Abstract

Real-world second-by-second vehicle driving cycle data is very important for vehicle research and development. A project solely dedicated to generating such information would be tremendously costly and time consuming. Alternatively, we developed such a database by utilizing two publicly available passenger vehicle travel surveys: 2004-2006 Puget Sound Regional Council (PSRC) Travel Survey and 2011 Atlanta Regional Commission (ARC) Travel Survey. The surveys complement each other - the former is in low time resolution but covers driver operation for over one year whereas the latter is in high time resolution but represents only one-week-long driving operation. After analyzing the PSRC survey, we chose 382 vehicles, each of which continuously operated for one year, and matched their trips to all the ARC trips. The matching is carried out based on trip distance first, then on average speed, and finally on duration. Of the total 509,158 trips made by the 382 PSRC vehicles, 496,276 trips (97.47%) were successfully matched to single original ARC trips. The remaining trips were matched to either ARC sub-trips or combined ARC trips. The resulting high-resolution year-long database can be used by drive cycle analysis tools such as the advanced vehicle simulator ADVISOR™ to investigate fuel economy, battery life, and vehicle emissions under various conditions. Our approach can be employed to produce other realistic databases from publicly available vehicle travel surveys.

Introduction

Improving vehicle fuel economy is a central part of international and regional efforts to reduce the risks of climate change. Consumers put significant weight on fuel/battery efficiency and cost [1]. In the United States, they rely on a vehicle window sticker that shows vehicle’s estimated fuel economy in miles per gallon (mpg) from the Environmental Protection Agency (EPA). This figure is produced by automaker’s laboratory tests and is regulated by federal laws. The fuel economy tests are performed in a laboratory using a dynamometer. They are based on controlled conditions using up to five standard EPA drive cycles to simulate standard vehicle trips in different environments. EPA standard drive cycles and some of European standard drive cycles are listed in Table 1. The EPA standard intentionally excludes some known on-road factors that affect fuel economy for the purpose of improved test repeatability. These factors include road and weather conditions, traffic, driving style, and geographical locations [2, 3], which are difficult to model accurately. After auto manufacturers report their results to EPA, the agency reviews the results and checks about 15%-20% of them through its own independent testing [4]. It has been reported that the on-road fuel economy can be markedly lower than the window sticker number [3], which has happened to conventional vehicles as well as to hybrid electric vehicles (HEVs) [6]. To improve the accuracy of the lab tests and to better evaluate new technologies’ potential impact on the fuel/battery usage, automakers desire to incorporate a broader range of the factors into their testing, analysis, and estimates. In order to achieve improved fuel economy estimates of on-road driving, a large set of on-road data must be used. A project solely dedicated to generating such a dataset can be prohibitively expensive and time consuming to execute.

Few options remain. One of them is to utilize travel survey data. Various vehicle fuel consumption models have been developed based on historical data collected from self-reporting travel surveys. The historical data is affined to the vehicle itself. Some types of this data, for example, are vehicle type and model, engine type, number of cylinders, transmission system type, horsepower, vehicle weight, displacement, and acceleration [7, 8, 9]. Jun [8], Slavin [10], and Wu [11] developed fuel prediction models that are entirely based on vehicle historical data using the regression modeling approach and the neural network approach. The proposed models showed reasonable fuel economy predictions. Rusiman [9] also proposed a fuel economy estimation model based on the...
The performance of different vehicle powertrains can significantly differ based on technologies because the characteristics of the new vehicles, such as start/stop, hybridization, and heat recovery, can cause different powertrain behaviors.

Numerous travel survey datasets are available to the public by various providers and organizations. Several studies have used this type of datasets to enhance vehicle fuel economy and emissions tests. Although the proposed techniques using historical data have been found to have good capabilities and reliability in predicting fuel consumption, they cannot represent the effects of driving patterns because they lack Global Positioning System (GPS) data. It is difficult to extend the predictions to cover new technologies because the characteristics of the new vehicles, such as start/stop, hybridization, and heat recovery, can cause significantly different powertrain behaviors.

The travel survey methods have continued to evolve with the introduction of GPS-enhanced travel survey techniques. An important direction in fuel economy studies is to use the data that were collected by these techniques. The use of GPS data collection has been found to have many advantages over traditional survey methods. The data collected can include trip distance, duration, average speed, and maximum speed besides time-speed points.

Numerous travel survey datasets are available to the public by various providers and organizations. Several studies have used this type of datasets to enhance vehicle fuel economy. In [12], GPS data collected over a 24-hour period was used to study the influence of real-world drive cycles on plug-in hybrid electric vehicles (PHEV) fuel efficiency and cost for different powertrain and battery characteristics. In [13], a dataset collected in California from 422 vehicles within seven days using in-vehicle GPS devices was used to predict fuel economy by customizing on-road drive cycles of the dataset. In [14], one-day-long GPS driving profiles for 783 vehicles operating in Texas were used to simulate and study the performance of different vehicle powertrains. Using the same dataset, opportunities of increasing fuel savings by adjusting the on-road drive cycles based on standard levels of acceleration rates and cruising speeds are studied in [15]. In [16], the data collected in a GPS-based travel survey was used to obtain a large set of real-world drive cycles from 227 vehicles in 24 hours with second-by-second time resolution in the St. Louis metropolitan region. The study used the GPS data to investigate the performance of different vehicle technologies. The study in [17] investigated the impact of on-road driving cycles on PHEVs using the GPS data from Southeastern Michigan collected from 11 vehicles in 26 days.

