Transportation Research Part D 64 (2018) 36–52

Contents lists available at ScienceDirect

Transportation Research Part D

journal homepage: www.elsevier.com/locate/trd

Data-driven decomposition analysis and estimation of link-level electric vehicle energy consumption under real-world traffic conditions

Xuewei Qi[a,b,*], Guoyuan Wu[b], Kanok Boriboonsomsin[b], Matthew J. Barth[a,b]

[a] Department of Electrical and Computer Engineering, University of California, Riverside, USA
[b] College of Engineering-Centre for Environmental Research and Technology (CE-CERT), University of California, Riverside, USA

ARTICLE INFO

Keywords:
Electric vehicles
Energy consumption estimation
Decomposition analysis
Data-driven models
Feature selection

ABSTRACT

Electric vehicles (EVs) have great potential to reduce transportation-related fossil fuel consumption as well as pollutant and greenhouse gas (GHG) emissions, due to their use of renewable electricity as the sole energy source. Therefore, the wide-spread deployment of EVs is regarded as an attractive means to mitigate the environmental problems (e.g., air pollution and climate change) resulting from transportation activities. Government agencies are trying to promote EV deployment by allocating considerable funding as well as promulgating supportive policies. However, the mass adoption of EVs is still impeded by the limited charging infrastructure and all-electric-range (AER). All these lead to a critical research topic: the EV energy consumption analysis and estimation under real-world traffic conditions, which is fundamental to various types of EV-centred applications that aim at improving the EV energy efficiency and extending the AER. For example, eco-routing systems for EVs rely on accurate link-level energy consumption estimation to calculate the EV energy consumption costs of the different route options. In this work, to obtain an accurate link-level energy consumption estimation model for EVs, the energy consumption under real-world traffic congestion is decomposed based on two proposed impact factors: positive kinetic energy (PKE) and negative kinetic energy (NKE). Upon this decomposition, a data-driven model is built to estimate EV energy consumption on each roadway link considering real-world traffic conditions. Finally, the model performance is evaluated by comparing with the performance of baseline model adapted from existing models. The results show that the proposed EV link-level energy consumption estimation model outperforms the existing models in terms of accuracy, implying that it is quite promising in various on-board EV applications.

1. Introduction

An efficient transportation system is critical to the economic growth of a country, especially for developing countries. However, the significant amount of transportation-related fossil fuel consumption and greenhouse gas (GHG) and other criteria pollutant emissions has garnered a great deal of public concern. In 2014, the total energy consumed by the transportation sector in the United States was as high as 23.70 Quadrillion BTUs which is approximately a 28% share of the total energy used in the country (BTS, 2015).

* Corresponding author at: College of Engineering-Centre for Environmental Research and Technology (CE-CERT), University of California, Riverside, 1084 Columbia ave. Riverside, 92507, USA.
E-mail address: xqi001@ucr.edu (X. Qi).

http://dx.doi.org/10.1016/j.trd.2017.08.008
Received 31 October 2016; Received in revised form 27 April 2017; Accepted 10 August 2017
Available online 18 August 2017

The U.S. Environmental Protection Agency (EPA) reported that nearly 33.4% GHG emissions resulted from fossil fuel combustion for transportation activities in 2014 (EPA, 2016) . Hence, the resulted air pollution and climate change impacts have motivated many researchers to propose different ways to reduce transportation-related fuel consumption and GHG emissions. One promising pathway is to develop and use cleaner alternative energy sources to replace fossil fuels, such as electricity from renewable resources (e.g., solar, wind) and hydrogen. With these alternative fuels, many new powertrain systems are being developed, such as those found in electric and fuel cell vehicles. In recent years, transportation electrification has been a very active research area with significant progress, including the development of very efficient hybrid electric vehicles (HEVs) and battery electric vehicles (BEVs or EVs). HEVs are able to achieve higher fuel efficiency than internal combustion engine (ICE) vehicles by taking advantage of the electric energy source. BEVs have the potential to eliminate the use of fossil fuels by using only renewable electricity. Gradually replacing conventional ICE-powered vehicles with electric vehicles (EVs) is a promising way to reduce the dependence of fossil fuel and pollutant and GHG emissions. From a transportation systems point-of-view, it is also known that a large amount of energy consumption is due to the inefficient movement of traffic. It is reported that nearly 7 billion hours of delay and more than 3 billion gallons of fuel were wasted in 2014 due to traffic congestion in U.S. (Schrank et al., 2015). Toward this end, various Intelligent Transportation System (ITS) technology has been applied to induce more energy efficient driving. An example is to reduce unnecessary stop-and-go maneuvers at signalized intersections by using connected and automated vehicle (CAV) technologies.

Among all these ITS applications, dynamic eco-driving technology has attracted considerable interest as a way to improve vehicle energy efficiency not only for conventional ICE-powered vehicles, but also for EVs. Recent research shows that these dynamic eco-driving technologies are capable of improving the energy (i.e., electricity) efficiency of EVs so that the AER can be extended. As an example, an Eco-Approach and Departure (EAD) system at traffic signals (Barth et al., 2011) has been developed to help vehicles travel through the signalized intersection smoothly and avoid unnecessary idling and acceleration/deceleration with the knowledge of signal phase and timing (SPaT) information.

For all of these ITS technology solutions, accurate estimation of EV energy consumption using real-world driving data (e.g., data that can be provided by ITS traffic management systems) is a critical prerequisite. For example, in an eco-routing system for EVs, the EV energy consumption on each road segment or link has to be estimated prior to selecting the most energy-efficient route. Thus far, several researchers have investigated different factors that impact EV energy consumption and have built different estimation models upon these impact factors (see Figs. 1 and 2). These existing EV energy consumption models can be classified by the following features:

a. *Granularity:* For different application purposes, EV energy consumption estimation models can be developed at different levels of granularity (i.e., macroscopic, mesoscopic and microscopic). For instance, some applications require detailed energy-efficient velocity profile planning (such as the EAD application reference above), therefore requiring *microscopic* models to estimate instantaneous energy consumption (e.g., on a second-by-second basis) (Zhang and Yao, 2015; Alvarez et al., 2014; Yao and Yang, 2014; Yao et al., 2013). On the other hand, applications such as eco-routing systems, *mesoscopic* models may be developed to estimate the total energy consumption over roadway network links. These mesoscopic models can be aggregated from the previously mentioned microscopic models and also aggregated further to a macroscopic model that estimates trip level energy consumption (Wu et al., 2015). For some sophisticated applications, both types of model are required to form a hybrid model that is capable of estimating energy consumption at different granularities simultaneously (Cauwer et al., 2015; Chang and Hong, 2014).
b. *External Impact Factors:* Various external factors (in contrast to internal factors associated with the vehicle itself, e.g., different powertrain and internal efficiency parameters) have been shown to be influential to EV energy consumption (Fig. 1). These external impact factors can be classified into the following categories: 1) factors related to traffic conditions that indirectly influence vehicle dynamic parameters such as speed and acceleration (Yao et al., 2013; Shankar and Marco, 2013); 2)

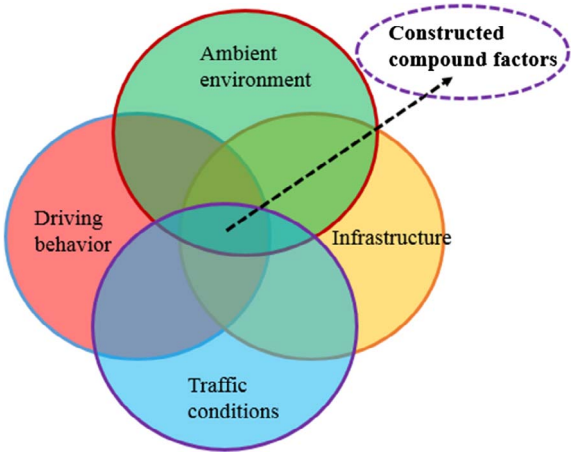


Fig. 1. Impact factors of EV energy consumption.

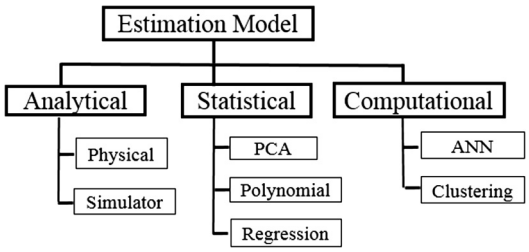


Fig. 2. Categories of EV energy consumption estimation models.

infrastructure-related factors such as road grade, road surface roughness (Yi and Bauer, 2015); 3) ambient environment factors such as ambient temperature and wind speed (Yi and Bauer, 2015); and 4) driving behavior factors, such as driver aggressiveness, brake/pedal position/pressure, driving mode selection (Alvarez et al., 2014) etc. Thus far, all the above mentioned factors are regarded as raw factors that can be obtained by (instantaneous) measurement. In most cases, it is quite difficult to measure and obtain all these types of factors due to limited resources. Therefore, there should be a new type of factors that can be constructed to aggregate the compound influences of several different types of raw factors. As shown in Fig. 1, these constructed compound factors are identified at the intersections of all types of raw impact factors, which are defined and described in this paper.

c. *Type of Model:* An EV energy consumption model explains the quantitative relationship between EV energy consumption rate (or factor) and various impact factors. Both analytical and data-driven models can be used to capture this relationship.

