Range Prediction for Electric Bicycles

Lukas Gebhard
Universität Passau
Innstrasse 41
94032 Passau, Germany
gebhard@fim.uni-passau.de

Lukasz Golab
University of Waterloo
200 University Avenue West
Waterloo, Ontario, Canada
lgolab@uwaterloo.ca

S. Keshav
University of Waterloo
200 University Avenue West
Waterloo, Ontario, Canada
keshav@uwaterloo.ca

Hermann de Meer
Universität Passau
Innstrasse 41
94032 Passau, Germany
Hermann.DeMeer@uni-passau.de

ABSTRACT

Thanks to their affordability and practicality, electric bicycles (e-bikes) are becoming popular, especially in urban areas. They are a zero-emission and zero-carbon alternative to cars and therefore have a high potential to mitigate climate change. However, range anxiety can be a significant barrier to the adoption of electric vehicles. To address this challenge, in this paper we focus on how an e-bike’s remaining range can be accurately predicted. Using real data from the University of Waterloo WeBike field trial, combined with OpenStreetMap data, we evaluate two range prediction methods that take riding behaviour and route characteristics into account. Surprisingly, we find that predicting range for a particular cyclist based on his or her past energy consumption works as well as more complex methods that include additional information such as the route being travelled. Our findings also reveal which additional hardware and sensors e-bike manufacturers should provide in the future to make it easier to implement on-board range prediction. To the best of our knowledge, this is the first study of range prediction specifically for e-bikes.

Keywords
Electric vehicles, e-bike, range prediction, map-matching GPS, distance-to-empty

1. INTRODUCTION

According to the fifth assessment report of the Intergovernmental Panel on Climate Change (IPCC), the global transportation sector accounts for 14 percent of all anthropogenic greenhouse gases [10]. Electric vehicles can play a key role in mitigating this problem, especially as electricity grids worldwide transition to renewable energy sources. In particular, electric bicycles (e-bikes) are a rapidly-growing mode of urban transportation, with over 200 million on the road in China [19] and annual sales expected to grow from 32 million in 2014 to 40 million in 2023 [11]. E-bikes provide a clean (zero-emission and zero-carbon) and relatively inexpensive option for urban transportation as compared to cars, while being less physically demanding and allowing greater range and speed than regular bicycles.

The problem we study in this paper is how to predict an e-bike’s remaining range. This is an important problem because range anxiety, i.e., the fear of running out of battery with no place to recharge it before reaching the destination, is known to be a major barrier to the adoption of electric vehicles (see, e.g., [3]). Furthermore, in some parts of the world, e-bikes have been displacing cars and motorcycles, and therefore are used for long trips [1]. Some e-bikes allow the rider to pedal, but the combination of weight (e-bikes can be much heavier than traditional bikes), usage patterns (some owners use their e-bikes for shopping trips and therefore carry additional weight), and drag resistance from the motor can make pedalling unpleasant or even impractical.

Currently, most e-bike manufacturers publish the maximum range of their models, but it is not clear how this number is derived or how accurate it is. Furthermore, while most e-bikes are equipped with a digital display such as that shown in Figure 1, the display usually shows battery voltage, which is only a rough approximation of the remaining range. Thus, a more accurate and principled approach is required and would serve to mitigate some barriers to the adoption of e-bikes.

There has been a great deal of work on range estimation for electric cars (details in Section 2), but we are not aware of any published research specifically for e-bikes, which raise new challenges. For example, to keep prices low, e-bikes are
not equipped with sophisticated sensors and electronics, and any additional hardware required for on-board range prediction must be inexpensive. Towards the goal of practical and accurate range estimation for e-bikes, we make the following contributions:

1. We analyze data from a fleet of 31 sensor-equipped e-bikes used in the University of Waterloo WeBike project [23], combining it with OpenStreetMap [13] data to derive route-specific features that may affect range.

2. We test two data-driven range prediction approaches on the above datasets: a simple model based on riding history and a more complex regression model that considers the characteristics of the anticipated route. Surprisingly, we find that both approaches have similar predictive power.

