

On the Need for Revisions of Utility Factor Curves for Plug-in Hybrids in the US

Author, co-author (Do NOT enter this information. It will be pulled from participant tab in MyTechZone)

Affiliation (Do NOT enter this information. It will be pulled from participant tab in MyTechZone)

[Now that this draft manuscript is no longer under the anonymity of the review process, we can include the authors' names, which are Hamza, K. and Laberteaux, K.]

Abstract

Plug-in hybrid electric vehicles (PHEVs) have the capability to drive an appreciable fraction of their miles on electric power from the grid, similar to battery-only electric vehicles (BEVs), and then automatically switch to gasoline power and operate similar to a regular (non-plug-in) hybrid electric vehicle (HEV) when the battery energy runs low. Part of the allure of PHEVs comes from being able to drive almost like a BEV but without range anxiety. Towards gauging the former, the concept of utility factor (UF) has been developed as a simplified metric that gauges the fraction of total distance travelled by a PHEV while in charge depletion (CD) mode, in which electric power from the battery is the primary source of energy for vehicle propulsion. Estimation of expected UF as function of the all-electric range (AER) is typically done via UF curves that have been codified in various standards. Until recently, UF curves from SAE J2841 standard appeared to hold well compared to real-world public datasets for PHEVs in the US. However, a study in 2022 analyzing two new datasets (BAR and Fuelyly) suggested a need for revisions and/or updates to the UF curves in SAE J2841.

In this work, we re-analyze both datasets (BAR and Fuelyly) cited in the 2022 study. Our findings suggest that the BAR dataset, while potentially useful in the future when it will contain a broader, more-representative sample of PHEVs operating in California, currently has biasing issues that are extremely difficult to filter out. This is because the rules surrounding BAR PHEV data collection currently excludes most Californian PHEVs, leaving only a small group of vehicles sampled due to unusual circumstances. In the future, while vehicles become required to join the BAR dataset, it may contain a much larger, more-representative sample. However, utilizing BAR dataset in the near-term will lead to a mischaracterization of PHEVs in California at large. On the other hand, the treatment of Fuelyly dataset in the 2022 study included both modeling and data filtering errors, which when corrected, leads to reasonable match with SAE J2841 UF curves within acceptable uncertainty limits. Thus, while there may be a need to update the standard to account for present-day or future PHEV usage patterns, such an update would require much more thorough work than what has thus far been conducted.

Introduction

Plug-in hybrid electric vehicles (PHEVs) combine some of the attractive traits of battery-only electric vehicles (BEVs), such as the capability to drive on energy from the electric grid [1], but without the drawback of vehicle owners having to plan stops for charging on long trips or worrying about getting stuck without access to charging, sometimes referred as “range anxiety” [2]. In terms of usability experience, when PHEVs run out of charge, they simply operate like a regular (non-plug-in) hybrid electric vehicle (HEV) [1]. Though operating similar to HEVs is already beneficial in terms of fuel economy and reduction of greenhouse gas (GHG) emissions or criteria pollutants compared to conventional internal combustion engine (ICE) vehicles, much of the allure of PHEVs comes from their ability to drive “like a BEV on most days”. For PHEVs, the Utility Factor (UF) is a simplified metric that’s been developed towards gauging “how close to BEV” the performance of the PHEV is. In formal terms, UF is defined per SAE J2841 standard [3] as the ratio of distance travelled by the PHEV while it is in charge depletion (CD) mode, to the total distance travelled by the PHEV. CD mode is defined as the portion of driving when the PHEV is utilizing electric energy from the battery as the main source of propulsion power [3], as opposed to charge sustaining (CS) mode, during-which fuel consumption in the ICE is the main source of propulsion power.

UF curves are the result of modeling efforts towards enabling the estimation of the expected UF of a PHEV as a function of its all-electric range (AER). Different standards around the world exist for UF curves [3-5], which include some differences in the assumptions about the driving and charging patterns, resulting in different UF curves, as shown in Fig. 1. Unlike some other standards, SAE J2841 recognizes differences between “individual” or multi-day individual utility factor (MDIUF) and “fleet” utility factor (FUF). Simply put, MDIUF is a simple average of the UF among a population of PHEVs, while FUF is the expected ratio of CD mode miles by the population of PHEVs to the total miles by the population of PHEVs. The distinction between MDIUF and FUF is important because the vehicles that drive longer distances are more likely to exceed the AER of the PHEV, and drive a larger fraction of their miles in CS mode, leading to lower UF values. Thus, given some value of AER, FUF is typically the lesser UF expected value among FUF and MDIUF.

Whether UF curves of some standard successfully approximate the real-world performance of PHEVs generally depends on whether the underlying assumptions that led to the development of the UF curves hold true. This generally comes down to three categories of possible discrepancy [6]: i) charging frequency and patterns, ii) actual achieved real world AER, and iii) statistical distribution of distance travelled per day. Several studies [7-9] have emerged that showed PHEVs real world UF performance fell short of the expectation per European Union (EU) standards. In contrast, analysis of publicly accessible datasets [10, 11] showed the performance of PHEVs in the US to be mostly well-aligned with SAE J2841 standard. However, a recent study in 2022 [12] showcased an analysis of two new datasets

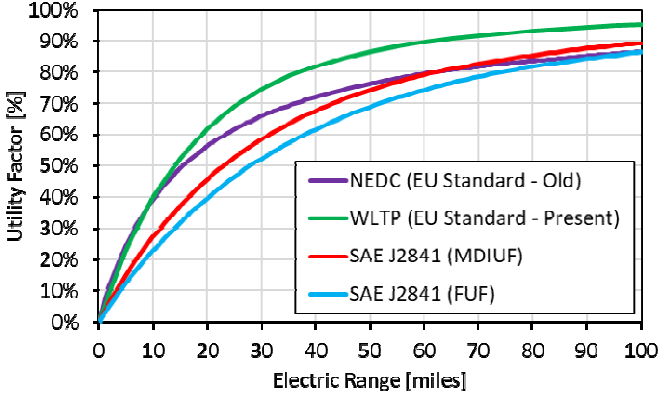


Figure 1. Example Utility Factor Curves from Different Standards

of US vehicles and drawing a claim that PHEVs in the US significantly under-performed their estimated UF values via SAE J2841 standard. The study in [12] was cited by the US Environmental Protection Agency (US-EPA) in the notice of proposed rule-making (NPRM) of April 2023 [13] as a reason for need of revisions to SAE J2841 standard and proposed a reduction to the FUF values utilized by the US-EPA for GHG emissions compliance. In this work, we re-analyze both newly introduced datasets in an effort to assess how well do the claims in [12, 13] hold.

