High-Fidelity Heavy-Duty Vehicle Modeling Using Sparse Telematics Data

Kyle Carow  Western Michigan University
Nathaniel Cantwell and Andrej Ivanco  Allison Transmission Inc.
Jacob Holden, Chad Baker, and Eric Miller  National Renewable Energy Laboratory
Zachary Asher  Western Michigan University


Abstract

Heavy-duty commercial vehicles consume a significant amount of energy due to their large size and mass, directly leading to vehicle operators prioritizing energy efficiency to reduce operational costs and comply with environmental regulations. One tool that can be used for the evaluation of energy efficiency in heavy-duty vehicles is the evaluation of energy efficiency using vehicle modeling and simulation. Simulation provides a path for energy efficiency improvement by allowing rapid experimentation of different vehicle characteristics on fuel consumption without the need for costly physical prototyping. The research presented in this paper focuses on using real-world, sparsely sampled telematics data from a large fleet of heavy-duty vehicles to create high-fidelity models for simulation. Samples in the telematics dataset are collected sporadically, resulting in sparse data with an infrequent and irregular sampling rate. Captured in the dataset was geospatial information, time series measurements, and vehicle-specific metadata from a subset of 96 vehicles from varied geographic regions across North America. A series of custom algorithms was developed to process vehicle data and derive both vehicle model input parameters and representative drive cycles. Derived models provide a basis on which to simulate real-world vehicles and iterate on vehicle aerodynamics, auxiliary power loads, transmission shift schedules, and other parameters to achieve reduced fuel consumption and increase energy efficiency. Notably, these models were developed without the use of expensive field data collection, using only data collected through fleet telematics. Processed representative drive cycles are used to validate the fuel economy of derived models. The models developed through this research allow for more representative vehicle simulations with increased flexibility regarding vehicle-to-vehicle variations.

Introduction

Heavy-duty vehicles (HDVs) consume a greater amount of fuel compared to light-duty vehicles because of their substantial mass, significant aerodynamic drag, and high auxiliary power loads. The greater consumption of fuel in HDVs leads to increased expenses for HDV fleet companies. Fueling accounts for 24% of motor carrier operating costs, second only to driver wages [1]. In addition, HDV fuel consumption contributes significantly to harmful greenhouse gas emissions. According to the U.S. Environmental Protection Agency (EPA), the combustion of fuel in freight trucks comprises the second largest source of CO2 emissions within the transportation sector, at 23.6% [2]. Due to this, the EPA has set targets through the Clean Trucks Plan which sets increasingly strict limits for HDV greenhouse gas emissions [3]. These factors prompt HDV corporations to explore optimization of fuel efficiency.

Energy efficiency optimization can be achieved through modeling and simulation of vehicles through software. While physical experimentation on vehicle parameters is expensive and time-consuming, simulation using validated models allows for rapid results through software alone. Modeling of HDVs presents a unique challenge, however, as unlike light-duty vehicles, HDVs possess a high degree of specialization as a result of vehicle vocation. Vehicle specialization causes great variations in mass, shape, and auxiliary power loads from vehicle to vehicle. In order to create sufficiently representative HDV models, model parameters must be tuned on a vehicle-to-vehicle basis, necessitating the use of a flexible and customizable vehicle simulation software, as well as an abundance of tuning data.

The Future Automotive Systems Technology Simulator (FASTSim), developed by the National Renewable Energy Laboratory (NREL), is a high-level advanced vehicle...
powertrain systems analysis tool that provides a simple yet extremely flexible vehicle simulation platform [4]. Due to FASTSim’s high-level nature, simulations execute very quickly, allowing rapid iteration. In addition, as FASTSim is written entirely in Python, it is highly transparent and customizable by nature. These features make FASTSim an excellent choice for highly customized vehicle modeling and simulation, as is necessary for HDVs.