3. We provide recommendations to e-bike manufacturers about the additional sensing hardware that is required to implement our range prediction approaches.

The remainder of this paper is organized as follows. Section 2 discusses related work on range prediction for electric vehicles; Section 3 defines the problems we want to solve; Section 4 discusses the required data processing steps; Section 5 presents our range prediction techniques and evaluates their performance; and Section 6 concludes the paper with lessons learned and directions for future work.

2. RELATED WORK

While we are not aware of research on range estimation for e-bikes, which are still a relatively new technology, there is a long history of work on range estimation for cars [20]. In particular, recent work on range estimation for electric cars can be classified into model-based, data-driven, and approaches combining both.

Model-based approaches employ physical modelling of the vehicle and the environment. For example, Hayes et al. have developed a general powertrain model with various parameters for different types of vehicles [5]. Then, given the planned route, they incorporate route topography into the model based on data from Google Earth. Oliva et al. propose another model-based approach employing Markov chains and particle filters to account for non-deterministic factors such as traffic situation and driving style [12]. To increase prediction accuracy, one can continue to model the vehicle and its environment in-depth. For instance, Valentina et al. present a detailed model of an electric vehicle’s air conditioning system [24].

In contrast, data-driven approaches do not require much knowledge of the internals of the vehicle, and aim to identify and model the factors affecting range from observable data. Vehicle sensor data can easily be accessed via the standardized Onboard Diagnostics port while smartphones provide GPS data. For example, Bolovinou et al. present a regression model that uses speed, route elevation and (short-term and long-term) past energy consumption as variables [2]. The model is dynamically updated over time as more data become available. Recently, Tseng et al. proposed a crowdsourcing approach, in which they collect route, speed and battery consumption data from multiple electric vehicles [21]. Based on these data, they identify drivers with similar habits and leverage similar drivers’ battery consumption to help determine the remaining range. This approach increases the amount of data that can be used to predict the range for any given driver and vehicle.

There are also range estimation approaches that combine powertrain modelling and data-driven techniques. For instance, Grubwinkler et al. present a statistical model derived from crowd-sourced power consumption data [7]. First, they model vehicle characteristics using a power consumption map which provides the amount of energy consumed for each operating point, that is, for each speed and acceleration value. Second, they distinguish between several road types defined by OpenStreetMap [13]. For each road type, they then derive a probability density function of operating points based on the crowd-sourced data. Given a specific road type, it describes how likely each operating point is. Together, this provides all the information needed to statistically predict the energy consumption for a particular route.

3. PROBLEM STATEMENT

We take a data-driven approach and investigate simple and practical models for range prediction for a given e-bike-rider pair. In the remainder of this paper, we discuss the challenges in identifying and extracting useful features from e-bike usage data, and we compare the accuracy of various models, from simple ones which only take prior riding behaviour into account, to more complex ones that also include route properties.

We consider two situations, as described in Table 1. In the first setting, we do not know the rider’s destination and therefore we can only extrapolate past energy consumption to predict distance-to-empty. At a minimum, we need a time series of battery consumption and odometer readings, from which we estimate the average battery consumption per kilometre. Using this input, we will build one model per rider that reflects his or her typical battery consumption. The model can then be used for on-board range estimation by periodically probing the battery consumption and estimating the number of kilometres that can be travelled.

In the second setting, the destination is also given, therefore we can add route-specific parameters to each rider’s model. This requires a richer dataset: in addition to a sequence of mileage and battery consumption values, we need timestamped GPS traces of the rider’s past trips, which we can cross-check against a map to extract route-specific information that can influence range, including average speed, the number of stop signs, and the road surface (most e-bikes...
Table 1: Two settings for range prediction studied in this paper.