Among all the above reviewed existing work, there are only a few focused on the link-level EV energy consumption estimation. A vehicle specific power (VSP) based link-level EV energy consumption estimation model was built and evaluated for different road types by Yao et al., 2013, considering the average speed and accelerations. Grubwinkler S., 2014 proposed another flexible link-level energy consumption estimation model with different degrees of accuracy and various requirements in respect of complexity, computational time and required input features. A single energy consumption estimation model was built based on the crowd-sourced energy consumption data from other vehicles on the road network, which might be not accurate due to the different vehicle models and energy efficiency. More than 18 link-based features are extracted for estimating the energy consumption on each link, such as the average speed, mean and standard deviation of accelerations. The major drawback of these existing efforts are the lack of analyzing the impact of regenerative braking power of EVs, a key feature differentiating from conventional ICE-powered vehicles, when operating in real-world traffic conditions.

In this paper, the EV energy consumption on a road link under real-world traffic conditions is first decomposed based on physical fundamentals. Then, with the real-world EV driving and energy consumption data, a data-driven energy consumption decomposition analysis is conducted by analyzing two compounding factors for estimating EV energy consumption: positive kinetic energy (PKE), and negative kinetic energy (NKE). The NKE is analyzed and related to the regenerative braking characteristics. The energy consumption rate is plotted against the link average speed and a “W” shape curve is identified and explained with real-world data. An accurate and simple EV energy consumption estimation model is built upon this decomposition and the subsequent feature selection analysis proves the effectiveness of these two compounding factors. The comparison with the existing models are also conducted to evaluate the performance of the proposed model.

The major contributions of this paper are as follows:

- NKE is used as a variable for capturing the regenerative braking effect in a link-level EV energy consumption estimation model;
- A systematic data-driven EV energy consumption decomposition analysis is conducted;
- A novel link-level EV energy consumption estimation model is built upon the decomposition analysis; and
- A “W”-shaped relationship between link-level EV energy consumption rate and link average speed is discovered and explained with the real-world EV driving data.

The remainder of the paper is organized as follows: In Section 2, an analytical model for EV energy consumption is deduced to provide a theoretical foundation for the data-driven analysis; Section 3 provides details about the real-world EV data collection; and a feature selection analysis is presented in Section 4. In Section 5, link-level EV energy consumption is decomposed with the proposed two compound indicators. Furthermore, the estimation model built upon these proposed effective indicators are fitted and compared with the existing models in Section 6. Conclusions and future work are stated in Section 7.

2. Analytical model of EV energy consumption

An analytical understanding of vehicle energy consumption is important for building a vehicle energy consumption estimation model. Fig. 3 shows an example of a vehicle moving from point A to point B on a road segment with length L_{link} and road grade Θ. The vertical displacement is H_{link}. The total energy consumption (from a power source) of the vehicle (with mass m) is calculated as:

$$E_{total} = E_{tractive} + E_{A/C} + E_{accessory} \quad (1)$$

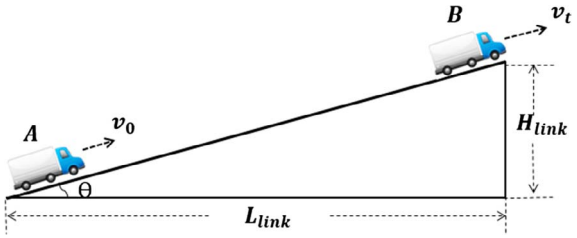


Fig. 3. Change of mechanical energy of vehicle movement.

$$E_{tractive} = \Delta E_{kinetic} + \Delta E_{potential} + E_{rolling} + E_{aerodynamic} + E_{loss} \tag{2}$$

and,

$$\Delta E_{kinetic} = \frac{1}{2}mv_t^2 - \frac{1}{2}mv_0^2 \tag{3}$$

$$\Delta E_{potential} = mgH_{link} = mgL_{link}tan(\Theta) \tag{4}$$

where $E_{tractive}$ is the energy consumed for traction; $E_{A/C}$ and $E_{accessory}$ represent the energy consumed by air conditioner and other accessories, respectively; $\Delta E_{kinetic}$ and $\Delta E_{potential}$ represent the changes in vehicle kinetic energy and potential energy between point A and point B, respectively; $E_{rolling}$ represents the energy consumed to overcome the friction on road surface; $E_{aerodynamic}$ denotes the energy consumed to overcome the air friction; and E_{loss} is the internal energy loss due to e.g., friction in the transmission system or heat loss from the motor and mechanical brake during the travel from point A to point B;

In this study, the tractive energy is of the major interest, while the energy consumption from air conditioner and other accessories are measured and excluded from the total energy consumption as shown in Eq. (1). It is also noted that for a specific road segment (or link), $\Delta E_{potential}$ and $E_{rolling}$ are independent of speed trajectories (under traffic conditions). For a passenger vehicle (i.e., NISSAN LEAF in this study), the energy consumed to overcome the air resistance, $E_{aerodynamic}$ is not so significant (around 5% (Holmberg et al., 2012)), especially when the vehicle is affected by the other traffic and the speed is not so high. It is very challenging to accurate model the internal energy loss, E_{loss}. For simplicity, we assume that it is proportional to the tractive energy. Furthermore, the sum of $E_{aerodynamic}$ and E_{loss} is written as

$$E_{aerodynamic} + E_{loss} \approx \mu \cdot E_{tractive} \tag{5}$$

where μ is a constant, independent of the driving cycle.

Along a specific road segment, therefore, the tractive energy can be approximated as

$$E_{tractive} \approx \Delta E_{kinetic} + \delta + \mu \cdot E_{tractive} \tag{6}$$

or,

$$E_{tractive}(v) \approx \alpha + \beta \cdot \Delta E_{kinetic}(v) \tag{7}$$

where $\delta \approx \Delta E_{potential} + E_{rolling}$, $\alpha = \delta/(1-\mu)$, and $\beta = 1/(1-\mu)$.

From a statistical point of view, if we are trying to build a regression model to estimate the vehicle energy consumpiton, then the change in kinetic energy (due to the speed variation) should be the most powerful predictor. This is the fundamental hypothesis of the data-driven analysis presented in the following section.

If the road segment is discretised by time step, then Eq. (3) can be expressed as:

$$\Delta E_{kinetic} = \frac{1}{2}m \cdot \sum_{i=1}^{N-1} (v_{i+1}^2 - v_i^2) \tag{8}$$

where N is the total number of time steps driving on the road segment; and v_i is instantaneous speed (mph). When we consider the positive portion and negative portion separately, then (9) is converted to

$$\Delta E_{kinetic} = \frac{1}{2}m \cdot \sum_{i=1}^{N-1} \max(v_{i+1}^2 - v_i^2, 0) + \frac{1}{2}m \cdot \sum_{i=1}^{N-1} \min(v_{i+1}^2 - v_i^2, 0) \tag{9}$$

Since we are estimating the energy consumption per unit distance, we divide Eq. (7) by the distance of the road segment,

$$\frac{E_{tractive}}{L_{link}} \approx \frac{\alpha}{L_{link}} + \beta \cdot \left[\frac{\frac{1}{2}m \cdot \sum_{i=1}^{N-1} \max(v_{i+1}^2 - v_i^2, 0)}{L_{link}} + \frac{\frac{1}{2}m \cdot \sum_{i=1}^{N-1} \min(v_{i+1}^2 - v_i^2, 0)}{L_{link}} \right]$$
$$\approx \frac{\alpha}{L_{link}} + \beta \cdot \left[\frac{\frac{1}{2}m \cdot \sum_{i=1}^{N-1} \max(v_{i+1}^2 - v_i^2, 0)}{\sum_{i=1}^{N-1}(d_{i+1} - d_i)} + \frac{\frac{1}{2}m \cdot \sum_{i=1}^{N-1} \min(v_{i+1}^2 - v_i^2, 0)}{\sum_{i=1}^{N-1}(d_{i+1} - d_i)} \right] \quad (10)$$

where d_i is cumulative travel distance up to the i-th time step; and $\frac{E_{tractive}}{L_{link}}$ is the energy consumption per unit distance on the road segment, which we call energy consumption rate (ECR). The term, $\frac{\sum_{i=1}^{N-1} \max(v_{i+1}^2 - v_i^2, 0)}{\sum_{i=1}^{N-1}(d_{i+1} - d_i)}$, is the cumulative positive change in kinetic energy rate (PKE), which was first developed by Watson (Watson et al., 1982) and used as a measure of change in kinetic energy per unit distance due to acceleration. The term, $\frac{\sum_{i=1}^{N-1} \min(v_{i+1}^2 - v_i^2, 0)}{\sum_{i=1}^{N-1}(d_{i+1} - d_i)}$, is newly defined in this study as the cumulative negative change in kinetic energy rate (NKE), which may account for the regenerative braking effect of an EV. Therefore, Eq. (11) can be simplified as:

$$\text{ECR} \approx \frac{\alpha}{L_{link}} + \frac{\beta}{2} \cdot m \cdot (\text{PKE} + \text{NKE}) \quad (11)$$

When Θ is small, α / L_{link} can be approximated as a constant. So, we can say that the energy consumption rate is (approximately) linearly correlated with PKE and NKE. In other words, PKE and NKE are strong predictors of the link-level EV energy consumpiton rate.