Parameters must be tuned through the use of real-world vehicle performance data to generate a model with sufficient fidelity to be useful. Traditionally, the threshold for vehicle model validation has been an absolute error of 3% between measured and simulated fuel economy. In this paper, the phrase “high-fidelity” will refer to models achieving absolute errors below this threshold, as the phrase has no consistently used definition other than a model that simulates results to a subjectively acceptable degree. Real-world fleet telematics is one potential source of tuning data, containing a wide range of useful data such as geospatial information, time series measurements, and vehicle-specific metadata. However, it does not allow for straightforward model creation, as measurements are often sparsely sampled or sometimes entirely absent. For this reason, a high degree of data processing is often required to derive useful model parameters from fleet telematics data.

Telematics data has previously been used to characterize and optimize fuel consumption and vehicle emissions. A study of HDVs in the Houston-Galveston Area successfully utilized telematics measurements to characterize idling and found an average idling time of 185 minutes per day for the analyzed dataset [5]. Work by Mane, Djordjevic, and Ghosh shows that HDV telematics data can be leveraged to construct a framework for incentivizing HDV drivers to adopt more fuel efficient driving behavior [6]. Telematics data has also been used to identify potential for fuel economy improvement through HDV “platooning,” where multiple vehicles drive in close proximity in order to minimize aerodynamic drag energy losses [7]. This research demonstrates that telematics data is an extremely useful resource for the characterization and optimization of HDV fuel economy.

Other research extracted fuel economy information from telematics data without the use of vehicle modeling and simulation. One paper studied utilizes Kalman Filters created from telematics data to predict fuel economy for different hypothetical departure times over a predetermined route [8]. Another paper explores the use of geospatial data and CAN bus measurements to predict fuel economy using machine learning [9]. Both of these rely on the usage of collected data to calculate fuel economy measurements, rather than extrapolation to hypothetical drive cycles, as is possible with vehicle simulation. In general, little published work has been done to use telematics data as source of vehicle model derivation and validation, despite the potential for low-cost simulation and fuel cost savings.

This paper seeks to remedy this by providing a detailed and comprehensive methodology for using a large fleet telematics dataset to derive validated HDV fuel economy models. Due to the preliminary nature of the research, the primary aim of this paper is to show that HDV models can be generated and validated using sparse and poor-quality telematics data from a large and varied fleet of vehicles. A dataset consisting of approximately 100 vehicles of many vocations from regions across North America was processed using a custom algorithm to derive model parameters. Representative drive cycles from the telematics dataset were used in FASTSim vehicle simulations to validate models against measured fuel economy.

Methodology

Telematics Data Overview

The telematics dataset used in this study consists of data from approximately 100 HDVs of multiple vocations from across North America. The dataset includes time series vehicle measurements, time series geospatial data, and vehicle metadata. Time series measurements encompass a majority of important signals such as velocity and fuel consumption. Time series geospatial data includes GPS coordinates over time. Vehicle metadata provides information on the types of vehicles captured in the dataset, mostly regarding engine identification codes and vehicle vocation. Each type of data comes with its own set of unique challenges for adaptation to HDV modeling.

A challenge common to all time series measurements in the dataset is the sparsity and uneven frequency of sample collection. Due to bandwidth limitations, the Ramer-Douglas-Peucker algorithm was utilized for data reduction in the sample collection process. Originally developed for image vector graphic optimization, this algorithm reduces the number of data points, maintaining overall curve detail at the expense of an acceptable level of error [10, 11]. The reduction in sample count results in an uneven sampling rate and loss of temporal resolution, making the application of traditional data processing algorithms or any other process dependent on evenly spaced samples more complicated.

During data collection, certain signals appear to be prioritized, resulting in great variations in sample counts measurement-to-measurement. The signal with the highest sample count was geospatial data, followed by measured vehicle velocity. Other signals with high sample counts included cumulative fuel consumption, engine speed, and miscellaneous engine fluid temperatures. Table 1 lists sample counts by measurement signal for the 15 most common time series measurements. The same data is shown graphically in Figure 1, with signal names removed for simplicity.