This paper started with introducing the motivation for the current work. The rest of the paper is organized as follows: the following section provides a broad-line overview of the re-analysis of the two newly cited datasets in [12, 13], followed by two sections dedicated to detailed analysis of each dataset. A general discussion of findings then leads to drawing out conclusion of the paper.

Overview of the Re-Analysis of Datasets

Prior to critiquing findings from the datasets, it is relevant to keep in mind that neither of the two newly cited datasets in [12, 13] had the purpose of evaluating real-world performance of PHEVs as part of their mission for data collection. The first dataset we re-analyze comes from California Bureau of Automotive Repair (BAR) [14]. Source of the data for BAR comes from mandatory Smog tests whose primary purpose is to ensure that the ICEs in vehicles are running within proper operating limits set by the California emissions standard for the respective vehicle's model and model-year. Starting with model-year 2019 PHEVs, additional PHEV-only data, which adhered to On-Board Diagnostic II (OBD2) standards, is collected and recorded. For the population of PHEVs required to submit to a

BAR Smog Test, the collected OBD2 data could be used to estimate several types of information, including UF. BAR data could be requested under the public records act (PRA) [15], but until the US EPA released the snapshot BAR dataset upon which they conducted their analysis, it was not readily available to the general public. For ease of reproducibility of our analysis, we have copied the files released by the US EPA to a cloud storage folder [16].

One major issue with BAR dataset, which will be analyzed in detail in the next section, is that California rules only began requiring the collection of PHEV-specific OBD2 data with model-year 2019. Further, newer vehicles were exempt from the Smog check, during which the PHEV-specific OBD2 data was collected. Per the information and explicit example on BAR website [17], model-year 2019 vehicles sold as used vehicles do not require a Smog check due until the calendar year 2023. Further, model-year 2019 vehicles that remain with their first owners do not require the biennial Smog Check until the year 2027. As such, in 2022, when the snapshot of BAR data analyzed by [12, 13] was captured, the vast majority of model-year 2019 (or later) vehicles purchased in California, either new or used, were exempt from smog checks, and thus did not contribute data to the BAR PHEV dataset. However, California does require a smog check for any vehicle that had re-located to California from outside of the state. Since conducting a smog check in California takes both time and money, and is more expensive (at \$30-\$60, per [18]) than a Smog abatement fee (at \$20 per [19]), it is unlikely that the BAR dataset analyzed in [12, 13] includes many PHEVs except those that had relocated from outside the state of California, which poses a strong likelihood for bias in the data.

Another issue with the analysis of BAR dataset in [12], is that although the data retrieved via OBD2 from the Smog checks of vehicles includes information that allows estimating the true UF, the study in [12] relies on an indirect approach for estimation of UF via the total amount of fuel consumed. The indirect approach, which will be referred as "Fuel-inferred UF" relies on an equation of the form:

$$FC_{Tot} = (1 - x) FC_{CS} \quad (1)$$

where FC is the average fuel consumption rate in units of fuel volume per unit travel distance, such as gal/mile or liter/km. The subscripts Tot and CS refer to "total", i.e. dividing the total amount of fuel consumed by the total distance travelled by a vehicle, and charge-sustaining respectively, while the symbol x in Eqn. (1) is the unknown that is desired to be estimated. While Eqn. (1) is attributed to the peer-reviewed work in [8] and has been repeated in other work such as [10], it is noted that: i) the equation only provides an approximate estimate, which is not needed when sufficient information is available for calculating the true UF, and ii) care must be exercised when utilizing the method due to potential sensitivity of the result to real-world variations in fuel consumption as well as ensuring the computed x is compared to the correct reference metric. For example, if the FC_{Tot} and FC_{CS} terms in Eqn. (1) were pre-averaged for all vehicle samples in a dataset (i.e. dividing total volume fuel consumed by all vehicles by the total distance travelled by all vehicles), then x would be an approximate estimate of FUF. On the other hand, if Eqn. (1) was applied to each vehicle sample individually, then the different values of x (one for each vehicle sample) were averaged as a simple average, then that average value of x would be an approximate estimate of MDIUF. In further sections of this paper however, unless otherwise stated, we adhere to the context of x being an approximate estimate of FUF.

Further distinctions when utilizing fuel-inferred UF depend on what value is used for the term FC_{CS} . If FC_{CS} in Eqn. (1) were to be assumed equal to the US-EPA Window-Sticker label ratings [20], then the estimate of UF is designated with an additional subscript as x_L ('L' subscript for "Label-value" CS mode FC). However, it is a well-known fact that fuel economy varies in the real world [21], so if sufficient information is available in a dataset (as is the case for BAR), a better approximation of fuel-inferred UF can be done by utilizing the true real-world CS mode FC (volume of fuel consumed in CS mode divided by distance travelled in CS mode) in Eqn. (1). In this case, the estimate of UF will be designated with an additional subscript as x_R ('R' subscript for "Real" CS mode FC).

An additional theoretical application of Eqn. (1) could be to use FC "while engine is running" for the value of FC_{CS} , which would result in x becoming an estimate of the fraction of driving distance with the engine off. However, Window-Sticker Label values for CS mode (hybrid-like operation) are an average that includes episodes of engine-on and engine-off, with the "FC while engine is on" not typically published. This theoretical case was one of the errors in [12], where FC_{CS} in Eqn. (1) was taken as EPA Window-Sticker label ratings but claimed that x was an estimate of "electric drive share", which was further explained in [12] as "fraction of distance travelled with the engine off".

The second dataset considered in [12, 13] is named "Fueled" [22]. The Fueled website, and an associated smart-phone app, allow users to track the running cost, including fuel consumption, of their vehicles, with the aim of allowing the users to make better educated decisions about their vehicle usage and possibly their next vehicle purchase. While nicely designed for its intended purpose, Fueled website and app appear to have been primarily (if not solely) intended for use by conventional ICE vehicle (and possibly HEV) owners. The data entry unit is the "Refueling Event", for which the user enters the odometer miles, (liquid) fuel amount and price of fuel, as well as some other indicators such as estimated fraction of miles in city driving conditions versus highway driving. There is only one entry field for "fuel amount", and the back-end code of the website always treats it as a volume of liquid fuel. Furthermore, the data is entered "manually" by the users into the app or the website, without connection or verification via data ports/connections on the vehicle. As such, there is nothing to prevent incorrect user data entries from being included in the Fueled dataset. For example, a screen shot from Fueled's website [23], shown here in Fig. 2.a, shows some BEV owners use Fueled's "fuel fields" to track charging costs, producing BEV efficiency results that are labeled as "miles per gallon" (MPG) when they are much more likely to be "miles per kWh". Fig. 2.b on the other hand shows a snapshot of the refueling events log of one PHEV with discernable kWh entries (both in terms of "miles per gallon" likely to be "mile per kWh", as well as unit price of the fuel).