3. Real-world EV data collection

According to the analytical model discussion, PKE and NKE are very good predictors for estimating energy consumption rate of EVs. To further identify the numerical evidence, real-world EV driving data are collected for data-driven energy consumption analysis in the following sections.

3.1. Data acquisition system and test vehicle

A 2013 NISSAN LEAF was used as the test EV for data collection in this study (Fig. 4.). To obtain second-by-second vehicle states (e.g., speed), energy consumption, and road topology (e.g., road grade) data, the following two data acquisition systems were used simultaneously:

A diagnostic tool was used to access data from the test vehicle's CAN bus, including vehicle speed, battery current (positive for charging while negative for discharging) and voltage, air conditioner (A/C) power, and accessory power. Therefore, the net propulsion power, P^{prop}, can be estimated as:

$$P^{prop} = -(I^{bp} \times V^{bp}) - (P^{AC} + P^{acc})(1) \quad (12)$$

where I^{bp} and V^{bp} represent the instantaneous current (in ampere) and voltage (in volt), respectively, from/to the battery pack. P^{AC} and P^{acc} are the power consumed by A/C and other accessories (e.g., radio), respectively.

In addition, it is well known that road grade is one of the major factors affecting a vehicle's energy consumption. We used a GPS data logger to collect latitude and longitude data, which are then map matched with a high-fidelity 3 D map to acquire the road grade of each data point.

3.2. Collection of real-world driving data

Using the data acquisition systems described above, two types of experiments were conducted to collect EV driving and energy

a. Test EV b. data acquisition system with OBD reader c. GPS data logger

Fig. 4. Test Vehicle and Data Acquisition Equipment.

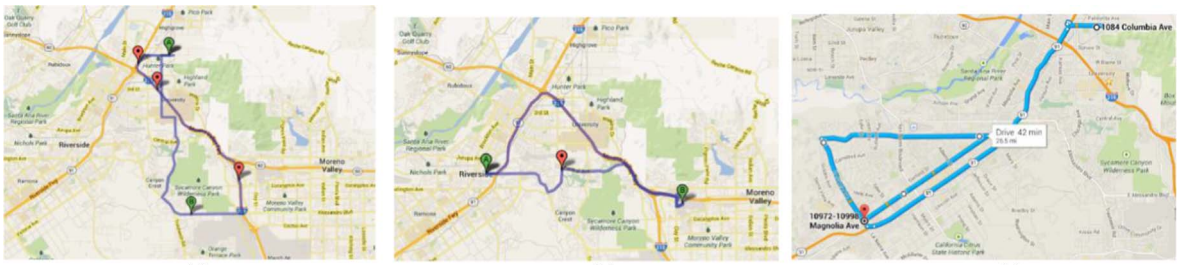


Fig. 5. Routes for real-world driving data collection.

consumption data in the field:

1) *Driving under real-world conditions*

To cover a variety of traffic conditions, road types and road grades, we carefully chose three pairs of origin-destination (shown in Fig. 5) and routes in Southern California to conduct real-world driving data collection for about four months. The vehicle was driven by different drivers at different traffic conditions during the field driving test. In total, more than 100 h of driving under real-world conditions were conducted, yielding a substantial amount of data.

2) *Driving under controlled environment*

In addition to the real-world driving experiments, we also collected data under a controlled environment (i.e., no interaction with other traffic) to better understand the energy consumption of the test EV at different cruise speeds. The cruise speeds vary from 5 mph to 50 mph with a 5-mph increment. These data will be used to build the baseline scenarios case for comparison.

3.3. Data fusion

Before the field data from the vehicle data acquisition system and the GPS data logger can be used for analysis, they have to be time-synchronized (see Fig. 6). The synchronized data then have to be associated with road grade values of road segments through map matching. More specifically, there are two steps in the data fusion process:

1) *Frequency alignment*

The raw data files from the GPS data logger are not aligned with the ones from the vehicle data acquisition system in terms of update rate. Hence, all raw data files were reprocessed into 1 Hz, which is suitable for the data synchronization as well as the energy consumption estimation to be done later.

3) *Trip data synchronization and fusion*

It is noted that the GPS data logger reports the Coordinated Universal Time (UTC) as a temporal reference. However, the vehicle data acquisition system only reports the relative time stamp (i.e., each run always starts from time “0”) for each run. To fuse these two data sources, a common feature needs to be identified. In this study, we selected the vehicle speed and applied the cross-correlation technique (Billings, 2013) to synchronize these two data sources. Fig. 7 presents an example of speed trajectories after synchronization.

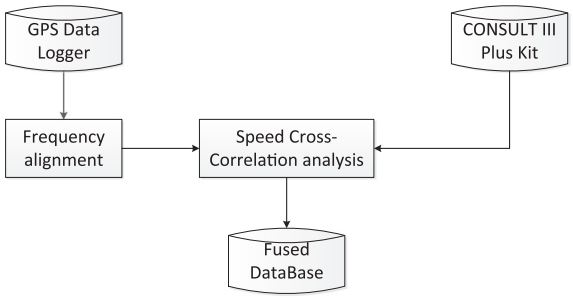


Fig. 6. Flowchart for fusing data from the vehicle data acquisition system and the GPS data logger.

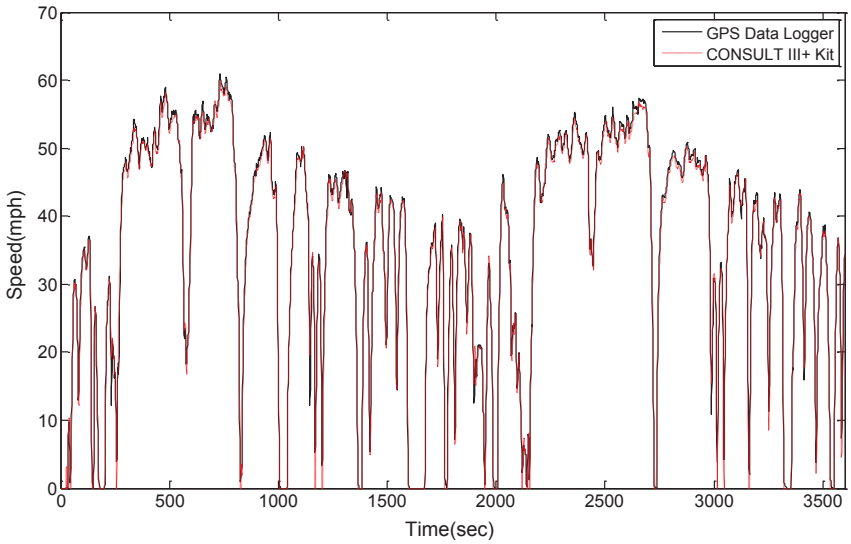


Fig. 7. An example of synchronized speed trajectories between GPS data logger and the vehicle data acquisition system.

3.4. Map matching

The map matching was conducted in this study by following the steps below:

- Conduct map matching onto the fused dataset, where each data point is matched to the associated roadway link on the digital map that has the least orthogonal distance to the data point (see Fig. 8);
- Break down the fused data stream into short driving snippets based on roadway link;
- Group link-based snippets by roadway type, average speed and grade etc; and
- Calculate the energy consumption factors (per mile) for each real-world driving snippet.

4. Feature selection and analysis

4.1. Data visualization

Using the collected EV driving data, PKE and NKE values were calculated on a link-by-link basis. Fig. 9 and Fig. 10 illustrate the plots of link-level ECR as a function of PKE and NKE, respectively. It can be seen that there is a clear linear trend between ECR versus PKE or NKE. This implies that PKE and NKE are good indicators of EV energy consumption rate and can be used in EV energy consumption models.

