Estimating an Electric Vehicle’s “Distance to Empty” Using Both Past and Future Route Information

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

Citation

As Published
http://dx.doi.org/10.1115/DETC2013-12880

Publisher
ASME International

Version
Final published version

Accessed
Thu Apr 04 11:07:58 EDT 2019

Citable Link
http://hdl.handle.net/1721.1/119235

Terms of Use
Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.

Detailed Terms
ESTIMATING AN ELECTRIC VEHICLE’S “DISTANCE TO EMPTY” USING BOTH PAST AND FUTURE ROUTE INFORMATION

Lennon Rodgers
Massachusetts Institute of Technology
Cambridge, MA, USA

Daniel Frey
Massachusetts Institute of Technology
Cambridge, MA, USA

Erik Wilhelm
Singapore University of Technology and Design, Singapore

ABSTRACT

An electric vehicle’s Distance to Empty (DTE) is defined as the actual distance the vehicle can be driven before recharging is required. A real-time estimate is commonly displayed on the vehicle’s electronic instrument cluster and is used by the driver to plan their route. It is proved in this paper that the challenge for any DTE estimation algorithm is to accurately predict the future energy consumption of a vehicle. Future energy can be predicted reliably if either (i) future energy consumption is sufficiently similar to the past or (ii) applicable information about the future is known beforehand. A stochastic simulation was used to show that the average energy use (Wh/km) measured over the past ~300 km often does an adequate job at predicting future energy use. A conventional DTE algorithm assumes this condition by “blending” both a long- and short-term average of past energy use. However, significant changes in driving conditions (e.g. traffic or auxiliary energy use) for sustained periods of time can cause large errors in DTE estimates. This paper showed that DTE error can be reduced if those future changes are detected beforehand by obtaining route information from the driver. A multivariate linear regression model was derived that adjusts a historical average of energy consumption based on estimated changes in speeds, traffic and temperature. This method utilizes the measuring ability of the vehicle and thus does not require complex physics-based models. The algorithm could be implemented as a cloud-based mobile phone application since it is computationally light and the model is fitted using historical driving data. Finally, the algorithms were compared using a stochastic vehicle simulation and it was shown that incorporating future route information can significantly reduce DTE error.

Keywords: Distance to Empty, range algorithms, EV range estimation, stochastic driving cycles, electric vehicle simulation, range anxiety

INTRODUCTION

The maximum DTE for electric vehicles (EV) is typically ~100 to 400 km less than gasoline vehicles and a full recharge usually takes hours instead of minutes [1]. Also, the energy consumption of EVs is more influenced by auxiliary loads (e.g. heating). For these reasons it is important to provide an accurate DTE estimate. Recent studies have shown that current DTE algorithms are insufficient and often cause “range anxiety” among drivers [2][3]. Predicting DTE is difficult because of the stochastic nature of driver behavior and the environment, the lack of a quantitative understanding for how energy use is affected, and the fairly basic algorithms being used.

A previous study used a basic vehicle model combined with static assumptions about auxiliary use to estimate the maximum DTE, commonly referred to as “range” [4]. Though their results closely matched the expected values published by the manufacturers, they did not take into account the stochastic nature of the driving conditions nor did they investigate the accuracy of DTE with distance. Some have approached the more general topic of estimating the energy consumed between two points [5][6]; they used commercially available simulation software with historical traffic and road data. There are similar approaches that require high fidelity models to estimate energy use and/or DTE [7]. Others describe a method where historical data is used to predict the energy required for future trips [8][9]. Many of the techniques related to energy estimation are applied to power-split control algorithms for plug-in hybrids [10] or gasoline vehicles [11]. Finally, there are many patents issued by automobile companies that describe various historical data averaging techniques used to estimate DTE[17][18][19].
None of the research described above show results of a $D_{TE}$ simulation or experimentation, nor do they explain how their approaches would be implemented. They also all stop short of measuring the quality of the $D_{TE}$ estimates and making comparisons to other methods. Finally, there has been little effort to understand the fundamental relationship between energy estimation and $D_{TE}$ error. This paper aims to contribute in these areas and start a more open and in-depth discussion of $D_{TE}$ algorithms.