Use of Fueled data to estimate UF (as was done in [12, 13]), which is unlikely to have been among the intended purposes of Fueled website creators, necessitates relying on fuel-inferred UF via Eqn. (1) because Fueled data only provides total miles and total fuel amount (or average overall MPG), without separate logs for electric powered miles or amount of kWh charged. Furthermore, user entry errors of the type illustrated in Fig. 2 (amount of kWh charged entered as gallons of fuel) will create refueling events with MPG values in the range of 2-4 MPG, which can significantly skew the term (FC_{Tot}) in Eqn. (1) and lead to completely incorrect estimates of the UF.

Rather than completely discarding Fueled dataset, with a bit of exploration, it became apparent that alongside displaying the overall average MPG for each <make, model, model-year>, Fueled website also allows website visitors to view individual vehicles and even individual refueling events for each vehicle. Though somewhat time-consuming, we took the time to create an offline snapshot summary (in January 2023) of each vehicle in the Fueled dataset. PHEVs whose refueling data logs showed a strong likelihood of kWh entries as gasoline were excluded. To allow reproducibility of results in our work, a copy of our cleaned-up version of the Fueled dataset is publicly accessible via cloud storage [16]. Note that, due to the time burden for conducting proper data cleaning, it was only conducted for a select set of popular PHEV models while still covering a broad range of AER among PHEV designs.

In the section focusing on detailed analysis of Fueled dataset in this paper, we will show how the estimated UF changes after proper data cleaning compared to simply using the nameplate average MPG into Eqn. (1) for <make, model, model-year> as was done in [12, 13]. Furthermore, keeping in mind the limitations of Eqn. (1), we utilize analysis of BAR dataset, which has sufficient information from OBD2 data to estimate the real-world values of FC_{CS} , to further adjust the results of Fueled dataset in the form of uncertainty bounds.



(a) Fuely summary for several Tesla Model-Y vehicles

#	Date	Miles	Gallons	MPG	Price	Note
538	1/4/23	37.00	13.200	2.80	0.130	
537	1/4/23	34.00	12.000	2.83	0.130	
536	1/3/23	49.00	17.300	2.83	0.130	
535	12/21/22	45.00	13.900	3.24	0.130	
534	12/21/22	191.00	5.232	36.51	2.549	
533	12/20/22	34.00	13.200	2.58	0.130	
532	12/20/22	34.00	13.200	2.58	0.130	
531	12/19/22	37.00	13.000	2.85	0.130	
530	12/19/22	221.00	6.195	35.67	2.749	
529	12/19/22	30.00	11.000	2.73	0.130	

(b) Section of Fuely logs of the “Refueling Events” for one Chevrolet Volt

Figure 2. Snapshots from Fuely web page [23] highlighting the occurrence of user errors in data entry

BAR Dataset

As discussed in the previous section of this paper, while keeping in mind that the available facts strongly point to a speculation that most PHEVs in the BAR dataset are vehicles that re-located to California from outside the state, we proceed to examine some of the data content. The original “raw” BAR dataset in [13], a copy of which is available in supporting documents in the Regulations.Gov docket as well as cloud storage at [16], contained 8369 vehicle records, with each record containing many data fields collected from the vehicles via OBD2. However, whatever the reason may be (data transmission failures, data retention protocols being new for a newly launched program), not all the vehicle records are complete and valid. In fact, out of the 8369 vehicles, only 2060 vehicles are cited in [13] as having been included in the analysis, with approximately 75% of the sample being “filtered out” due to data quality issues involving incomplete vehicle data fields, large mismatch between odometer miles reading and OBD2 data field for vehicle lifetime miles, odometer miles being too low (minimum of 3000km was reported in

the analysis in [13]), or a combination of such reasons to filter out the vehicle records.

There is a broader question of how reliable a data source is if 75% of its contents need to be filtered out due to obvious errors and inconsistencies, whether the remaining 25% of entries accurately represent reality, or are they also inaccurate (but were not excluded because their inaccurate values were deemed plausible), and whether the non-excluded samples had something in common (such as a particular geographic sub-zone whose emissions test station had newer and better equipment) that could cause unknown biases. However, those type of concerns are difficult to verify and will not be further explored in this manuscript.

The remaining 2060 vehicles were further split in [13] into 30 groups of <make, model, model-year> data points. However, when examining the plot in (Figure 13 of [13]), one may note that some of the <make, model, model-year> groups contain only an extremely small number of vehicle samples that should have been excluded in the analysis per typical good practices in data analytics. In the analysis in this section, and henceforth, we re-started at the “raw” BAR dataset of the 8369 vehicle records, reciprocated the condition that records must be complete, but set the data filtering criteria per Table 1. This resulted in only twenty-five <make, model, model-year> groups of PHEVs (as opposed to thirty in [13]) being available for analysis. We then compare the following in Fig. 3:

- Reference FUF per SAE J2841 for the AER of the <make, model, model-year> PHEV
- True real-world FUF (fraction of miles travelled in CD mode) from the OBD2 data in BAR dataset
- Estimated FUF via the approach in [12] (using Eqn. 1, along with Window-sticker label CS mode MPG)
- Estimated FUF via Eqn. (1) but using true real-world CS mode MPG from the OBD2 data in BAR dataset

Table 1: Data quality filtering criteria for BAR dataset

Criteria	Value in [13]	Value in this work	Reasoning
Minimum vehicle lifetime travel	3000 km	3000 miles	3000 km (1865 miles) is practically a brand-new vehicle
Maximum mismatch between Odometer and OBD2 lifetime travel distance	20%	5%	20% uncertainty about travel distance could lead to large errors
Min. num. vehicles per <make, model, model-year> group	N/A	20	Statistical relevance

While it is notable in Fig. 3 that only two PHEV <make, model, model-year> groups (BMW i3 and BMW X5) among twenty-five others, have better real-world FUF than estimated value via SAE J2841 standard FUF curve, such result is not necessarily representative of PHEVs in the US due to bias in the sample of vehicles that will be further highlighted in this section of the paper. Also notable in Fig. 3, is that utilizing Eqn. (1), per the approach followed in [12] or with the real-world CS MPG, almost universally (except for the BMW i3) under-estimates the true UF, which is an important observation that will be revisited when analyzing Fuely dataset (since the Fuely dataset does not contain the number of miles in CS Mode, estimates of UF must rely on Eqn. 1).