4.2. Regenerative braking and NKE

As aforementioned, the change in kinetic energy of the vehicle on a road segment is decomposed into PKE and NKE. As shown in Fig. 11, NKE is used to measure the negative change in kinetic energy per unit distance due to deceleration, which can be correlated with the electricity energy gain from EV regenerative braking (RB) on a flat road. This is because the vehicle kinetic energy is

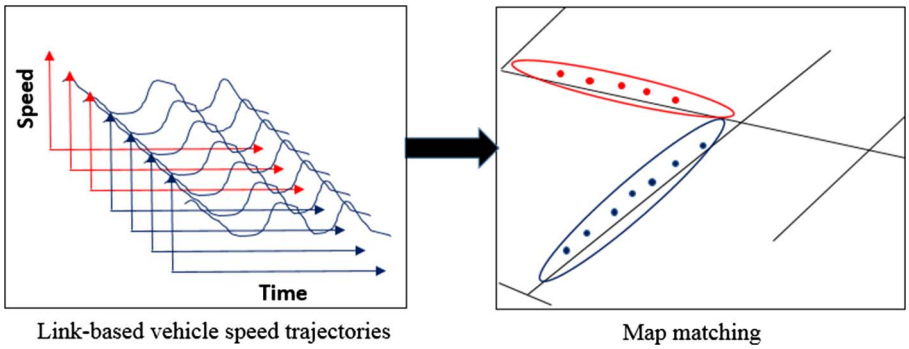


Fig. 8. Trajectory snippets and map matching.

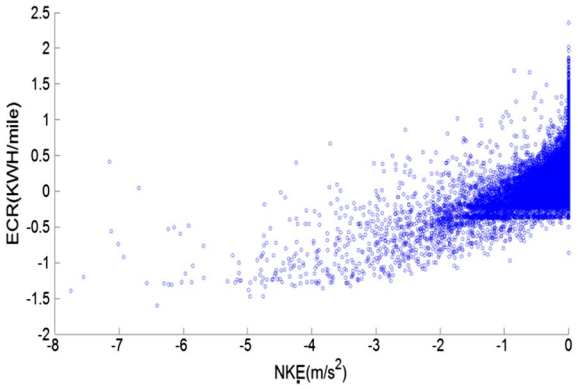


Fig. 9. ECR vs. NKE.

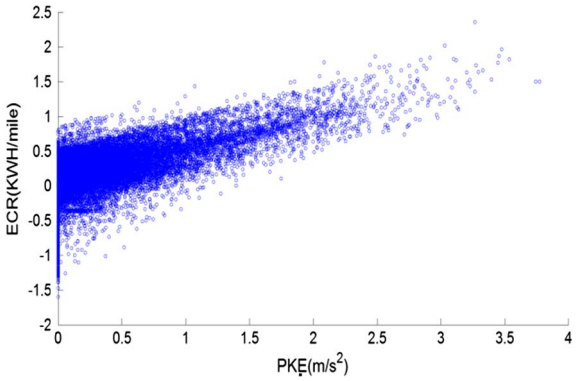


Fig. 10. ECR vs. PKE.

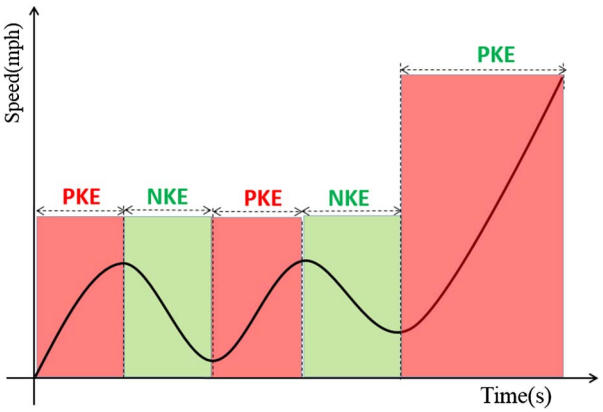


Fig. 11. PKE and NKE along a trip.

converted to electricity through the motor acting as a generator when the vehicle decelerates. Fig. 12 shows the electricity flow of EV regenerative braking system.

The NKE and energy gain from RB over different average speeds are plotted in Fig. 13 and Fig. 14, respectively, using the collected field data. As can be seen in these figures, both NKE and energy gain from RB are relatively constant when the average speed is over 30 mph but a surge is observed in both figures under the average speed of 30 mph. These observations indicate that NKE is a good indicator of energy gain from RB.

4.3. Principal components analysis

To analyse the importance of different predicting variables for estimating EV link-level energy consumption, and to provide further convincing data-driven evidence for our previous discussions, principal component analysis (PCA) (Armstrong, 1985) is

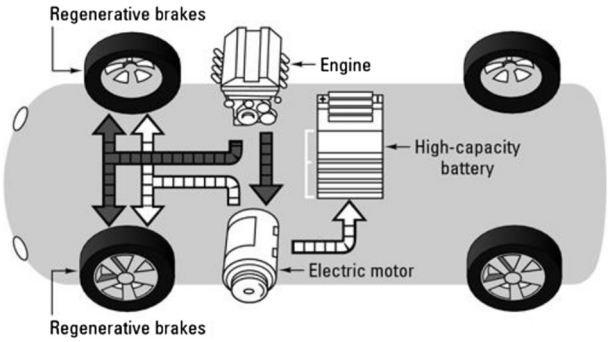


Fig. 12. Regenerative braking (RB) in EV.

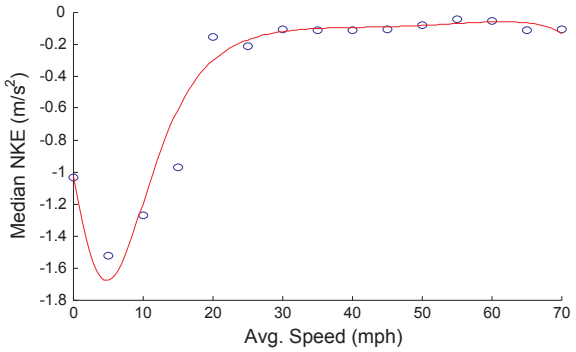


Fig. 13. Polynomial fit of median NKE by average speed.

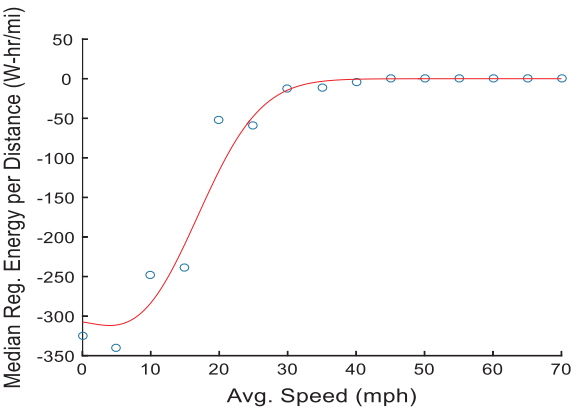


Fig. 14. Polynomial fit of collected regenerative power by average speed.

utilized to numerically study the significance of each variable. PCA is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components (PCs). Through such transformation, the first principal component normally has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance out of the rest of components. The resulting vectors are an uncorrelated orthogonal basis set. PCA is sensitive to the relative scaling of the original variables; hence, the original values for all variables have been normalized into the same scale.

Suppose $\boldsymbol{X}^T = (X_1, X_2 \ldots, X_p)$ is an random vector that represents all the predicting variables. p is the number of predicting variables. $\sum = \mathrm{Var}(\mathbf{X})$ is the $p \times p$ variance-covariance matrix. (λ_i, e_i) for $i = 1, \ldots p$ are the eigenvalue and the associated eigenvector, and $\lambda_1 \geqslant \ldots \geqslant \lambda_p \geqslant 0$. For p variables, we can construct p PCs, $Y_1 \ldots Y_p$, each of which is a linear combination of the original set of variables:

$$Y_1 = \boldsymbol{\alpha}_1^T \boldsymbol{X} = a_{11}X_1 + \ldots + a_{1p}X_p \tag{13}$$

$$Y_2 = \boldsymbol{\alpha}_2^T \boldsymbol{X} = a_{21}X_1 + \ldots + a_{2p}X_p \tag{14}$$

...

Table 1
coefficients of principle components.