We point out that the maximum collection period for the GPS data in the above studies is only 26 days, which would be too short for a comprehensive fuel economy or battery life study that needs to cover conditions in different circumstances and seasons throughout an entire year.

Another important application of GPS-enhanced travel surveys is to study greenhouse gas emissions. For instance, the study in [18] used the GPS data collected from over 15,000 taxi vehicles to predict air pollution and emissions from vehicles for Singapore. A model was implemented to predict the microscopic emissions of carbon dioxide (CO₂), nitrogen oxide, volatile organic compounds, and total suspended particles. The model was based on the velocity and acceleration parameters extracted from the GPS data. In [19], the GPS data was used besides the traffic density and CO₂ concentration data to construct an estimation model to infer and predict instantaneous emission rates.

The GPS-enhanced travel surveys can also be used to study travel behavior and demand [20]. For example, the study in [21] used GPS data and combined it with Geographic Information System to observe and explore travel activity patterns and activity scheduling behavior.

The U.S. National Renewable Energy Laboratory (NREL) makes some travel survey datasets of passenger vehicles accessible to the public [22]. Among them are the PSRC and ARC datasets. The two datasets included different types of passenger vehicles, such as sedans, mini-vans, pick-up trucks, and sport utility vehicles. The PSRC is the only dataset provided by NREL that covers customer full driving operation for over one year with full detailed trips including the soak time between trips (the time length between the end of a trip and the start of the next trip), and this was the main reason to choose it for our study. A higher resolution and a longer period of the GPS data collection (compared to the datasets provided by NREL), less data errors, and a larger number of vehicles are among the reasons that we chose the ARC dataset. The two datasets were used by other researchers. In [2], the researchers used them to understand the effect of travel time on auto travel choices by developing a method to calculate observed trip-level and household-level reliability measures. In [23] the PSRC survey data was used to explore the relationship between the population and employment densities and CO₂ equivalent (CO₂e) emissions taking residential self-selection into the account, while in [24] the ARC dataset was used to measure driving volatility.

In this study, we use the PSRC and ARC datasets in a different and innovative way. We introduce a method to produce a new dataset of vehicle driving profiles from the two datasets. The longer period of the PSRC survey captures driver

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Distance (miles)</th>
<th>Average speed (mph)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UDDS</td>
<td>Urban Dynamometer Driving Schedule - EPA city cycle</td>
<td>7.45</td>
<td>19.59</td>
</tr>
<tr>
<td>FTP-75</td>
<td>Federal Test Procedure - EPA city cycle (EPA75)</td>
<td>11.04</td>
<td>21.20</td>
</tr>
<tr>
<td>US06</td>
<td>EPA high acceleration aggressive driving schedule</td>
<td>8.01</td>
<td>48.37</td>
</tr>
<tr>
<td>SC03</td>
<td>EPA air-conditioning driving schedule</td>
<td>3.58</td>
<td>21.55</td>
</tr>
<tr>
<td>HWFET</td>
<td>EPA Highway Fuel Economy Test Driving Schedule. It represents highway driving conditions under 60 mph</td>
<td>10.26</td>
<td>48.30</td>
</tr>
<tr>
<td>ARTEMIS Urban</td>
<td>European ARTEMIS (Assessment and Reliability of Transport Emission Models and Inventory Systems) urban driving schedule</td>
<td>3.03</td>
<td>11</td>
</tr>
<tr>
<td>ARTEMIS Rural</td>
<td>European ARTEMIS rural driving schedule</td>
<td>10.74</td>
<td>35.74</td>
</tr>
<tr>
<td>ARTEMIS Highway</td>
<td>European ARTEMIS highway driving schedule</td>
<td>17.86</td>
<td>60.22</td>
</tr>
</tbody>
</table>
Generating New TRIP Profile Dataset from the PSRC and ARC Datasets

Overview of the PSRC Dataset

The Puget Sound Regional Council ran a travel survey between 2004 and 2006 to collect data with GPS devices from 484 passenger vehicles [26]. The data collection was a part of a pricing study sponsored by the Federal Highway Association. The main objective of the project was to investigate the travel behavior diversities (numbers, modes, routes, and times of vehicle trips) in response to variable charges for road use (variable or congestion-based tolling) [26]. NREL makes a portion of the GPS survey data accessible to the public after handling the privacy issues. The data provided by NREL is in raw format and may contain errors [26, 22]. The data is provided at the household and trip levels. About, 750,000 trips are available from the 484 passenger vehicles recorded over 18 months, and logged more than 4.5 million vehicle miles travelled. The data contains 38 variables, and among these variables are the actual trip distance, duration, and average speed. The long time period and the large number of trips make the PSRC dataset useful representing the driver’s vehicle usage patterns over a complete year.

