Kivekäs, Klaus; Vepsäläinen, Jari; Tammi, Kari

Stochastic driving cycle synthesis for analyzing the energy consumption of a battery electric bus

Published in:
IEEE Access

DOI:
10.1109/ACCESS.2018.2871574

Published: 20/09/2018

Please cite the original version:
Stochastic driving cycle synthesis for analyzing the energy consumption of a battery electric bus

Klaus Kivekäs¹, Student member, IEEE, Jari Vepsäläinen¹, Student member, IEEE and Kari Tammi¹, Member, IEEE
¹Department of Mechanical Engineering, Aalto University, 00076 Aalto, Finland

ABSTRACT
Previous driving cycle generation studies have focused on constructing a single cycle from a large data set. This paper takes the opposite approach by presenting a novel driving cycle synthetization method that can generate a large number of varying cycles and passenger numbers for a chosen bus route based on only a handful of measured cycles. The validity of the method was confirmed by comparing the statistical properties of synthesized cycles to those of measured cycles on a selected route. The method was applied to analyze the impact of driving cycle and passenger load variations on energy consumption variations of a battery electric bus on a suburban route in Espoo, Finland. Ten-thousand synthetic cycles were run with a validated electric bus simulation model, and the energy consumption during each cycle was recorded. The mean consumption was 0.914 kWh/km, the standard deviation was 0.043 kWh/km, and the range of the consumption variation was 0.331 kWh/km. The new cycle synthetization and consumption distribution forecasting methods are useful tools for public transport authority, route operators, and bus manufacturers, as they could be used in optimizing bus powertrain dimensions as well as bus line schedules.

INDEX TERMS
Driving cycle, electric bus, energy consumption, passenger load, uncertainty

I. INTRODUCTION
City buses are an obvious target for electrification due to the plannable nature of bus routes and the low range requirement between charging points on the route. Technological advancements have reduced the gap in life-cycle costs between battery electric buses (BEB) and diesel buses (DB) [1]. Consequently, several governments have started to take measures to accelerate the electrification of bus fleets in major cities in order to cut down on greenhouse gas emissions. For example, the bus fleets in Paris and Amsterdam are expected to be fully electric by 2025 [2].

The production costs of BEBs are currently significantly higher than those of DBs, and the lower operating costs are not enough to offset them in most cases [1]. Hence, there is a need for better tools to be developed for optimizing the components of BEBs in order to further reduce the costs [3]. One way to improve the optimization process is to develop a method for incorporating the statistical uncertainties in the driving cycles of the bus routes into the design process. Such a method would allow for more optimal component sizing because the driving cycle is one of the most significant factors influencing the energy consumption of a BEB [4].

Driving cycles from various different regions have been used in previous BEB simulation studies, which are discussed in section II.B. The energy consumption variations on a single bus route caused by driving cycle uncertainty have not been analyzed. Furthermore, the driving cycle construction methods developed until now, which are discussed in section II.A, have only focused on creating a single cycle that represents the average driving characteristics in the specified region. Driving cycle uncertainty also provides challenges for the optimization of plug-in hybrid electric vehicle (PHEV) energy management [5]. Various solutions to this problem have been proposed such as using a probability distribution of driving cycles in predicting power demand [6] as well as creating a computationally efficient on-line optimization algorithm that accounts for trip length uncertainty [7].
The purpose of this research is to develop a method for synthesizing a multitude of driving cycles for a single bus route based on limited measurement data. The method is then used to analyze the energy consumption distribution of a BEB in a case example. In the novel driving cycle synthesis method, new cycles are synthesized from measured cycles by combining segments and varying the stations at which the bus stops. Additionally, the number of passengers boarding and alighting from the bus at each stop is varied. By implementing the probabilities of the stops and passenger variations based on real data, a realistic forecast for the energy consumption distribution of the bus on that specific route can be obtained with simulation studies.

The ability to forecast the energy consumption variations caused by driving cycle uncertainty and passenger load variations on a specific route benefits the public transport authority (PTA), route operator, bus manufacturer, and eventually also public transport users. That is because it allows for the powertrain dimensioning to be optimized better for the route and thus improve cost efficiency. In particular, the battery can be scaled accurately, reducing costs and improving the energy efficiency of the bus due to lower weight. Despite the constantly falling costs of batteries, they are still the single most significant factor in the cost of electric vehicles [8].

Developing new methods for optimizing BEBs is particularly relevant currently due to the large-scale electrification of city bus fleets going on around the world. Furthermore, the cycle synthetization could be used to optimize the timetables of bus lines operated with BEBs, as the energy consumption directly influences the charging times. The cycle synthesis method could also prove beneficial for optimizing PHEV bus energy management for a specific route when used in conjunction with the cycle distribution based control strategy optimization method presented in [6].

