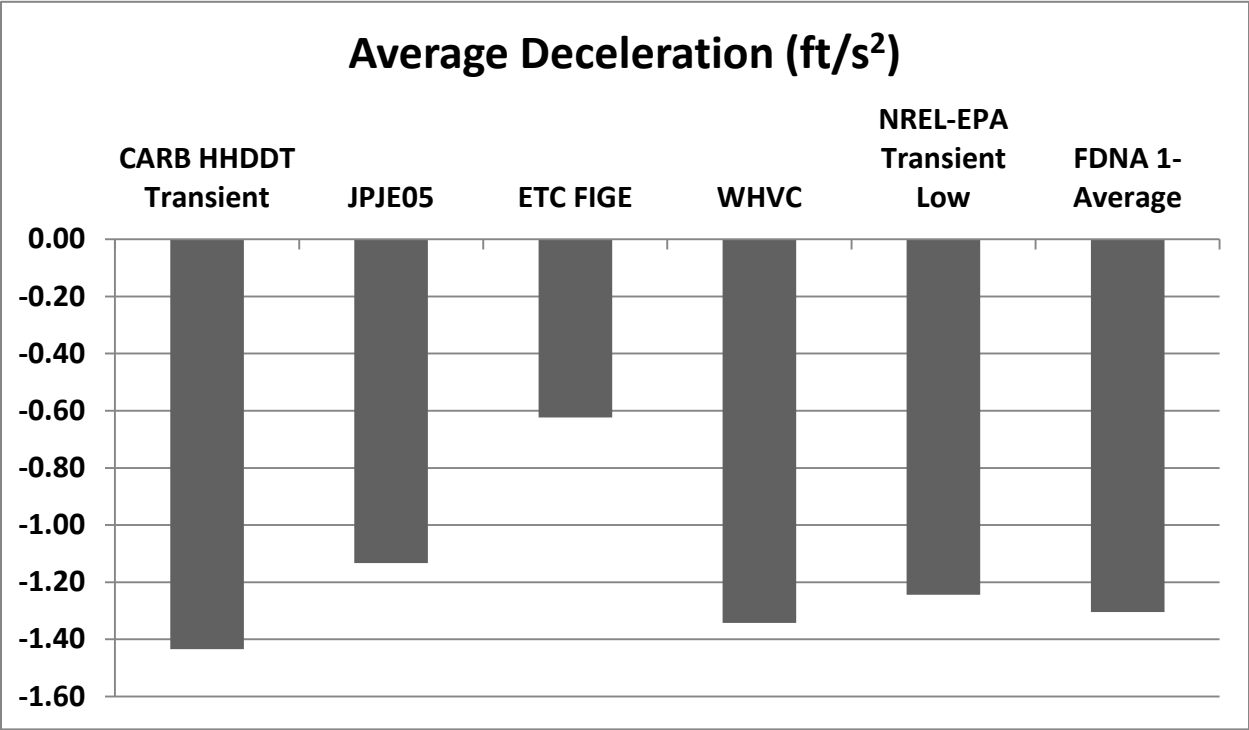


Figure 34: Comparison of standard cycle average deceleration rates

## 5 Development of Representative Drive Cycle Weights

Following a similar approach as was used to develop the representative transient cycles, the top 50 drive cycles from each of the 3 clusters in the cluster analysis ranked by RMS distance from medoid were used to compute cluster averages for each of the drive cycle metrics examined in Section 4. These averages represented the typical drive cycle behavior captured by the cluster. Using these calculated average values, drive cycle weights for zero speed time (% of total cycle duration), percentage of total mileage accumulated below 55 mph, percentage of total mileage accumulated between 55 and 65 mph, and percentage of total mileage accumulated above 65 mph could be determined for each cluster in alignment with the existing proposed phase 2 GHG rule framework.

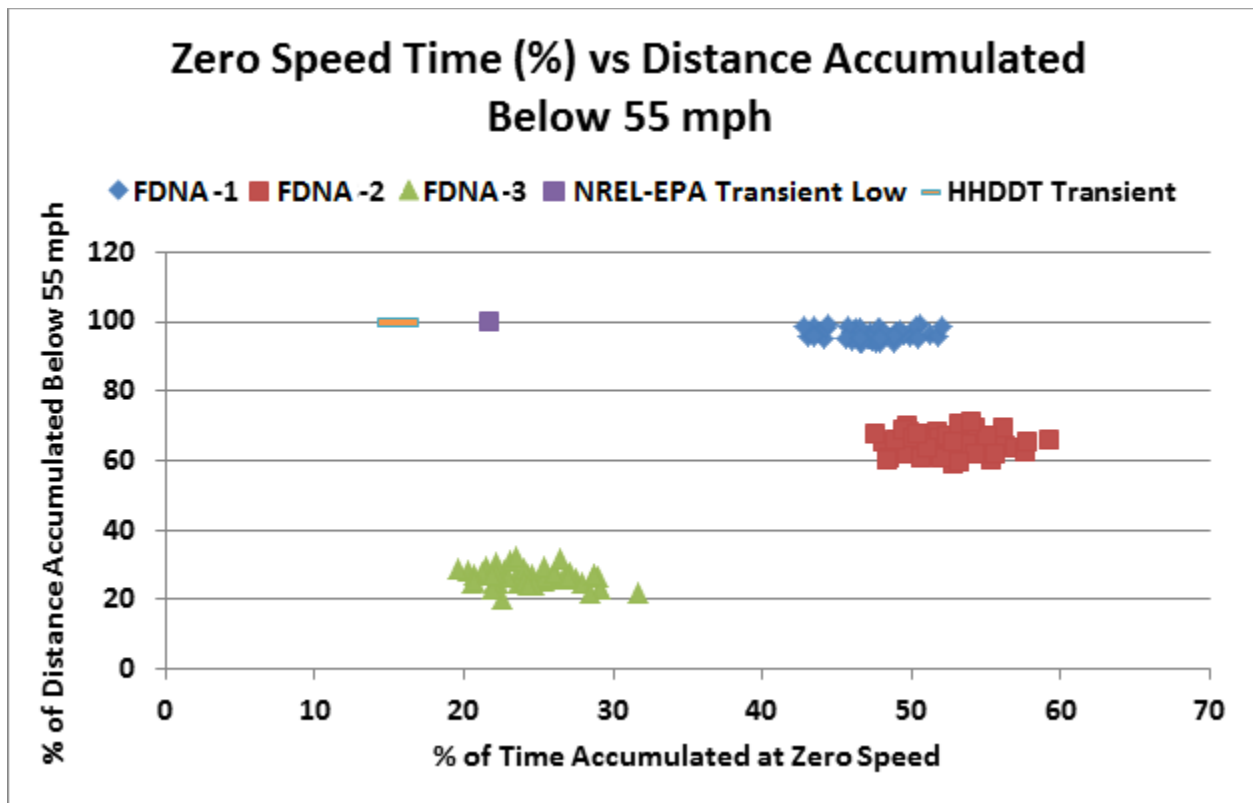
### 5.1 Comparative Drive Cycle Characterization

To explore the differences in vehicle behavior captured by each of the vehicle segments identified by the cluster analysis, a detailed drive cycle characterization was performed using DRIVE. The same metrics as used in the drive cycle cluster analysis were used to compare the different clusters and representative transients. Table 14 shows the results of the drive cycle characterization.

**Table 14. Comparison of Representative Cycle and Cluster Drive Cycle Metrics**

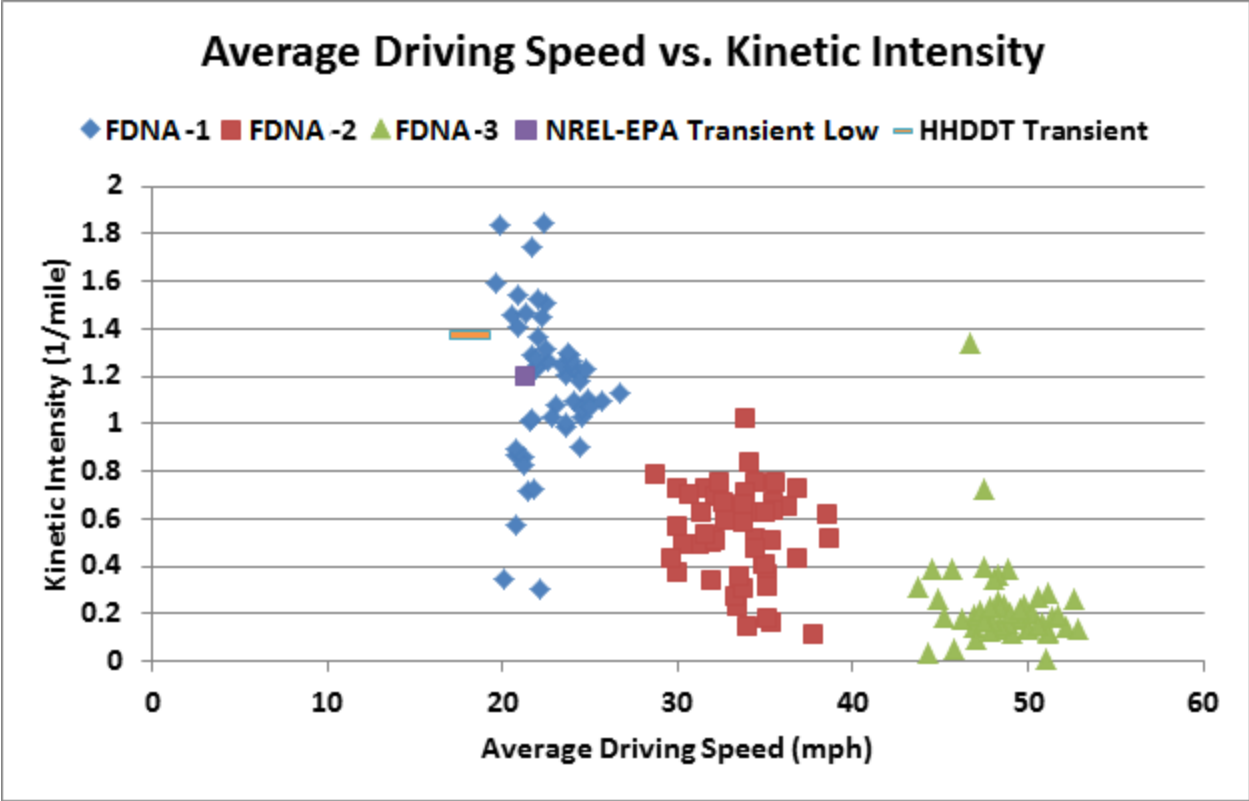
Cluster	Aerodynamic Speed (ft/s)	Characteristic Acceleration (ft/s <sup>2</sup> )	Percent of mileage accumulated below 55 mph	Percentage of Cycle Time at Zero Speed	Number of Stops Per Mile	Average Driving Speed (mph)	Maximum Driving Speed (mph)	Standard Deviation of Driving Speed (mph)
FDNA 1 - Average	50.59	0.55	96.45	47.30	1.71	22.62	62.39	13.74
FDNA 2 - Average	69.14	0.46	64.64	52.40	0.76	33.64	66.85	19.25
FDNA 3 - Average	83.40	0.29	26.47	24.50	0.21	48.57	70.81	19.17
NREL - EPA Transient Low	50.99	0.59	100.00	21.69	2.98	21.31	51.67	16.16
CARB HHDDT Transient	43.36	0.49	100.00	15.57	1.40	18.20	47.50	13.38
FDNA 1 - Medoid	48.88	1.22	97.85	47.87	1.22	24.12	62.61	13.03

Examining Table 14, one can see that the NREL-EPA representative transient cycle closely matches the FDNA 1 – Average cycle values for Characteristic Acceleration, Aerodynamic Speed, and average driving speed, while the CARB HHDDT is a better match for standard deviation of driving speed and stops per mile. This means that the NREL-EPA cycle is better matching the cluster characteristics on an energy consumption level, while the CARB cycle is better matching on a “stop and go” basis. Both cycles demonstrate less zero speed time than is necessary according to cluster 1 statistics; however this is of little concern given the ability to add idle time to modeling/testing. To further illustrate the differences in vehicle operating behavior and the three distinct operating modes defined by the cluster analysis (Section 3), data from Table 14 were graphed in Figures 35–37.