Variables	PCA coefficients						
	PC1	*PC2*	*PC3*	*PC4*	*PC5*	*PC6*	*PC7*
Road type	−0.08936	0.180285	0.120264	0.376948	**0.89605**	0.007005	−0.00216
Road grade	−0.00903	0.001795	−0.07008	**0.785232**	−0.13896	7.30E−05	0.000419
Avg. speed	−0.13645	0.42776	**0.99019**	0.011401	−0.4206	0.00016	−0.00017
PKE	0.117852	**0.885204**	0.007689	−0.44944	0.021687	−0.00024	−0.00076
NKE	**0.979502**	−0.03045	0.00941	0.197959	0.019273	−0.00056	0.000278
Acum Acc.*	−0.00088	0.001518	0.000231	0.001664	0.004372	−0.42387	**0.905709**
Acum.Dec.*	0.000946	−0.00054	−0.00088	−0.00227	−0.00478	**0.905695**	0.423894

Note
* See Reference (Hemmerle and Hermanns, 2014).

$$Y_p = \boldsymbol{\alpha}_p^T \boldsymbol{X} = a_{p1}X_1 + \ldots + a_{pp}X_p \tag{15}$$

then

$$\mathrm{Var}(Y_i) = \boldsymbol{\alpha}_i^T \sum \boldsymbol{\alpha_i} \tag{16}$$

$$\mathrm{Cov}(Y_i, Y_k) = \boldsymbol{\alpha}_i^T \sum \boldsymbol{\alpha_k} \tag{17}$$

All these constructed PCs are orthogonal to each other such that $\|\boldsymbol{\alpha}_i\|^2 = \boldsymbol{\alpha}_i^T \boldsymbol{\alpha}_i = 1$. And a larger weight indicates a greater importance of that variable in a PC. Each PC is obtained by:

$$Y_i = \boldsymbol{\alpha}_i^T \boldsymbol{X} = \text{such that } \mathrm{Var}(Y_i) = \boldsymbol{\alpha}_i^T \boldsymbol{\alpha}_i = max_{\|\alpha\|} = 1\boldsymbol{\alpha}^T \sum \boldsymbol{\alpha}, \text{with } \boldsymbol{\alpha}_i^T \sum \boldsymbol{\alpha_j} = 0 \textbf{ for } j = 1,\ldots,\text{i}-1$$

Actually, $\boldsymbol{\alpha}_i = \boldsymbol{e}_i$, *for* $i = 1, \ldots, p$ and $Y_i = \boldsymbol{e}_i^T \boldsymbol{X}$ and $\mathrm{Var}(Y_i) = \lambda_i$.

In this work, we take 7 predicting variables (shown in Table 1) as the original variables and conduct the principal components analysis. The following table gives coefficients of all the PCs. It can be observed that NKE and PKE are the first two PCs with the largest variance.

The trace of a symmetric matrix is the sum of its eigenvalues $\lambda_1 \geqslant \ldots \geqslant \lambda_p$. Thus the total variance is

$$\mathrm{tr}(\sum) = \lambda_1 + \ldots + \lambda_p \tag{18}$$

Since covariance matrix $\sum$ is positive-semidefinite, the total variance is non-negative and the first k PCs make up

$$\left(\frac{\lambda_1 + \ldots + \lambda_k}{\lambda_1 + \ldots + \lambda_p} * 100\%\right) \tag{19}$$

of the total variance (in particular, they make up 100% of the total variance when $k = p$). As we can see in Fig. 15, the first two PCs make up over 80% of total variance. This indicates that NKE and PKE are the most significant predicting variables for estimating link-based EV energy consumption.

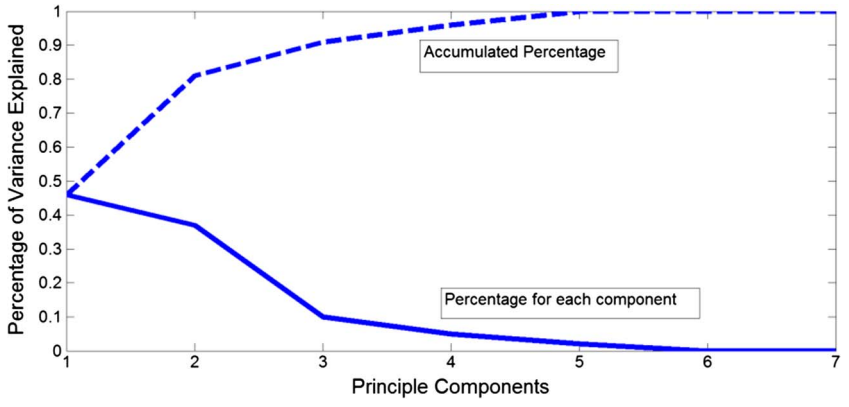


Fig. 15. Percentage of variance explained by different impact factors.

4.4. Decision tree analysis

Decision tree builds regression or classification on models in the form of a tree structure (Quinlan, 1992, 1993). It breaks down a dataset into smaller and smaller subsets while an associated decision tree is incrementally developed. The final result is a tree with *leaf nodes* and *branches*. In these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. The topmost decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data. Decision trees where the target variable can take continuous values (typically real numbers) are called regression trees. In this work, a regression tree model is used to analyse the energy consumption predicting variables.

There are two important process in building a regression tree. One is the optimal partition of each variable in the continuous range, and the other one is to choose the best variable for splitting the data set at each *leaf node.* In this study, the entropy based information gain is adopted as the criteria for optimally splitting the dataset on different attributes (i.e., predicting variables). And the Gini index (Breiman et al., 1984) is used for choosing the best variable at each *leaf node* to obtain the largest standard deviation reduction. In regression tree, the continuous variables are first discretised into multiple intervals, and then an optimal partition is selected based on information gain criteria. Given a set of samples S, if S is partitioned into two intervals $S1$ and $S2$ using boundary T, Entropy is calculated based on class distribution of the samples in the set. Given m classes, the entropy before the partition is

$$Entropy(S) = -\sum_{i=1}^{m} p_i log_2(p_i) \tag{20}$$

the entropy after partitioning is:

$$I(S,T) = \frac{S_1}{S} Entropy(S_1) + \frac{S_2}{S} \text{Entropy}(S_2) \tag{21}$$

where p_i is the probability of class i in S. Therefore, the information gain is obtained by:

$$Gain(T) = Entropy(S) - I(S,T) \tag{22}$$

To select the best attribute for each *leaf node,* the Gini index is used in this study and calculated by

$$\text{Gini}(D) = 1 - \sum_{j=1}^{n} p_j^2 \tag{23}$$

where D is the data set which contains examples from n classes, p_j is the relative frequency of class j in D. If a data set D is split on A into two subsets D_1 and D_2, the Gini index is defined as

$$Gini_{After}(D) = \frac{D_1}{D} Gini(D_1) + \frac{D_2}{D} Gini(D_2) \tag{24}$$

The reduction in impurity is

$$\Delta Gini(D) = Gini(D) - Gini_{After}(D) \tag{25}$$

The variable providing the smallest reduction in impurity is chosen to split the node. Fig. 16 presents the first three layers of the constructed regression tree. The values on the branches are the optimal split on the predecessor node. It is noted that the most important 2 variables are NKE, PKE. The 3-D plot of ECR on PKE/NKE (Fig. 17a) and partition (Fig. 17b) based on the decision tree analysis is provided in Fig. 17. It can be seen that the ECR distribution is partitioned into four major regions. Region A (average 0.02) represents the driving situation where PKE and NKE are both not too high. This region corresponds to the low speed driving scenarios. Also, a larger variation is witnessed in this region, which can be explained by the frequent fluctuation in driving under low (average) speed traffic conditions. Region B (average 0.71) where PKE is high but NKE is low represents the set of driving scenarios when acceleration maneuvers dominate. In contrast to Region B, Region C (average −0.31) is mainly characterized by deceleration maneuvers. Region D (average 0.24) is just in the middle between region B and C. Please note that the boundary between D and C is

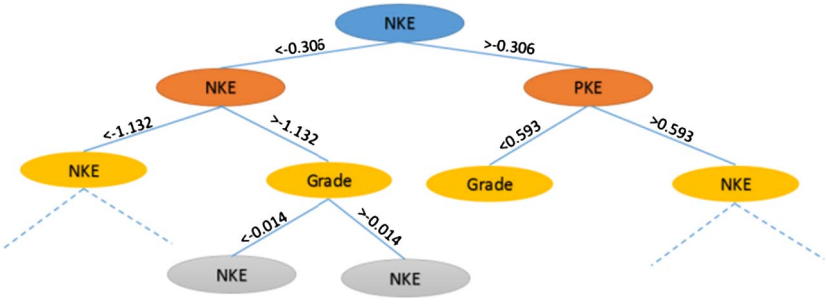
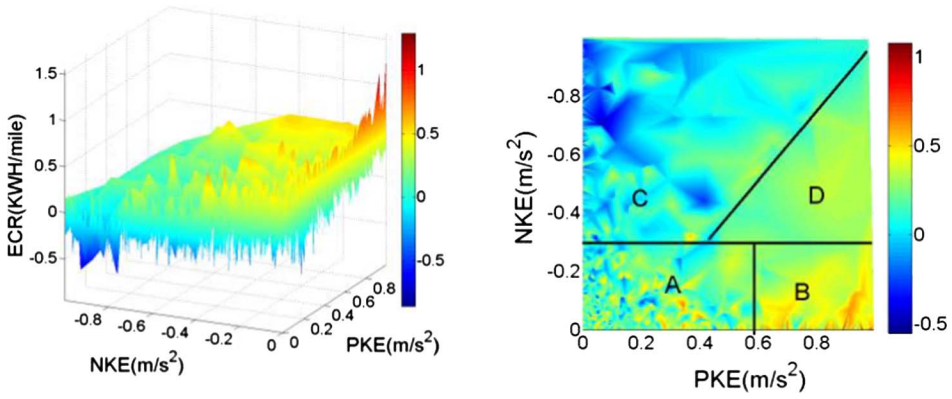


Fig. 16. Decision tree built from the real-world driving data (top 3 layers).



a. 3-D plot of ECR by PKE/NKE b. ECR partition based on decision tree

Fig. 17. ECR partition based on decision tree analysis.

because of the road grade (cut by -0.014 as shown in Fig. 22). And also this boundary does not have to be exactly a straight line but it reflects the basic difference between region C and D.