While the results of the case example are not universal for all routes, the results are nonetheless considered to provide novel insights into the magnitude of the energy consumption variations on a single bus route, as such an exhaustive consumption distribution analysis for a specific bus route has not been performed previously. The results should also provide novel information about how significantly the different cycle parameters influence the consumption. Additionally, previous publications have not quantified the influence that stops and passengers have on the energy consumption of a BEB. That is another research gap this paper aims to fill.

In this paper, we begin by examining previously developed driving cycle construction methods and previously conducted BEB energy consumption analyses. In chapter three, the data collection methods, the novel driving cycle construction method, and the implementation of the passenger load variations are discussed. The cycle construction method is also validated. The BEB simulation model used in the research is presented and its validation demonstrated. In the fourth chapter, the case example simulation results are shown, analyzed, and discussed. In the fifth and final chapter, conclusions are drawn, suggestions for improvements are given, and future developments are considered.

II. STATE-OF-THE-ART

A. DRIVING CYCLE CONSTRUCTION METHODS

Driving patterns vary from region to region. Standard driving cycles intended to represent driving patterns across larger regions, countries, and continents are thus often not suited to describing the typical driving characteristics of a specific city or smaller region [9], [10]. Consequently, numerous methods have been developed for constructing region-specific custom driving cycles for buses and passenger cars based on measured speed-time data. The measurements can be conducted with the chase car method, on-board measurement method, or a combination of the two. In the chase car method, an instrumented vehicle is used to follow a target vehicle in a predetermined area. The driver of the instrumented vehicle attempts to replicate the driving maneuvers of the car in front as accurately as possible. In the on-board measurement method, selected vehicles are instrumented in order to collect data from a specific area. This method requires more resources but produces more accurate speed data [11].

Synthetic driving cycle construction methods can be broadly divided into four categories: micro-trip based, segment-based, pattern classification, and modal cycle construction method [12]. In the micro-trip based method, the measured speed data are divided into micro-trips, which are defined as excursions between two successive time points at which the vehicle has been stationary. Candidate cycles are synthesized from the micro-trips, which are chosen either randomly or based on specific modal characteristics. The candidate cycles are assessed based on target parameters that are defined according to the measured data, and the best candidate is then chosen as the new driving cycle. Hung et al. developed urban, suburban, and highway driving cycles for Hong Kong by randomly combining measured micro-trips into full cycles and evaluating the resulting candidate cycles based on their performance values (PV) and speed-acceleration probability distributions [13]. The PVs were calculated by comparing the statistical acceleration and velocity properties of the synthesized cycles to those obtained from the entire measurement data set. Kamble et al. developed a driving cycle for Pune, India by evaluating the percentage of time spent accelerating, decelerating, cruising, and idling, as well as the average velocity in candidate cycles that had been formed by combining measured micro-trips [10]. A similar approach with a higher number of target parameters was used in creating a bus driving cycle for Chennai, India [14]. Lai et al. developed bus driving cycles for rapid transit, express, and regular lines in Beijing by assessing candidate cycles based on their vehicle-specific power requirements [15]. The Singapore Driving Cycle was constructed from selected micro-trips such...
that its proportions of arterial road and expressway driving, as well as low and high-traffic driving corresponded to the proportions found in driving statistics obtained with surveys [16].

The segment-based cycle construction method is similar to the micro-trip method, but the measured data are partitioned based on not only the stops but also the road characteristics and traffic conditions. Hence, segments can begin and end at any speed, and constraints on speed and acceleration must be used in chaining the segments together when synthesizing a new cycle [12]. The Australian Composite Urban Emissions Drive Cycle was developed using the segment-based method by segmenting the speed data based on road type and micro-trips [17]. The cycle was then synthesized from the segments by setting target parameters based on the measured data.

In the pattern classification method, the measured speed data are divided into kinematic sequences similar to micro-trips [12], [18]. Using statistical methods, the sequences are classified into heterogeneous classes based on chosen characteristic parameters. A new driving cycle is constructed by combining kinematic sequences based on the statistical properties of the classes. In several studies, new cycles have been synthesized from kinematics sequences with the aid of principal component analysis (PCA) and cluster analysis [18]–[24]. Jing et al. created a passenger car driving cycle for Tianjin, China by analyzing kinematic sequences with two-class Fisher discriminant analysis instead of PCA and clustering [25].

The modal cycle construction method uses maximum likelihood estimation clustering to partition the measured speed data into snippets and to classify them into modal bins [12], [26]. The new cycle is then created from selected snippets as a Markov chain with the aid of a transition matrix containing the succession probabilities between different modes.

While various driving cycle generation methods have been presented in previous papers, they have all centered around the concept of constructing a single statistically representative cycle for a region based on a large quantity of measured data. The cycle synthetization method presented in this paper takes the opposite approach by creating a large number of varying cycles for a single bus route based on a limited number of measured cycles.