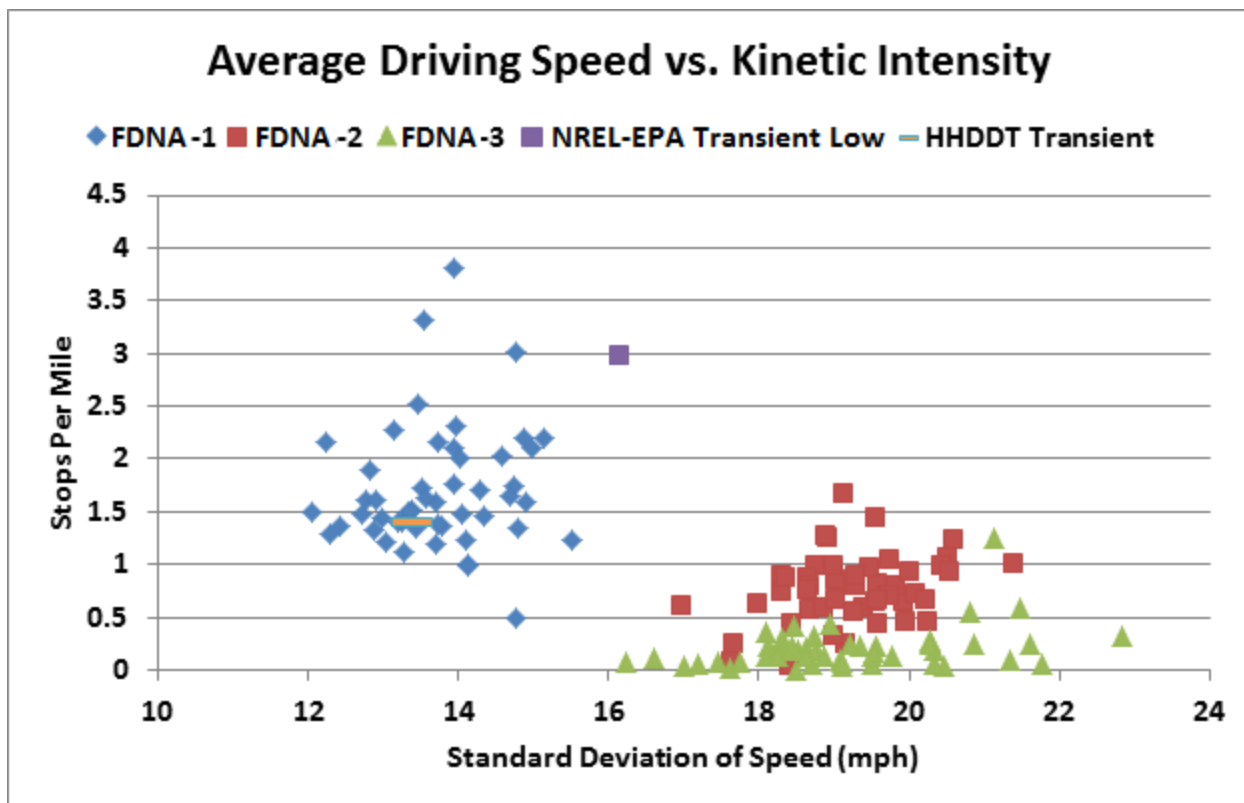
**Figure 35: Comparison of cluster and representative cycle zero speed time vs. mileage accumulated below 55 mph**

Looking closer at Figure 35, one can see that there is a significant difference between the three clusters in mileage accumulation. One can see that cluster 3-1 accumulates nearly all of its mileage below 55 mph, while the cluster 3-2 accumulates roughly half of its mileage at 55 mph or below, and the cluster 3-3 accumulates approximately 30% of its mileage at speeds below 55 mph. It can also be seen that there is a significant difference in zero speed time between the NREL-EPA representative transient test cycle and the low speed urban cluster (3-1). This difference in zero speed time can be adjusted through the addition of extra zero time in a supplemental idle test, allowing the transient cycles to match the final zero speed time percentages of any of the three clusters.



**Figure 36: Comparison of cluster and representative cycle average driving speed and kinetic intensity**

Similarly, illustrated in Figure 36 is the distinct separation of the drive cycles when looking at Kinetic Intensity and average driving speed. We can see that cluster 3-1 is centered on 20 mph average driving speed, cluster 3-2 around 35 mph, and cluster 3-3 at 50 mph. These represent urban driving, mixed driving, and highway driving cases. The HHDDT cycle is similar to the NREL-EPA transient low speed cycle with a slightly higher Kinetic Intensity (1.4) and slightly lower average driving speed (18 mph).



**Figure 37: Comparison of cluster and representative cycle standard deviation of speed and stops per mile**

Shown in Figure 37 are a comparison of stops per mile and standard deviation of driving speed. These two metrics combine to describe how much transient driving behavior a given drive cycle possesses. Higher values for stops per mile are indicative of congested urban driving with frequent stops in traffic and at stop lights/signs. Lower values for stops per mile are indicative of constant speed cruising drive cycles reflective of highway or rural driving behavior. Standard deviation of speed when coupled with stops per mile helps describe the relative variation within a daily drive cycle. Cycles possessing high standard deviation of speed show high levels of variation, or significant transient behavior. Lower standard deviations of speed reflect more constant driving behavior as would be displayed on highway or rural drive cycles. Comparing the two clusters with a low number of stops per mile (clusters 3-2 and 3-3) we can see that they display similar standard deviations of speed. This means that the two clusters possess similar driving behavior in terms of transient operation, with the primary difference between the clusters in this case being driving speed and zero speed time.

## 5.2 Analysis of Mileage Accumulation

As part of the drive cycle analysis, a detailed characterization of mileage accumulation was performed to further aid in understand the driving behavior represented by each of the representative cycles and identified clusters. The mileage accumulation for each of the drive cycles examined can be found in Table 15 and Table 16, where mileage is accumulated in 5 mph speed bin increments as a percentage of total miles traveled over the cycle. By splitting the data into 5 mph speed bins, researchers were able to examine transient speed mileage accumulation

(<45 mph), rural (45-55 mph), and highway type (55 mph) driving behavior at a finer level of detail than could be described by any single high level drive cycle metric. The Fleet DNA cluster 1 medoid data was included in Tables 14 and 15 to help illustrate the tightness of the cluster and whether or not the average skewed to the left or right of the median for each mileage accumulation bin. Additional even finer resolution data (2 mph bins) can be found in Appendix D.

**Table 15. Sub 45 mph Mileage Accumulation Comparison for Cluster and Representative Cycles**

<b>Cluster</b>	<b>0+ - 5 mph distance (%)</b>	<b>5+ - 10 mph distance (%)</b>	<b>10+ - 15 mph distance (%)</b>	<b>15+ - 20 mph distance (%)</b>	<b>20+ - 25 mph distance (%)</b>	<b>25+ - 30 mph distance (%)</b>	<b>30+ - 35 mph distance (%)</b>	<b>35+ - 40 mph distance (%)</b>	<b>40+ - 45 mph distance (%)</b>
FDNA 1 - Average	1.39	4.35	7.11	8.90	11.05	12.95	15.15	14.06	10.69
FDNA 2 - Average	0.68	1.98	2.74	3.36	4.46	6.06	7.54	8.75	9.05
FDNA 3 – Average	0.23	0.57	0.81	1.03	1.39	2.02	2.74	3.10	3.82
NREL - EPA Transient Low	2.66	4.40	5.42	7.08	8.37	8.30	15.15	18.69	12.55
CARB HHDDT Transient	1.25	12.72	3.97	9.02	18.20	10.20	15.72	11.96	8.29
FDNA 1 – Medoid	0.87	3.60	5.94	8.37	10.96	14.02	17.95	15.61	11.69

Examining the data shown in Tables 15 and 16, one can see that the average and medoid values for cluster 1 are in good agreement and closely match normal distributions centered around 35 mph, while the NREL-EPA representative transient drive cycle accumulates mileage in a similar normal distribution while being skewed slightly towards higher speed mileage accumulation. We can see that for all 3 clusters identified in the segmentation analysis, there is mileage accumulation in excess of 50 mph, demonstrating one of the weaknesses of the currently prescribed CARB HHDDT Transient cycle when used in certification and testing. Nearly 8% of cluster 1’s mileage is accumulated at speeds in excess of 50 mph.

**Table 16. 45+ mph Mileage Accumulation Comparison for Clusters and Representative Cycles**

<b>Cluster</b>	<b>45+ - 50 mph distance (%)</b>	<b>50+ - 55 mph distance (%)</b>	<b>55+ - 60 mph distance (%)</b>	<b>60+ - 65 mph distance (%)</b>	<b>65+ - 70 mph distance (%)</b>	<b>70+ - 75 mph distance (%)</b>
FDNA 1 - Average	6.57	4.22	2.52	1.00	0.04	0.00
FDNA 2 - Average	9.00	11.00	17.95	14.28	3.09	0.04
FDNA 3 - Average	4.27	6.49	17.57	44.90	10.88	0.16
NREL - EPA Transient Low	12.75	4.64	0.00	0.00	0.00	0.00
CARB HHDDT Transient	8.66	0.00	0.00	0.00	0.00	0.00
FDNA 1 – Medoid	5.41	3.44	1.70	0.45	0.00	0.00

Illustrated in Figure 38 is the mileage accumulation of the Fleet DNA cluster 1 average and medoid as well as the NREL-EPA transient. As was mentioned earlier when examining the data stored in Tables 15 and 16, the mileage accumulation for both the average and medoid of cluster 1 is roughly normal, with the average and medoid cycles being in very close agreement. The generated represented cycle is skewed slightly towards higher speed transient mileage accumulation with a dearth of mileage accumulated between 25 and 30 mph when compared to the cluster values of approximately 14%.



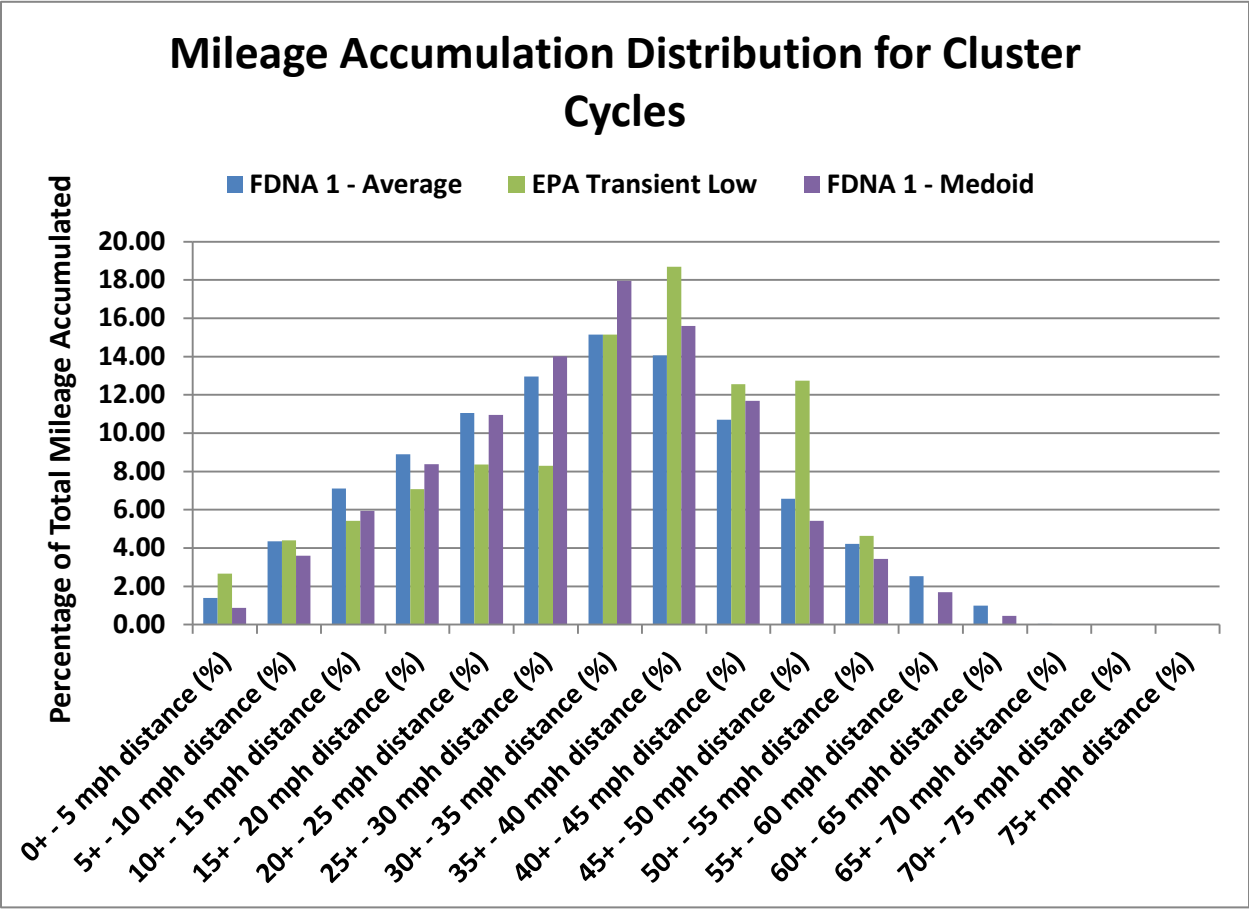


Figure 38: Comparison of cluster and representative cycle mileage accumulation

### 5.3 Weighting of Clusters by Mileage Accumulated in Broad Speed Categories

Compiling the results of the mileage accumulation analysis into the form defined in Phase 1 of the EPA Greenhouse Gas regulations and rounded to the nearest percent as shown in Table 17, one sees that there is significant differences between the Fleet DNA proposed weightings and the existing regulation. In particular, one key difference is the identification of three unique driving modes, urban, mixed urban, and highway driving for vocation vehicles. In addition, there is no differentiation between conventional and hybrid vocational vehicle operation.

**Table 17. Final Weights for Identified Clusters**

<b>Cluster</b>	<b>Percent of mileage accumulated below 55 mph</b>	<b>Percentage of mileage accumulated between 55 and 65 mph</b>	<b>Percentage of mileage accumulated above 65 mph</b>	<b>Percentage of Cycle Time at Zero Speed</b>
FDNA 1 – Urban	96	4	0.0	47
FDNA 2 – Mixed Urban	65	32	3.0	52
FDNA 3 - Highway	27	62	11	25
Phase 1 Vocational	42.0	21.0	37.0	NA
Phase 1 Vocational Hybrids	75.0	9.0	16.0	NA

Examining the statistics in greater detail, one can see that the observed mileage accumulation at speeds in excess of 65 mph for the Fleet DNA database data is far smaller than that of the existing regulations (by approximately 1/3 for the highway cycle compared to the vocational classification in phase 1). Conversely much higher mileage is accumulated at lower transient speeds below 55 mph as well as transitional highway speeds between 55 and 65 mph.