Overview of the ARC Dataset

The Atlanta Regional Commission conducted its travel survey to perform a comprehensive study of the demographic and travel behavior characteristics of residents within the 20-county study area [27]. The objective of this survey was to improve ARC travel demand forecasts. The survey was run during a two-month period (March-May and July-September 2011) to collect demographic and trip data from a minimum of 10,000 households, including a subsample of over 1,000 households that also provided GPS data with a maximum of seven days recording. The final dataset included 10,278 households that completed the travel diary survey and 1,653 households that completed the GPS part of the survey with about 40,000 trips [27]. The NREL removed personal identification information and made the resulting datasets available to the public along with the raw GPS speed traces used to validate survey responses [22]. In addition to the survey results, all vehicle GPS data was applied to a processing routine prepared by NREL to filter vehicle speed traces, link vehicles to the streets of travel, and add instantaneous grade/elevation [22]. The NREL processing routine was a six-step process resulting in 350-plus variables indicating the type of roads, drive cycle characteristics that incorporated the filtered speed and elevation, and trip type classifications (home, work, school) when available [22, 27]. The vehicle GPS subsample includes 1,653 passenger vehicles, 1,651 of which completed the processing and are included in the NREL-processed results. The quality and high time resolution of the data makes the ARC appropriate for second-by-second analysis using vehicle simulation models or other techniques.

Processing PSRC Dataset

1. Analysis of PSRC Data

We analyzed each of the datasets to investigate their characteristics and determine their suitability for the intended study. In the analysis process, all the trips for every PSRC vehicle were investigated. The variables used were distance, average speed, maximum speed, and duration of a trip. These are the key variables in vehicle driving pattern and fuel economy studies that use travel variation encountered throughout the year. The ARC survey was run in a shorter period but had a second-by-second GPS resolution. The two datasets are complementary to each other and a new dataset generated by combining them can have both year-round trip representation and second-by-second drive cycle traces. Such a dataset can be very useful in vehicle on-road researches such as fuel economy or battery life studies, tailpipe emissions, and driving pattern analysis. Our approach is innovative as no work in the literature has produced such a dataset.

NREL performed some data cleansing and error correction procedures to the raw data of both the surveys. However, we found data errors in some sections and missing data in others, especially in the PSRC dataset. Therefore, we needed to analyze and eliminate GPS-related data problems as exemplified by location drift and signal dropouts that can cause errors, missing values, and inconsistency. We developed error-correcting algorithms for both the datasets. After the analysis and error-correcting steps, we entered the main stage of the study and developed procedures to match the trips of PSRC vehicles to the trips of ARC vehicles using key data variables. Our findings, including a variety of characteristics of the datasets (e.g., speed and distance distributions of the trips), matching results, and matching error statistics, are presented. The resulting dataset contains trips with second-by-second drive data representing each trip made by the PSRC vehicles during a one-year period. Our new dataset is more comprehensive than the other second-by-second datasets used in the literature for fuel economy and other transportation-related studies because it covers a full year. It represents not only the driving style of the driver, with respect to speed and acceleration of the vehicle, but also the trip patterns, including times when the trips occurred and times when the vehicle is parked. The dataset is ready to be used by vehicle drive cycle analysis tools, such as ADVISOR.

Our study is innovative with respect to the literature. It is different from the efforts reported in [7, 23] where no error-correction process was applied to the PSRC dataset. While the GPS data of the trips were used in [7] to identify the start and end of a trip, only a 3-month period of the PSRC Travel Survey was utilized.

We presented our initial findings in a conference paper [23]. The results were preliminary as they were limited to only the phase before the main study stage. The error-correcting algorithms were also less comprehensive than the ones used in this paper.
survey data collected by GPS devices. The distance-average speed relation was analyzed. Also, the distribution of maximum speed of all PSRC trips was analyzed. The PSRC vehicles that had trips in a period of one year or more were the focus of this study. A filtering process was applied only to those PSRC vehicles that have such a one-year window. The filtering process comprises removal of trips with either zero distance or zero or negative duration, average speed recalculation, and maximum speed correction. Figures 1 to 3 indicate that a pre-processing was necessary. Figure 1 shows the distribution of the distance-average speed relation of the PSRC trips for the 382 targeted vehicles, while Figures 2 and 3 show the distribution of the maximum speeds of the same trips.

In Figure 1, the distance of some trips was reasonable, but the average speed was unreasonable. In some cases the average speeds were greater than 100 mph for short distances, while in other cases the average speeds were less than 25 mph for long distances, which indicate erroneous recordings. Similarly, in Figure 3, it can be seen that, in some trips, the maximum speed was unreasonable. In all the cases, the maximum speed was greater than 100 mph. Some of these trips are short trips considering their distances. Compared to the average speed of any short trip, its provided maximum speed is apparently incorrect, which indicates that the recordings of these specific cases encountered errors.

2. Correction of PSRC Data  
Because of the sensitivity of downstream applications to both the quality and integrity of GPS source data, the operating behavior and errors inherently associated with GPS devices have fostered the need for a correction process. The data collected by GPS instruments are prone to errors, such as location drifts and signal dropouts. Thus, the correction process was required to make the PSRC dataset suitable for the study.