OVERVIEW

The first section, Fundamental Equations, shows that the future energy use must be predicted in order to accurately estimate $D_{TE}$. This section also quantifies how errors in predicting the future energy use relate to errors in $D_{TE}$. To serve as an example of the algorithms being used today, a Conventional $D_{TE}$ Algorithm is derived that blends together a long-term and short-term average of past energy use. This algorithm works well except in cases when significant changes in driving conditions (e.g. traffic or auxiliary energy use) occur for sustained periods of time. Next, the section titled A Regression-based $D_{TE}$ Algorithm shows that $D_{TE}$ error can be reduced if those future changes are detected beforehand. A multivariate linear regression-based model is derived that adjusts a historical average of energy consumption up or down based on estimated changes in driving conditions (temperature and traffic). Both the conventional and regression-based algorithms are compared in the remaining sections using stochastic vehicle simulations.

FUNDAMENTAL EQUATIONS

This section derives equations that quantify $D_{TE}$ error and aid in deriving $D_{TE}$ algorithms. The energy stored in the battery, $E_b$, is consumed as the vehicle is driven a distance $x$ (Figure 1a). An important metric used frequently in this discussion is the average energy use, $\bar{\rho}$, which is defined as the quantity of energy, $\Delta E_B$, consumed over a distance $\Delta x$ and has units of Wh/km:

$$\bar{\rho} \equiv \frac{\Delta E_B}{\Delta x}$$  

The future average energy use, $\bar{\rho}_f(t)$, can be determined by evaluating Equation 1 between the current time, $t$ and final time, $t_f$:

$$\bar{\rho}_f(t) = \frac{E_b(t)}{x(t_f) - x(t)}$$  

Where $t \rightarrow \mathbb{R} \in \{t_0, t_f\}$, $t_0$ is the time when the battery is fully charged and $t_f$ is the time when the battery is fully discharged. $E_b(t)$ is the battery energy remaining at time $t$. By definition, $D_{TE}$ can be written as (Figure 1a):

$$D_{TE}(t) = x(t_f) - x(t)$$  

Combining Equations 2 and 3 yields:

$$D_{TE}(t) = \frac{E_b(t)}{\bar{\rho}_f(t)}$$  

It is useful to evaluate Equation 2 at $t_0$:

$$\bar{\rho}_f(t_0) = \frac{E_b(t_0)}{D_{TE}(t_0)} = \frac{E_b(t_0)}{x(t_f)}$$  

Now solving for $x(t_f)$:

$$x(t_f) = \frac{E_b(t_0)}{\bar{\rho}_f(t_0)}$$  

Combining Equations 3 and 6 yields a different equation for predicting $D_{TE}$:

$$D_{TE}(t) = \frac{E_b(t_0)}{\bar{\rho}_f(t_0)} - x(t)$$  

Conceptually, both Equations 4 and 7 show that $D_{TE}$ can be determined if the current battery energy and the future energy consumption are known. An onboard Battery Management System (BMS) measures $E_b(t)$ and it will be assumed to be known perfectly. Thus the task of a $D_{TE}$ algorithm is to estimate the future energy consumption, $\bar{\rho}_f$. Also, Equation 7 reveals that a perfect algorithm would predict a linear relationship for the “Actual” $D_{TE}$ (Figure 1b).

\[\text{Figure 1: A schematic of the vehicle’s battery energy (a) and } D_{TE} \text{ (b) with distance.}\]

Error Analysis

Errors in $D_{TE}$ are caused by an algorithm’s inability to perfectly predict $\bar{\rho}_f$. Thus this section quantifies how errors in estimating $\bar{\rho}_f$ cause errors in $D_{TE}$. Assume that a $D_{TE}$ algorithm estimates $\bar{\rho}_f(t)$, and the associated error is defined as:

$$e_{\bar{\rho}_f}(t) \equiv \bar{\rho}_f(t) - \bar{\rho}_f(t)$$  

Where $^\wedge$ designates an estimate of the actual value. The corresponding error in the $D_{TE}$ estimate is then (Figure 1b):

$$e_{D_{TE}}(t) \equiv D_{TE}(t) - D_{TE}(t)$$  

Substituting Equations 4 and 8 into Equation 9 yields:

$$e_{D_{TE}}(t) = E_b(t) \left[ \frac{1}{\bar{\rho}_f(t)} + e_{\bar{\rho}_f}(t) - \frac{1}{\bar{\rho}_f(t)} \right]$$  