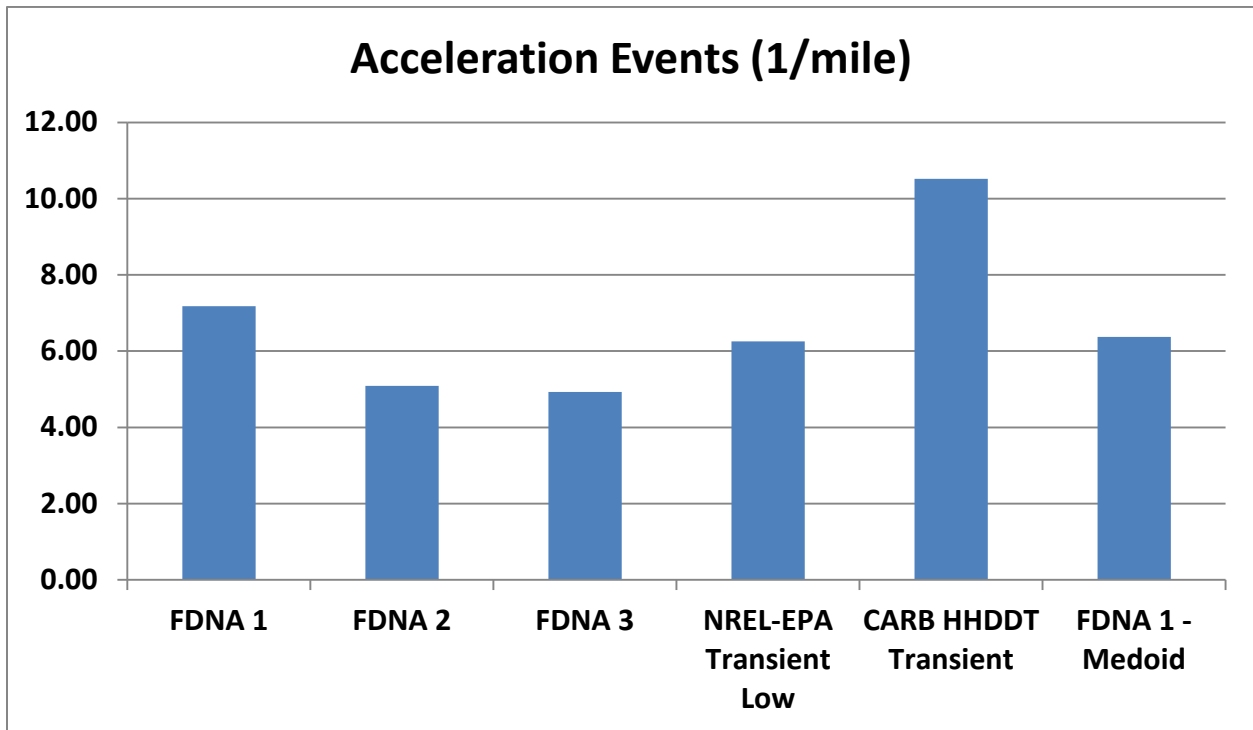
## **5.4 Characterization of Acceleration/Deceleration Behavior**

A detailed analysis of the acceleration and deceleration behavior of both the representative low speed transient cycle and the CARB HHDDT was performed to compare the behavior of the two transient cycles to the clusters identified in the segmentation analysis. Metrics describing the frequency (per mile), and aggressiveness (average and maximum rates) were calculated and displayed in Table 18.

**Table 18. Comparison of Acceleration Behavior for Clusters and Representative Cycle**

Cluster	Acceleration events (1/mile)	Deceleration events (1/mile)	Max_accel (ft/s <sup>2</sup> )	Max_decel (ft/s <sup>2</sup> )	Avg accel (ft/s <sup>2</sup> )	Avg decel (ft/s <sup>2</sup> )
FDNA 1 – Average	7.18	7.19	9.90	-9.98	1.25	-1.43
FDNA 2 – Average	5.09	5.09	8.97	-10.28	1.01	-1.13
FDNA 3 - Average	4.93	4.93	10.74	-8.88	0.58	-0.62
NREL - EPA Transient Low	6.26	6.26	5.34	-5.66	1.41	-1.34
CARB HHDDT Transient	10.52	10.17	4.25	-4.06	0.94	-1.24
FDNA 1 - Medoid	6.37	6.37	9.01	-8.80	1.18	-1.30

Examining the results shown in Table 18, one can see that the proposed NREL-EPA low speed transient cycle displays fewer more aggressive acceleration events than the CARB HHDDT Cycle. However, when compared to the Cluster 1 Fleet DNA average values representative of urban driving, the NREL-EPA transient cycle more closely matches in acceleration/deceleration events per mile while still not being aggressive enough to match the overall maximum acceleration/deceleration values observed for each of the clusters which are quite aggressive. Table 18 results have been visualized in Figures 40 through 45.



**Figure 39. Comparison of cluster and representative cycle acceleration frequency**

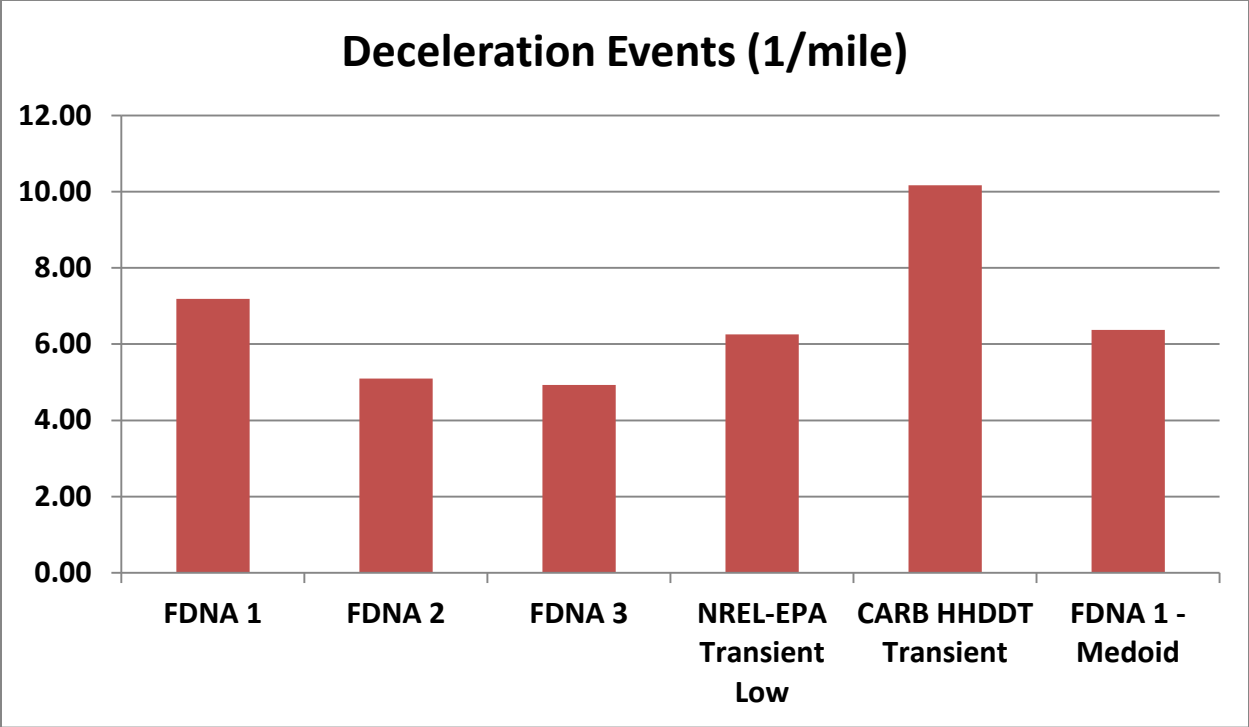


Figure 40. Comparison of cluster and representative cycle deceleration frequency

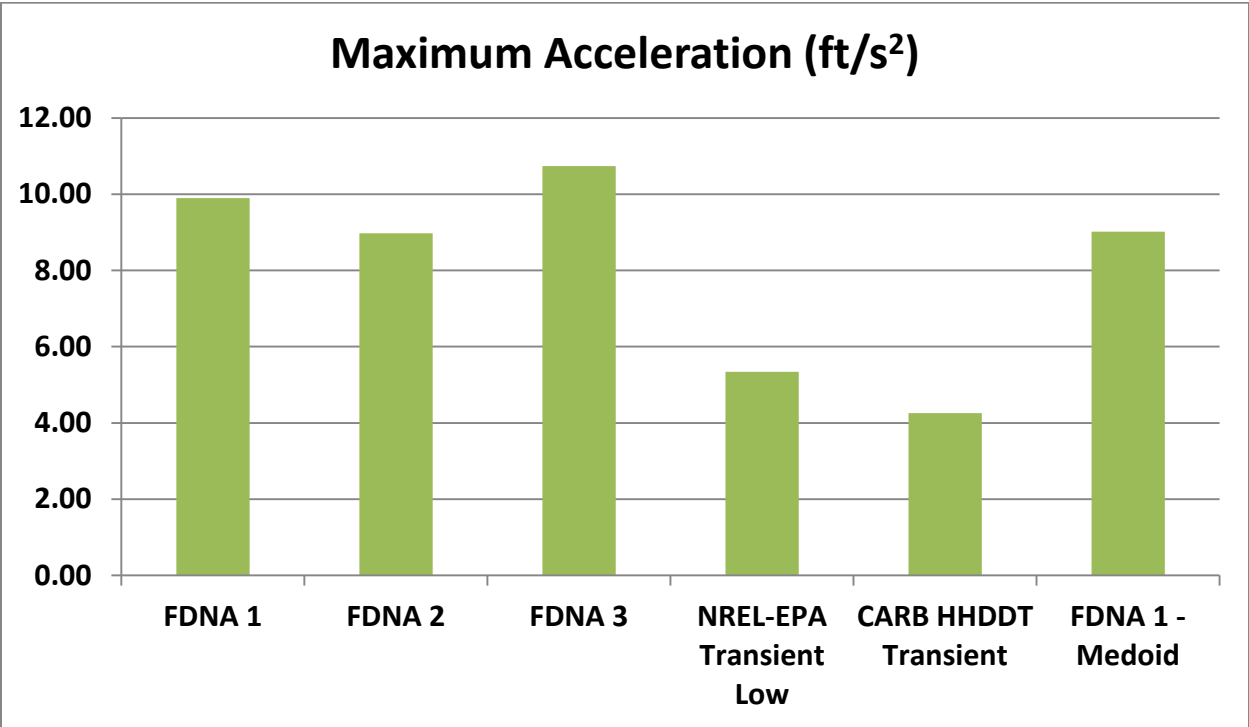


Figure 41. Comparison of cluster and representative cycle peak acceleration rates

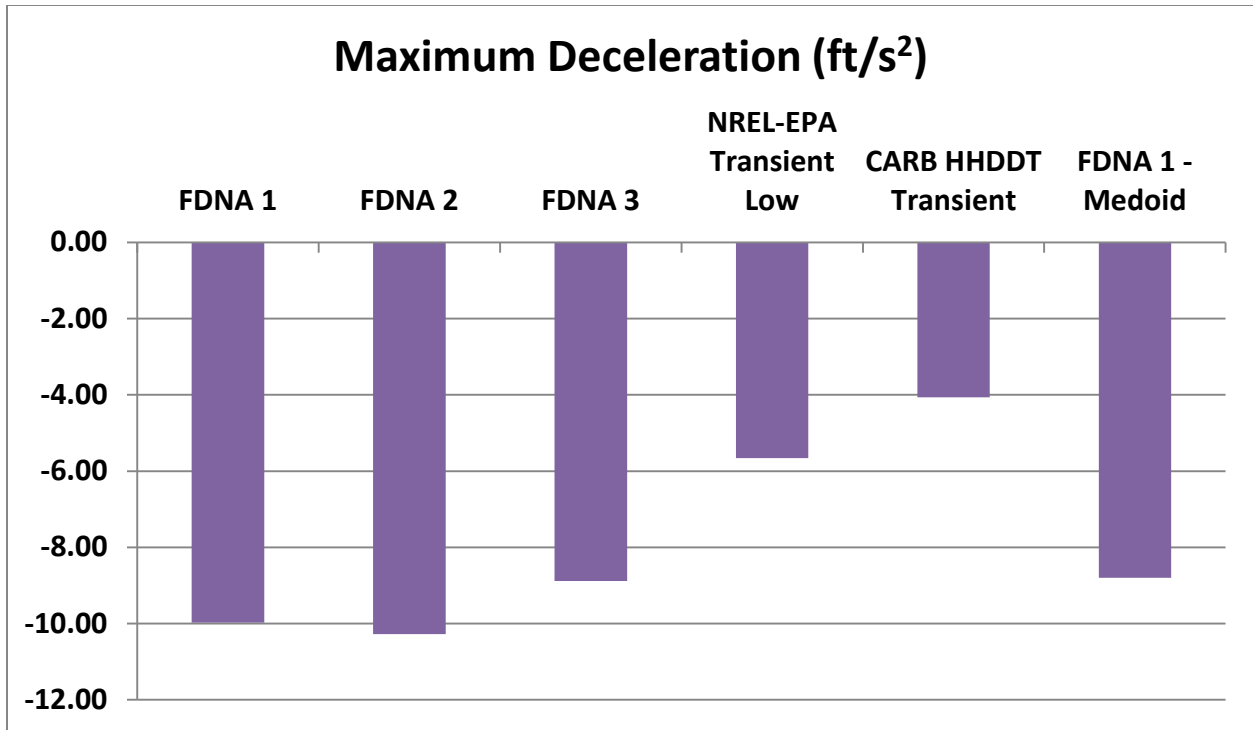


Figure 42. Comparison of cluster and representative cycle peak deceleration rates

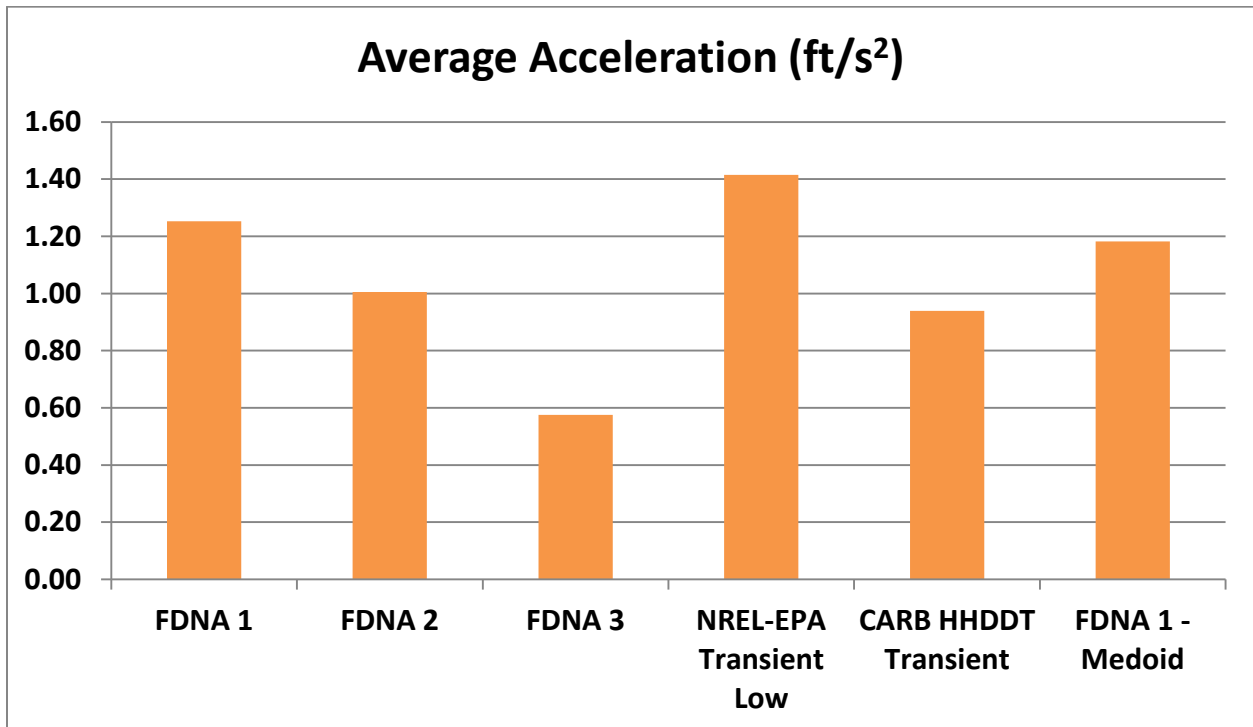


Figure 43. Comparison of cluster and representative cycle average acceleration rates