In the correction process, all the trips for each PSRC vehicle were analyzed for any errors or invalid data range. Some of the data errors could be corrected by recalculating some variables or using GPS speed-time data points that were provided by the travel survey source. In some trips, the errors could not be corrected or the trip’s data could not be validated due to lack of information. When this was the case, the trip was excluded from further processing. The following four steps were applied to every PSRC vehicle to filter the data and correct errors.

**Step 1: Removing trips with zero distance, zero duration, or negative duration:** After analyzing PSRC trips, it was noticed that some trips had zero distance and some other trips had zero or negative duration (trip start time and date equal to trip end time and date or trip end time is less than trip start time). These types of trips were specifically marked in the original dataset as “true trips,” but they were

---

**FIGURE 1**  Distance-average speed distribution of 653,312 PSRC trips of the 382 vehicles before applying the data correction process.

**FIGURE 2**  Maximum speed distribution of 648,376 uncorrected PSRC trips of the 382 vehicles whose maximum speed was less than or equal to 100 mph.

**FIGURE 3**  Maximum speed distribution of 730 PSRC trips of the 382 vehicles whose maximum speed was between 100 mph and 300 mph. Note that the maximum speed of about 4,206 PSRC trips was greater than 300 mph.
Step 2: Average speed recalculation: The average speed in the original data was calculated using the calculated distance and duration. According to the information provided by the data source [9], the calculated distance was computed from the GPS speed-time data points and might have (large) errors. Therefore, using actual trip distance led to more accurate average speed calculations. Hence, the average speed of all the PSRC trips was recalculated using the actual distance and duration of the trip:

\[
\text{Average speed (mph)} = \frac{\text{Actual trip distance (miles)}}{\text{Trip duration (hours)}}
\]

Eq. (1)

Step 3: Maximum speed correction: The maximum speeds for some PSRC trips were corrected. As shown in Figures 2 and 3, in some trips unrealistic maximum speeds were provided. It is clear that the maximum speed calculations encountered errors, and these errors can be corrected by referring to the GPS speed-time data points of the vehicle. Based on our investigations, the maximum speed of the trip was the maximum speed value in the GPS speed-time data points (neglecting all unreasonable speed points). The procedure for correcting the maximum speed errors is shown in Figure 4.

An upper threshold of 100 mph for the maximum speed was assumed and used to validate every trip including the trips with maximum speed that was less than or equal to 100 mph. We check the maximum speed for all the PSRC trips and, whenever feasible, correct any trips and, whenever feasible, correct any unmatched (i.e., when reported maximum speed does not equal to the actual maximum speed from GPS speed-time data points) or unrealistic maximum speeds (i.e., when maximum speed is over 100 mph). For every PSRC trip, if the GPS speed-time points were provided and not all of the speed values are greater than 100 mph, the trip maximum speed is calculated as the maximum speed value of the speed points after neglecting all speed points that are more than 100 mph. Then, the trip provided maximum speed is replaced with the new calculated maximum speed if the two maximum speeds do not match. The maximum speed of the PSRC trip will not be corrected if the GPS speed-time points are not provided for the PSRC trip or the GPS speed-time points are provided but all of the speed values are greater than 100 mph.

In some PSRC trips, the maximum speed was not provided by the data source. These trips were excluded from the maximum speed correction step because it was noticed that the GPS speed-time data points were also missing in the data source. These trips are still used in this study since the trip’s maximum speed is not our focus.

Step 4: Validation of distance-speed relation of trips: From Figure 1 it can be seen that, for some trips, the relationship between the distance and the average speed may not be correct (e.g., small distance with high average speed). To check for this error and validate the PSRC trips, we subject the trips to a simple trip model that we devised for the purpose of validation. To the best of our knowledge, this type of validation was not applied in any of the existing studies in the literature. PSRC trips with actual distance less than or equal to 0.05 miles are excluded from the model test. They were considered as “zero distance trips” and excluded from the matching process. In the trip-testing model shown in Figure 5 that acts as a screener, the acceleration was set to 0.35 g and the deceleration 1 g, representing a vehicle achieving 60 mph from 0 mph in 7.82 seconds and a
DEVELOPING A REAL-WORLD, SECOND-BY-SECOND DRIVING CYCLE DATABASE

stopping distance of 120 feet when decelerating from 60 mph to a complete stop. According to the real-
world testing results provided in the latest Auto Issue of the Consumer Reports magazine [28], the average 
time required to accelerate from 0 to 60 mph for 255 of 2018 passenger cars was 7.96 seconds and the 
average dry-braking distance was 133 feet. Given that 
the vehicles in the PSRC and ARC surveys were made 
before 2007 and 2012, respectively, it is reasonable to 
to assume that, as a group, these vehicles’ acceleration 
and breaking performances were worse than those of 
the 2018 vehicles. Therefore, using 0.35g for the 
acceleration and 1g for the deceleration in the testing 
model is expected to be able to perform effective 
screening - excluding those trips that are likely false 
and keep the trips that are genuine.