5. Energy consumption decomposition analysis

5.1. Impact of real-world congestion

Based on the fundamentals in vehicle dynamics, the instantaneous power of EV is determined by vehicle speed, acceleration and roadway grade. It has been derived by Wu et al., 2015 that EV's instantaneous power can be estimated by:

$$P = \frac{rR^2}{(k_a D_d)^2}(ma + kv^2 + f_{rl}mg + mgsin\Theta)^2 + v(kv^2 + f_{rl}mg + mgsin\Theta) + mav \tag{26}$$

where r is the resistance of the motor conductor; R is the radius of the tire; k_a is the armature constant of the motor; D_d is the magnetic flux; f_{rl} is the rolling resistance constant; a and v are the instantaneous acceleration and velocity; Θ is the road grade; m is the vehicle mass; g is the gravitational acceleration (9.81 m/s^2). From Eq. (16), we know that the EV instantaneous power is proportional to the 4th order of velocity.

To understand the impact of traffic congestion on EV energy consumption, the relationship between EV energy consumption per unit distance and average vehicle speed on a road link is investigated since average vehicle speed is a good indicator of the level of real-world traffic congestion. We first created boxplots of distance-based ECR vs. average speed (every 5 mph) for different levels of road grade. Fig. 18 presents the ECR at different speed levels. A 4th order polynomial fit is applied to the median values of ECR by different speed bins:

$$f_k = \sum_{i=0}^{4} \alpha_i \cdot v_k^i \tag{27}$$

where i is the order of polynomial; v_k is (link-level) average vehicle speed (mph); α_i's are regression coefficients; and f_k is (link level) distance-based energy consumption rate (W-hr/mi). The fitted curve is shown in Fig. 19. Unlike the typical "U-shaped" ECR curve for

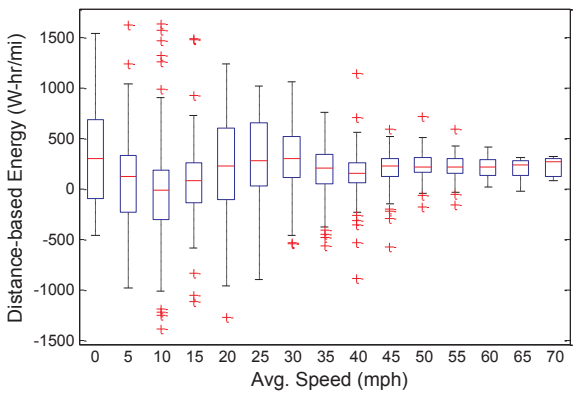


Fig. 18. Boxplot of ECR for different average speed.

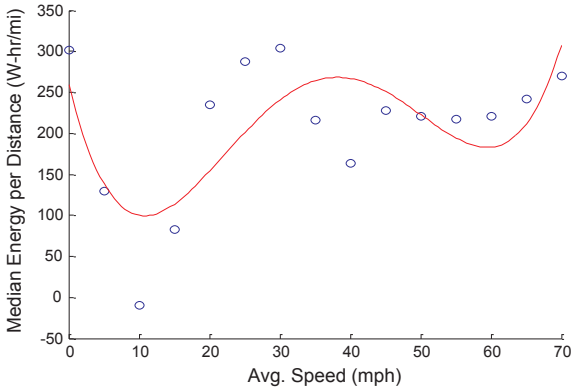


Fig. 19. Polynomial fit of median ECR by average speed.

gasoline light-duty vehicles found in our previous work (Barth and Boriboonsomsin, 2008) as shown in the Fig. 19, the ECR curve for the test EV displays a "W-shape". The explanation for this difference is given in the following section.

As described in the data collection section, to fully investigate the impact of traffic congestion on EV link based energy consumption, EV energy consumption data under different constant speeds were also collected in a controlled environment, where vehicles were running without interaction with other traffic and always kept in a constant or close constant speed. Based on these data, we further compared the EV energy consumption at different constant speeds with that at different average speeds (in traffic). Fig. 20 shows the fitted ECR curves by speed. As can be seen in the figure, ECR in real traffic with stop-and-go driving is lower than that of constant speed driving when the average speed is less than about 10 mph (region A in Fig. 20). This suggests that the test EV is very energy efficient in low-speed driving on urban routes considering the real-world vehicle activity patterns. Region B in Fig. 20 shows potential energy savings of the test EV from a smoother traffic flow that can be achieved by different ITS or intelligent vehicle strategies (e.g., speed harmonization). When the average is higher than 60 mph, there is little difference in the ECR due to the lack of traffic congestion in both driving scenarios at high average speeds on freeways. It is also noteworthy that there is an intersection between two curves for EV but no intersection for gasoline light-duty vehicles (See Fig. 21). This is mainly due to the regenerative braking that collects power when EVs are decelerating, but the energy is dissipating in the form of heat for gasoline vehicles.

5.2. Energy consumption decomposition

To explain the interesting "W-shaped" relationship between median ECR and average vehicle speed in Fig. 20, PKE, NKE, and PKE + NKE values by different average speeds are further investigated using the collected EV data (see Fig. 22). Please note that the y axis for PKE is positive and NKE is negative. As we can see from these three sub figures, the curve in Fig. 22c is the sum of curve in Fig. 22a and b. Therefore, the trend in Fig. 22c can be explained by combining the trend in Fig. 22a and b And the trend in Fig. 22c is quite similar to that in Fig. 19, which is also a "W" shape (Note: all these 3 figures are fitted separately from raw data, hence they are not likely to be exactly the same). This implies that PKE + NKE is a good indicator for EV energy consumption under real world congestion. In other words, ECR by average speed under real-world congestion can be decomposed by PKE and NKE. This is an evidence for the previously proposed hypothesis presented by Eq. (15) based on fundamental physics.

More specifically, as shown in Fig. 22b, there is drop in NKE when the average speed is lower than 10 mph, which coincide with the first drop in Fig. 19 around the average speed of 10 mph. This drop in NKE actually means the increase of collected energy from RB, which is due to the high frequency of stop-and-go in this low speed range. We can also see that PKE deceases after 30 mph. This can be explained by lower speed fluctuation over 25 mph (see Fig. 18). After 60mph, although the speed fluctuation is still at a low level, but the high speed cruise results in higher power demand and therefore the ECR is increased again as shown in Fig. 20. In

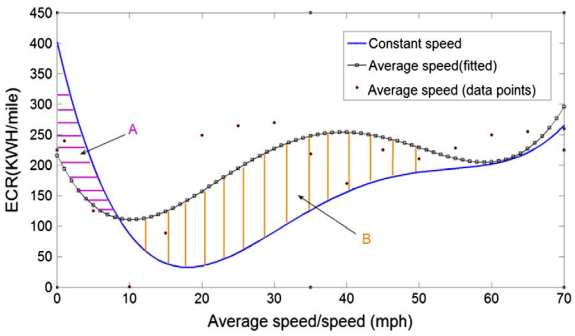


Fig. 20. ECR by constant speed vs. average speed.

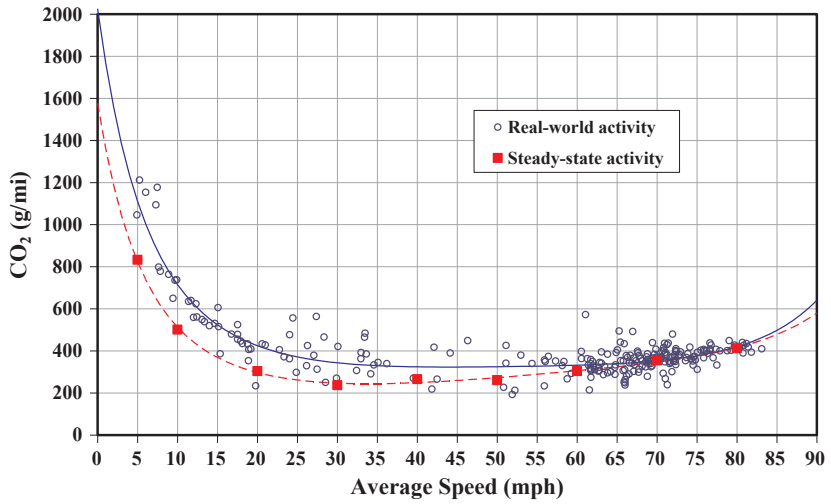


Fig. 21. ECR by constant speed vs. average speed [21].