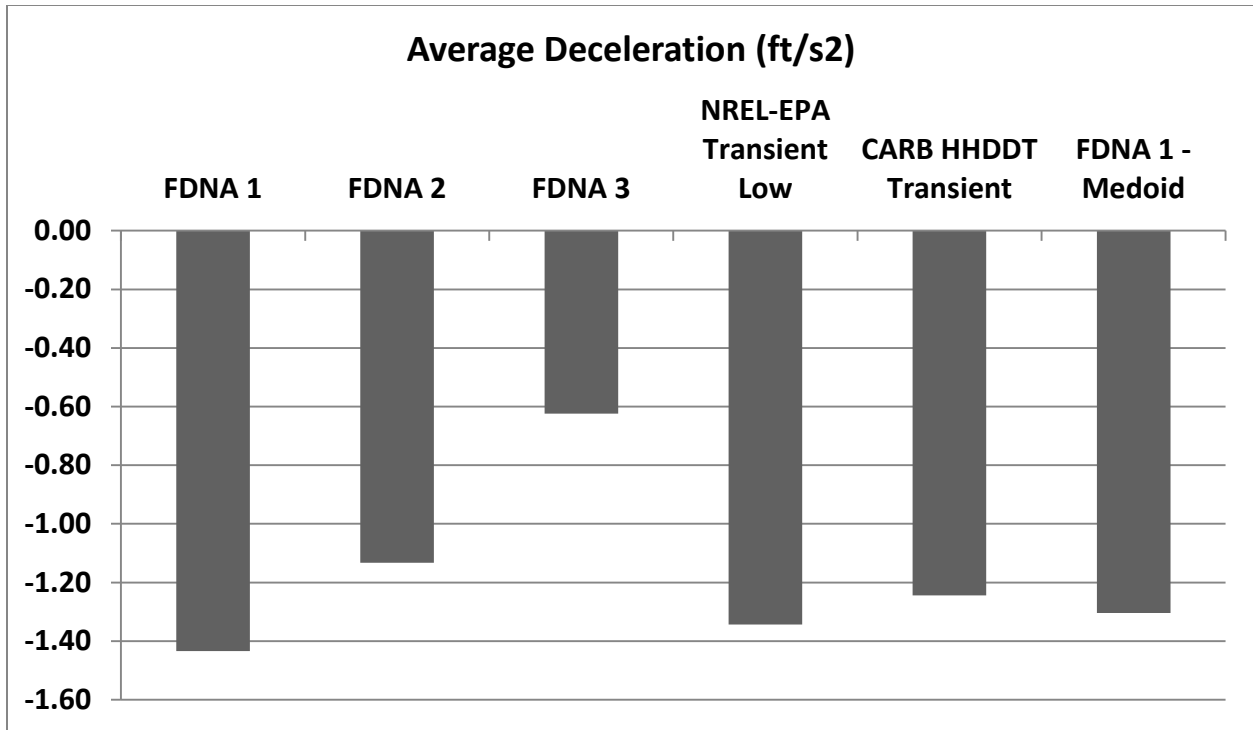


Figure 44. Comparison of cluster and representative cycle average deceleration rates

## 5.5 Analysis of Idle Behavior

Working with a subset of the full Fleet DNA dataset comprised of the 128 class 8 vehicles possessing combined engine speed, vehicle speed, and brake switch data, NREL researchers analyzed 3,448 days of operating data to characterize the cumulative amount of in-gear and out-of-gear operating time for each cluster identified as a result of the segmentation efforts. The breakdown of this subset can be seen in Table 19.

**Table 19. Vocational Class 8 Breakdown**

<b>Vocation</b>	<b>Number of vehicles</b>	<b>Number of days</b>
Beverage Delivery	7	116
Drayage	31	606
Food Delivery	11	226
Freight	20	876
Local Delivery	3	104
Parcel Delivery	4	44
Refrigerated Truck	6	221
Regional Haul	1	40
Tanker	21	940
Transfer truck	24	275
Grand Total	128	3448

To separate in-gear from out-of-gear operation, raw data were examined in detail and a pair of logic based conditional statements were developed and then applied.

Out-of-gear idle operation was defined as the time periods under which all four of the following conditions were achieved:

- Engine Speed  $\leq$  Engine Idle Speed
  - This logical component bounds conditions under which the vehicle engine is idling and not operating at zero speed under additional loads such as PTO.
- Engine Speed  $> 0$ 
  - This logical component ensures that the data examined occurred when the engine was on and not during a parked, key-off event.
- Vehicle Speed = 0
  - This logical operator requires that the vehicle be stopped.
- Brake Switch Signal = 0
  - As defined in the SAEJ1939 under SAE SPN597, the Brake switch signal indicates whether or not the driver operated brake foot pedal is being pressed. A brake switch signal = 0 indicates that the brake switch is released, indicating that under the combined conditions above the data being examined describes a vehicle stopped and out of gear rather than stopped at a stop light or in traffic.

In-gear idle operation was defines as the time periods under which all four of the following conditions were achieved:

- Engine speed  $\leq$  Engine Idle Speed

- Similar to the out-of-gear idle conditional logic, this statement requires that any in-gear-idle occur when the vehicle engine is operating at idle speeds and not under loading such as PTO
- Engine Speed > 0
  - This logic requires that in-gear engine idle possess engine on operation and not include any time period where the vehicle would be parked and keyed off.
- Vehicle Speed = 0
  - The vehicle must be stopped.
- Brake Switch Signal = 1
  - Per SAE J1939 SPN 597 definition, this condition requires that the driver operated brake foot pedal be depressed or otherwise engaged. Under this condition it is assumed that under this condition the vehicle is in gear and stopped under normal operating conditions such as traffic, stops lights, and daily activity.

Based on these the conditions, the results of the supplemental zero speed analysis can be seen in Table 20. Note the values shown in Table 20 have been rounded to the nearest whole percent in alignment with the values contained in Table 17.

**Table 20. Proposed In-Gear and Out-of-Gear Idle Weights**

Cluster	Operating Mode	Percent of Workday
1	Out-of-Gear Idle	28
1	Total Zero Speed Time (Both In- and Out-of-Gear)	47
2	Out-of-Gear Idle	22
2	Total Zero Speed Time (Both In- and Out-of-Gear)	52
3	Out-of-Gear Idle	25
3	Total Zero Speed Time (Both In- and Out-of-Gear)	25

Examining the results, one can see that for both cluster 1 and 2 vehicles, roughly half of the zero speed time consists of out-of-gear idle operation, while for high speed cluster 3 100 percent of zero speed time is considered out-of-gear. It is important to note when examining these findings that the results for this subset of 128 class 8 vehicles may have idle operation that differs from the broader Fleet DNA subset of 754 vehicles which contains a variety of vehicle vocations and weight classes.

These results make sense intuitively as the types of behavior characterized by the three individual clusters would have different percentages of in-gear idle behavior as a result of driving conditions. For example and as illustrated by the data shown, one would expect that urban and mixed drive cycles would possess a greater percentage of in-gear idle due to stop lights and traffic conditions than vehicles operating under highway conditions.



## 6 Known Limitations and Potential for Future Work

As mentioned in previous reports, the need for coupled road grade and drive cycle information is an area of ongoing interest that provides the opportunity for future development and refinement of both the high speed transient and constant high speed cruise drive cycle components [36]. In addition, exploring the opportunity for simplified testing through the development of a dual mode test cycle similar to the one shown in Figure 46 to replace the existing 3 component test procedure is a topic of future interest which could result in faster less costly certification. Additional information regarding this effort can be found in Appendix E.

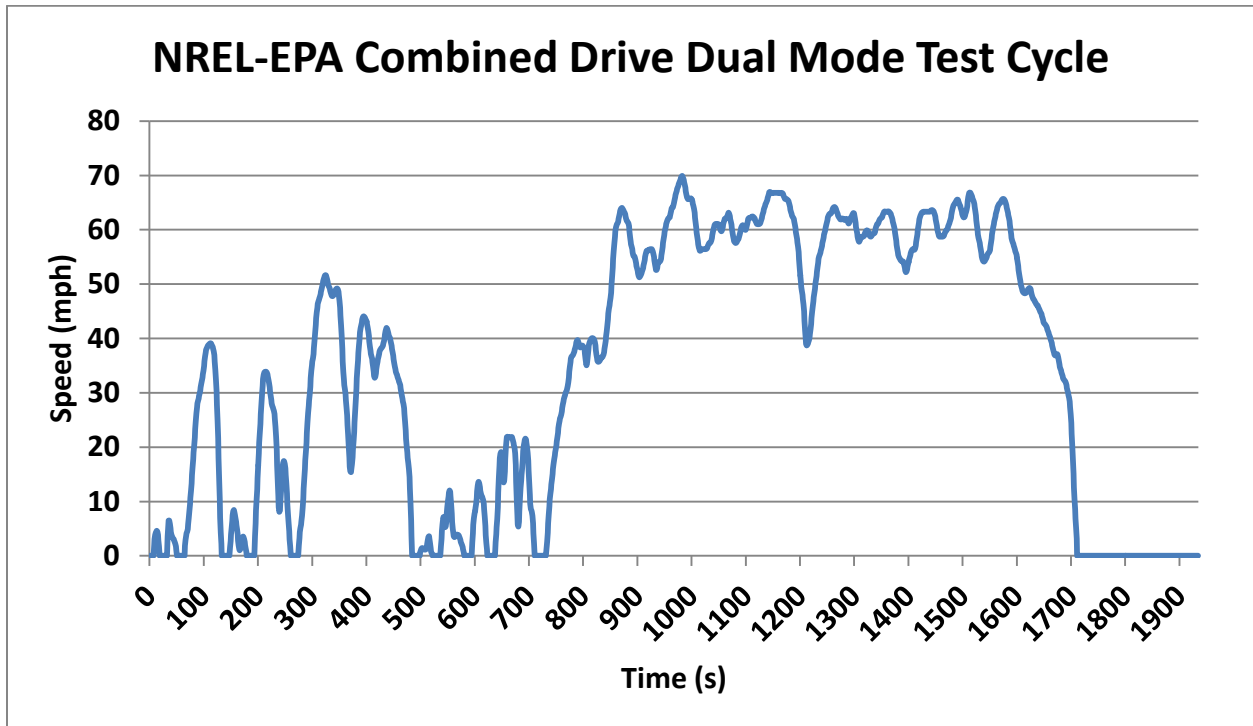


Figure 45: Sample potential dual mode test cycle

## References

- [1] U.S. Environmental Protection Agency. (2011). *EPA and NHTSA Adopt First-Ever Program to Reduce Greenhouse Gas Emissions and Improve Fuel Efficiency of Medium- and Heavy-Duty Vehicles*. EPA-420-F-11-031. Office of Transportation and Air Quality. Accessed February 6, 2016. <http://www.epa.gov/otaq/climate/documents/420f11031.pdf>
- [2] U.S. Environmental Protection Agency. (2011). *Final Rulemaking to Establish Greenhouse Gas Emission Standards and Fuel Efficiency Standards to Medium- and Heavy-Duty Engines and Vehicles: Regulatory Impact Analysis*. EPA-420-R-11- 901. Office of Transportation and Air Quality and National Highway Traffic Safety Administration. Accessed January 13, 2016. <http://www.epa.gov/otaq/climate/documents/420r11901.pdf>
- [3] Federal Register. (2011). “Greenhouse Gas Emissions Standards and Fuel Efficiency Standards for Medium- and Heavy-Duty Engines and Vehicles: Final Rule.” Environmental Protection Agency and Department of Transportation. (76:179); Accessed February 16, 2016. <http://www.gpo.gov/fdsys/pkg/FR-2011-09-15/pdf/2011-20740.pdf>
- [4] U.S. Environmental Protection Agency. (February 10, 2015). “Transportation and Climate: Regulations & Standards: Heavy-Duty.” Office of Transportation and Air Quality. Accessed February 16, 2015. <http://www.epa.gov/otaq/climate/regs-heavyduty.htm>
- [5] The White House, Office of the Press Secretary. (February 18, 2014). “Remarks by the President on Fuel Efficiency Standards of Medium and Heavy-Duty Vehicles.” Accessed February 16, 2015. <http://www.whitehouse.gov/the-pressoffice/2014/02/18/remarks-president-fuel-efficiency-standards-medium-and-heavyduty-vehicl>
- [6] The White House, Office of the Press Secretary. (February 18, 2014). “Fact Sheet: Opportunity for All: Improving the Fuel Efficiency of American Trucks – Bolstering Energy Security, Cutting Carbon Pollution, Saving Money and Supporting Manufacturing Innovation.” Accessed February 6, 2016. <http://www.whitehouse.gov/the-press-office/2014/02/18/fact-sheet-opportunity-allimproving-fuel-efficiency-american-trucks-bol>
- [7] “EPA and NHTSA Propose Standards to Reduce Greenhouse Gas Emissions and Improve Fuel Efficiency of Medium- and Heavy-Duty Vehicles for Model Year 2018 and Beyond,” EPA Fact Sheet EPA-420-F-15-901, 2015, <http://www3.epa.gov/otaq/climate/documents/420f15901.pdf>
- [8] Walkowicz, K., Kelly, K., Duran, A., and Burton, E. (2014). “Fleet DNA Project Data.” National Renewable Energy Laboratory. <http://www.nrel.gov/fleetdna>