By using the provided maximum speed (after recalcula-
tion with errors detected and corrected) and the duration of 
the trip, the trip model as shown in Figure 5 was constructed, 
where d is the trip duration in seconds, \( \upsilon_{\text{max}} \) is the trip 
maximum speed in mph, \( b = \frac{\upsilon_{\text{max}}}{0.35g} \) (seconds), and 
c = \( d - \frac{\upsilon_{\text{max}}}{g} \) (seconds). Then the distance and the average 
speed for the assumed trip model can be calculated from the 
constructed pattern using the following equations:

\[
\text{Distance} = \frac{c-b+d}{7200} \times \upsilon_{\text{max}} \text{ (miles)} \quad \text{Eq. (2)}
\]

\[
\text{Average speed} = \frac{\text{Distance}}{d} \times 3600 \text{ (mph)} \quad \text{Eq. (3)}
\]

The distance and average speed calculated from the trip 
model represent a reasonable upper bound on the distance 
and average speed of an actual trip, given only the duration. 
If the distance and/or the average speed that was calculated 
from the constructed pattern is less than the trip provided 
distance or average speed, then the trip will not be counted 
and will be dropped from this study. Also, the trip will 
be dropped from this study if it is not long enough to reach 
the maximum speed (i.e., \( c < b \) in Figure 5). Note that in the 
case of the maximum speed not being provided for a PSRC 
trip, the trip is excluded from the trip model test, but is still 
used in the study because the most important information is 
the trip distance and average speed. A trip with missing 
maximum speed but having all the other information can still 
be useful in many vehicle research studies.

3. Selection of the Best One-Year Window for 
Trips Research such as fuel economy studies should consider 
all driving circumstances, which implies using the driving 
data from the trips that took place through the entire year. 
Thus, we analyze every PSRC vehicle and choose only vehicles 
that made trips throughout a period of one year or more. For 
every chosen PSRC vehicle, a one-year sliding window is 
applied to all the trips made by the vehicle, starting with the 
first trip and moving on until reaching the last possible 
one-year window of trips. This generates multiple consecutive 
one-year windows of trips. Then, among these one-year 
windows, the window that has the number of trips closest to 
the average number of trips of one-year windows of the vehicle 
is selected as the best one-year window for the vehicle:

\[
\text{Average number of trips for one-year windows} = \frac{\text{Sum of number of trips in one-year windows}}{\text{Number of one-year windows for the vehicle}} \quad \text{Eq. (4)}
\]

3. Selection of the Best One-Year Window for 
Trips Research such as fuel economy studies should consider 
all driving circumstances, which implies using the driving 
data from the trips that took place through the entire year. 
Thus, we analyze every PSRC vehicle and choose only vehicles 
that made trips throughout a period of one year or more. For 
every chosen PSRC vehicle, a one-year sliding window is 
applied to all the trips made by the vehicle, starting with the 
first trip and moving on until reaching the last possible 

**Figure 6** Distance-average speed distribution of 39,433 original ARC trips of the 1651 vehicles after data processing.
consistent with the PSRC trips whose average speed was calculated using trip’s actual distance and duration, we recalculated the average speed for all the ARC trips based on the actual trip duration (including the idle time) and distance.

Figure 6 shows the distance-average speed distribution of the original ARC trips that we obtained after adding the removed time points at zero speed and recalculating the average speed. Compared to the distribution of the PSRC trips shown in Figure 1, the distribution of the ARC trips looks more normal where none of the ARC trips had low distance and high average speed, or high distance and low average speed. Figure 7 shows the distribution of their maximum speeds which also looks more normal comparing to the distribution of the PSRC trips shown in Figure 2.

Generating New ARC Trips from the Original ARC Trips

The main goal of our work is to use the PSRC and ARC trips to develop a driver-based second-by-second driving cycle database for vehicle usage patterns covering an entire year. Such a driving cycle database may be produced by using the second-by-second data of the ARC trips and the validated PSRC trip data after applying certain procedures to these trips. This can be accomplished by matching the PSRC trips to the ARC trips based on trip information such as the distance, average speed, and duration. For good matching, we required that the two trip datasets involved have very similar distance-average speed distributions. As shown in Figures 6 and 12, the distributions of the PSRC and ARC trips have some similarity, but some regions of the PSRC distributions are not included in the ARC distributions. To address this problem, and more specifically, to include the PSRC trips that had short duration and high average speed in ARC trips’ distributions, we generate new ARC trips from the original ARC trips by considering every driving cycle between each two idle time portions as a trip (i.e., micro-trip [27]). As a result, we can have trips that have small distances and high average speeds.

Matching the PSRC Trips by the Original and ARC Micro-Trips

After pre-processing the two datasets as described above, the PSRC trips are matched to the trips of the ARC dataset as follows. The trips of every PSRC vehicle are matched to the trips of ARC vehicles based on the key variables of the trip, which were distance, average speed, and a specific duration condition. Generally, the distance of the PSRC trip is first matched to the distances of all the ARC trips with an error band of ±3%. Then, an average-speed error band of ±3% is applied to the average speeds of the resulting ARC trips. Finally, all ARC trips that passed the previous two matching steps are subject to a duration test - the duration of the ARC trip should not be greater than the duration of the PSRC trip plus the soak time. If more than one ARC trip meets these criteria, then one of them was randomly chosen as the final ARC trip that matched the PSRC trip. Figure 8 shows the procedure for matching the PSRC trips to single original ARC trips using the abovementioned strategy.