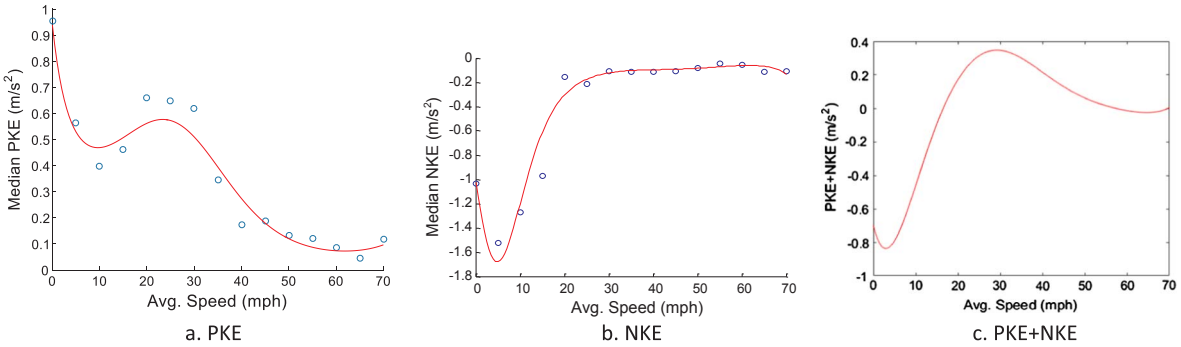


Fig. 22. Polynomial fit of PKE, NKE and PKE + NKE by average speed.

summary, the "W" shape in EV energy consumption comparing to that of gasoline vehicle is mainly due to the NKE drop at the low average speed which results from the regenerative braking.

6. Energy consumption estimation model

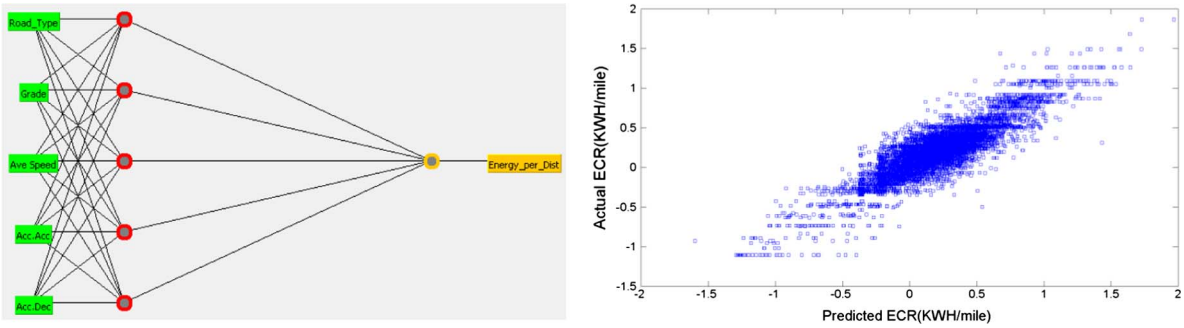
With the proposed predictors, the EV energy consumption estimation models are built and compared with the existing models.

6.1. Model fitting and evaluation

Based on the abovementioned factor analysis, three different types of link-level EV energy consumption models with different levels of information availability are proposed and compared (see Table 2). Specifically, Model type 1 takes the compound factors PKE and NKE as the only predicting variables; Model type 2 uses a 4th order polynomial of average speed and Model type 3 uses all the possible raw factors as predictors, such as average speed, PKE, NKE, road type, road grade, accumulated accelerations and

Table 2
Estimation models.

Model type	Predictors	Regression model	Goodness of fit(R^2)
1	PKE, NKE	ECR = 0.3394*PKE + 0.3153*NKE + 0.179	0.8911
2	Avg. speed (up to 4th order)	ECR = $-0.00011*v^4 + 0.024v^3 - 1.836*v^2 + 53.37*v + 250.2$	0.7737
3	Avg. speed, Acc.Acc, Acc.Dec, road grade, road type	ECR = $-0.0158*$road type $+ 0.186*roadgrade + 0.0016*avg.\ speed + 0.4095*Acc.\ Acc + 0.2897*Acc.\ Dec. + 0.1045$	0.6013



a. Network structure of model type 3 b. Plot of actual/predicted ECR

Fig. 23. Network structure of model type 3 with ANN.

accumulated decelerations. To better capture the different relationship (linear or nonlinear) between the predictors and energy consumption rate, for each type of model, both a linear regression fitted model and an artificial neural network (ANN) fitted model are built. Please note that Model type 1 is proposed in this study and model type 2 & 3 are regarded as baseline models to represent the similar existing models that discussed in (Yao et al., 2013; Grubwinkler et al., 2014). ANN is selected because of its good performance in fitting the nonlinear relations between input variables and output variables.

The performance of the 3 models was evaluated based on the estimation error that is defined in terms of symmetric mean absolute percentage error (Barth and Boriboonsomsin, 2008), or SMAPE:

$$SMAPE = \frac{\sum_{t=1}^{n} |E_t - A_t|}{\sum_{t=1}^{n} |A_t + E_t|} \tag{28}$$

where n is sample size, which is also the total number of road link that data were collected from; A_t is the actual energy consumption on road link t; and E_t is the estimated energy consumption on road link t. When building the model, 70% of the data set is used for training and 30% for model testing. For ANN models, 1 hidden layer and 5 hidden neuros as well as a 0.1 learning rate are chosen after a parameter tuning process. An example of the network structure of model type 3 is provided in Fig. 23a and a plot of actual and predicted ECR is provided in Fig. 23b. Table 3 summarizes the error statistics, where the best estimation is achieved by model type 1 with less than 5% of SMAPE, in spite of their simplicity. This result further verifies the findings in previous discussion that PKE and NKE are the most powerful compound indicators of link-level EV energy consumption when link-level typical vehicle speed trajectories are known (or just PKE and NKE is known). In addition, ANN based model is not performing significantly better than a simple linear regression model is because of the linear relation in nature between the predicting variables and predicted variable. This is also an important evidence to further validate the strong linear correlation between ECR and PKE/NKE that is discussed in previous sections. The polynomial regression in type 2 model achieves the second best performance (both fall into the error range of 10%) indicating that average speed is also a strong and effective indicator when only link-level average speed is available (e.g., based on measurements from the loop detectors). Type 3 models have the worst performance which is due to the fact that more prediction variables may introduce more noise. In addition, this model in real-world applications may require high-resolution data which are usually not available.

6.2. Model application and trip level validation

The proposed link-level EV energy consumption estimation model 1 is applied to a crowd-sourcing data based eco-routing system as shown in Fig. 24. All the EVs equipped with the routing system are connected through the mobile network. For each vehicle, the proposed best model (the regression model of Type 1) is trained with its own driving data collected on different road links using the same technology as described in Section 3. Please note that for each link, a model 1 can be built if the data is available. Once the model is built, the energy consumption on a selected route will be estimated using the estimation model with the crowd sourced PKE

Table 3
Estimation performance.

Model type	Regression		Neural Network	
	R-square	Error	R-square	Error
1	0.8019	4.93%	0.8076	4.97%
2	0.6162	6.94%	0.6099	9.81%
3	0.5909	15.11%	0.5801	12.55%

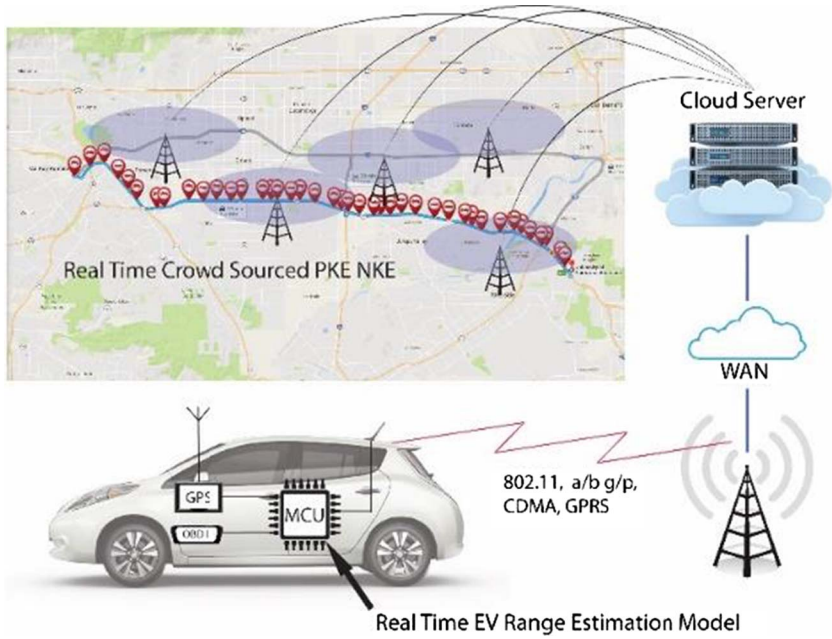


Fig. 24. crowed-sourced EV link level energy consumption estimation.

and NKE data from other vehicles on the road links at the time when selecting the route.