- [9] Transportation Secure Data Center. (2014). National Renewable Energy Laboratory. Accessed October 31, 2014: [www.nrel.gov/tsdc](http://www.nrel.gov/tsdc)
- [10] National Renewable Energy Laboratory. (2013). "DRIVE Analysis Tool Generates Custom Vehicle Drive Cycles Based on Real-World Data." NREL/FS-5400-54507. Golden, CO: National Renewable Energy Laboratory. Accessed February 16, 2015. <http://www.nrel.gov/docs/fy13osti/54507.pdf>
- [11] *NREL Vehicle Drive Cycle Tool, User Guide*. (2009). Alliance for Sustainable Energy, LLC. All Rights Reserved.
- [12] National Renewable Energy Laboratory. (2014) *Contributing Data to the Fleet DNA Project*. NREL/BR-5400-62771. Golden, CO: National Renewable Energy Laboratory. Accessed February 2, 2015. [www.nrel.gov/docs/fy14osti/62771.pdf](http://www.nrel.gov/docs/fy14osti/62771.pdf)
- [13] Duran, A., and Earleywine, M., (2012). "GPS Data Filtration Method for Drive Cycle Analysis Applications." SAE Technical Paper 2012-01-0743, doi:10.4271/2012-01-0743.
- [14] O'Keefe, M., Simpson, A., Kelly, K., and Pedersen, D., "Duty Cycle Characterization and Evaluation Towards Heavy Hybrid Vehicle Applications." SAE Technical Paper 2007-01-0302, 2007, doi:10.4271/2007-01-0302.
- [15] Kaufman, L., and Rousseeuw, P.J. (1990). *Partitioning Around Medoids (Program PAM)*, in *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons, Inc., Hoboken, NJ, USA. doi: 10.1002/9780470316801. Chapter 2.
- [16] Kaufman, L., and Rousseeuw, P. J. (1990). *Clustering Large Applications (Program CLARA)*, in *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons, Inc., Hoboken, NJ, USA. doi: 10.1002/9780470316801. Chapter 3.
- [17] Rousseeuw, Peter J., (1987). "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis." *Journal of Computational and Applied Mathematics* 20:53-65, ISSN 0377-0427, [http://dx.doi.org/10.1016/0377-0427\(87\)90125-7](http://dx.doi.org/10.1016/0377-0427(87)90125-7).
- [18] U.S. Environmental Protection Agency. (December 1, 2015). "MOVES (Motor Vehicle Emissions Simulator)." Office of Transportation and Air Quality. Accessed February 4, 2016. <http://www3.epa.gov/otaq/models/moves/#generalinfo-2014a>
- [19] Draper, N.R., Smith, H., (1998.) *Applied Regression Analysis*. New York, John Wiley & Sons, Inc., 706 p.
- [20] Diggle, P.J., Liang, K.-Y., Zeger, S.L., (1994.) *Analysis of Longitudinal Data*. Oxford, Clarendon Press, 253 p.
- [21] Sharer, P., Leydier, R., and Rousseau, A., (2007). "Impact of Drive Cycle Aggressiveness and Speed on HEVs Fuel Consumption Sensitivity." SAE Technical Paper 2007-01-0281, doi:10.4271/2007-01-0281.

- [22] Fellah, M., Singh, G., Rousseau, A., Pagerit, S., Nam, E., and Hoffman, G., “Impact of Real-World Drive Cycles on PHEV Battery Requirements.” SAE Technical Paper 2009-01-1383, 2009, doi:10.4271/2009-01-1383.
- [23] Sun, Z., and Andreae, M., (2012). “Vehicle Duty Cycle Characteristics for Hybrid Potential Evaluation.” SAE Technical Paper 2012-01-2023, doi:10.4271/2012-01-2023.
- [24] Stichter, J., (2012). “Investigation of Vehicle and Driver Aggressivity and Relation to Fuel Economy Testing.” (Master’s Thesis). Accessed February 6, 2016. <http://ir.uiowa.edu/cgi/viewcontent.cgi?article=3542&context=etd>
- [25] Berry, I., (2010). “The Effects of Driving Style and Vehicle Performance on the Real-World Consumption of U.S. Light-Duty Vehicles.” (Master’s Thesis). Accessed February 2, 2016 [http://web.mit.edu/sloan-auto-lab/research/beforeh2/files/IreneBerry\\_Thesis\\_February2010.pdf](http://web.mit.edu/sloan-auto-lab/research/beforeh2/files/IreneBerry_Thesis_February2010.pdf)
- [26] Reinhart, T. E., (2015). *Commercial Medium- and Heavy-Duty Truck Fuel Efficiency Technology Study - Report #1*. (Report No. DOT HS 812 146). Washington, DC: National Highway Traffic Safety Administration.
- [27] Lee, T., and Filipi, Z., (2011). “Synthesis of Real-World Driving Cycles Using Stochastic Process and Statistical Methodology.” *Int. J. of Vehicle Design* 57(1):17-36, 2011, doi:10.1504/IJVD.2011.043590.
- [28] Tong, H., and Hung, W., (2010). “A Framework for Developing Drive Cycles with On-Road Driving Data.” *Transport Reviews* 30(5):589-615. doi:10.1080/01441640903286134.
- [29] Park, J., Lee, J., and Lee, J., (2013). “Development of Driving Cycle for CO<sub>2</sub> Emission Test of Heavy-Duty Vehicles.” SAE Technical Paper 2013-01-2520, doi:10.4271/2013-01-2520.
- [30] Moore, W., Finch, T., and Sutton, M., (2013.)”Development of Heavy Duty Diesel Real World Drive Cycles for Fuel Economy Measurements.” SAE Technical Paper 2013-01-2568, doi:10.4271/2013-01-2568.
- [31] Kulkarni, A., Sapre, R., and Sonchal, C., “GPS Based Methodology for Drive Cycle Determination.” SAE Technical Paper 2005-01-1060, 2005, doi:10.4271/2005-01-1060.
- [32] Dieselnets.com. (2016). “Emission Test Cycles.” Accessed December 16, 2015 <http://dieselnets.com/standards/cycles/#us>
- [33] Dieselnets.com. (2016) “Heavy-Duty Vehicles: GHG Emissions and Fuel Economy.” Accessed December 16, 2015 <http://dieselnets.com/standards/cycles/#us>
- [34] Barlow, T.J., Latham, S., McCrae, I.S., Boutler, P.G. (2009). *A Reference Book of Driving Cycles for Use in the Measurement of Road Vehicle Emissions*. (Report PRP354). Workingham, United Kingdom: TRL Limited.

- [35] “Development of on Track Chassis Dynamometer Drive Cycles and Subsequent Evaluation of Sensors Europe On Board Emissions Sampling System.” (2003). UK Department For Transport, Report Number MBK 03/0592. <http://webarchive.nationalarchives.gov.uk/+http://www.dft.gov.uk/pgr/roads/environment/research/cqvcf/ardemissionsmeasurements3837.pdf>
- [36] Wood, E., Duran, A., Burton, E., Gonder, J., and Kelly, K. (2015). *EPA GHG Certification of Medium- and Heavy-Duty Vehicles: Development of Grade Profiles Representative of US Controlled Access Highways*. NREL Technical Report TP-5400-63853. <http://www.nrel.gov/docs/fy15osti/63853.pdf>
- [37] Wefky, A., Espinosa, F., Prieto, A., Garcia, J.J., and Barrios, C. (2011). “Comparison of Neural Classifiers for Vehicles Gear Estimation.” *Applied Soft Computing* 11(4): 3580-3599, ISSN 1568-4946, <http://dx.doi.org/10.1016/j.asoc.2011.01.030>.

## Appendix A: Standard Fleet DNA Data Channels

When performing data collection through either manual vehicle instrumentation or automated data transfer via installed onboard vehicle telemetric devices, NREL researchers collect a standard baseline of vehicle operating data for analysis and storage in the Fleet DNA database. The specific channels chosen for data collection have been developed with a focus on characterization of vehicle drive cycle and engine/drivetrain operation. Shown in Table 21 is an example of the standard Global Positioning Satellite (GPS) channels collected and stored in Fleet DNA.

**Table 21. Example GPS File**

Data Check	Latitude	Direction	Longitude	Direction	Time	Date	Speed (mph)	Heading (degree)	Altitude	HDOP	Satellites
A	44.99508	N	93.21397	W	74920	150710	8.9	122	889	2.1	7
A	44.99528	N	93.21172	W	74938	150710	12.7	92	932	1.9	5
A	44.99528	N	93.21165	W	74939	150710	10.2	89	932	2	5
A	44.99528	N	93.21158	W	74940	150710	10.2	89	932	2	5
A	44.99528	N	93.21152	W	74943	150710	6.3	94	945	1.7	5
A	44.99527	N	93.21148	W	74944	150710	6.3	94	945	1.7	5
A	44.99527	N	93.21145	W	74945	150710	0	94	945	2.1	5

Of particular interest are the latitude, longitude, elevation, speed, and time channels. Latitude and Longitude information can be coupled with elevation data and additional datasets stored in Fleet DNA such as the Tom Tom National Road Network database to provide Geospatial context for route analysis. GPS and time channels provide the requisite source data necessary for drive cycle characterization and vehicle idle time analyses.

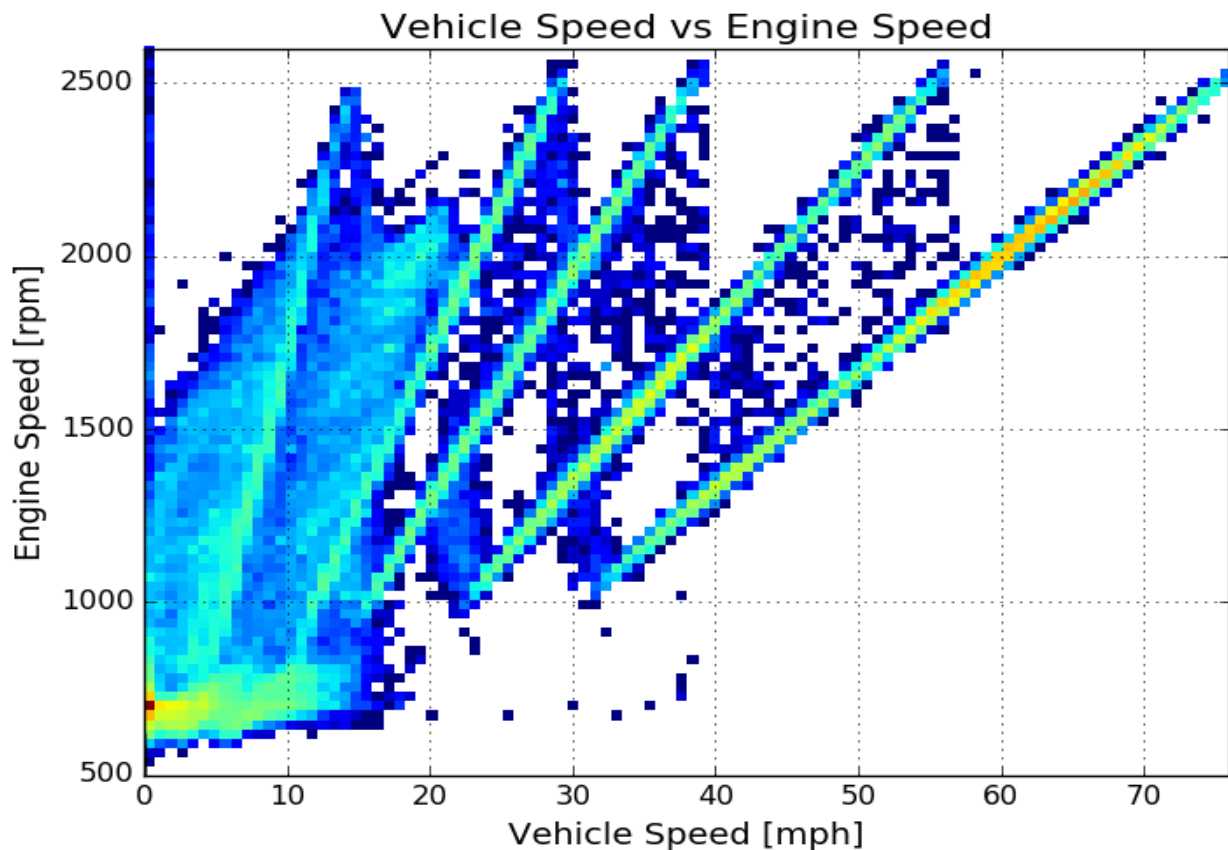
To enable deeper analysis of the influence of drive cycles on vehicle fuel consumption, performance, and emissions, a number of additional data channels publically product over the vehicle Controller Area Network (CAN) bus according to SAE J1939 standards have been identified and are collected at 1Hz in conjunction to standard GPS telematics data. Table 22 details the list of basic vehicle engine, powertrain, and emissions channels collected and stored in the Fleet DNA database.