The presence and frequency of time series measurement signals varies greatly between vehicles in the data, where some measurements are sampled very sparsely or not at all in certain vehicles. Low sample counts present a challenge for use in modeling, as for some vehicles critical information like odometer readings can be too infrequently sampled to be usable. In this study, “sparse data” will refer to data that is sampled too infrequently and irregularly to be useful for straightforward vehicle modeling. In addition, some samples had gaps, where samples were uncaptured for an extended period of time. These gaps are especially problematic for signals like vehicle velocity, where signal details critical to fuel
economy may be missing. Moreover, information about power takeoff loads, ambient environment, auxiliary loads, and some other operating conditions influential to fuel economy were missing or insufficiently captured in the dataset for use in modeling.

The dataset consists of vehicles from across North America, with most geospatial data points located in the United States and some in southern Canada. Geospatial data consisted of time series latitude and longitude measurements, and had the highest sample count of all time series signals in the dataset. Due to this, the primary challenge presented by the geospatial data was the amount of processing required to derive useful information rather than sample sparsity.

Obtained by augmenting available vehicle information with National Highway Traffic Safety Administration (NHTSA) data [12], the metadata includes details on the HDVs in the telematics dataset, providing engine parameters, transmission parameters, gross vehicle weight rating, and other useful information. Vehicles from a wide variety of vocations were captured in the dataset, including coach and school buses, construction vehicles, refuse trucks, emergency service vehicles, and utility company service vehicles. This provided a diverse set of HDV telematics data to work with.

### Table 1

Sample counts for the 15 most common telematics measurement signals. See Figure 1 for graphical representation.

<table>
<thead>
<tr>
<th>Signal</th>
<th>Total Sample Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Geospatial coordinates</td>
<td>1,150,367</td>
</tr>
<tr>
<td>2 Measured velocity</td>
<td>730,650</td>
</tr>
<tr>
<td>3 Fan drive state</td>
<td>477,016</td>
</tr>
<tr>
<td>4 Engine speed</td>
<td>360,669</td>
</tr>
<tr>
<td>5 Total cumulative fuel consumption</td>
<td>246,658</td>
</tr>
<tr>
<td>6 Ambient air temperature</td>
<td>205,106</td>
</tr>
<tr>
<td>7 Longitudinal acceleration</td>
<td>172,060</td>
</tr>
<tr>
<td>8 Acceleration (up-down)</td>
<td>161,135</td>
</tr>
<tr>
<td>9 Acceleration (side-to-side)</td>
<td>129,050</td>
</tr>
<tr>
<td>10 Transmission oil temperature</td>
<td>87,265</td>
</tr>
<tr>
<td>11 Acceleration (forward/braking)</td>
<td>78,714</td>
</tr>
<tr>
<td>12 Engine coolant temperature</td>
<td>76,173</td>
</tr>
<tr>
<td>13 Odometer</td>
<td>74,358</td>
</tr>
<tr>
<td>14 Cranking voltage</td>
<td>73,601</td>
</tr>
<tr>
<td>15 Torque converter lockup count</td>
<td>46,079</td>
</tr>
</tbody>
</table>

### Figure 1

Sorted bar plot showing counts for the 15 most common telematics measurement signals. Note that the y-axis is in millions.

### Table 2

Velocity signal processing summary table. Visualization of the results at each step can be found in Table 2.

<table>
<thead>
<tr>
<th>Step</th>
<th>Step Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Preprocessing • Organize data for further processing</td>
</tr>
<tr>
<td></td>
<td>• Derive secondary velocity signal from geospatial data</td>
</tr>
<tr>
<td>1</td>
<td>Stop Capturing • Apply discontinuity correction algorithm and set small velocities to zero</td>
</tr>
<tr>
<td>2</td>
<td>Signal Fusion • Fuse both velocity signals into one using custom algorithm</td>
</tr>
<tr>
<td>3</td>
<td>Upsampling • Resample and interpolate signal to regular 1 Hz frequency</td>
</tr>
<tr>
<td>4</td>
<td>Smoothing • Apply Savitzky-Golay filtering to smooth signal while retaining fine detail</td>
</tr>
</tbody>
</table>

### Figure 2

Map of all telematics geospatial data.