A brief recap of real-world UF estimation from BAR dataset is provided as follows:

$$UF_k = \frac{CDD_k}{TD_k} \quad (2)$$

$$MDIUF_i = \frac{1}{|J_i|} \sum_{k \in J_i} UF_k \quad (3)$$

$$FUF_i = \frac{\sum_{k \in J_i} CDD_k}{\sum_{k \in J_i} TD_k} \quad (4)$$

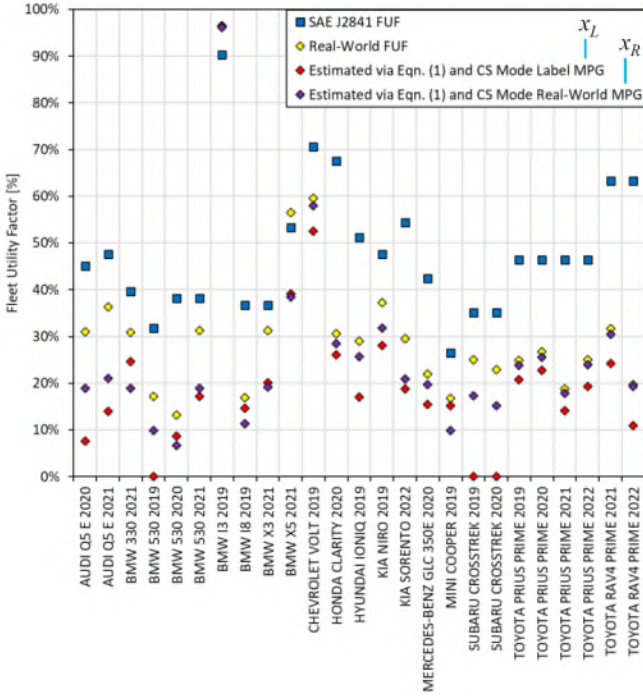


Figure 3. Re-analyzed FUF in BAR dataset

where the index i refers to a group of <make, model, model-year> PHEVs, while the index k refers to an individual vehicle sample and the set J_i is the collection of vehicle indices belonging to group i of <make, model, model-year> PHEVs. The symbol CDD represents the vehicle total lifetime CD mode distance travelled, which per SAE J2841 standard, includes both CD mode with engine off and CD mode with engine On, while the TD represents the vehicle total lifetime distance travelled.

In order to quantitatively gauge how much the UF results in BAR dataset are sensitive to one biasing element in the data, namely the relocation event, which many vehicles in the dataset may have experienced, we introduce the parameter (δ) that represents an amount of miles accumulated by the vehicle due to relocation trip (that most likely had very little access to charging), as well as possible additional miles while the vehicle was in temporary housing (where the vehicle owner may have not yet settled into their new permanent residence in California). We then calculate the sensitivity of FUF to the relocation event by considering the what-if scenario for

an amount of δ miles not being among the vehicle total lifetime distance travelled. For an individual vehicle sample, the adjusted UF for the what-if scenario is calculated as:

$$\bar{UF}_k(\delta) = \frac{CDD_k}{\max(CDD_k, TD_k - \delta)} \quad (5)$$

While the adjusted FUF for a group of vehicles (of PHEV variant i) is calculated as:

$$\bar{FUF}_i(\delta) = \frac{\sum_{k \in J_i} CDD_k}{\sum_{k \in J_i} \max(CDD_k, TD_k - \delta)} \quad (6)$$

Lastly, we calculated a normalized sensitivity metric for FUF by considering the impact of the change due to the relocation event as:

$$\phi_i(\delta) = \frac{\bar{FUF}_i - FUF_i}{FUF_i} \quad (7)$$

In equations (5-7), the ' $\sim$ ' marker on top of UF quantities represents the estimated UF for the what-if scenario of a relocation event, where an equivalent amount of miles δ had not occurred in the history of the vehicle. ϕ in Eqn. (7) is then our metric of interest as a sensitivity indicator.

Another possible source of bias in BAR dataset could be that some of the vehicles have unusually high annual mileage compared to typical vehicles in the US. Though annual distance (AD) travelled is not explicitly included in BAR dataset, we approximately estimate annual distance by dividing the vehicle total lifetime distance travelled (TD) by the number of years difference between the vehicle model-year and the time signature of the Smog check in BAR dataset. For vehicles whose Smog check occurred in the same year as the model-year, we estimate an approximate duration of 6 months. For vehicles whose model-year is ahead of the year of their Smog check, e.g. a Model Year 2021 vehicle tested in Calendar Year 2020, we approximately estimate an annual distance as 3 months. We then calculate a normalized metric (ψ) for the deviation from average annual mileage of vehicles in the US as:

$$\psi_i = \left(\frac{\sum_{k \in J_i} AD_k}{\beta \times |J_i|} \right) - 1 \quad (8)$$

where β is the annual mileage of vehicles in the US, whose value is taken as 11,467 miles/year based on estimates stemming from the National Highway Travel Survey of 2017 (NHTS-2017) [24].

In order to quantitatively explore the extent of bias in the data, an estimate for the value of δ is needed, which is not an easy task because BAR dataset does not include location history of the vehicles. With the only piece of information being that the vehicles "likely relocated from out of state into California", we examine some city-center to city-center shortest route driving distance: Phoenix, Arizona to San Diego is approx. 360 miles, Portland, Oregon to San Francisco is approx. 650 miles, while further out location such as

Denver, Colorado to San Francisco is approx. 1320 miles or Dallas, Texas to Los Angeles is approx. 1440 miles, and even further out locations on the East coast such as New York City or Boston can exceed 3000 miles. As such, we consider a value of 1600 miles as a reasonable value for δ that considers a mid-range relocation event. Plot of ϕ (sensitivity of FUF to relocation event) versus ψ (deviation of annualized miles from US average) is shown in Fig. 4.

For each of the <make, model, model-year> groups of PHEVs, the horizontal axis in Fig. 4 shows the deviation from US average annual miles travelled (ψ), with negative values implying “less than average US annual vehicle miles” and positive values implying “greater than average US annual vehicle miles”. The vertical axis in Fig. 4 shows the sensitivity (ϕ) of FUF calculation from BAR dataset with respect to a relocation event into California with $\delta=1600$ miles. In the subplot of Fig. 4.a, which shows ϕ versus ψ for the baseline vehicles samples filtering criteria summarized in Table 1, one may observe that for the majority of PHEV groups, FUF sensitivity ranges from high ($\phi > 10\%$) to extremely high ($\phi > 40\%$), which agrees with the general perception that for some low mileage vehicles (which filtering criteria setting a minimum of 3000 miles will allow), the inclusion of approx. 1600 miles with little or no access to charging would indeed skew the UF result.