**Table 22. List of Additional Logged Controller Area Network Channels**

Data Channel Name	PGN#	SAE SPN#	Channel Description
Actual Engine - Percent Torque	61444	513	The calculated output torque of the engine. The data is reported as a percent of reference engine torque.
Aftertreatment 1 Outlet NO <sub>x</sub>	61455	3226	The amount of combined NO and NO <sub>2</sub> in the exhaust exiting the aftertreatment system measured by a NO <sub>x</sub> sensor at the aftertreatment outlet, represented in NO <sub>x</sub> molecules parts per million non-NO <sub>x</sub> molecules.
Aftertreatment 1 Intake NO <sub>x</sub>	61454	3216	The amount of combined NO and NO <sub>2</sub> in the exhaust entering the aftertreatment system measured by a NO <sub>x</sub> sensor at the aftertreatment outlet, represented in NO <sub>x</sub> molecules parts per million non-NO <sub>x</sub> molecules.
Engine Fuel Rate	65266	183	Instantaneous amount of fuel consumed by the engine per unit of time.
Engine Intake Air Mass Flow Rate	61450	132	Mass flow rate of fresh air entering the engine air intake, before exhaust gas recirculator mixer is used, if applicable. Flow rate of fresh air conducted to the engine cylinders to support combustion.
Engine Speed	61444	190	Actual engine speed calculated over a minimum crankshaft angle of 720 degrees divided by the number of cylinders.
Estimated Engine Parasitic Losses - Percent Torque	65247	2978	The calculated torque that indicates the estimated amount of torque loss due to engine parasitics, such as cooling fan, air compressor, air conditions, etc.
Nominal Friction - Percent Torque	65247	514	The calculated torque that indicates the amount of torque required by the engine including additional losses via pumping torque loss, fuel, oil, and cooling pumps, and frictional and thermodynamic engine losses.
Referenced Torque	65251	544	This channel is the 100% reference torque value for all defined indicated engine torque paramters. It is only defined once and does not change if a different engine torque map becomes valid.
Wheel-Based Vehicle Speed	65265	84	Speed of the vehicle as calculated from wheel or tailshaft speed.

## Appendix B: Determining Engine to Vehicle Speed Ratios via Kernel Density Estimation

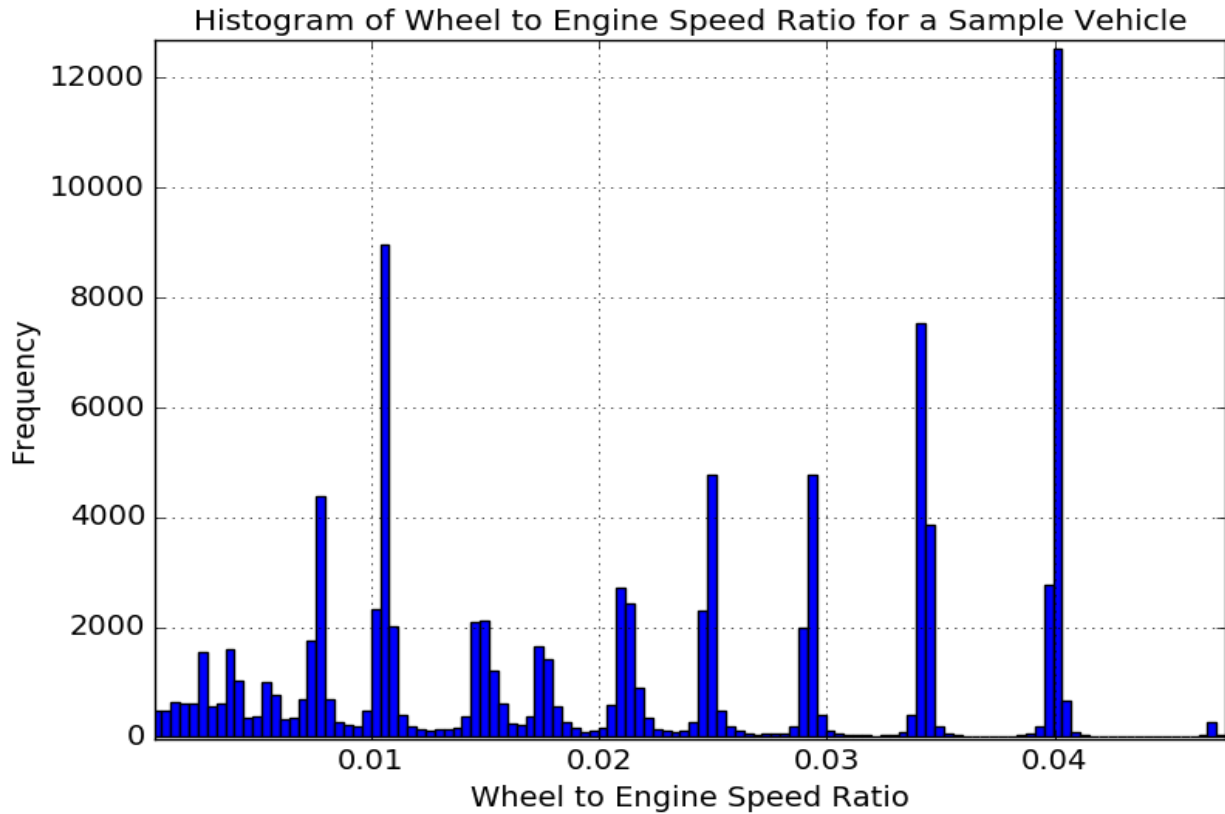
Given the presence of limited vehicle chassis specification information stored in the Fleet DNA database, NREL researchers developed a novel kernel density estimator method to infer vehicle gear ratios using existing captured 1-Hz Fleet DNA data such as instantaneous vehicle and engine speed. Similar approaches to estimating vehicle gearing have been developed as part of previous research and have been adapted and expanded upon in this study [37]. Starting with filtered raw vehicle speed and engine data as shown in Figure 47, the first step in the process of identifying individual operating gears is to combine the two individual metrics into a single variable by dividing vehicle speed in mph by engine speed in rotations per minute to compute a ratio of vehicle speed to engine speed.



**Figure 46: Sample heat plot of vehicle speed vs. engine operating speed**

Shown in Figure 48 are the results of this transformation. Visually inspecting the results, it is apparent to see that there are a series of common peaks at which the ratio of vehicle speed to engine speed is most frequently observed in the dataset.

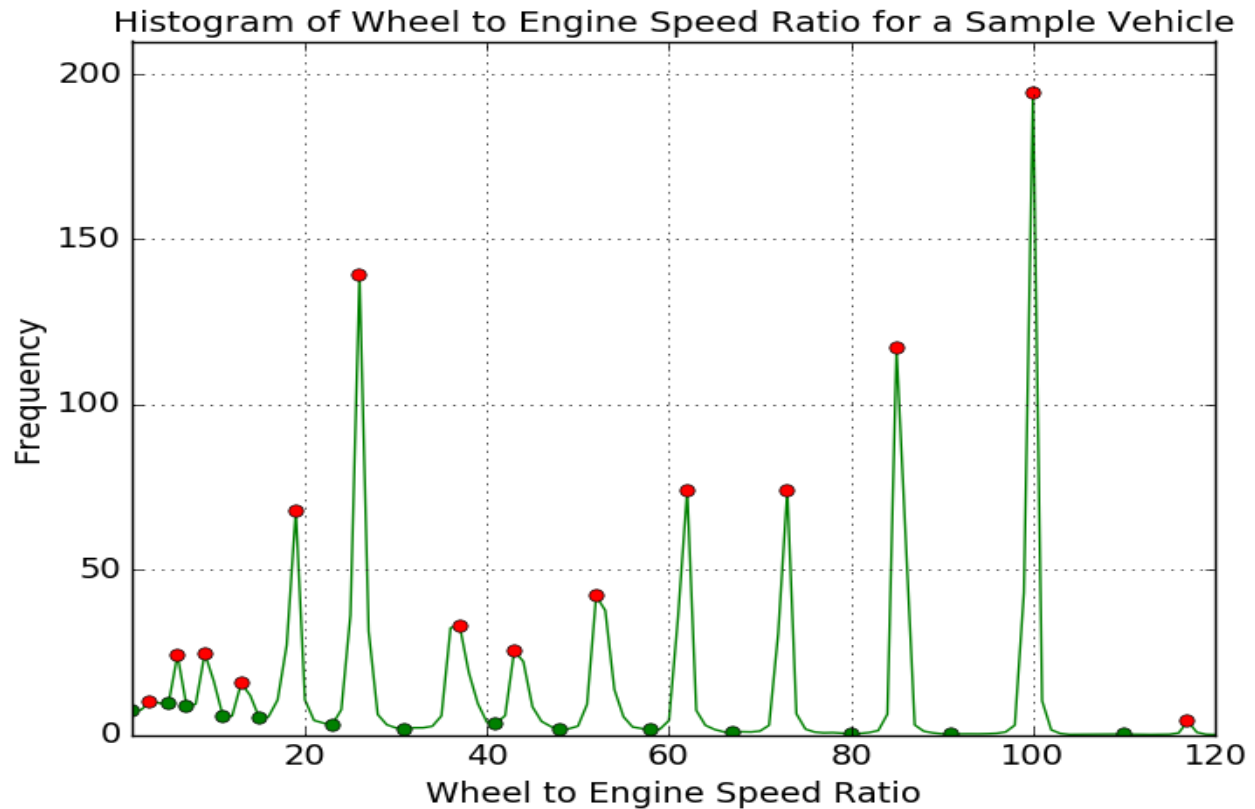




**Figure 47: Single vehicle sample distribution of vehicle speed to engine speed ratio data**

Higher vehicle speed to engine speed ratios are easier to both visualize and statistically separate, as the noise (time spent operating between gears) to signal (time spent operating in a given gear) ratio is much smaller at these high vehicle speeds.

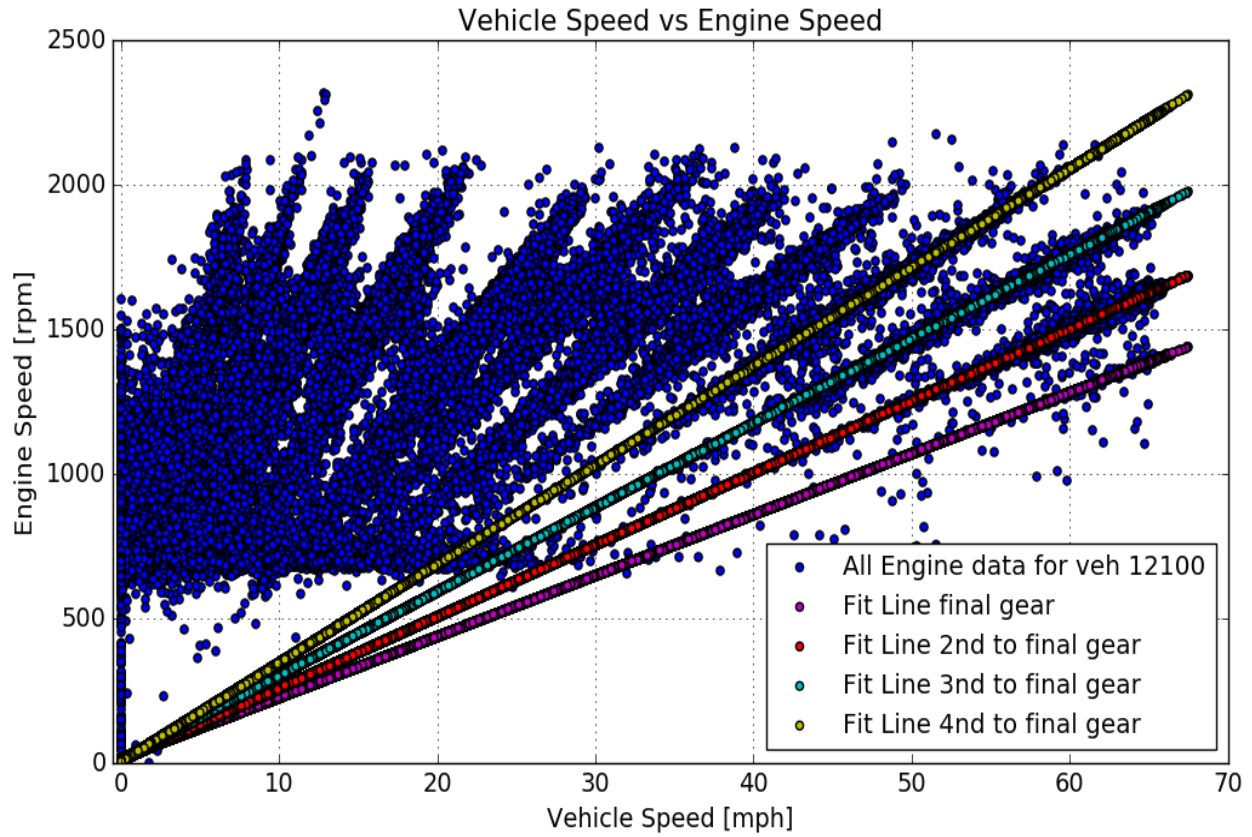
Having calculated vehicle speed to engine speed ratios, a kernel density estimator can be applied to the collection of ratios to convert the histogram shown in Figure 48 from a discrete set of binned data into a continuous function as shown in Figure 49.



**Figure 48: Single vehicle sample distribution of vehicle speed to engine speed ratio data with peaks identified**

Once the set of ratios have been converted to a continuous function, application of any number of peak detections algorithms can be performed to produce estimated vehicle speed to engine speed ratios.

Once the gear ratios have been identified, plotting linear fit curves using the identified ratios for slope allows researchers to visually validate that appropriate ratios have been selected. This process is shown in Figure 50.