**B. ELECTRIC BUS CONSUMPTION ANALYSES**

A limited number of studies have been conducted that have included analyzing the energy consumption of BEBs with simulation models. Lajunen examined the energy consumption and conducted a cost-benefit analysis for a conventional diesel bus, parallel and series hybrid buses, and a BEB [27]. Six different driving cycles were used in the simulation studies. The results showed that the cost efficiency contributions of hybrid and electric buses depend on the operation route and schedule. Additionally, the plug-in series hybrid and battery electric city buses appeared to be impacted by the driving cycle less than the other tested technologies. A study conducted by Zeng et al. similarly indicated that the energy consumption and range of a BEB are influenced by the driving cycle less than those of a traditional internal combustion engine bus [28]. The phenomenon was further confirmed in a study that was conducted with real diesel and battery electric buses in Macau [29]. Furthermore, the results of the study indicated that BEBs are not as sensitive to passenger load variations as diesel buses. The results of a simulation study showed that the influence of passenger load on the energy consumption of a BEB depends on the driving cycle and the aggressiveness of the driving maneuvers [4].

In another simulation-based study, Lajunen showed that the effect the driving cycle has on the efficiency and energy consumption of a BEB depends significantly on the powertrain configuration [30]. Halmeaho et al. found that the approach used in modeling the electric motor affects the efficiency of the BEB model considerably, with a resistance model providing more realistic results than a simple motor efficiency map based model [31].

Although different driving cycles have been used in previous BEB energy consumption studies, an exhaustive analysis of the consumption variations caused by driving cycle uncertainty and passenger variations on a single bus route has not been conducted. The research gap is addressed in this paper by employing a novel cycle synthetization algorithm to simulate a large quantity of varying cycles on a suburban bus route with a validated BEB simulation model.

**III. METHODS**

**A. DATA COLLECTION AND PROCESSING**

The suburban bus route examined in this research is line 11 in Espoo, Finland. The direction from Friisilä to Tapiola was considered, including the drive from the terminus at Tapiola to the fast-charging station nearby (Fig. 1). The total distance of the route is approximately 10.42 km. The line is a typical Finnish suburban bus line, as it commutes between smaller centers and features short urban-type sections as well as sections with slightly higher speed limits.

A battery electric bus operating on the route was equipped with an on-line data acquisition system. The speed and GPS coordinates of the bus, along with their respective timestamps, were obtained via this system. Additionally, the battery current and voltage as well as the state-of-charge (SOC) estimate were acquired for one driving cycle in order to validate the bus simulation model. The power drawn by the auxiliary devices was also measured for the validation cycle. The sampling rates of the measured signals are presented in Table I. The speed signal was obtained from the anti-lock brake system. The battery voltage and current were measured at the inverter, and the SOC estimate was acquired from the battery management system.
Passenger numbers were not tracked with the on-line data acquisition system. Hence, passenger data were collected separately from 37 runs of the route over a period from March 2016 to February 2017. The number of passengers in the bus throughout the cycle was logged manually by travelling in the bus. These measurements were also conducted only during weekdays.

The instants of time when the vehicle stopped at or passed by each bus stop during each cycle were estimated using Google Maps and the measured GPS and speed data. In the case of the bus passing by a stop, the GPS location of the stop was projected onto a line connecting the two nearest measured GPS points, allowing for the measured speed data to be used to determine the time at which the bus passed by the stop. These timestamps are needed for the synthetization process.

### B. STOCHASTIC DRIVING CYCLE SYNTHESIS

The developed driving cycle synthetization algorithm, which is an expansion of the concept presented in [32], [33], is based on dividing the bus route into segments. In this study, a segment is defined as the speed profile between two bus stops. Traffic-related stops (e.g., traffic lights and pedestrian crossings) were not included as segment start and end points in order to avoid excessively short segments. The route under examination features 25 bus stops, which is also the number of segments, as the drive from the terminus to the fast-charging station is included in the synthetization. The synthesized driving cycles consist of randomly selected segments of the measured cycles. Only the respective segments are used, i.e., for the \(i\)-th segment in the synthesized cycle, only the \(i\)-th segment in the measured cycles can be selected. A respective segment in each measured cycle is equally likely to be selected. The measured cycle used for a particular segment is referred to as base cycle. When the base cycle is changed, the two different cycles must be connected to one another. Throughout the rest of this paper, the base cycle used on the segment prior to a bus stop is referred to as pre-stop cycle and the one used after the stop as post-stop cycle.

A flow chart of the synthetization process is presented in Fig. 3. At the start of the process, the algorithm randomizes not only the base cycle for each segment but also which bus stops the vehicle will stop at and which ones it will pass by. The probability of stopping was defined individually for each bus stop based on the measured stop frequency.
When the algorithm connects the pre-stop cycle to the post-stop cycle, there are eight possible different scenarios that can occur based on three variables:

- Does the vehicle stop at this bus stop in the synthetic cycle?
- Did the vehicle stop at this bus stop in the pre-stop cycle?
- Did the vehicle stop at this bus stop in the post-stop cycle?