<table>
<thead>
<tr>
<th>Goal</th>
<th>Destination known</th>
<th>Destination unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction based on</td>
<td>Past riding behaviour</td>
<td>Past riding behaviour; Anticipated route profile</td>
</tr>
</tbody>
</table>

can be ridden on paved roads and off-road trails). These parameters can capture more information about riding habits such as how quickly a rider accelerates from a stop sign. Then, when the rider specifies a destination, we can calculate the shortest path to it and estimate the state-of-charge upon reaching the destination using the characteristics of the planned route.

In this paper, we seek simple and interpretable models using inputs that bike manufacturers can reasonably expect to obtain, e.g., battery state of charge, battery temperature, speed, route profile. An interesting direction for future work is to study the effects of other types of factors, such as wind speed, tire pressure and rider’s weight, on e-bike range.

Of course, any data-driven model assumes that past history is a good predictor of the future. If a rider changes his or her riding style, our models must be recomputed.

4. DATA PROCESSING AND ANALYSIS

4.1 Data Sources

Before presenting our range estimation techniques, we discuss the challenges involved in obtaining the required data. The main problem is that in contrast to cars, e-bikes are not equipped with an on-board diagnostics port and therefore there is no built-in way of collecting data. Usually, the battery is sealed and the digital display (recall Figure 1) has no data storage or transmission capabilities.

Fortunately, field trials are being conducted, in which participants ride e-bikes that are outfitted with cameras and sensors [4, 18, 14, 23]. We use data from the University of Waterloo WeBike project [23]. In this project, 31 participants have been selected to ride instrumented e-bikes. Prospective riders filled out a questionnaire and the final 31 participants were selected based on their anticipated usage of the e-bike. Roughly half the riders are male and half are female, and one-third each are University of Waterloo students, staff and faculty.

Each participant was given an eProdigy Whistler e-bike, shown in Figure 2 equipped with a custom-made hardware kit shown in Figure 3. The manufacturer-specified range of the bike is 45 km. The hardware kit is mounted on the battery and includes battery voltage, battery current and battery temperature sensors, as well as a Samsung Galaxy III smart phone which provides GPS, acceleration and gyroscope data. The phone battery is charged by the bike battery. To save power, the phone is in sleep mode for 56 seconds out of every minute, and takes four measurements from each sensor during the four seconds it is active. When feasible, the collected data are uploaded to a server over WiFi.

The collected dataset spans from June to October 2015. Since all the participants are members of the University of Waterloo, most of the data are generated from commuting trips to and from the university and surrounding areas. We refer the interested reader to [17] for more details about the collected data.

In addition to obtaining the e-bike usage data, we created a routable street graph using OpenStreetMap [13]. The graph includes bike trails and footpaths as well as traffic lights and stop signs.

In the remainder of this section, we explain how to process and analyze the above data to obtain the variables we need for our range prediction models. Recall that if the destination is not given, we only require a sequence of distance and battery consumption measurements. There is no direct access to the odometer on the bike, but we can calculate distance by reconstructing the routes travelled using GPS traces and OpenStreetMap data (Section 4.2). Furthermore, we can calculate battery consumption from voltage and current measurements (Section 4.4). Route properties can also be obtained from the OpenStreetMap data (Section 4.3).

4.2 Map-Matching GPS Traces

The first data processing step is to reconstruct trip routes from GPS data. Recall that this is necessary to determine distance travelled and to link route-related features such as road surface or the number of stop signs with battery consumption. This task, known as map-matching GPS, amounts to matching GPS data points, referred to as GPS fixes, to an underlying road network.

There are many existing techniques that can solve this problem, from relatively simple to highly sophisticated ones; see [15] for a review. Most algorithms “are designed for use with high frequency positioning data (i.e. 1 Hz or 1 second sampling interval)” [16]. In addition, there are some approaches (e.g., [16, 25]) that specifically account for sampling intervals of 1 minute and higher. Since the WeBike dataset contains bursts of up to four GPS readings ever

Some data are lost due to the time taken to obtain a GPS
Figure 3: Hardware kit used to collect e-bike data. An aluminum box attached to the battery (upper-right corner) contains a smartphone and several sensors. An assembly guide can be found at [22].