**Figure 49: Identified gear ratios plotted atop sample vehicle speed vs. engine operating speed data**

## Appendix C: Comparison of Fleet DNA and MOVES Resampled Distributions by Cluster

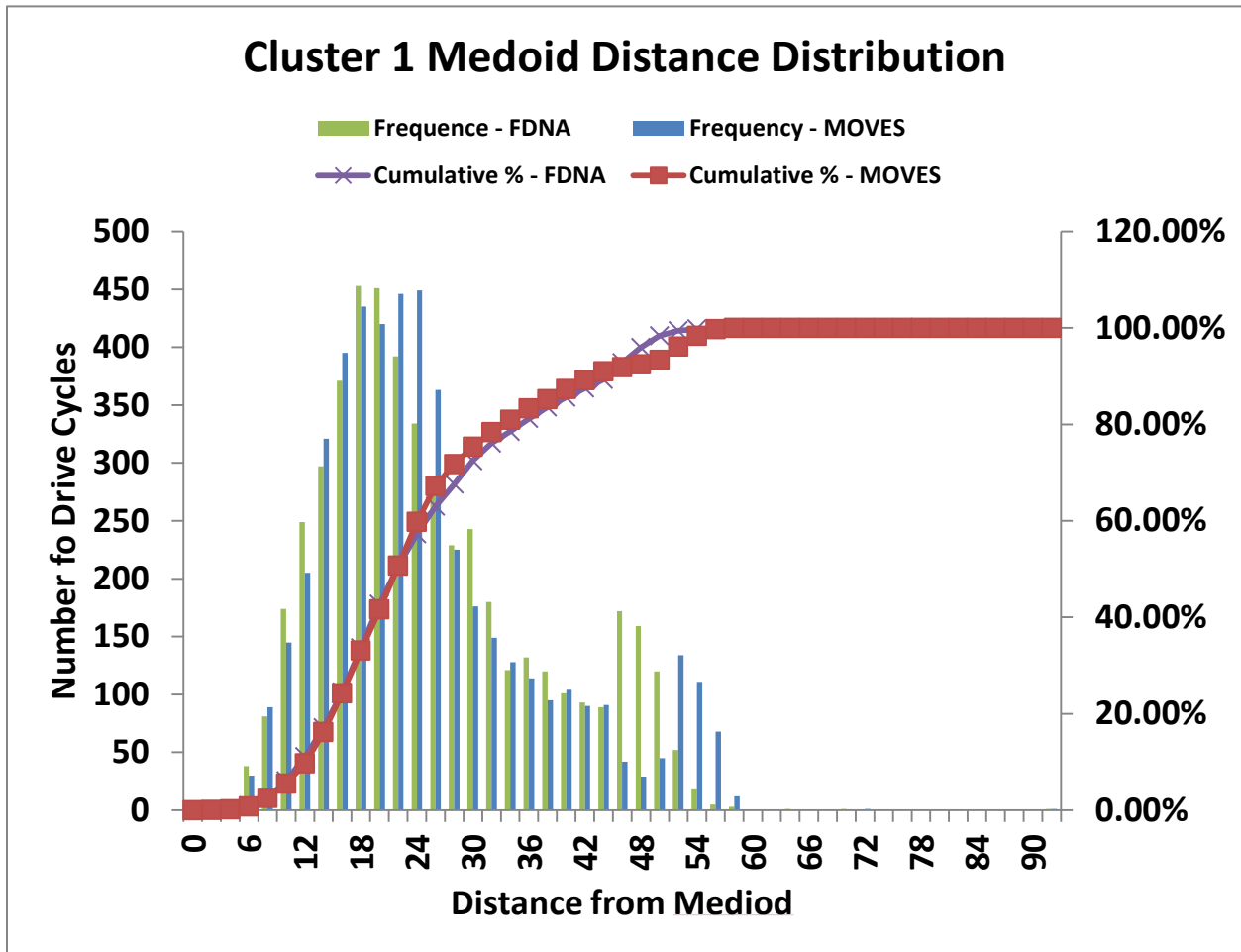


Figure 50: Comparison of multidimensional RMS Distance distributions for full fleet DNA Dataset and MOVES resampled population for Cluster 1

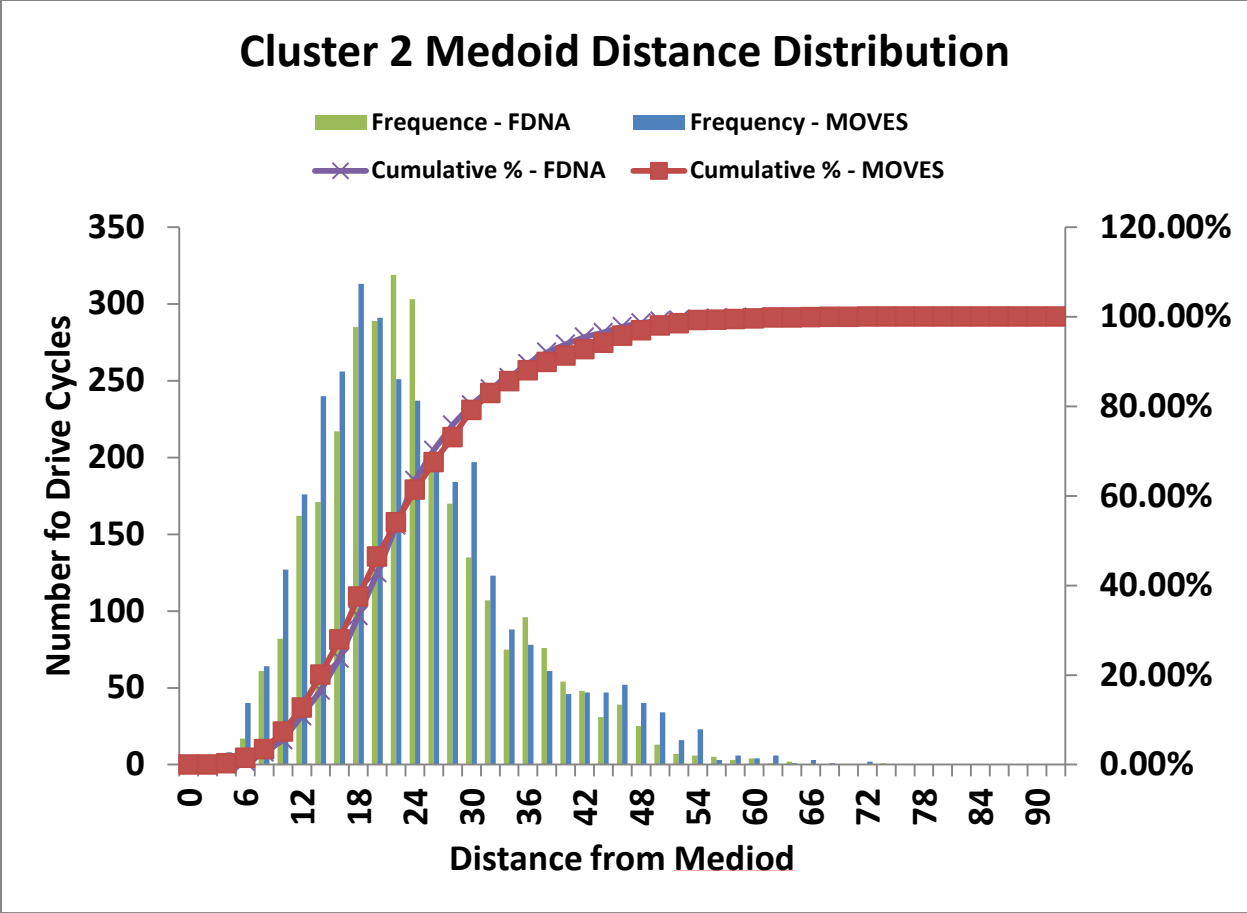


Figure 51: Comparison of multidimensional RMS distance distributions for full Fleet DNA dataset and MOVES resampled population for Cluster 2

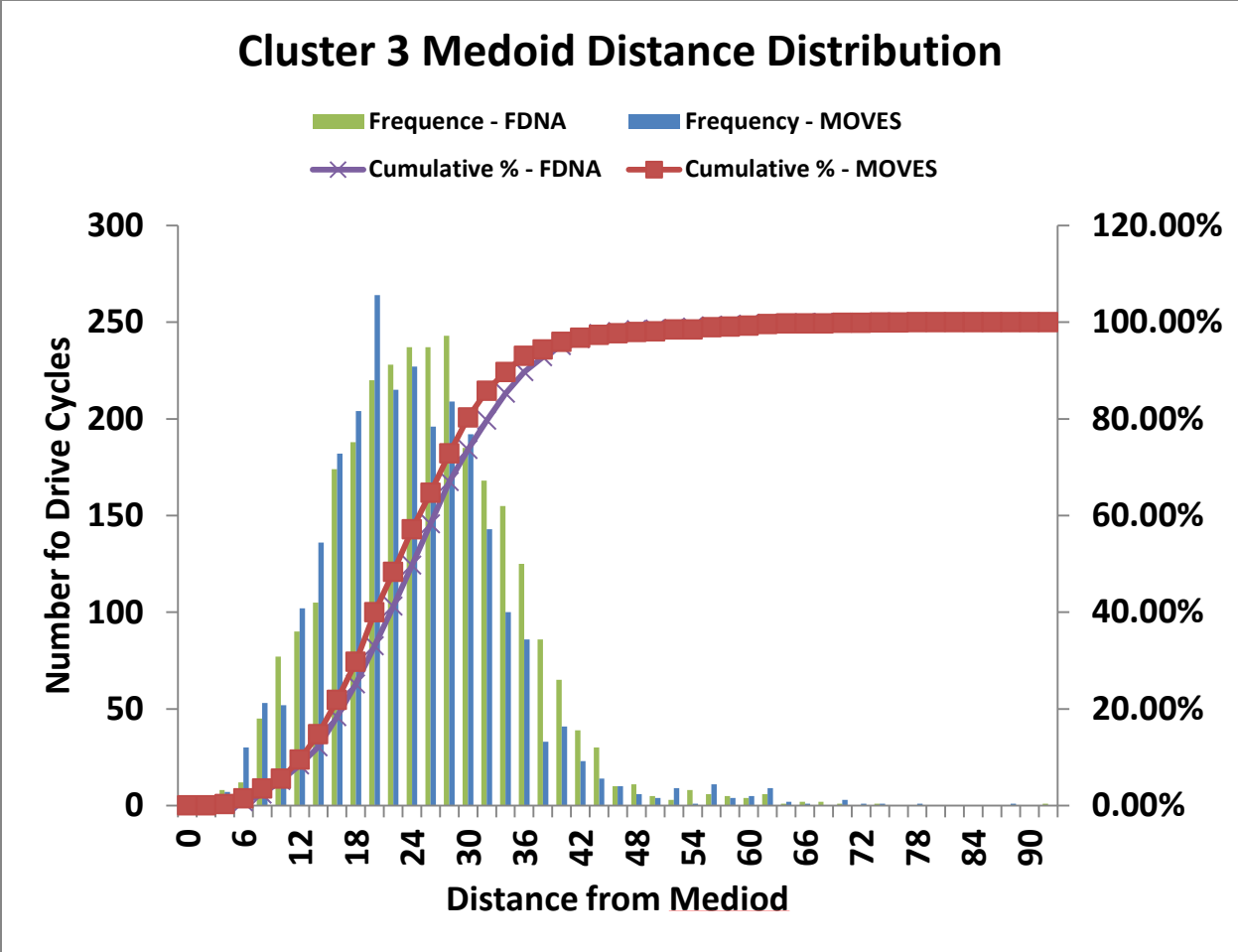


Figure 52. Comparison of multidimensional RMS distance distributions for full Fleet DNA dataset and MOVES resampled population for Cluster 3

## Appendix D: Detailed Mileage Accumulation Analysis for NREL-EPA Low Speed Transient Cycle

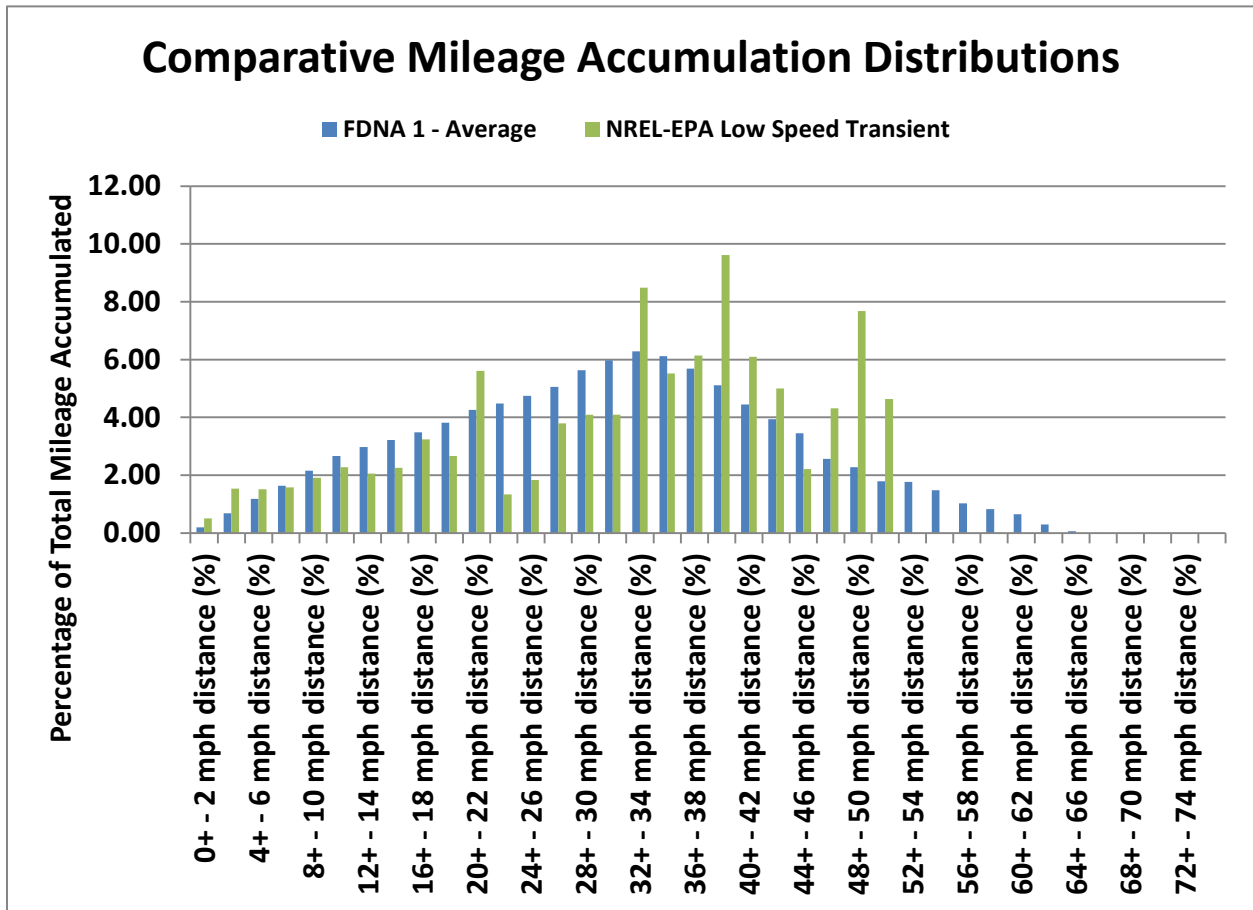


Figure 53. Detailed comparison of cluster and representative cycle mileage accumulation using 2-mph speed bin increments