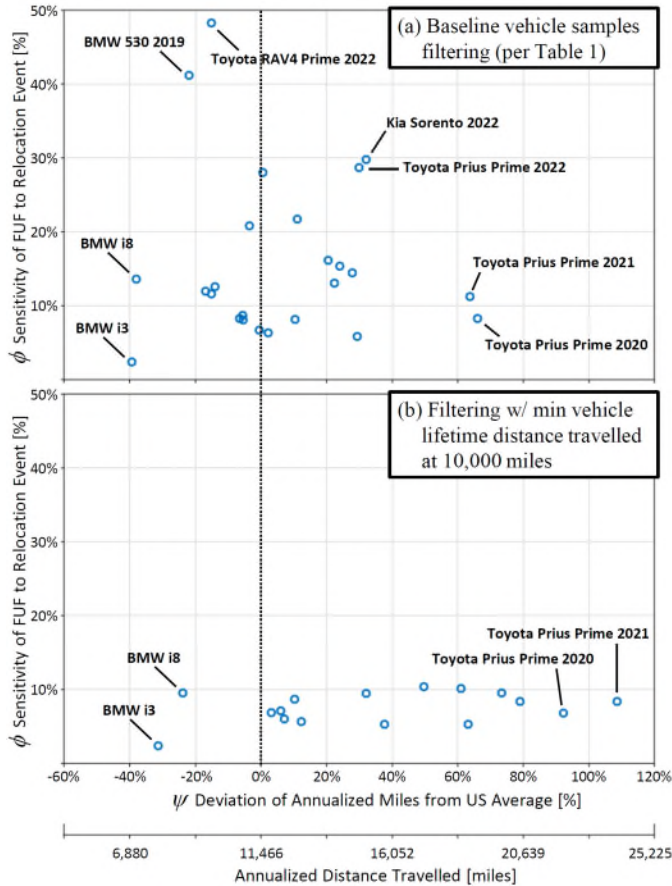


Figure 4. Sensitivity analysis of FUF estimation via BAR dataset to plausible relocation event to California from out of state

One potential counter measure to reduce the sensitivity of FUF estimation to a relocation event in BAR data, would be to adjust the

filtering criteria for the analyzed vehicle samples to consider only vehicles that have accumulated a larger amount of vehicle lifetime distance travelled at the time of the Smog check. We explore setting such a threshold at 10,000 miles, with the updated sensitivity result shown in Fig. 4.b. While noting that a US vehicle that had just accumulated 10,000 miles is still in the early stage of its lifetime, and additionally that such filtering successfully reduces FUF calculation sensitivity for all PHEV groups in Fig. 4.b, this filter setting brings forth unintended consequences that include:

- The remaining number of analyzable <make, model, model-year> PHEV groups gets reduced from twenty-five to sixteen
- Majority of the PHEV groups become biased towards high annual mileage vehicles that are not representative of typical US average driving, further skewing the UF of the remaining vehicles downward.

From such observations, it could be said that the currently available data for PHEVs in the BAR dataset is practically at a no-win situation. Setting a very low threshold for vehicle mileage causes the results to become uncertain due to high sensitivity to a relocation event, while increasing the mileage threshold skews that data into a different type of bias, that of high annual mileage vehicles. In fact, if one were to check the FUF values for the remaining vehicle samples in Fig. 4.b (which was done, but not shown due to space constraints), the FUF would appear worse than it does in Fig. 3 for the majority of PHEV groups. In other words, the biasing effect of high annual mileage vehicles is just as bad if not worse than the biasing effect of a relocation to California event, and we are unaware of countermeasures, outside the passage of time when non-relocated vehicles will be required to Smog check, that allow the BAR dataset to produce meaningful unbiased results.

Fuelly Dataset

As previously discussed in an earlier section of this paper, user data-entry errors in the Fuelly app or website can be detected by manually inspecting the data on the website at vehicle-by-vehicle, refueling-event-by-refueling-event. A copy of the cleaned-up data for several popular PHEV models across a broad span of AER has been uploaded to publicly accessible cloud storage [16]. However, since inferring UF from Fuelly data requires relying on Eqn. (1), we consider additional corrections with the help of observations from BAR dataset.

Obviously, if an inaccurate fuel consumption is assumed in Eqn. (1), the resulting estimate of UF will also be inaccurate. While Fuelly does not provide an estimate of real-world CS Mode fuel consumption, the BAR PHEV dataset does. As described in the preceding section, the PHEVs contained in the BAR dataset are far from representative of the larger PHEV population. Still, we can use the BAR dataset to quantify inaccuracies generated by relying on Eqn. (1) to estimate the real world UF. Doing so will inform our understanding of likely errors occurring when Eqn. (1) is used on other PHEV datasets, as was done in [12, 13] with the Fuelly data.

We attempt to correct two separate but related issues with using fuel-inferred UF via Eqn. 1. Both of these issues are observable in Fig. 3, with the first issue being that the “true” mileage-based UF is a larger number compared to x_R (via Eqn. 1, even when FC_C s is from the real-world data and not the window-sticker label) for the majority of PHEVs. This inaccuracy can mostly be attributed to the implicit, but

inaccurate, assumption of Eqn. (1) that CD Mode has no engine-on driving (otherwise known as blended mode). Examining the relative error due to this issue in Fig. 5, which shows wide span from almost no error (for BMW i3) to about 50% error (for BMW 530 2020).

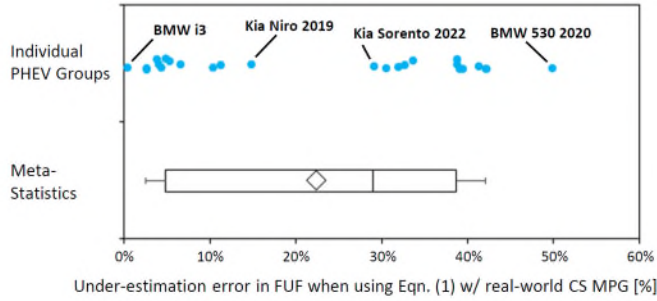


Figure 5. Under-estimation error of FUF in BAR dataset (x_R to real-world FUF in Fig. 3). Box plot indicates the 25th, 50th, and 75th percentile values, the whiskers indicate the 5th and 95th percentiles, while the diamond marker indicates the average value

The second issue occurs due to information about the real-world CS mode fuel consumption not being available in the data, thereby necessitating the use of label value CS mode fuel consumption. This means the fuel-inferred UF via Eqn. (1) will compute x_L rather than x_R . As previously noted, by the construction of Eqn. (1), if the label value CS mode fuel consumption under estimates the real-world CS mode fuel consumption, this will result in x_L under-estimating x_R . To quantify the extent of this second issue, we examine the real-world CS mode fuel consumption in BAR dataset (total fuel amount in CS mode divided by total miles in CS mode) for each vehicle sample then compare it to the corresponding label value CS mode fuel consumption. We present the statistics of this comparison for each <make, model, model-year> PHEV in Fig. 6.