### Model Derivation

As a starting point, the two pre-validated HDV models provided with FASTSim were used to create a base model. Many model parameters remained the same between the two and were reused in the base model. Some parameters were not the same for both models so the average value of the two parameters was used in the base model instead, where applicable. The result was the creation of a generic HDV model, but simulation results were poor. This was expected due to the wide variety of HDV characteristics.

Dynamic model variation was implemented using telematics data where possible to provide more representative parameter values. Vehicle mass, number of wheels, and other parameters were varied automatically when appropriate metadata was available, resulting in much higher fidelity simulation performance, as detailed in the ‘Results and Discussion’ section.

The use of vehicle metadata was enough to produce high-fidelity models. However, there are vehicles where metadata is incomplete or insufficient for deriving models. This will be the subject of future work, as described in the ‘Conclusions’ section.

### Drive Cycle Derivation

Along with vehicle models, speed-by-time drive cycles are critical inputs to a backward-looking simulation, establishing
the driving conditions and serving as the framework for iterative calculation. Drive cycles derived from vehicle speed measurements can be used to perform a simulation in order to validate vehicle models, allowing measured and simulated vehicle performance to be compared directly. These vehicle speed measurements, however, must adequately represent the true driving behavior for use as a model validation drive cycle. This necessitates extensive drive cycle processing for sparse velocity signals. A summary of all velocity signal processing steps described in this section, along with the associated simulation results, can be found in Table 3. A flowchart of velocity signal processing depicting all steps in detail is shown in Figure 5.

As the telematics measurements include many driving sessions over the course of multiple months, slices of data were selected for simulation. The two requirements to select data slices were a duration of around 30 to 60 minutes, and having at least 5 samples of each signal important to fuel economy derivation (odometer readings and cumulative fuel consumption).

Initially, the measured velocity signal was used as drive cycles for model validation. It quickly became apparent that large gaps in the velocity signal were causing issues with simulation results. While FASTSim allows unevenly spaced drive cycles, large time steps from one data point to another cause breakdowns in simulation fidelity. In addition, the use of the Ramer-Douglas-Peucker algorithm and other data loss led to loss of nuanced detail in the velocity signal. Velocity signal events and details such as accelerations, velocity peaks, stops, and overall signal smoothness are highly influential in the overall vehicle fuel economy [13]. If any of these events or details are not adequately captured due to missing data, simulation fidelity will be negatively impacted.

To remedy issues caused by gaps in data, a secondary velocity signal was derived using geospatial data. Iterating through the data, geodesic distances and timestamp differences were calculated between points. By dividing calculated distances by time differences, a new geospatial velocity signal was generated. The average timestamp between each data slice was selected for simulation. The two requirements to select data slices were a duration of around 30 to 60 minutes, and having at least 5 samples of each signal important to fuel economy derivation (odometer readings and cumulative fuel consumption).