Algorithm 1: Matching GPS fixes with road graph edges

```plaintext
1: function MapMatch(G, [f1, f2, ..., fn]) ⋄ n ≥ 2
2: u ← GetClosestEdge(f1, G) ⋄ Start edge
3: s ← GetClosestNode(f1, u) ⋄ Start node
4: M ← [] ⋄ Result list
5: i ← 2 ⋄ i
6: while i ≤ n do
7: v ← GetClosestEdge(fi, G) ⋄ Target edge
8: t ← GetClosestNode(s, v) ⋄ Target node
9: M.append(GetShortestPath(s, t, G))
10: i ← i + 1
11: s ← t ⋄ Shift start node
12: end while
13: return M
14: end function
```

minute, we use an approach similar to that from Quddus et al. [16]. The main idea is to connect the observed GPS fixes via shortest paths consistent with the vehicle’s trajectory. In addition, Quddus et al. take bearing information into account and introduce a weighing mechanism. We show that even without these advanced features, our algorithm provides results that meet our needs.

Algorithm 1 shows the details. As input, we take in a routable road network graph G and a sequence of GPS fixes [f1, f2, ..., fn]. The output is a list of matched road graph edges M that represent the reconstructed route. The algorithm is divided into an initialization part (lines 2-5) and a loop (lines 6-12). The initialization part identifies a node s from the road graph G representing the trip’s start position. This can be achieved by first searching for the edge u that is closest to f1 (line 2). We define the distance between a node and an edge in a standard geometric sense, as the length of the shortest straight line connecting both. As an optimization, we may only consider edges within a certain radius. Second, we look at the two nodes defining u. We select the one that is closer to f1 and call it our start node s (line 3).

Next, the algorithm iteratively considers the section from the current to the next GPS fix and then moves on to the next pair of fixes. Let fi be the current fix and s the (recently found) corresponding node of the road graph. Similarly to the initialization step, we first search for the edge v that is closest to fi+1. Second, we compare the two nodes defining v, select the one that is closer to s and call it t. Third, we append all edges that belong to the shortest path between s and t to our result list M. Finally, we let t be the new start node s and move on to the next fix.

Figure 4 shows the results of the algorithm on three GPS traces from three different e-bikes. The red dots indicate the GPS fixes. Note the shortest-path interpolation in Trip C, which has missing GPS data in the middle.

4.3 Trip Segmentation

The next step is to divide the reconstructed trip routes into trip segments, i.e., consecutive road graph edges that correspond to the likely route taken by an e-bike between two GPS fixes. We will then build a model that correlates battery consumption with various properties of a segment (length, number of stop signs, etc.). In order to be included in the set of segments that will be used to build the model, a segment must satisfy the following three properties:

Minimum segment length α We denote the minimum length of a trip segment by α. We may need to combine multiple consecutive road graph edges until their total length is at least α.

Minimum riding duration β We denote the minimum time of travel for a segment by β. This is important to ensure an accurate speed calculation, i.e. combined segment lengths divided by combined riding duration.

Maximum GPS time gap γ The maximum duration for which GPS may be unavailable is denoted by γ. If GPS data are missing for a long period of time, the shortest path assumed by Algorithm 1 may not necessarily be the path taken by the e-bike, so we want to exclude such trip sections.

Since the data set we are using contains bursts of measurements every minute, we set β = 60 s. We tested different values of γ ∈ [80 s, 300 s] and found that the total number of road segments we identified did not change significantly, so we conservatively set γ = 80 s. Increasing α leads to a decreasing total number of obtained trip segments but on the other hand results in higher variance of energy consumption per segment. As a trade-off, we select α = 400 m.

Given these parameters, the trip segmentation algorithm produced 3619 segments from the WeBike dataset. Figure 5 shows a segmentation of a sample trip; just for this illustration, we increase α to 700 m to make the segments more visible. The blue, red and pink lines represent the created segments. They all have approximately equal length. In between the blue and red segments, no GPS data were logged, probably due to missing satellite connection. As the duration of GPS being unavailable exceeds γ, this section is not included in our set of 3619 trip segments.
Finally, for each segment, we obtain its properties from OpenStreetMap data, such as road surface, the number of stop signs, etc.