Rearranging yields:

$$e_{D_{TE}}(t) = -\frac{E_b(t)}{\bar{\rho}_f(t)} \left( e_{\bar{\rho}_f}(t) \frac{\bar{\rho}_f(t)}{\bar{\rho}_f(t)} + 1 \right)$$  

Combining with Equation 4 yields:

$$e_{D_{TE}}(t) = -D_{TE}(t) \left[ e_{\bar{\rho}_f}(t) \frac{\bar{\rho}_f(t)}{\bar{\rho}_f(t)} + 1 \right]$$  

Copyright © 2013 by ASME
Equation 12 shows that when Equation 4 is used, the error in estimating $D_{TE}$ will attenuate to zero towards the end of the discharge. Equation 12 can be rearranged as:

$$\frac{e_{D_{TE}}(t)}{D_{TE}(t)} = -\left[\frac{e_{pf}(t)}{\bar{pf}(t)} \right] \frac{e_{pf}(t)}{\bar{pf}(t) + 1}$$  \hspace{1cm} 13$$

Now defining the percentage error, $E$, as:

$$E_{D_{TE}}(t) = \frac{e_{D_{TE}}(t)}{D_{TE}(t)}$$

and

$$E_{pf}(t) = \frac{e_{pf}(t)}{\bar{pf}(t)}$$

Combining Equations 13 and 14:

$$E_{D_{TE}}(t) = -\left[ \frac{E_{pf}(t)}{\bar{pf}(t) + 1} \right]$$

Equation 15 is plotted in Figure 2 and shows the relationship between errors in estimating $\bar{pf}$ and the resulting $D_{TE}$ errors. It can be seen that the relationship is linear for small percentage errors (~10%) but becomes increasingly non-linear for larger errors. The plot also shows that $D_{TE}$ error is greater when $\bar{pf}$ is underestimated versus when it is overestimated (e.g. when $\bar{pf}$ is underestimated by 30% the corresponding $D_{TE}$ error is ~43% while when $\bar{pf}$ is overestimated by 30% the corresponding $D_{TE}$ error is ~23%).

![Figure 2: A plot of Equation 15 shows how errors in estimating $\bar{pf}$ result in $D_{TE}$ errors.](image)

**CONVENTIONAL $D_{TE}$ ALGORITHMS**

Conventional algorithms assume that the future energy use will be similar to the past:

$$\bar{pf} \approx \bar{p}_p$$  \hspace{1cm} 16$$

Where $\bar{p}_p$ is the average energy use of “past driving,” which can be determined by using past (historical) driving energy data (e.g. 1km, running, or blended averages as defined below). The $D_{TE}$ can then be estimated by combining Equations 4 and 16:

$$\bar{D}_{TE}(t) = \frac{E_b(t)}{\bar{pf}(t)}$$  \hspace{1cm} 17$$

The values of $\bar{D}_{TE}$ will deviate from $D_{TE}$ according to Equation 12 and the amount depends on the validity of Equation 16 (Figure 1b). This is considered a “conventional” approach since it is likely very similar to the methods being used in EVs today, based on the limited amount of related details in patents and literature [17][18][19].

The energy use averaged over the past 1km is defined as:

$$\bar{p}_{1km}(t) = \frac{E_b(t_{1km}) - E_b(t)}{x(t) - x(t_{1km})}$$  \hspace{1cm} 18$$

where $t$ is the current time and $t_{1km}$ is the time 1 km in the past.

And the running average is defined as:

$$\bar{p}_{\text{running}}(t) = \frac{E_b(t_{0}) - E_b(t)}{x(t) - x(t_{0})}$$  \hspace{1cm} 19$$

The “blended” algorithm uses a long-term average of energy use, $\bar{p}_{\text{long}}$, during the beginning but as the battery is discharged the algorithm blends to a more recent short-term average of energy use, $\bar{p}_{\text{short}}$. This can be written as:

$$\bar{p}_{\text{ blends}}(t) = \bar{p}_{\text{long}}(t) - b(\bar{p}_{\text{long}}(t) - \bar{p}_{\text{short}}(t))$$  \hspace{1cm} 20$$

Where $b \in [0, 1]$. The value of $b$ is typically chosen based on a linear function that changes with State of Charge (SOC):

$$b(t) = 1 - SOC(t)/100$$  \hspace{1cm} 21$$

SOC is defined as the percentage of battery energy remaining:

$$SOC(t) \equiv \frac{E_b(t)}{E_b(t_{0})} \cdot 100$$  \hspace{1cm} 22$$

These historical averages work well as long as the future conditions are similar to the past. However, there are cases when changes in energy use will cause significant errors. For example, assume that a vehicle has been consuming energy at a constant rate of 210 Wh/km for the long- and short-term past. When the vehicle is ~50 km into a full discharge, a heater load is turned on, which causes a 30% increase in energy consumption1 (Figure 3a and b). Figure 3c shows the actual and estimated $D_{TE}$ when the running average or blended algorithms are used. The corresponding error for the blended algorithm is shown in Figure 3d.