<table>
<thead>
<tr>
<th>Step</th>
<th>Signal</th>
<th>Plots</th>
<th>FE Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Preprocessing</td>
<td>Measured</td>
<td>Simulated FE: 6.77 MPG (10.55% error)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Geospatial</td>
<td>Simulated FE: 8.24 MPG (8.83% error)</td>
</tr>
<tr>
<td>1</td>
<td>Stop Capturing</td>
<td>Measured</td>
<td>Simulated FE: 6.77 MPG (10.57% error)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Geospatial</td>
<td>Simulated FE: 8.24 MPG (8.83% error)</td>
</tr>
<tr>
<td>2</td>
<td>Signal Fusion</td>
<td></td>
<td>Simulated FE: 7.34 MPG (3.05% error)</td>
</tr>
<tr>
<td>3</td>
<td>Upsampling</td>
<td></td>
<td>Simulated FE: 7.34 MPG (3.05% error)</td>
</tr>
<tr>
<td>4</td>
<td>Smoothing</td>
<td></td>
<td>Simulated FE: 7.44 MPG (1.67% error)</td>
</tr>
</tbody>
</table>
point was used as the timestamp for the newly created velocity signal. Points derived from data with large timestamp differences or unrealistically high accelerations were removed from the derived signal. The presence of a secondary signal provides details not captured in the measured velocity signal and allows for a single velocity signal to be derived. Before this is possible, velocity signal features influential to fuel economy such as stops must be accurately represented. In both velocity signals, vehicle stops are occasionally captured inadequately due to signal gaps. At many of these gaps, both velocity signals “jump” from a moving velocity to a stopped velocity of approximately zero, or from a stopped velocity to a moving velocity, leaving expansive straight lines across gaps. These features lead to derivation of inaccurate drive cycles and must be remedied.

To better capture vehicle stops, a custom algorithm was developed to detect discontinuities in velocity. The algorithm is applied on the entirety of a velocity signal, iteratively identifying samples that meet all of the following criteria: (1) there is a sufficient gap between analyzed sample and the previous sample, (2) the velocity at the analyzed sample is at least a threshold value, and (3) the velocity at the previous sample is at most a secondary threshold value. The custom discontinuity detection algorithm also detects downward velocity drops by looking forward at the next sample, with the third criterion satisfied if the velocity at the next sample is at most a secondary threshold value. Examples of both upward and downward velocity discontinuities are shown in Figure 3, outlined with dashed boxes.

After detection of velocity discontinuities, zero-velocity samples are imputed in the velocity signal. For upward velocity discontinuities, where a near-zero velocity is followed by a gap in samples before a nonzero velocity, a zero-velocity point is created in the signal just before the nonzero sample. The temporal location of this data point is found in a similar way as in upward velocity discontinuities, but uses the acceleration before the nonzero velocity if it is negative. Otherwise, the timestamp of the imputed zero-velocity point is found using a reasonable negative acceleration value. The timestamp of imputed zero-velocity points is found by adding the time difference calculated with Equation 1 to the analyzed sample timestamp. In addition to discontinuity correction, velocities close to zero were set to zero to better capture vehicle stops.

\[
\Delta t = -\frac{\nu}{a}
\]  

Where,
\(\Delta t\) = time relative to analyzed sample to place zero-velocity value
\(\nu\) = velocity of analyzed sample
\(a\) = acceleration, selected as detailed above.

Once major problems with the velocity signals had been addressed, the two signals needed to be fused into a single velocity signal to be used as a drive cycle. Due to the existence of gaps, straightforward signal averaging and other simplistic methods could not be used. Instead, a custom algorithm was developed to iteratively fuse the measured and geospatial velocity signals into a single output velocity signal. The algorithm first splits the signals into 60-second time intervals. For each time interval, the algorithm chooses one of two options to construct an output signal depending on interval sample counts. If one signal has at least twice the number of samples compared to the other signal, the signal with more samples is the output for that interval. If no signal has at least twice the number of samples than the other, the two signals are interpolated and averaged together. This algorithm fuses two signals together in a novel method robust to signal gaps, preserving detail from either signal where possible. A demonstration of the custom signal fusion methodology is shown in Figure 4.

For optimal simulation performance in FASTSim, drive cycles should not have large time steps between samples. Typically, a regular sampling rate of 1 Hz is most common for fuel economy evaluation. To implement this sampling rate, nonlinear interpolation was initially used to upsample the
fused signal, but was found to produce undesirable features such as velocity peaks and valleys when applied over large gaps, so linear time-based interpolation was used instead. Upsampling to a regular frequency also enables the utilization of traditional signal processing techniques, such as Savitzky-Golay filtering.