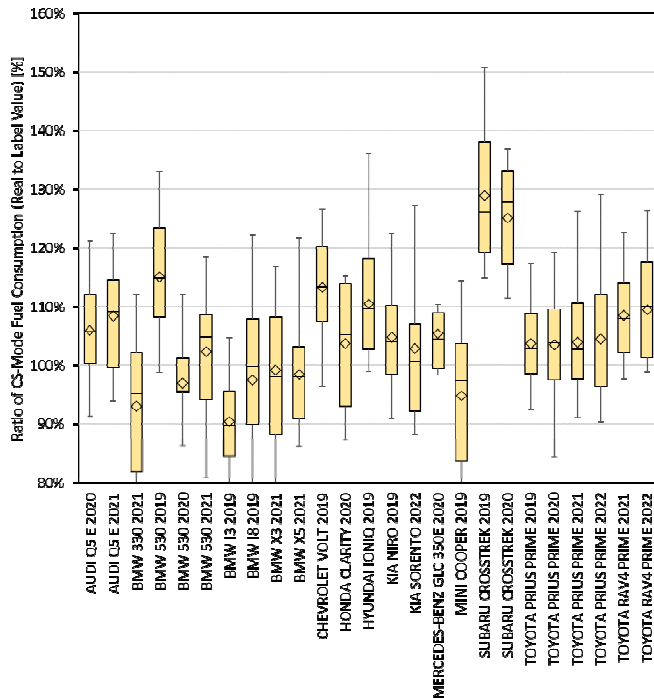


Figure 6. Ratio of real-world fuel consumption in BAR dataset while in CS mode compared to US-EPA window-sticker label value

While it is generally not surprising to see some variation in fuel consumption for real-world vehicles, the wide range of variation in Fig. 6 may be attributed to the relatively small sample size for some of the PHEV groups in the BAR dataset. Nonetheless, since the vehicles sample size in Fuely dataset is not significantly larger than that in the BAR dataset, similar variation in the Fuely vehicles could be expected. Accordingly, in what follows as a correction to Fuely results (from x_L to x_R), we will consider comparable uncertainty bounds in real-to-label fuel consumption of $\{-10\%, +10\%$ and $+30\%\}$ as observed in Fig. 6, notably with the median value of CS mode fuel consumption shifted by $+10\%$, corresponding to how the median value among medians in Fig. 6 is approx. at “110% of CS mode label value. One exception to this correction is for the BMW i3 which appeared to have less real-world CS mode fuel consumption than its label value CS mode fuel consumption, and its fuel-inferred UF (x_L) coincided well with its true mileage-based UF in Fig. 3.

Summary of UF estimation via Fuely dataset, along with different stages of corrections is shown in Fig. 7. Initially, fuel-inferred UF estimates from Fuely data were shown in [12] without proper data cleaning to remove vehicles with kWh entries as gasoline and against a backdrop of SAE J2841 MDIUF curve (when it should have been FUF curve), as in Fig. 7.a. It also bears to note that aside from correction of the back-drop curve to FUF, the Fuely results from [12] were utilized as-is without corrections in the NPRM by the US EPA [13] and this uncorrected Fuely result became a justification for the newly proposed FUF curve for GHG compliance of PHEVs.

After only removing the data entry errors of vehicle samples that had kWh values entered as gallons of gasoline, and combining PHEV vehicle samples for different model years that have the same powertrain, the result of Fuely dataset, which at that stage corresponds to x_L estimates, becomes the one shown in Fig. 7.b. We note that UF results in Fig. 7.b already appear much closer to SAE J2841 standard than they did in Fig. 7.a.

Applying uncertainty limits (at -10% and $+30\%$ of label value CS mode fuel consumption) and median value correction ($+10\%$ of label value CS mode fuel consumption) per the observations from Fig. 6, a correction from x_L to x_R is calculated and shown in Fig. 7.c. Further adjustment is then done in Fig. 7.d in order to correct for the difference between x_R and true (mileage based) UF per observations from Fig. 5. We scale the lower extension lines of Fig. 7.c by a factor of 1.05 (corresponding to 5% increase, which aligns with a cluster of PHEV groups towards the lower end in Fig. 5), scale the diamond marker of Fig. 7.c by a factor of 1.15 (corresponding to 15% increase, which is approximately the median value of Fig. 5 if the PHEV groups at values higher than 40% are excluded), and we scale the upper extension lines of Fig. 7.c by a factor of 1.3 (corresponding to 30% increase, which is approximately the median value of Fig. 5). These scaling updates (from Fig. 7.c to Fig. 7.d) are perceived to be better estimates of the true real world FUF, where UF includes the blended operation within CD mode. However, whether one takes the result in Fig. 7.c or Fig. 7.d as “the result” of Fuely dataset, either way, the result is reasonably matching with the FUF curve from SAE J2841 standard and does not seem to warrant a necessary revision to the standard.

Discussion

Having re-analyzed both BAR and Fuely datasets, we note that several <make, model, model-year> PHEVs in BAR dataset are

observed to under-perform in terms of their real-world FUF (not as badly as in [12], but still under-performing) compared to SAE J2841. However, such observation from BAR dataset is inconclusive towards making judgements about PHEVs in the US at large due to the multitude of issues noted about BAR dataset. Setting aside that nearly 75% of the original BAR dataset was removed from further consideration due to data quality issues, with high likelihood that

introducing a different bias towards high annual mileage vehicles. Either way, the accumulated mileage and driving/charging patterns of the PHEVs in the BAR dataset would not be representative of normal PHEVs usage in the US. In the future, as the non-relocated vehicles are required to report for Smog tests, the BAR PHEVs have the potential to provide valuable insight into how PHEVs (post Model Year 2019) were used in the past. However, the data quality issues that disqualified 75% of the collected vehicles must be understood and addressed before confidence is restored. Other issues that remain to date unidentified are also possible.