### 4.4 Obtaining and Analyzing Energy Consumption

The last piece of the puzzle is to estimate the battery consumption for each trip segment, call it $\Delta E$. We do this by linearly interpolating the product of voltage and current over time. Let $t_{\text{start}}$ and $t_{\text{end}}$ be the times when the rider entered and exited the trip segment, respectively. Let $U(t)$ and $I(t)$ denote voltage and current at time $t$, respectively. Moreover, let $t_0, t_1, \ldots, t_k$ be the times at which voltage and current were measured. Note that our trip segmentation algorithm ensures that $t_{\text{start}} = t_0$ and $t_{\text{end}} = t_k$. We approximate

$$\Delta E = \int_{t_{\text{start}}}^{t_{\text{end}}} U(t)I(t) \, dt$$

by computing the Riemann sum

$$\sum_{i=1}^{k} U(t_i)I(t_i)(t_i - t_{i-1}).$$

Finally, we divide $\Delta E$ by the length of the segment to obtain a battery consumption value in units of Wh/km.

Figure 6 shows a histogram of the distribution of energy consumption for all 3619 trip segments (across all riders). We observe a broad range of per-segment consumption values, most of which range between zero and 30 Wh/km. The spike at 0 Wh/km corresponds to trip segments where riders did not use the motor. This happens after the riders have reached their desired speed and are coasting, usually downhill—the rider always spins whenever the rider is pedalling—or after they have reached the e-bike’s maximum speed of 32 km/h and the motor has disengaged automatically. The per-segment average works out to 7.9 Wh/km. The eProdigy Whistler e-bikes have a battery capacity of 330 Wh, which corresponds to an observed average range of $\frac{330}{7.9}$, or about 42 km. This is quite close to the manufacturer-specified maximum range of 45 km.

Next, we analyze the energy consumption of different riders. Of the 31 total riders, only 13 had at least 80 road segments in the collected dataset. Figure 7 shows a clustering of the average energy consumption of these 13 rid-
Figure 6: Energy consumption distribution for the 3619 trip segments identified from the data.

Figure 7: Clustering riders by normalized mean energy consumption: “economical” riders consuming 5 to 6 Wh/km, “average” riders consuming around 8 Wh/km, and “aggressive” riders consuming over 9 Wh/km. Note that the rider with the highest consumption (9.9 Wh) uses, on average, more than twice as much energy per kilometre as the most economical rider (4.8 Wh). This motivates the need to develop per-rider range prediction models rather than using the manufacturer-specified maximum range for everyone.

Upon further inspection of the data, we found that economical riders were much more likely to make trips with average speed under 20 km/h. On the other hand, aggressive riders had few trips with average speed under 20 km/h. The Whistler eProdigy e-bikes provide five motor-assist modes—each mode corresponding to a target speed—that are selected via a thumb switch integrated into the digital display. Mode 1 provides little power and has a target speed of about 15 km/h, and mode 5 has a target speed of about 35 km/h, and provides enough assist that the rider does not need to provide much additional power under most conditions. While the dataset, unfortunately, did not include assist level, we hypothesize that aggressive riders may use mode 5 more often than economical riders. Moreover, all four of the aggressive riders are staff or faculty members and none are students, suggesting that perhaps older riders are more likely to use higher assist levels.

Figure 8: Energy consumption distributions: Cluster of “economical” riders

In Figure 8, Figure 9 and Figure 10, we show the consumption histograms for the 13 riders with at least 80 road segments, split by cluster. Each histogram shows noticeable spread, meaning that the same rider may use different amounts of energy in various trip segments.