Savitzky-Golay filtering is often used for signal smoothing due to its simplicity and appropriate handling of signal endpoints. However, it requires evenly spaced samples, like many other digital filters. After upsampling, the velocity signal could be smoothed using this algorithm. Smoothing parameters such as window size and polynomial order were carefully selected to preserve as many fuel-economy-critical details (stops, peaks, and accelerations) as possible.

The end result of velocity signal processing was transformation from one measured velocity signal and geospatial data into a single, more reliable velocity signal, easily applied as a FASTSim simulation drive cycle. A summary of velocity signal processing at each step is shown in Table 2. A flowchart visualizing signal processing at each step is shown in Figure 5.

Model Validation

To validate derived models, simulated fuel economy over processed drive cycles must be compared to measured fuel economy, with a target absolute error between the two being within 3%. Because fuel economy is calculated over the entire drive cycle, rather than at one instantaneous point, measured fuel economy needed to be calculated from sparse odometer readings and fuel consumption measurements.

Measured fuel economy can be derived by comparing samples of odometer readings and cumulative fuel consumption from the start and end of data slices. Distance traveled can be calculated by subtracting the first odometer reading in the drive cycle from the last odometer reading. Fuel consumption over a data slice can be calculated similarly by subtracting the first value of the total fuel consumption from the last fuel consumption value. Dividing distance traveled by fuel consumption provides a measured fuel economy value. However, due to sparsity of data, the first and last signal values within the data slice available may not be temporally close to data slice endpoints, which could cause measured fuel economy to be inaccurate. That is, if the available odometer and fuel consumption values are far from drive cycle endpoints, it is possible the fuel economy derived could differ from its true value. To minimize this risk, before deriving measured fuel economy, the odometer and fuel consumption signals were both linearly interpolated and linearly extrapolated using available signal processing libraries to provide measurement samples close to data slice endpoints. While alternative methods of odometer and fuel consumption interpolation were considered, linear interpolation was selected as the data was essentially linear. A minimum sample count for both signals used to derive measured fuel economy was enforced to ensure measured data adequately captured real-world behavior.

As distance traveled is not an input to FASTSim drive cycles, simulated fuel economy is derived by dividing a distance derived from the integral of the velocity by the total simulated fuel consumption. This is useful for applications where total distance is unavailable, but the telematics data provides a better indication of the true total distance traveled. This measured distance traveled was divided by simulated fuel consumption to provide an alternative simulated fuel economy with error from FASTSim distance calculation minimized.

Results and Discussion

Overall Results

Due to the vast amount of vehicles and trips in the dataset, a single vehicle was selected for result analysis, however, significant reductions in absolute errors occurred for the vast majority of vehicles tested. The application of model and drive
cycle derivation described in this study resulted in a simulated fuel economy with an absolute error of 1.67% when compared to measured fuel economy. This is below the traditional fuel economy model validation threshold of 3%. This shows the methodology presented in this paper can be used to successfully derive and validate HDV models. The cumulative fuel consumption over time from measurements and simulation are compared in Figure 6.

**Velocity Signal Processing Results**

Derivation of a secondary geospatial velocity provided a secondary source of velocity data, however both signals had issues with sparsity and missing features. This is shown in Figure 7 through comparison of the two signals. In sections where the two signals follow each other closely, measured-versus-simulated fuel economy absolute errors are relatively low. In sections where a feature is missing from one signal due to inadequate sampling, errors are larger. This is shown in Figure 6.

Missing features are undesirable because they negatively affect drive cycle fidelity. This can be seen in the high absolute errors when unprocessed signals are used as simulation drive cycles. Use of unprocessed (but upsampled, for FASTSim performance) measured velocity resulted in an error of 10.55%. Similarly, use of geospatial velocity resulted in an error of 8.83%.