When re-analyzing Fuely dataset, it is necessary to clean the data from vehicles whose owners made obvious data entry errors (e.g., kWh of electricity entered as gallons of gasoline). Further, accounting for uncertainty bounds of fuel-inferred UF estimation via Eqn. (1), the Fuely FUF result for the analyzed PHEVs appears mostly in-line with SAE J2841 FUF curve, within reasonable bounds of “mileage will vary”.

Combining our findings from the re-analysis of BAR and Fuely datasets, we find no strong evidence to support an imminent need to revise the UF curves in SAE J2841 standard. It is true that this standard was last updated in 2010, and due to the large economic disruptions that started in 2008, the National Highway Travel Survey of 2001 (NHTS-2001), and not its 2009 revision, served as the standard of “typical US driving”. However, the US EPA have stated in [13] that they found no significant difference between UF curves derived from the most recent NHTS-2017 version, compared to the UF curves derived from the original NHTS-2001. Thus, for now, the available public data, both from the corrected Fuely analysis, as well as the five previous datasets examined in [10, 11] suggest that PHEVs in the US are, for the most part, achieving their expected UF per SAE J2841 standard.

Should a new investigation of this issue be conducted in the future, a good starting point would be reconsidering the three main categories of possible mismatch between the standard assumptions and the real-world [6, 25], namely: i) charging frequency and patterns, and an example of theoretical modeling work in that domain may be found in [25], ii) estimated AER, which is certified via the same procedures used for AER of BEVs, for which studies such as [26] have shown discrepancy compared to the real-world AER, and iii) statistical distribution of distance travelled per day. Furthermore, the combined effects of all three categories of possible UF mismatch is, in the authors opinion, mostly uncharted territory that is worthy of exploration. Such exploration however, ought to be done thoroughly via utilizing proper mathematically modeling that is grounded and verified via suitable real-world PHEVs data. BAR and Fuely datasets are at present, not suitable for such a task.

Conclusion

This paper re-examined recent claims that PHEVs in the US were significantly under-performing in terms of utility factor compared to their estimated performance via SAE J2841 standard. When re-analyzing two datasets (BAR and Fuely) that were cited as the source for such claims, we find that several PHEVs in BAR dataset do under-perform in terms of their FUF compared to SAE J2841, but the PHEVs in BAR dataset are not representative of general PHEVs in the US. When re-analyzing Fuely dataset after proper data cleaning and accounting for uncertainty bounds, the resulting FUF for analyzed PHEVs appears mostly in-line with SAE J2841 standard. As

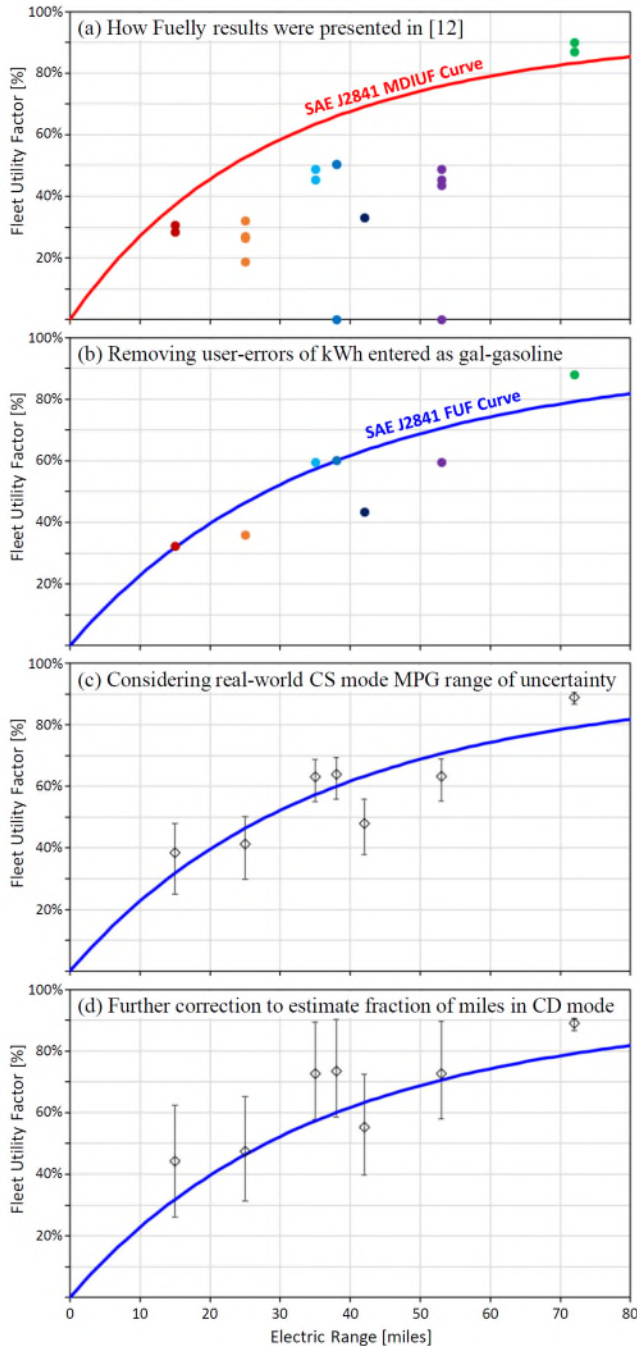


Figure 7. Re-analysis stages of Fuely dataset

most/entire set of PHEVs in the data are ones that relocated to California from outside the state, it practically impossible to de-bias the data to reduce sensitivity to the relocation event without

such, we find no strong evidence of an imminent need to revise the UF curves in SAE J2841 standard. However, with improvements in data and analytic tools available at present-day compared to 2010, in the spirit of continuous improvement, the motivation for re-visiting the standard could well be justified. However more thorough research is needed in order to properly facilitate such an update.