Table 23. Detailed Comparison of Cluster and Representative Cycle Mileage Accumulation Using 2-mph Speed Bin Increments

	FDNA 1 - Average	FDNA 2 - Average	FDNA 3 - Average	NREL-EPA Low Speed Transient
0+ - 2 mph distance (%)	0.20	0.10	0.03	0.50
2+ - 4 mph distance (%)	0.69	0.33	0.11	1.54
4+ - 6 mph distance (%)	1.18	0.55	0.19	1.51
6+ - 8 mph distance (%)	1.64	0.77	0.21	1.58
8+ - 10 mph distance (%)	2.16	0.91	0.26	1.92

	<b>FDNA 1 - Average</b>	<b>FDNA 2 - Average</b>	<b>FDNA 3 - Average</b>	<b>NREL-EPA Low Speed Transient</b>
10+ - 12 mph distance (%)	2.66	1.03	0.30	2.28
12+ - 14 mph distance (%)	2.98	1.12	0.34	2.06
14+ - 16 mph distance (%)	3.22	1.20	0.36	2.26
16+ - 18 mph distance (%)	3.48	1.34	0.40	3.24
18+ - 20 mph distance (%)	3.82	1.41	0.44	2.66
20+ - 22 mph distance (%)	4.26	1.60	0.53	5.61
22+ - 24 mph distance (%)	4.48	1.84	0.56	1.33
24+ - 26 mph distance (%)	4.75	2.11	0.65	1.83
26+ - 28 mph distance (%)	5.06	2.40	0.77	3.80
28+ - 30 mph distance (%)	5.63	2.58	0.91	4.09
30+ - 32 mph distance (%)	5.98	2.77	1.04	4.09
32+ - 34 mph distance (%)	6.29	3.11	1.13	8.49
34+ - 36 mph distance (%)	6.11	3.41	1.16	5.52
36+ - 38 mph distance (%)	5.69	3.50	1.19	6.14
38+ - 40 mph distance (%)	5.11	3.52	1.31	9.61
40+ - 42 mph distance (%)	4.45	3.51	1.45	6.10
42+ - 44 mph distance (%)	3.94	3.67	1.55	5.00
44+ - 46 mph distance (%)	3.45	3.69	1.59	2.21
46+ - 48 mph distance (%)	2.57	3.58	1.68	4.32
48+ - 50 mph distance (%)	2.28	3.60	1.82	7.68



	<b>FDNA 1 - Average</b>	<b>FDNA 2 - Average</b>	<b>FDNA 3 - Average</b>	<b>NREL-EPA Low Speed Transient</b>
50+ - 52 mph distance (%)	1.79	3.69	2.01	4.64
52+ - 54 mph distance (%)	1.77	4.57	2.69	0.00
54+ - 56 mph distance (%)	1.48	5.98	4.01	0.00
56+ - 58 mph distance (%)	1.02	7.07	6.16	0.00
58+ - 60 mph distance (%)	0.83	7.65	9.19	0.00
60+ - 62 mph distance (%)	0.65	7.24	10.03	0.00
62+ - 64 mph distance (%)	0.30	4.75	16.96	0.00
64+ - 66 mph distance (%)	0.06	3.80	23.61	0.00
66+ - 68 mph distance (%)	0.01	1.33	4.63	0.00
68+ - 70 mph distance (%)	0.00	0.24	0.55	0.00
70+ - 72 mph distance (%)	0.00	0.04	0.11	0.00
72+ - 74 mph distance (%)	0.00	0.00	0.03	0.00
74+ mph distance (%)	0.00	0.00	0.04	0.00

## Appendix E: Generation of Representative High Speed Transient Component Using DRIVE

An approach similar to the one employed for the low speed transient cycle was also used to develop a high speed transient cycle using Fleet DNA data. The same top 50 drive cycles from the three clusters identified in Section 3 were fed into DRIVE and used as source data for the generation of the high speed transient cycle. Inputs such as the desired and minimum cycle duration and the set of drive cycle metrics to match were then swept to generate hundreds of unique drive cycles. These cycles were then compared using a non-weighted least squares approach. In the case of the high speed transient however, modifications were made to the DRIVE constraints to ensure the generation of a high speed transient cycle.

The key DRIVE constraints for the high speed transient cycle included:

- Targeted 2,050-second duration to match the CARB HHDDT Cruise Cycle duration.
- A minimum allowable cycle duration of 900 seconds was established to ensure sufficient test duration.
- Exclusion of any drive cycles with a maximum speed below 55 mph to ensure generated drive cycle is representative of high speed transient operation only

Shown in Figure 54 is the resulting representative high speed transient cycle developed from the Fleet DNA data. As one can see, the resulting transient is significantly shorter than the CARB HHDDT tcruise segment, possessing approximately half the duration.

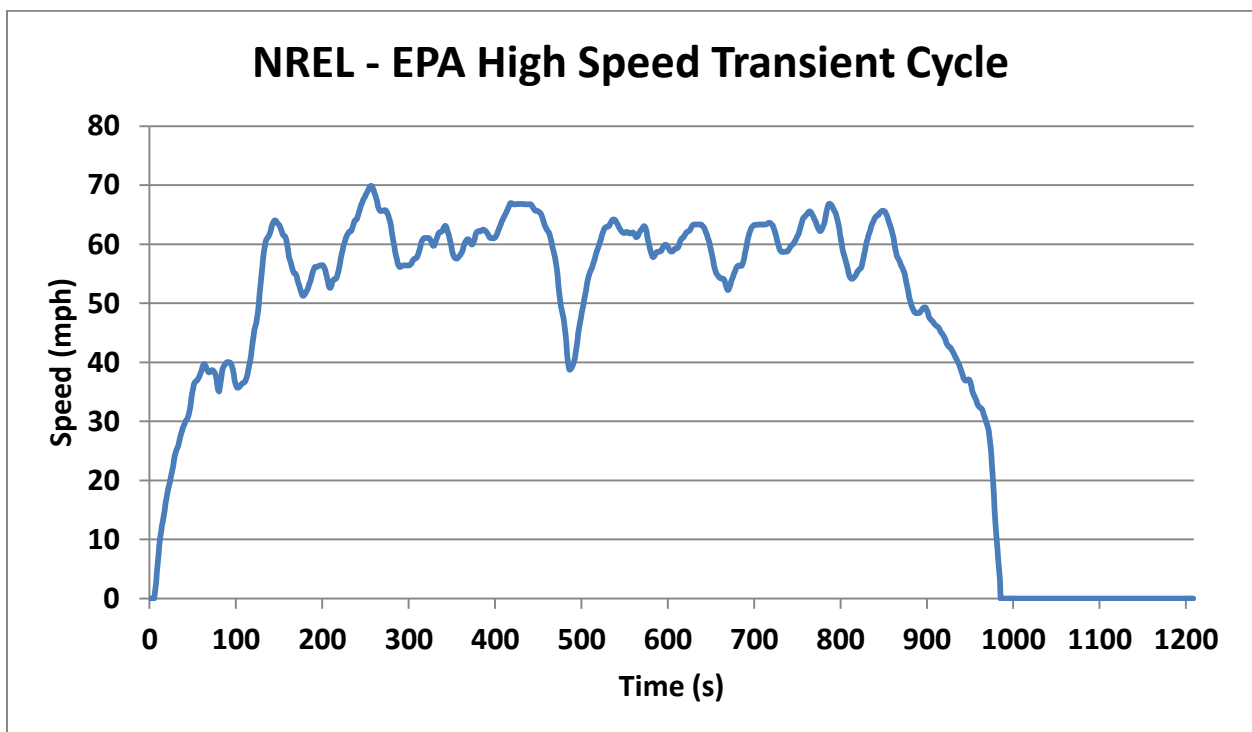


Figure 54. Speed–time trace for representative high-speed EPA transient cycle

Key statistics for the final high speed transient cycle include:

- 1,208 seconds in duration
- Total of 1 microtrip
- Maximum speed of ~70 mph
- Average driving speed of ~54 mph
- ~19% of total cycle duration is at zero speed.

Similar to the low speed transient, a target of 24.5% zero speed time as an upper bound was selected for the generation of the high speed transient. Since this cycle consists of a single high speed transient microtrip and was generated using real world data, the options for zero speed time collection are at either the beginning or end of the cycle. Due to the definition of microtrip (driving to next driving event including any zero speed time) zero speed time for this cycle is accumulated at the end of the cycle rather than the beginning.

Comparing the high speed transient to the complete WHVC, JE05, and ETC FIGE cycles as shown in Tables 24–26, one can see that the maximum speed achieved by the representative cycles closely match those of the standard test cycles, with the EPA high speed transient possessing a higher maximum driving speed than any of the standard drive cycles.

**Table 24. Comparison of Standard Cycle Drive Cycle Metrics**

Cluster	Aerodynamic Speed (ft/s)	Characteristic Acceleration (ft/s <sup>2</sup> )	Percent of mileage accumulated below 55 mph	Percentage of Cycle Time at Zero Speed	Number of Stops Per Mile	Average Driving Speed (mph)	Maximum Driving Speed (mph)	Standard Deviation of Driving Speed (mph)
CARB HHDDT Transient	43.36	0.49	100.00	15.57	1.40	18.20	47.50	13.38
WHVC	61.00	0.28	100.00	13.61	1.07	25.98	54.56	19.03
JE05	51.31	0.40	100.00	25.19	1.62	22.70	54.43	15.94
ETC FIGE	68.71	0.23	94.46	0.00	0.00	36.65	56.61	17.83
NREL - EPA Transient High	85.01	0.25	25.13	19.09	0.07	54.28	69.92	24.13
FDNA 1 - Average	50.59	0.55	96.45	47.30	1.71	22.62	62.39	13.74

**Table 25. Sub-45 mph Mileage Accumulation Comparison for Standard Cycles**

<b>Cycle</b>	<b>0+ - 5 mph distance (%)</b>	<b>5+ - 10 mph distance (%)</b>	<b>10+ - 15 mph distance (%)</b>	<b>15+ - 20 mph distance (%)</b>	<b>20+ - 25 mph distance (%)</b>	<b>25+ - 30 mph distance (%)</b>	<b>30+ - 35 mph distance (%)</b>	<b>35+ - 40 mph distance (%)</b>	<b>40+ - 45 mph distance (%)</b>
CARB HHDDT Transient	1.25	12.72	3.97	9.02	18.20	10.20	15.72	11.96	8.29
WHVC	1.07	3.87	6.11	7.47	4.34	6.34	9.30	7.42	8.09
JE05	0.97	4.68	8.98	9.53	9.32	13.53	12.14	12.10	2.06
ETC FIGE	0.19	0.60	2.40	4.24	3.81	2.36	1.31	3.50	17.49
NREL - EPA Transient High	0.02	0.08	0.16	0.29	0.42	0.84	1.46	6.19	2.87
FDNA 1 - Average	1.39	4.35	7.11	8.90	11.05	12.95	15.15	14.06	10.69

**Table 26. 45+ mph Mileage Accumulation Comparison for Standard Cycles**

<b>Cycle</b>	<b>45+ - 50 mph distance (%)</b>	<b>50+ - 55 mph distance (%)</b>	<b>55+ - 60 mph distance (%)</b>	<b>60+ - 65 mph distance (%)</b>	<b>65+ - 70 mph distance (%)</b>	<b>70+ - mph distance (%)</b>
CARB HHDDT Transient	8.66	0.00	0.00	0.00	0.00	0.00
WHVC	7.28	38.71	0.00	0.00	0.00	0.00
JE05	9.46	17.23	0.00	0.00	0.00	0.00
ETC FIGE	15.41	43.14	5.54	0.00	0.00	0.00
NREL - EPA Transient High	4.77	8.03	23.23	38.74	12.89	0.00
FDNA 1 - Average	6.57	4.22	2.52	1.00	0.04	0.00

Examining Table 26 in greater detail, we can see that the high speed EPA transient cycle is the only cycle analyzed with more than 5% of its mileage accumulated at speeds in excess of 60 mph. The bulk of the ETC (43.14%) and WHVC (38.71%) cycle mileage is accumulated at speeds between 50 and 55 mph. Looking at the unweighted combined drive cycle mileage accumulation results, the differences in high speed mileage accumulation between the cycles becomes apparent. None of the standard cycles examined accumulate mileage at speeds in excess of 60 mph, while the ETC FIGE is the only standard chassis test cycle which accumulates mileage at speeds in excess of 55 mph.

**Table 27. Comparison of Cycle Acceleration Statistics**

<b>Cluster</b>	<b>Acceleration events (1/mile)</b>	<b>Deceleration events (1/mile)</b>	<b>Max accel (ft/s<sup>2</sup>)</b>	<b>Max decel (ft/s<sup>2</sup>)</b>	<b>Avg accel (ft/s<sup>2</sup>)</b>	<b>Avg decel (ft/s<sup>2</sup>)</b>
CARB HHDDT Transient	10.52	10.17	4.25	-4.06	0.94	-1.24
WHVC	13.28	13.10	3.84	-4.99	0.71	-0.81
JPJE05	9.50	9.38	5.22	-6.01	1.03	-1.08
ETC FIGE	11.95	11.73	12.58	-13.21	0.63	-0.72
NREL-EPA Transient High	2.64	2.64	2.51	-4.57	0.52	-0.60
FDNA 1 - Average	7.18	7.19	9.90	-9.98	1.25	-1.43

Finally, examining the acceleration and deceleration behavior of the high speed transient developed by NREL shown in Table 27, one observes that the NREL-EPA high speed transient cycle possesses significantly fewer acceleration/deceleration events per miles (~20%) of comparable standard test cycles, while displaying similar acceleration/deceleration rates. This makes sense intuitively as the high speed transient was designed to represent high speed transient driving behavior while the standard cycles examined are all composites except for the CARB HHDDT Transient. This type of high speed transient cycle mimics the type of driving consistent with mid and high speed highway vocational vehicle operation, illustrative of vehicles with fewer and gentler accelerations cruising at highway speeds.