One might ask whether the same rider uses the same amount of energy whenever he or she travels on the same route, or whether different riders use different amounts of energy when travelling on the same routes. Unfortunately, we cannot answer these questions due to the low frequency of GPS data (and missing GPS data) in the WeBike dataset. Since data are collected roughly once per minute, the route between two consecutive GPS fixes (and therefore two consecutive battery current/voltage measurements) typically consists of several OpenStreetMap road graph edges. This means that we do not have accurate battery consumption data at the granularity of road graph edges. Furthermore, trips that follow the same route are typically split into different trip segments (recall Section 4.3), depending on the timing of the GPS fixes.

5. RANGE PREDICTION

We now evaluate two methods for range prediction corresponding to the two cases mentioned in Section 3. For situations where the destination is not given, we study mean prediction, i.e., we use the average historical consumption (recall Figure 7) of the rider to estimate his or her range. If the destination is given, we consider a linear regression model that uses route-specific details. We reiterate that to account for individual riding behaviour, we fit a separate model for each rider.

We use the first 120 km of each rider’s trips as the training data and we use the remaining data to test the models. To prevent overfitting, in this section we only use the top four most frequent riders, summarized in Table 2. The other riders have fewer than 300 trip segments and not much more than 100 km of riding data. To evaluate prediction perfor-
Figure 9: Energy consumption distributions: Cluster of "average" riders

Figure 10: Energy consumption distributions: Cluster of "aggressive" riders

Table 2: Characterizing the four riders used for model fitting

<table>
<thead>
<tr>
<th>Rider</th>
<th>Mean consumption</th>
<th>Average trip length</th>
<th>Number of trip segments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.7 Wh/km</td>
<td>3.5 km</td>
<td>317</td>
</tr>
<tr>
<td>2</td>
<td>8.3 Wh/km</td>
<td>2.8 km</td>
<td>752</td>
</tr>
<tr>
<td>3</td>
<td>8.4 Wh/km</td>
<td>6.3 km</td>
<td>306</td>
</tr>
<tr>
<td>4</td>
<td>9.9 Wh/km</td>
<td>4.6 km</td>
<td>653</td>
</tr>
</tbody>
</table>

5.1 Mean Prediction

The mean prediction model is simple to compute and easy to implement in practice on an e-bike display. It suffices to periodically record the battery voltage and current, and the odometer reading and maintain the mean consumption. Figure 11a shows the distribution of the per-trip normalized residuals. For 89 percent of trips, the predicted consumption deviates from actual consumption by at most ten percent of the battery capacity. Depending on the rider, the normalized root mean squared errors (RMSE) range from 4.9 to 7.4 percent, as listed in Table 3. Plotting the normalized residuals versus the corresponding test trip lengths (see Figure 11b) reveals that short and long trips appear to be predicted equally well (and that there are more short trips under 4km than long trips).

The accuracy of the simple mean prediction technique may seem surprising given what we have seen in Figure 8, Figure 9 and Figure 10. This may be explained as follows. Clearly, using mean prediction to estimate battery consumption for a single trip segment is inaccurate because the same rider may use a different amount of energy on different trip segments. However, when predicting the remaining range at the end of a trip, as we do in this paper, some segments are underestimated while others are overestimated, and the per-segment errors roughly cancel each other out.

5.2 Linear Regression

5.2.1 Model Definition

We now investigate the predictive power of incorporating all available data into the model. To predict the energy consumption for a given trip, we first split it into segments as described in Section 4.3. For each road segment, we model the energy consumption \( \hat{y} \) as a linear function of four variables: the average speed \( v \), the number of traffic lights, stop signs and give-way signs \( z \), the off-road fraction of the segment (i.e., foot paths and bike paths where cars are not allowed) \( r \), and the average battery temperature \( \theta \):

\[
\hat{y} := m_v v + m_z z + m_r r + m_\theta \theta + b.
\]
For each rider, we then use ordinary least-squares linear regression to obtain the coefficients $m_i$ for $i \in \{v, z, r, \theta\}$. $b$ is the intercept term.