If the matching procedure shown in Figure 8 does not result in any original ARC trip, different matching strategies are then applied. First, the PSRC trip will be matched to a combination of up to four original ARC trips using the procedure shown in Figure 9. In this procedure, the average speed of the PSRC trip is first matched to the average speeds of all the original ARC trips with an error band of ±3%. Then,
among the resultant original ARC trips whose distances are less than the distance of the PSRC trip, the trip with the closest distance to the PSRC trip’s distance is selected. Finally, the selected original ARC trip is combined with up to four trips of the resultant original ARC trips to achieve the satisfaction of error band of ±3% for the distance and average speed and the satisfaction of the duration condition for the combined trip. If the combination procedure failed to generate a matching trip for the PSRC trip, the ARC combination procedure shown in Figure 9 will be repeated, but using both original and/or ARC micro-trips with as many original trips to be utilized as possible.

**Distance error (%)**

\[
\text{Distance error} = \left( \frac{\text{Distance of current PSRC trip} - \text{Distances of all ARC trips}}{\text{Distance of current PSRC trip}} \right) \times 100 
\]

Eq. (5)

**Average speed error (%)**

\[
\text{Average speed error} = \left( \frac{\text{PSRC trip’s average speed} - \text{Average speed of matched ARC trips}}{\text{PSRC trip’s average speed}} \right) \times 100 
\]

Eq. (6)

If the PSRC trip still could not be matched, the ARC combination procedure shown in Figure 9 will be repeated, but using both original and/or ARC micro-trips with as many combinations as needed preferring the original ARC trips. By combining ARC trips (original and/or micro trips) as shown in Figure 9, the distance of the generated ARC trip can be increased while maintaining the average speed. Finally, in the case when all the previous procedures failed to find a matching ARC trip for the PSRC trip, a procedure that modifies an original ARC trip to match the PSRC trip will be applied as shown in Figure 11. In this procedure, the distance of the PSRC trip is first matched to that of all the original ARC trips with an error band of ±3%. Then, among the resultant ARC trips, the trip whose average speed is greater than the average speed of the PSRC trip by the least amount is selected. Then, the idle time that is required to reduce the average speed of the selected ARC trip to make it within an error band of ±3% of the average speed of the PSRC trip is added to this ARC trip, provided that the added idle time is less than one-third of the PSRC trip’s duration. Finally, if the resultant ARC trip satisfies the duration condition presented in (7), a new vehicle ID and trip ID are generated for the resultant ARC trip, and this trip is assigned as the matching trip for the PSRC trip. By using this procedure, the length of the idle periods of the trip can be increased to reduce the average
speed of the trip without changing distance. In the case when all the four procedures failed to find a matching ARC trip for a PSRC trip, that specific PSRC trip is excluded from this study. The matching procedure produced a dataset that inherits the usage patterns of the PSRC dataset (the long time period covering the entire year) and the advantages of the ARC dataset (the high time resolution of the clean GPS speed-time data points).

Results

Results of Processing the PSRC Dataset

The results of filtering and processing the PSRC trips are shown in Table 2. It can be observed that after the correction process, the trips have a reasonable distribution. From Table 2, 3.43% of the PSRC trips had either a maximum speed over 100 mph or an erroneous maximum speed. Most of these trips were correctable and were fixed. Only 0.15% of the trips could not be corrected because of the limitations of the GPS speed-time data points.

Figure 12 shows the distance-average speed distribution of the corrected PSRC trips of the 382 vehicles that passed the one-year window process (we call them the targeted vehicles) excluding “zero distance trips.” Compared with Figure 1, the distribution is improved and the new distribution is similar to the distribution of the ARC trips. The trips with a long distance and an unreasonably low average speed were corrected, so were the trips with a short distance and an unreasonably high average speed. Figure 13 provides a closeup examination of the distance-average speed distribution of the PSRC trips shown in Figure 12. The theoretical upper limit boundary imposed by the trip model (Figure 5) is also exhibited. All the corrected PSRC trips passed the model because they were all within the boundary. Figure 14 shows that the maximum speeds of the PSRC trips of the targeted vehicles tend to follow a bimodal distribution, with a large number of trips with low maximum speed.

The results of applying the best one-year window process are shown in Figures 15 and 16. Figure 15 illustrates the distribution of the number of one-year windows of the 382 targeted PSRC vehicles after the filtering process. The mean of the number of the one-year windows was 58 windows. Figure 16 shows the distribution of the number of the trips in the best one-year windows of the targeted vehicles after the filtering. The average number of the trips in the best one-year window was 1396 trips, while the minimum number of the trips in a best one-year window was 294 trips.