![Figure 3: A simple example showing how auxiliary use can cause errors in the $D_{TE}$ estimate when the running average or blended algorithms are used.](image)

1 Figure 3a shows that a 2 kW constant auxiliary load would cause a 30% increase in energy consumption when driving at a constant speed of 45 km/hr. The curves shown are based on simulation of a sedan-sized vehicle traveling at constant speeds and meant only for approximate values.
shown in Figure 4c and d, and for this case there is no change in auxiliary use or traffic. For these examples $\bar{p}_{\text{long}}$ is defined as the energy use averaged over the previous 300 km. All three of these examples support the following conclusion: taking averages and blends of past data and assuming that $\bar{p}_t \approx \bar{p}_p$ works well as long as there are no sustained changes in driving conditions (e.g., sudden use of heater for a sustained period of time). The following section describes a new $D_{TE}$ algorithm that utilizes these results. In other words, it adjusts a historical measurement of energy use based on estimated changes in driving conditions.

A REGRESSION-BASED $D_{TE}$ ALGORITHM

This section introduces the concept of using estimates of future driving conditions to more accurately predict $D_{TE}$. For example, if the route, traffic and weather are known before the start of a trip, this information could be fed into a model that predicts the future energy consumption. Figure 5 attempts to capture all of the factors that affect energy use in EVs. Modeling each of the factors using physics-based models is difficult. Physics-based models are also undesirable for broad application since they require a significant amount of vehicle-specific calibration and can be computationally intensive for real-time applications. To avoid these shortcomings, this section will derive a method that uses a model that can be learned and improved as the vehicle is driven.

As discussed in the previous section, $\bar{p}_{\text{long}}$ is often a good estimate for future energy consumption. An improved algorithm is one that could adjust the value of $\bar{p}_{\text{long}}$ up or down to provide a better estimate for $\bar{p}_t$ based on estimates of future driving conditions (e.g., traffic, outside temperature). Specifically, let us multiply $\bar{p}_{\text{long}}$ by an adjustment factor, $y$, that makes the following true:

$$p_t(t) = y(t) \bar{p}_{\text{long}}(t)$$  \hspace{1cm} 23

Combing Equations 4 and 23 yields:

$$D_{TE}(t) = \frac{E_p(t)}{y(t) \bar{p}_{\text{long}}(t)}$$  \hspace{1cm} 24

The advantage to this approach is that there is no need for a detailed physical model in order to use future information. In other words, instead of precisely predicting the future energy consumption using detailed models, a past measurement is simply adjusted up or down based on estimates of future conditions.

Since $y$ cannot be determined perfectly beforehand, an estimate $\hat{y}$ can be made when an unknown residual error, $\varepsilon$, is included:

$$y(t) = \hat{y}(t) + \varepsilon(t)$$  \hspace{1cm} 25

Rearranging:

$$\hat{y}(t) = y(t) - \varepsilon(t)$$  \hspace{1cm} 26

Then an estimate for $D_{TE}$ can be written as:

$$\hat{D}_{TE}(t) = \frac{E_p(t)}{\hat{y}(t) \bar{p}_{\text{long}}(t)}$$  \hspace{1cm} 27

Rewriting Equation 27 in discrete form yields:

$$\hat{D}_{TE}(t_i) = \frac{E_p(t_i)}{\hat{y}(t_i) \bar{p}_{\text{long}}(t_i)}$$  \hspace{1cm} 28

Finally, it is best to discretize the problem based on State of Charge (SOC), since the range of values is constant {0,100} between the various training sets (unlike min/max time and distance, which change based on driving conditions). In discretized form: SOC$_i$ → $\mathbb{Z}$ ∈ {0,100}:

$$SOC(t_i) = SOC_i$$  \hspace{1cm} 29

For example, $t_{50}$ is the time when SOC = 50%.