Table 4 describes the features of the model in more detail. We calculate average speed by dividing the length of a segment by the time taken to ride it. We obtain $z$ and $r$ from OpenStreetMap, and we get $\theta$ from the temperature sensor that is included in the hardware kit. Note that our dataset does not include any information about the riders themselves (such as weight) or any weather information (such as wind or precipitation). Also, even though the change in elevation throughout each trip segment can be computed based on satellite raster images (available from e.g., [9]), the topology of the area surrounding the University of Waterloo is relatively flat. As a result, most of the trip segments we identified have little or no change in elevation. Thus, we do not use this parameter in our regression model.

The linear regression model is somewhat more difficult to implement in practice. We require a navigation-like interface on the e-bike to input the planned destination, and a copy of a routable road graph to determine the shortest path to the destination, and to extract the features we need for each segment of the anticipated route. To determine whether there is sufficient range to reach the destination, we then sum up the predicted battery consumption for each segment of the planned route. We also need to train the regression model using past history, and perhaps periodically re-compute the model as more data become available. This can be done aboard the bike or remotely at the server that collects the data. For the training dataset, we need battery voltage and current sensors, as before, plus GPS data and battery temperature.

A subtle issue arises when predicting the average speed $v$ for each road segment along the planned route. In the training data set, the average speed can easily be calculated by dividing the length of a trip segment by the time it took the rider to travel through it. However, we do not know what the speed will be on the planned route. In our analysis, we can “cheat” since even the trips on which we are testing the model have already taken place. Thus, for this analysis, we input the actual average speed per segment to our model. This way we can directly examine the correlation of observed speed and energy consumption. However, in practice, we need a way to predict what the speed is going to be. This can be done by examining this rider’s average speed on (similar road segments in) previous trips, or by searching for similar riders who have already travelled these particular road segments, similar to the crowd-sourcing approach of Tseng et al. [21] that we mentioned in Section 2.

### 5.2.2 Results

Before fitting the regression model, we analyze the distribution histograms of the four features in Figure 12 to ensure that the data look “reasonable” and there are no obvious data quality problems; these histograms include data from all four riders. First, note that some road segments are showing an unusually high average speed, considering that the e-bike motor shuts off at 32 km/h. There are at least
two possible explanations. One is that the OpenStreetMap data do not include all possible routes that an e-bike could take (e.g., there may be unknown shortcuts through a park or a playing field). As a result, our map-matching algorithm (recall Section 4.2) cannot do any better than to output a long roundabout route for certain trip segments. Another possibility is that some riders have taken their e-bikes on a bus or a train for some portions of their trips. As a result, we remove all trip segments with an average speed above 32 km/h.

As for the other features, the number of traffic lights and stop signs is often zero and rarely more than four, the average battery temperature is around 25 degrees Celsius, and the off-road percentage is usually zero (e.g., a regular city street) or 100 percent (e.g., a dedicated bike path).

Now, after using the ordinary least squares method to fit four regression models, one for each rider, we obtain a coefficient of determination $R^2$ of zero. By definition, this means the estimator does not perform better than one that simply predicts the overall mean consumption across all riders. Accordingly, the residual plots resemble those obtained by mean prediction (Figure 11). Table 5 lists the per-trip prediction errors for each rider. They turn out to be slightly higher than the mean prediction errors (see Table 3) which clearly indicates overfitting.

To check if the features correlate with energy consumption in a non-linear fashion, we perform residual analysis. For each feature, we plot a subset of the residuals where the other features are approximately constant. For example, when plotting residuals versus $r$, we only consider trip segments with

$$z = 0$$

and $|v - \bar{v}| < \frac{5 \text{ km}}{h}$

and $|\theta - \bar{\theta}| < 5^\circ C$,

where $\bar{v}$ and $\bar{\theta}$ denote average speed and temperature, respectively. Figure 14 shows the (normalized) residual plots. We see no obvious evidence of non-linearity. Thus, it appears that although the features we have chosen seem reasonable, they do not improve range prediction beyond the simple mean prediction model.