Table 2

<table>
<thead>
<tr>
<th>Description</th>
<th>Mean (%)</th>
<th>STD (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trips with zero duration</td>
<td>5.86</td>
<td>5.62</td>
</tr>
<tr>
<td>Trips with zero actual distance</td>
<td>5.76</td>
<td>5.64</td>
</tr>
<tr>
<td>Trips failing the trip model test</td>
<td>5.30</td>
<td>2.75</td>
</tr>
<tr>
<td>Trips with incorrect maximum speeds</td>
<td>3.43</td>
<td>1.06</td>
</tr>
<tr>
<td>Trips with corrected maximum speeds</td>
<td>3.37</td>
<td>0.99</td>
</tr>
<tr>
<td>Trips removed due to erroneous maximum speeds</td>
<td>0.06</td>
<td>0.15</td>
</tr>
<tr>
<td>Trips with negative duration</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

FIGURE 11 The procedure for matching the PSRC trip by a single modified original ARC trip.

FIGURE 12 Distance-average speed distribution of 509,158 PSRC trips (excluding zero-distance trips) made by the 382 vehicles that fitted the one-year window.
Figure 17 presents the distribution of the number of driving days in the best one-year window for the targeted PSRC vehicles. We notice that most of the targeted vehicles had more than 200 driving days in their best one-year windows. The average of the driving days in the best one-year windows was 283 days, which means the average driver in the PSRC dataset used their vehicles for 77.5% of days in the year.

Some drivers in the PSRC dataset used their vehicles nearly every day, but the most typical usage pattern is around 300 days per year as Figure 17 shows. There is also a large variation in the number of trips taken per year, from as little as 250 trips to nearly 3000 trips per year as described in Figure 16. From Figure 18, it can be seen that there is a strong correlation between the number of yearly trips and the number of days per year that a vehicle is used.

Also, by looking at the data histograms, it can be seen that neither of these distributions is normal. Rather than attempting to fit statistical distributions to these datasets, our method uses the actual data itself to empirically describe driver usage patterns.

Results of Processing the ARC Dataset

From Figures 19 and 20, it can be seen that the ARC microtrips increased the regions of the distance-average speed distribution of the ARC trips and improved the ability of the ARC trips to match the PSRC trips. 143,905 trips were created from the 39,433 original ARC trips.
Results of Matching the PSRC Trips by the ARC Trips

The fraction of PSRC trips matched to the ARC trips was very high. Figure 21 shows that 99.718% of the targeted PSRC trips were successfully matched to the ARC trips. Only 0.278% of the PSRC trips could not be matched to the ARC trips because of average-speed mismatch, and only 0.004% of the PSRC trips failed to meet the duration condition. None of the targeted PSRC trips failed to be matched to the ARC trips due to distance mismatch.

As shown in Figure 22, the overlap region of the distance-average speed distributions of the PSRC trips and the ARC trips increased because of the inclusion of the ARC micro-trips. As will be seen in the next section, the ARC micro-trips contributed to about 1% of the matching results. In other words, 1% of the PSRC trips failed to be matched to the original ARC trips and were matched to the ARC micro-trips.

Table 3 shows the details of the successful matching results. Note that 97.47% of the matched trips were achieved by using only one original ARC trip (i.e., no combination of ARC trips was needed), while 1% of the matched trips were obtained by using only one ARC micro-trip.

There were 1.301% and 0.22% of the matched trips that were respectively produced by combining four or less and more than four of the original and ARC micro-trips. Only 0.009% of the PSRC matched trips were attained by modifying the original ARC trips by adding or removing their idle times.

Figure 23 shows the distribution of the matching percentage of the targeted PSRC vehicles, which was
Each and every trip of around 30% of the 382 targeted vehicles (114 vehicles, to be exact) was completely matched. The mean of the matching percentage for the 382 PSRC targeted vehicles was 99.72%, and the minimum matching percentage was 95.25%. In Figure 24, the distribution of the distance differences and the distance errors of the targeted PSRC trips are presented. An ideal matching process should not result in a matched ARC trip whose distance is substantially different from the distance of the PSRC trip being matched. Hence, we restricted the matching distance error to be ±3%. We calculated the percent of the total yearly distance error between the targeted PSRC trips and the matched ARC trips. We found that that percentage was a mere

![Figure 21](image1.png)

**FIGURE 21** Distance-average speed distribution of 183,338 ARC trips composed of both the original and ARC micro-trips and 522,731 PSRC trips made by the 382 vehicles fitting the one-year window.

![Figure 22](image2.png)

**FIGURE 22** Results of matching the 507,722 PSRC trips by the ARC trips subject to a distance error band of ±3% and an average-speed error band of ±3%.

![Figure 23](image3.png)

**FIGURE 23** Matching percentage distribution of the 382 PSRC vehicles subject to a distance error band of ±3% and an average-speed error band of ±3%.

![Figure 24](image4.png)

**FIGURE 24** (a) Distribution of distance differences of the PSRC trips after matched by the ARC trips. (b) Distribution of distance errors for the same PSRC trips after they were matched by the ARC trips.