Nomenclature

AD	(Vehicle) Annual Distance travelled
AER	All-Electric Range [units of distance]
BAR	(California) Bureau of Automotive Repair
BEV	Battery-only Electric Vehicle
CD	Charge Depletion
CDD	(Vehicle lifetime) CD mode Distance travelled
CS	Charge Sustaining
FC	Fuel Consumption [units of volume per distance]
FUF	Fleet Utility Factor [dimensionless]
GHG	Greenhouse Gas
HEV	Hybrid Electric Vehicle
MDIUF	Multi-Day Individual Utility Factor [dimensionless]
MPG	Miles per Gallon
NPRM	Notice of Proposed Rule-Making
OBD	On-Board Diagnostics
PHEV	Plug-in Hybrid Electric Vehicle
TD	(Vehicle lifetime) Total Distance travelled
UF	Utility Factor [dimensionless]
x	Unknown quantity in Eqn. (1), an estimate of UF referred in this paper as “Fuel-inferred UF”. With subscript ‘L’ implies Label CS mode FC used in the equation, or with ‘R’ subscript when real-world CS mode is used in the equation [dimensionless]
β	Vehicle average annual travel distance in the US [units of distance]
δ	Travel distance incurred during home relocation event, a sensitivity parameter [units of distance]
ϕ	Metric for sensitivity of FUF to a home relocation event [dimensionless]
ψ	Metric for deviation in annual travel distance from the US average [dimensionless]

References

- [1] US Department of Energy (2011) <http://energy.gov/sites/prod/files/2014/05/f15/52723.pdf>; [accessed November 2023]
- [2] P. Chakraborty et. al. (2022) Addressing the range anxiety of battery electric vehicles with charging en route, *Nature Scientific Reports* **12**: 5588
- [3] Society of Automotive Engineers (2010) *Utility Factor Definitions for Plug-In Hybrid Electric Vehicles Using Travel Survey Data*. SAE J2841_201009
- [4] Regulation No 101 of the Economic Commission for Europe of the United Nations (2004) Uniform provisions concerning the approval of passenger cars equipped with an internal combustion engine with regard to the measurement of the emission of carbon dioxide and fuel consumption and of categories M1 and N1 vehicles equipped with an electric power train with regard to the measurement of electric energy consumption and range, [http://data.europa.eu/eli/reg/2004/101\(2\)/oj](http://data.europa.eu/eli/reg/2004/101(2)/oj) [accessed November 2023]
- [5] A. Eder et. al. (2014) Development of a European Utility Factor Curve for OVC-HEVs for WLTP. https://circabc.europa.eu/sd/a/92324676-bd8c-4075-8301-6caf12283beb/Technical%20Report_UF_final.pdf [accessed November 2023]
- [6] S. Raghavan, G. Tal (2020) Plug-in hybrid electric vehicle observed utility factor: Why the observed electrification performance differ from expectations. *Int. J. Sustainable Transportation* **16**(2): 105
- [7] P. Plötz, C. Moll, G. Bieker, P. Mock, Y. Li (2020) *Real-World usage of Plug-in Hybrid Electric Vehicles Fuel Consumption Electric Driving and CO2 Emissions*, <https://theicct.org/wp-content/uploads/2021/06/PHEV-white-paper-sept2020-0.pdf> [accessed May 2022]
- [8] P. Plötz, C. Moll, G. Bieker, P. Mock (2021) From lab-to-road: real-world fuel consumption and CO2 emissions of plug-in hybrid electric vehicles, *Environmental Research Letters* **16**: 054078
- [9] P. Plötz, J. Jöhrens (2021) *Realistic Test Cycle Utility Factors for Plug-in Hybrid Electric Vehicles in Europe*. Karlsruhe: Fraunhofer Institute for Systems and Innovation Research ISI, https://www.isi.fraunhofer.de/content/dam/isi/dokumente/cce/2021/BMU_Kurzpapier_UF_final.pdf [accessed November 2023]
- [10] K. Hamza, K. Laberteaux, K.C. Chu (2022) On inferred real-world fuel consumption of past decade plug-in hybrid electric vehicles in the US, *Environmental Research Letters* **17**: 104053
- [11] P. Plötz, C. Moll, G. Bieker, P. Mock (2023) Corrigendum: From lab-to-road: real-world fuel consumption and CO2 emissions of plug-in hybrid electric vehicles, *Environmental Research Letters* **18**: 099502
- [12] A. Isenstadt, Z. Yang, S. Searle, J. German (2022) *Real world usage of plug-in hybrid vehicles in the United States*, International Council on Clean Transportation Report
- [13] US Environmental Protection Agency (2023) Multi-Pollutant Emissions Standards for Model Years 2027 and Later Light-Duty and Medium-Duty Vehicles 2023, <https://www.regulations.gov/search?documentTypes=Proposed%20Rule&filter=EPA-HQ-OAR-2022-0829> [accessed October 2023]
- [14] California Bureau of Automotive Repair (2023), <https://www.bar.ca.gov/> [accessed November 2023]
- [15] California Bureau of Automotive Repair (2023) Records Request, <https://www.bar.ca.gov/records-requests> [accessed November 2023]
- [16] Cloud Storage on Google Drive (2023) Data from BAR and Fuelly, https://bit.ly/EPA_BAR_Fuelly [accessed November 2023]
- [17] California Bureau of Automotive Repair (2023) Vehicles Subject to Smog Check, <https://www.bar.ca.gov/consumer/smog-check-program/faq> [accessed November 2023]
- [18] Samsara (2021) The Complete Guide to California Smog Check Requirements, <https://www.samsara.com/guides/the-complete-guide-to-california-smog-check-requirements/> [accessed November 2023]
- [19] The Orange County Registrar (2023) Your car might not need a smog test, but the state still wants money, <https://www.ocregister.com/2021/03/19/your-car-might-not-need-a-smog-check-but-the-state-still-wants-money/> [accessed November 2023]
- [20] US Department of Energy (2023) Plug-in Hybrid Electric Vehicles: Learn More About the Label,

- <https://www.fueleconomy.gov/feg/label/learn-more-PHEV-label.shtml> [accessed November 2023]
- [21] US Department of Energy (2023) Your Mileage Will Vary, https://www.fueleconomy.gov/feg/why_differ.shtml [accessed November 2023]
- [22] Fueily (2022) Track, Share, and Compare your Vehicle, <https://www.fuely.com/> [accessed January 2023]
- [23] Fueily (2023) Tesla Y MPG, <https://www.fuely.com/car/tesla/y> [accessed November 2023]
- [24] US Department of Transportation (2019) Highway Statistics 2017, <https://www.fhwa.dot.gov/policyinformation/statistics/2017/vm1.cfm> [accessed November 2023]
- [25] K. Hamza, K. Laberteaux (2023) Utility Factor Curves for Plug-in Hybrid Electric Vehicles: Beyond the Standard Assumptions,

World Electric Vehicle Journal **14**: 301, <https://www.mdpi.com/2032-6653/14/11/301> [accessed November 2023]

- [26] G. Pannone, D. VanderWerp (2023) Comparison of On-Road Highway Fuel Economy and All-Electric Range to Label Values: Are the Current Label Procedures Appropriate for Battery Electric Vehicles?, *SAE Technical Paper* 2023-01-0349

Contact Information

Toyota Motor North America-R&D (FRD)
1555 Woodridge Avenue, Ann Arbor, MI 48105
Tel: 734-546-2423 Fax: 734-995-4200
Email: karim.hamza@toyota.com