### 6. CONCLUSIONS AND FUTURE WORK

In this paper, we initiated the study of range prediction for electric bicycles. Using real e-bike usage data, we analyzed two techniques: one that only uses average battery consumption from past trips, and one that uses a linear regression model with battery temperature and route-specific information such as the number of stop signs and traffic

---

**Figure 12:** Distributions of average speed, number of traffic signals, average battery temperature and off-road percentage.

**Figure 13:** Evaluating the linear regression approach

---

**Table 5:** Per-trip root mean squared errors for linear regression

<table>
<thead>
<tr>
<th>Rider</th>
<th>RMSE</th>
<th>RMSE normalized by battery capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17.5 Wh</td>
<td>5.3%</td>
</tr>
<tr>
<td>2</td>
<td>18.7 Wh</td>
<td>5.7%</td>
</tr>
<tr>
<td>3</td>
<td>31.8 Wh</td>
<td>9.7%</td>
</tr>
<tr>
<td>4</td>
<td>20.8 Wh</td>
<td>6.3%</td>
</tr>
</tbody>
</table>
lights. Our main findings are as follows:

1. As we showed in Figure 6 and Figure 7, riding behaviour matters. The most economical rider in our dataset was over twice as efficient, on average, as the most aggressive one. Thus, the manufacturer-specified maximum range is not an accurate predictor for everyone.

2. Even though a particular rider may use a different amount of energy under different circumstances (Figure 8, Figure 9 and Figure 10), we found that in flat regions, range prediction based on historical mean energy consumption works well at the granularity of trips. Using real trip data, our predictions were usually within ten percent of the actual remaining range at the end of a trip.

3. Surprisingly, we found that a linear regression model using average speed per route segment, the number of traffic signs, off-road percentage and battery temperature performed no better than mean prediction. One possible explanation is that average speed per route segment does not vary significantly (as compared to, e.g., cars) and that accelerating an e-bike from full stop is not as costly as accelerating a heavier car.

4. To save power, e-bike sensor data may need to be collected with lower frequency than desired. As we discussed in Section 4.2 map-matching GPS data generated by e-bike usage required an algorithm that can handle sparse data. Moreover, as we mentioned in Section 5.2 e-bikes (and bikes in general) are not restricted to travelling on city streets and well-marked paths, meaning that some errors in interpolating routes between GPS fixes are difficult to avoid.

Our results should be of interest to e-bike manufacturers. Our findings suggest that a simple on-board prediction technique can be implemented by measuring battery voltage, battery current and mileage. Most e-bikes are already equipped with an odometer, so by making odometer data accessible, and deploying additional sensors at the battery, our simple technique can be implemented inexpensively. Since range anxiety is a major barrier to the adoption of electric vehicles, adding such a range estimator can help widen the appeal and increase the sales of e-bikes.

We acknowledge that our conclusions are based on only one dataset, and therefore further work on different datasets (including those with data sampled more often than once a minute) should be undertaken to validate our findings. Additionally, while we discovered that route-specific information was not helpful for range prediction in our relatively flat region, the results from Figure 8, Figure 9 and Figure 10 indicate that the same rider consumes battery power at a different rate for different trip segments. Thus, the impact of other factors, such as the actual battery capacity (which depends on its age and temperature), change in elevation, wind speed and rider’s weight, should be studied. Finally, we are interested in extending our range prediction algorithms to report confidence intervals rather than point predictions.

7. ACKNOWLEDGEMENTS

Funding for this research was provided by Cisco Systems and the Natural Science and Engineering Research Council of Canada (NSERC).

References


[22] University of Waterloo ISS4E group. Assembly guide for the WeBike hardware kit. URL: http://blizzard.cs.uwaterloo.ca/issue/?page_id=4945 (visited on 01/07/2016).

[23] University of Waterloo ISS4E group. WeBike project homepage. URL: http://blizzard.cs.uwaterloo.ca/issue/?page_id=3661 (visited on 12/10/2015).