**Figure 5:** The above diagram shows all of the factors that influence a vehicle’s energy consumption and specifically how various driver decisions and options (purple) lead to energy losses (red) and storage (green).

**Multivariate Regression Model**

The goal is to minimize the residual error written in Equation 26. This is accomplished using a multivariate linear regression model. A regression model uses historical data to learn the relationship between a set of measured explanatory
Variables, \( \chi \), and the response variable, \( \gamma \). Assuming a linear model of the form [14]:

\[
f(t) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m
\]

Where \( \beta_i \) is a set of \( m \) unknown coefficients that are determined from historical data (training set). The variables \( \chi \) must be measurable and predictable factors that cause differences in energy use between the past and future. Equation 30 can be written in matrix form:

\[
\gamma = \chi \beta
\]

Once \( \gamma \) is known, Equation 28 can be used to determine \( \tilde{D}_{TE} \).

Rewriting Equation 25 in matrix form:

\[
\gamma = \gamma + \epsilon
\]

Combining Equations 31 and 32:

\[
\gamma = \chi \beta + \epsilon
\]

Solving for the residual error:

\[
\epsilon = \gamma - \chi \beta
\]

The residual error can be minimized through a least squares estimator, which can be written as a function \( S(\epsilon) \):

\[
S(\epsilon) = \sum \epsilon^2 = \epsilon^T \epsilon = (\gamma - \chi \beta)^T (\gamma - \chi \beta)
\]

Expanding the right side of the equation:

\[
S(\epsilon) = \gamma^T \gamma - \gamma^T \chi \beta - \beta^T \chi^T \gamma + \beta^T \chi^T \chi \beta
\]

Setting the derivative of \( S(\epsilon) \) to zero solves for the minimum residual error:

\[
\frac{dS}{d\epsilon} = -2\chi^T \gamma + 2\chi^T \chi \beta = 0
\]

Rearranging Equation 37 yields the Normal Equation:

\[
\chi^T \gamma = \chi^T \chi \beta
\]

Solving for \( \beta \):

\[
\beta = (\chi^T \chi)^{-1} \chi^T \gamma
\]

The strategy is to “learn” values of \( \beta \) using known historical values of \( \gamma \) and \( \chi \) and Equation 39. Then the learned values of \( \beta \) and the real-time values of \( \chi \) can be used to calculate \( \tilde{y}(t) \) using Equation 30. The following subsection will explain how the explanatory variables \( \chi \) are defined and calculated.

Explanatory Variables: \( \chi \)

The variables \( \chi \) must be (1) measurable and predictable factors that (2) cause differences in energy use between the past and future. The past is defined by \( \tilde{P}_{long} \). For example, if \( \tilde{P}_{long} \) is the energy use over the past 300 miles, then the past is defined as the past 300 miles worth of data. The value of \( \tilde{P}_{long} \) compresses a large amount of microstructure driving data into a single measurement that quantifies how energy has been used historically. When using \( \tilde{P}_{long} \) to estimate the future, the value needs to be adjusted up or down if the future is different from the past. The future is the data contained between \( x(t) \) and \( x(t+1) \). (1) The Explanatory Variables must be factors that indicate a difference in energy use between the past (\( \tilde{P}_{long} \)) and future (\( \tilde{P}_{f} \)): For example, assume that the energy and distance values used to determine \( \tilde{P}_{long} \) were collected while the ambient temperature was on average 15°C, while it is predicted that the future will contain an average ambient temperature of 5°C. Since auxiliary heater use at 15°C ambient will be less than that at 5°C, it can be assumed that the change in ambient temperature may cause a change in future energy use. Thus the change in average ambient temperature can be one of the explanatory variables. (2) The Explanatory Variables should be measured from past data and estimated (predicted) into the future. The ambient temperature serves as a good example of a factor that can be easily measured using a temperature sensor. The future temperature can be predicted using weather forecasts and/or by assuming that the temperature at \( t_0 \) will be equal to the future average temperature.