Discontinuities at vehicle stops originally missing from the velocity signals were recaptured using a custom algorithm. For the specific drive cycle tested, absolute errors did not noticeably decrease, likely due to lack of discontinuities. For a vehicle with more occurrences of discontinuities, application of the correction algorithm to measured velocity resulted in a reduction of absolute error by 0.22 percentage points. Application of the algorithm to geospatial velocity resulted in a reduction of absolute error by 0.84 percentage points. Examples of stop capturing algorithm results are shown in Figure 8 and Figure 9.

After stops were captured, a custom signal fusion algorithm was applied to transform the two velocity signals into one through consideration of sample counts in 60-second time intervals. Application of the signal fusion algorithm resulted in an absolute error of 3.05%, a considerable improvement from the absolute errors achieved in the previous signal processing step, 10.55% and 8.83%.

To achieve optimal simulation performance, signals must be upsampled before use in FASTSim. For this reason, all fuel economy results shown are simulated using upsampled
velocity signals. This is also why fuel economy simulation fidelity does not increase due to upsampling.

As fuel economy results are impacted considerably by signal smoothness, Savitzky-Golay filtering was applied to the fused velocity signal. This resulted in a reduction in measured-versus-simulated fuel economy absolute error from 3.05% to 1.67%. A summary of velocity signal processing and associated simulation results is shown in Table 3.

Conclusions

This study showcases the methodology and results from utilizing real-world telematics data for heavy-duty vehicle (HDV) modeling and simulation in the Future Automotive Systems Technology Simulator (FASTSim). Models were derived using FASTSim pre-validated HDV models as a starting point, from which a basic HDV model was constructed and dynamically varied using vehicle metadata available. Models were simulated using vehicle velocity signals from both velocity measurements and geospatial data. Simulated fuel economy was compared to a derived measured fuel economy value and was found to achieve an absolute error of 1.67%, within the traditional fuel economy model validation threshold of 3%. It has been shown that sparse telematics data can be used to generate validated HDV models that allow for the extrapolation of fuel economy performance to hypothetical drive cycles and the examination of the impact of different HDV characteristics on fuel economy.

When working with a fleet of vehicles across large regions under many different vocations, instrumentation of vehicles beyond simple telematics data recorders quickly grows prohibitively expensive. The methodology presented in this paper can be used to derive and validate fuel economy models when physical experimentation is unacceptably impractical, as is the case with fleets of diverse HDVs with widely varied vocations. However, before wide-scale application of this methodology, additional work must be performed to expand the methodology to more vehicles in the dataset and further improve performance.

Much of the work in this study focused around the restoration of data quality lost during data collection. The amount of data processing could be cut down substantially while increasing the fidelity of data by increasing sample count. While sample count seems to be prioritized by measurement, with important signals such as geospatial coordinate, vehicle velocity, and fuel consumption sampled much more often than other signals—as shown in Figure 1—there often was not enough data within a slice to validate a vehicle model. Vehicle stops were often inadequately sampled, resulting in sudden velocity signal discontinuities, requiring rectification. Where possible, the sparsity of important measurements should be minimized.

To improve on the work done in this study, it is recommended that focus be placed on additional methods of deriving model parameters. Time series measurements such as engine torque and speed could provide ways of deriving vehicle parameters through theoretical relationships.

References

Contact Information

Kyle Carow
Mechanical & Aerospace Engineering Dept.
1903 W Michigan Ave.
Kalamazoo, MI 49008-5314 USA
kyle.j.carow@wmich.edu

Andrej Ivanco, Ph.D.
Group Lead—Research and Data Analytics
One Allison Way
Indianapolis, IN 46222-3271 USA
andrej.ivanco@allisontransmission.com

Zach Asher, Ph.D.
Mechanical & Aerospace Engineering Dept.
1903 W Michigan Ave.
Kalamazoo, MI 49008-5314 USA
zach.asher@wmich.edu

Definitions/Abbreviations

HDV - heavy-duty vehicle
FASTSim - Future Automotive Systems Technology Simulator
NREL - National Renewable Energy Laboratory
NHTSA - National Highway Traffic Safety Administration