<table>
<thead>
<tr>
<th>PSRC trips matched to</th>
<th>Fraction of trips (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>One of the original ARC trips</td>
<td>97.470</td>
</tr>
<tr>
<td>One of the ARC micro-trips</td>
<td>1.000</td>
</tr>
<tr>
<td>Combination of two original ARC trips</td>
<td>1.160</td>
</tr>
<tr>
<td>Combination of three original ARC trips</td>
<td>0.100</td>
</tr>
<tr>
<td>Combination of four original ARC trips</td>
<td>0.041</td>
</tr>
<tr>
<td>Combination of more than four original ARC trips</td>
<td>0.220</td>
</tr>
<tr>
<td>Modified ARC trips</td>
<td>0.009</td>
</tr>
</tbody>
</table>

**TABLE 3** Results of matching the targeted PSRC trips to the ARC trips.
0.0557%. The yearly percent distance errors were randomly distributed around a mean of 0.03%, which is near zero. This indicates that overall the change in the total yearly distance was very small for the matching trips.

From Figures 25 and 26, the overall results indicate that the matching distance error and the average speed error were approximately normally distributed. The mean of the distance errors was 1.48%, whereas for the average speed errors the mean was 1.49%. Notice that the standard deviations of these two errors were quite low - 1.72% for the matching distance error and 1.73% for the matching average speed error, confirming the distributions to be close to uniform.

Figure 27 shows some examples of the driving cycles of some PSRC trips after being matched to different ARC trips using the different matching procedures. Figure 27(a) shows the driving cycle of a PSRC trip with 2.01 miles distance, 22.27 mph average speed, and 325 seconds duration that is matched to an original ARC trip. Figure 27(b) shows the driving cycle of a PSRC trip with 0.98 miles distance, 2.08 mph average speed, and 1698 seconds duration matched to a combination of four original ARC trips. Figure 27(c) shows the driving cycle of a PSRC trip with 0.28 miles distance, 19.33 mph average speed, and 53 seconds duration, which is matched to an ARC micro-trip. Finally, Figure 27(d) shows the driving cycle of a PSRC trip with 0.25 miles distance, 2.27 mph average speed, and 1310 seconds duration that is matched to a combination of more than four original ARC trips.

The information for all the targeted PSRC trips including the ARC matching trips IDs were stored in a Microsoft Access database. Also, the vehicle IDs and trip IDs for the ARC matching trips for every PSRC trip that was matched to multiple ARC trips were saved in the same database so that they can be presented as different driving cycles for the same PSRC trip. Additionally, for every PSRC trip that was matched to a combination of several ARC trips, the vehicle IDs and trip IDs for these ARC trips were saved in the same database so they can be combined in a different order to generate an ARC trip that matches the same PSRC trip with a different driving cycle. The final product is a MS Access database that can be searched based on desired query criteria.

The structure of our resulting new database is based on individual vehicles. The database can be searched to find individual vehicles or trips that satisfy a user-specified criteria. The driving cycle (second-by-second speed-time data points) of a
trip is accessible as part of the search outcome. The dataset is ready to be used by analyzing tools such as ADVISOR for various studies, including fuel economy and battery range/life.

Figure 28 shows three different examples of the cycle of the average trip of the generated dataset. The distance of the average trip was 6.8 miles and the average speed was 26.8 mph. The average trip of the generated dataset was determined by averaging the distance of the dataset and choosing trips with an error band of ±3%. Then, an average-speed error band of ±3% is applied to the average speeds of the chosen trips to choose a pool of trips that match the average trip of the dataset. A total of 667 trips of the dataset matched the average trip, and in Figure 28 we plot the cycles of only three trips as examples.

As an illustration, Figure 29 shows the distance-average speed distribution of the 507,680 PSRC trips that were successfully matched to the ARC trips, including the average trip of the dataset, and the distance-average speed distribution of EPA standard drive cycles and European Artemis drive cycles that are listed in Table 1. It can be seen that the distance-average speed distribution range of the new dataset covers all of EPA and European ARTEMIS different drive cycles. This indicates that in addition to representing different real-world driving cycles and patterns, the generated dataset still takes into the account and covers the standard driving cycles that are used for fuel economy and gas emissions tests.

Conclusion

The PSRC and ARC datasets are complementary in terms of time resolution, traffic and environmental conditions, and variables such as distance, average speed, and duration. The two datasets were analyzed and processed. The filtering and cleansing of the PSRC data improved the data reliability. A combination of the ARC micro-trips from the original ARC dataset and the original dataset increased their coverage and produced better matching results for the PSRC trips. We achieved high degrees of matching between the PSRC trips and the ARC trips, original and micro-trips. The final product is a new driving cycle database in MS Access format that can be easily queried to produce diverse second-by-second realistic driving cycles, that also includes the yearly usage patterns for 382 drivers. Coupled with analyzing tools such as ADVISOR, this database can be very useful in studying vehicle fuel economy, battery life, and tailpipe emissions impact due to real-world driving scenarios.

More recently, NREL and other organizations has made more vehicle travel surveys available to the public. Our approach can be applied to these surveys as well.

Contact Information

Hao Ying
Department of Electrical and Computer Engineering
Wayne State University
Detroit, MI, 48202, USA
hao.ying@wayne.edu

Acknowledgment

This work was supported in part by a grant from the University Research Program of Ford Motor Company.

References


2. Posada, F., and German, J., “Measuring In-Use Fuel Economy in Europe and the US: Summary of Pilot Studies,”


27. Livingston, J., “Atlanta Regional Commission - Regional Travel Survey - Final Report,” PTV NuStata, Austin, TX, 2011.