There are other factors that might cause a difference in energy use but few can be directly measured and predicted. For example, the driver’s mood might cause an increase in future energy use though it cannot be easily measured and predicted. Given these constraints, the following explanatory variables were used:

1. Change in Ambient Temperature: As described previously, the change in average ambient temperature is defined as:

\[
\Delta T_a(t_i) = |T_{af}(t_i) - 20| - |T_{a,p}(t_i) - 20|
\]

2. Change in Traffic Conditions: Traffic conditions have a significant influence on energy use and recent advances in traffic sensing techniques make traffic measurements and predictions possible. For example, Google Maps is able to provide a quantitative measure of the upcoming (future) traffic delay and corresponding estimates of average speeds over distance segments [20]. This information is accessible in real time via their internet-based Application Programming Interface (API). If the future route is specified, an estimate for future traffic conditions can be made. However, to reduce the scope of this project, it was assumed that the traffic conditions can be captured through the percentage of time spent at idle conditions (i.e. zero speed):

\[
\Delta t_{idle}(t_i) = t_{idle,f}(t_i) - t_{idle,p}(t_i)
\]

Where \( t_{idle,p}(t_i) \) is the average percentage of time spent at idle in the past and \( t_{idle,f}(t_i) \) is the estimated average percentage of time spent at idle in the future.

3. Change in Average Speed: It is well known that high speeds require more energy [15]. It will be assumed that the future average speed could be estimated using route information from Google Maps (or equivalent). Thus a difference in average speed will be used as an explanatory variable:

\[
\Delta v_{ave}(t_i) = v_{f}(t_i) - v_{p}(t_i)
\]

Where \( v_{p}(t_i) \) is the average speed in the past and \( v_{f}(t_i) \) is the average speed estimated (or known) of the future.

Creating a Training Dataset

Assume we are analyzing multiple historical discharges that each start at 100% SOC and data is available to calculate both \( \tilde{P}_{long} \) and \( \tilde{P}_{f} \). The actual values of \( y(t_i) \) can be determined by rewriting Equation 23:

\[
y(t_i) = \tilde{P}_{f}(t_i) / \tilde{P}_{long}(t_i)
\]

Copyright © 2013 by ASME
Since there are \( n \) discrete values of \( t_i \), the values of \( y(t_i) \) can be written as a \( nx1 \) vector:

\[
\mathbf{y} = \begin{pmatrix}
y(t_1) \\
y(t_2) \\
\vdots \\
y(t_n)
\end{pmatrix}
\]

The values of \( \Delta T_a \), \( \Delta t_{idle} \) and \( \Delta v_{ave} \) for each time \( t_i \) form the explanatory matrix:

\[
\mathbf{X} = \begin{pmatrix}
1 & \Delta T_a(t_1) & \Delta t_{idle}(t_1) & \Delta v_{ave}(t_1) \\
1 & \Delta T_a(t_2) & \Delta t_{idle}(t_2) & \Delta v_{ave}(t_2) \\
\vdots & \vdots & \vdots & \vdots \\
1 & \Delta T_a(t_n) & \Delta t_{idle}(t_n) & \Delta v_{ave}(t_n)
\end{pmatrix}
\]

This process is repeated for multiple discharges to build a training set with a sufficient regression “fit” and to avoid extrapolation [14].

Validating the Assumptions and Fit of the Regression Model

The assumptions of a linear regression model can be summarized as follows [14]:

1. The linear model adequately describes the behavior of the data.
2. The random error, \( e \) (Equation 32), is an independent and normally distributed random variable, with a zero mean and variance \( \sigma^2 \).

There are various checks commonly used to evaluate these underlying assumptions for a given training dataset. An observation with Cook’s distance larger than three times the mean Cook's distance might be an outlier [16]. The fit might be improved if the outliers are removed. Figure 6a shows that < 3% of the data points for this training dataset are considered outliers. The second is to perform a significance test on each of the coefficients, \( \beta_i \) (Equation 30), to ensure that the model is not over-specified. This is done by calculating the pValues associated with the null hypothesis that each coefficient is zero (Table 1). The pValues can also be used along with the coefficient of determination, \( R^2 \), to check the fit of the model (Table 2); the small pValues (<< 0.05) and large \( R^2 \) shown in the tables indicate that the linear model is valid.

To verify the conditions in Assumption 2, a frequency distribution of the residual error is plotted to check for normality. It can be seen from Figure 6b that the distribution has a normal shape.

Table 1: Metrics used to determine the significance of each coefficient.