Therefore, we are also interested in knowing the performance of the proposed models in estimating the total energy consumption of complete trip. The estimated total energy consumption of a trip is simply a sum of estimated energy consumption of all the road segments in that route. Fig. 25 shows an example trip with the plotted actual energy consumption and estimated energy consumption with regression model of type 1 for each road link. For comparison purpose, results from models with better performance of each type (see Table 3) are plotted in Fig. 26. It is observed that type 1 is the closest curve to the actual energy consumption.

7. Conclusions and future work

In this work, the EV energy consumption on a road link in real-world traffic congestion is first decomposed based on fundamental physics. Then, with the real-world EV driving data, a data-driven energy consumption decomposition analysis is conducted by investigating two newly constructed compound factors: positive kinetic energy (PKE), and negative kinetic energy (NKE). The NKE is analysed and related to the regenerative braking power which is a unique feature for EVs (compared to ICE vehicles). The energy consumption rate curve along the average speed is constructed and a “W” shape curve is discovered and explained with real-world driving data. An accurate and simple EV energy consumption estimation model is built upon this decomposition and the feature selection analysis also proves the significant effectiveness of these two constructed factors. The comparison to the existing models shows that the proposed estimation model outperforms the existing models. As illustrated, the proposed models can be used to support the energy consumption estimation in eco-routing systems for electric vehicles as well as many other on-board and low cost eco-driving applications.

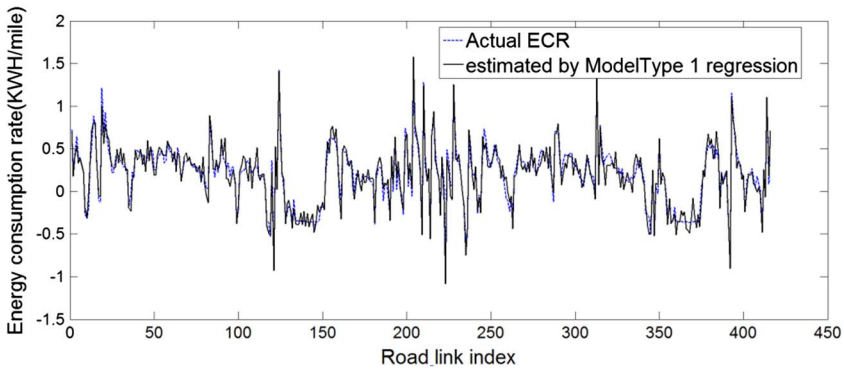


Fig. 25. comparison between actual energy consumption and estimated.

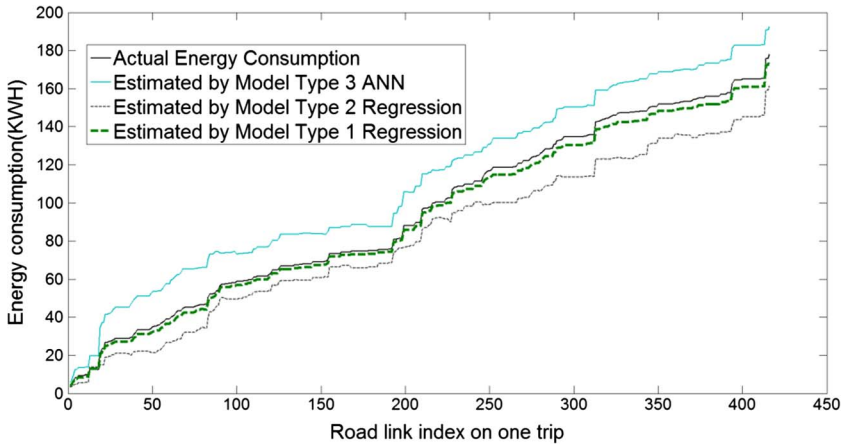


Fig. 26. Accumulated energy consumption on a trip.

Acknowledgement

This work is funded by California Energy Commission (Grant #: 11-01TE).

References

Alvarez, A.D., Garcia, F.S., Naranjo, J.E., Anaya, J.J., Jimenez, F., 2014. Modeling the driving behavior of electric vehicles using smartphones and neural networks. IEEE Intell. Transp. Syst. Mag. 6 (3), 44–53.

Armstrong, J.S., 1985. Long-range forecasting: from crystal ball to computer, 2nd. ed. Wiley.

Barth, M., Boriboonsomsin, K., 2008. Real-world CO2 impacts of traffic congestion. Transp. Res. Record No. 2058 163–171.

Barth, M., Mandava, S., Boriboonsomsin, K., Xia, H., 2011. Dynamic ECO-driving for arterial corridors. In: Integrated and Sustainable Transportation System (FISTS), 2011 IEEE Forum on, Vienna, 2011, pp. 182–188.

Billings, S.A., 2013. Nonlinear system identification: Narmax methods in the time, frequency, and spatio-temporal domains. Wiley.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth & Brooks/Cole Advanced Books & Software, Monterey, CA978-0-412-04841-8.

Bureau of Transportation Statistics (BTS).2015, Available at: http://www.bts.gov/publications/national_transportation_statistic.

Cauwer, C., Van Mierlo, J., Coosemans, T., 2015. Energy consumption prediction for electric vehicles based on real-world data. Energies 2015 (8), 8573–8593.

Chang, N., Hong, J., 2014. Power consumption characterization, modeling and estimation of electric vehicles. In: IEEE/ACM International Conference on Computer-Aided Design (ICCAD), San Jose, CA, 2014, pp. 175–182.

U.S. Environmental Protection Agency (EPA). 2016, DRAFT Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990–2014. Final report, Feb 2016.

Grubwinkler, S., Brunner, T., Lienkamp, M., 2014. Range prediction for EVs via crowd-sourcing. 2014 IEEE Vehicle Power and Propulsion Conference (VPPC). Coimbra 2014, 1–6.

Hemmerle, P., Hermanns, G., 2014. Macroscopic consumption matrix for on-line energy-efficient route guidance. In: Proc. Transp. Res. Record 2014. Washington D.C..

Holmberg, K., Andersson, P., Erdemir, L., 2012. Global energy consumption due to friction in passenger cars. Tribol. Int. 47, 221–234.

Quinlan, J., 1992. Learning with continuous classes. In: Proc. 5th Australian Joint Conf. Artificial Intell. pp. 343–348.

Quinlan, J., 1993. Combining instance-based and model-based learning. Proc. Tenth Int. Conf. Machine Learning 236–243.

Schrank, D., Eisele, B., Lomax, T., Bak, J., 2015. Urban mobility scorecard. http://d2dtl5nnlpfr0r.cloudfront.net/tti.tamu.edu/documents/mobility-scorecard-2015.pdf.

Shankar, R., Marco, J., 2013. Method for estimating the energy consumption of electric vehicles and plug-in hybrid electric vehicles under real-world driving conditions. IET Intel. Transport Syst. 7 (1), 138–150.

Watson, H. C. et al. 1982. Development of the Melbourne Peak Cycle. Paper #82148, SAE of Australia.

Wu, X.K., Freese, D., Cabrera, A., Kitch, W.A., January 2015. Electric vehicles' energy consumption measurement and estimation. Transp. Res. Part D: Transp. Environ. 34, 52–67.

Yao, E.J., Yang, M.J., 2014. Estimating energy consumption based on microscopic driving parameters for electric vehicle. Transp. Res. Record (J. Transp. Res. Board) 2454 (1), 84–91.

Yao, E., Yang, Z., Song, Y., Zuo, T., 2013. Comparison of electric vehicle's energy consumption factors for different road types. Discret. Dyn. Nat. Soc. 328757 (1–328757), 7.

Yi, Z.G., Bauer, P.H., 2015. Sensitivity analysis of environmental factors for electric vehicles energy consumption. In: 015 IEEE Vehicle Power and Propulsion Conference (VPPC), Montreal, QC, 2015, pp. 1–6.

Zhang, R., Yao, E.J., 2015. Electric vehicles' energy consumption estimation with real driving condition data. Transp. Res. Part D: Transp. Environ. 41, 177–187.