<table>
<thead>
<tr>
<th>Estimate of ( \beta_i )</th>
<th>Standard Error</th>
<th>tStat</th>
<th>pValue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.08</td>
<td>0.00381</td>
<td>283</td>
</tr>
<tr>
<td>( \Delta T_a )</td>
<td>0.00868</td>
<td>0.000161</td>
<td>53.9</td>
</tr>
<tr>
<td>( \Delta t_{idle} )</td>
<td>0.0306</td>
<td>0.000746</td>
<td>41.1</td>
</tr>
<tr>
<td>( \Delta v_{ave} )</td>
<td>0.0291</td>
<td>0.000629</td>
<td>46.34</td>
</tr>
</tbody>
</table>

\( \Delta T_a \), \( \Delta t_{idle} \) and \( \Delta v_{ave} \) are derived from simulations. The explanatory matrix:

\[
\mathbf{X} = \begin{pmatrix}
1 & \Delta T_a(t_1) & \Delta t_{idle}(t_1) & \Delta v_{ave}(t_1) \\
1 & \Delta T_a(t_2) & \Delta t_{idle}(t_2) & \Delta v_{ave}(t_2) \\
\vdots & \vdots & \vdots & \vdots \\
1 & \Delta T_a(t_n) & \Delta t_{idle}(t_n) & \Delta v_{ave}(t_n)
\end{pmatrix}
\]

STOCHASTIC VEHICLE SIMULATIONS

Driving data is needed to understand the relationship between past and future energy consumption of vehicles (e.g. Figure 4a) and to compare \( D_{TE} \) algorithms (e.g. Figure 4b). Driving data could be obtained by operating an instrumented EV and collecting the data required by the \( D_{TE} \) algorithm (e.g. speed, time, temperatures, energy consumption, etc.). The collected data could then be used to retrospectively simulate \( D_{TE} \) algorithms under the specific conditions. The challenge in this experimental approach would be to capture a variety of driving scenarios large enough to have confidence in the algorithm. For example, an algorithm might work well only under specific conditions (e.g. moderate temperatures with no traffic) though poorly under all other conditions. Obtaining such a large data set would be extremely time intensive and complex to obtain.

An alternative option is to use a vehicle simulation, which is the approach used in this paper. Ideally each factor shown in Figure 5 would be included in a system level simulation. However, it is only practical to include a subset of factors. An overview of the parametric-based “backward facing” stochastic simulation is shown in Figure 7. The simulation environment consists of a set of subsystem models that attempt to capture the most significant variations in vehicle energy use. For example, the stochastic nature of vehicle speed is simulated using a mode-based Markov model combined with real driving data. Additionally, an auxiliary energy model (e.g. heating and cooling of the cabin) is based on driving data collected using a fleet of EVs. The details of the models and approach are discussed separately [12].
Nissan Leaf and BMW ActiveE. The stochastic speed profile is generated using random combinations of city and highway conditions with and without traffic. The intent is to explore the space of possible speed profiles so that the $D_{TE}$ algorithms can be evaluated for a large number of scenarios. An example speed profile is shown in Figure 8a with the corresponding energy consumption in Figure 8b. All of the information from the past, present and future is calculated (speed, energy, temperature, etc.).

The conventional (Blend) and new (Regression) $D_{TE}$ algorithms were simulated using the three previously described datasets and the results are shown in Figure 8c and d. The speed and battery energy profiles are shown in Figure 8a and b. This simulation assumed that the discharge dataset occurred at 10°C as to simulate a drop in temperature from the past. The average $D_{TE}$ error, as defined in Equation 9, provides a quantitative comparison of the algorithms (Figure 8d). The averages are taken over three separate sections (start, mid, end) to observe if the algorithm is better at estimating a certain section of discharge (e.g. an algorithm might be accurate during the middle of a discharge but not at the start). Also, the positive and negative errors are shown separately to better understand if the algorithm is over or under estimating $D_{TE}$.

Figure 8d shows that, for this example simulation, the regression-based approach reduced the error by ~15% at the beginning of discharge. It is important to note that the ability to accurately estimate $D_{TE}$ at the start of a full charge is likely the most critical task of an algorithm since the driver uses this estimate to plan their route.

To better understand how the algorithms perform for a wider variety of conditions, 1000 full battery discharges were simulated using stochastic speed profiles and randomly generated ambient temperatures. To simplify the analysis, only the $D_{TE}$ error at the beginning of the discharge (“key-on”) was used for comparison. Figure 9a shows that the distribution of error associated with the blend algorithm has a larger mean and deviation. A t-test was performed to ensure that the difference is significant.

The algorithms were also compared by measuring the reduction in $D_{TE}$ error at the beginning of discharge, which is defined as:

$$\text{Reduction in } D_{TE} \text{ Error (key-on)} = \frac{|e_{D_{TE},f}(t_0) - e_{D_{TE},b}(t_0)|}{e_{D_{TE},b}(t_0)} \times 100$$

Where $e_{D_{TE},f}(t_0)$ and $e_{D_{TE},b}(t_0)$ are the $D_{TE}$ error at key-on for the regression and blended algorithm, respectively. This measures the percentage reduction that the regression algorithm has over the blended algorithm. For example, a -100% reduction in error corresponds to a situation where the regression-based algorithm eliminated essentially all of the error. The error reduction was calculated for the 1000 full battery discharges and a frequency distribution of the results is shown in Figure 9b. The regression-based algorithm performed better than the blend algorithm for cases when the error reduction is in between -100% and zero, and worse when the reduction is greater than zero. Overall it can be seen that the regression-based algorithm reduced the error, and thus improved the $D_{TE}$ estimate, ~90% of the time.

### RESULTS

In practice, the values of $T_{a,f}(t_1)$, $T_{idle,f}(t_1)$, and $\bar{v}(t_1)$ would be estimated using Google Maps, weather information, etc. However, the simulations shown in this paper have perfect knowledge of the future.
estimates. This paper showed that an average energy use (Wh/km) from the past is often similar to about the future is known beforehand. Simulations showed that sufficiently similar to the past or (ii) applicable information future energy consumption of a vehicle. Future energy can be sustained periods of time can cause large errors in driving conditions (e.g. traffic or auxiliary energy use) for average of past energy use. However, significant changes in (temperature and traffic). In practice the driver would specify down based on estimated changes in driving conditions those future changes are detected beforehand and incorporate (i.e. city driving when cold). Finally, it would be useful to log data from an EV fleet and run similar simulations using real data.

SUMMARY AND CONCLUSIONS

It was shown that a $D_{TE}$ algorithm must predict the future energy consumption of a vehicle. Future energy can be predicted reliably if either (i) future energy consumption is sufficiently similar to the past or (ii) applicable information about the future is known beforehand. Simulations showed that an average energy use (Wh/km) from the past is often similar to future energy use. A conventional $D_{TE}$ algorithm assumes this condition by “blending” both a long-term and short-term average of past energy use. However, significant changes in driving conditions (e.g. traffic or auxiliary energy use) for sustained periods of time can cause large errors in $D_{TE}$ estimates. This paper showed that $D_{TE}$ error can be reduced if those future changes are detected beforehand and incorporate into the algorithm. Instead of using a complex parametric physics-based model, a multivariate linear regression-based model was derived that adjusts the historical value ($P_{long}$) up or down based on estimated changes in driving conditions (temperature and traffic). In practice the driver would specify the route and a service such as Google Maps would provide an estimate of the future driving conditions.

There are two additional advantages to the proposed regression-based algorithm: first is that it is computational light and thus can be run in real-time on a variety of processor speeds. The second is that it does not require vehicle specific calibration and validation. In other words, the algorithm can be “learned” over time by simply capturing data that is already available on most EVs CAN-bus. These unique advantages make it conducive to mobile phone and cloud-based computing services currently being developed [21].

FUTURE WORK

It would be useful to add noise to the estimates of the future driving conditions (e.g. temperature, idle time, average speed) to simulate more realistic and less predictable conditions. It is also important to better understand the accuracy and utility of real traffic and route data (e.g. provided by Google Maps). There are other regression methods that could be used to adjust $P_{long}$ based on estimates of future conditions. For example, historical data could be used to track energy use of repeated routes and/or conditions. Trip types could be categorized and $P_{long}$ could be adjusted accordingly (e.g. city driving when cold). Finally, it would be useful to log data from an EV fleet and run similar simulations using real data.

ACKNOWLEDGMENTS

This work would not have been possible without the support of the Singapore University of Technology and Design (SUTD) and the MIT-SUTD International Design Center (IDC). The author would also like to thank Sanjay Sarma, Warren Seering, Radu Gogoana, Stephen Zoepf, Don MacKenzie, Shane Colton, Mark Jeunnette, Sriram Krishnan and John Heywood for their technical feedback.

REFERENCES