The possible scenarios are listed in Table III.

The synthetization algorithm was designed to ensure that all the generated driving cycles have equal total trip distances. A target distance was calculated for each of the 25 segments based on the average distances of the segments found in the measured cycles. The algorithm compares the distance of the base cycle segment to the target distance, and it then compensates for the difference. The compensation is always done in the following segment, i.e., after the bus stop in the post-stop cycle, allowing for individual segments to vary in length while keeping the overall distance of the trip constant (10421.66 m) for all synthesized cycles. If it should happen that the base cycle remains the same and either scenario 1 or 8 occurs, the distance correction is then not performed until the next bus stop (unless the same situation occurs there again) or at the very latest at the terminus. The handling of the different scenarios and the distance correction are further elaborated on in sections III.B.1 and III.B.2. The artificial driving portions were chosen to feature linear acceleration, which, despite being a simplification, was found to maintain the statistical properties of the original data set with reasonable accuracy, as will be shown in section III.B.3.

### Table III

**BASE CYCLE SWITCHING SCENARIOS.**

<table>
<thead>
<tr>
<th>Scenario no.</th>
<th>Synthesized cycle</th>
<th>Pre-stop cycle</th>
<th>Post-stop cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>stop</td>
<td>no stop</td>
<td>no stop</td>
</tr>
<tr>
<td>2</td>
<td>stop</td>
<td>stop</td>
<td>no stop</td>
</tr>
<tr>
<td>3</td>
<td>stop</td>
<td>no stop</td>
<td>stop</td>
</tr>
<tr>
<td>4</td>
<td>stop</td>
<td>stop</td>
<td>stop</td>
</tr>
<tr>
<td>5</td>
<td>no stop</td>
<td>stop</td>
<td>stop</td>
</tr>
<tr>
<td>6</td>
<td>no stop</td>
<td>stop</td>
<td>no stop</td>
</tr>
<tr>
<td>7</td>
<td>no stop</td>
<td>no stop</td>
<td>stop</td>
</tr>
<tr>
<td>8</td>
<td>no stop</td>
<td>no stop</td>
<td>no stop</td>
</tr>
</tbody>
</table>

### 1) Creating an Artificial Stop

Scenarios 1 to 4 involve creating a stop in the synthesized cycle. Fig. 4 shows the creation of a stop in scenario 1. The distance covered during the artificial deceleration is dictated by the following equation:

\[
s_{AD} = \int_{t_A}^{t_B} v_s \, dt = \int_{t_A}^{t_E} v_{pre} \, dt
\]

where \( t_i \) is the instant of time corresponding to symbol \( i \) in Fig. 4, \( v_s \) is the synthesized driving cycle, and \( v_{pre} \) is the pre-stop cycle. The distance covered during the artificial acceleration is:

\[
s_{DE} = \int_{t_D}^{t_E} v_s \, dt = \int_{t_B}^{t_C} v_{pst} \, dt + s_{cor}
\]

where \( v_{pst} \) is the speed profile of the post-stop cycle and \( s_{cor} \) the distance-to-be-corrected. The distance \( s_{cor} \) is positive if distance needs to be increased and negative if distance needs to be reduced.

**Figure 4.** Creating a stop in scenario 1. Point A is where the deceleration starts, line B represents the instant of time when the bus stop is passed in the pre-stop and post-stop cycles, C and E illustrate the point at which the synthesized cycle joins the post-stop cycle, and D indicates the stop.

The constant acceleration and deceleration values of the artificial stop maneuver are chosen randomly from appropriate distributions. These distributions were created based on the measured cycle data by calculating the average acceleration values during the 50 meters preceding and following a halt at a bus stop. The 50-meter distance value was chosen by...
comparing the overall acceleration and speed distributions of a large set of synthesized cycles to those of the measured cycles and finding the value that would minimize the differences. A constant acceleration was considered reasonable for this route, as the highest speed limit is only 60 km/h. Thus, the power demand won’t become excessive even at the highest speeds. On routes with higher speed limits, the acceleration would need to be limited at high speeds.

The fitted distributions are presented in Fig. 5. For the deceleration distribution, normal distribution was used. Generalized maximum value distribution was chosen for the acceleration distribution. These distribution types were selected using a distribution-fitting tool developed by Sheppard [34].

![Fig. 5. Bus stop acceleration and deceleration histograms and the fitted probability density functions. The left-side y-axes (blue) represent the probability density function values, and the right-side y-axes (red) represent the histogram bin percentages.](image)

The algorithm creates the deceleration the same way in scenario 3 as in scenario 1. In scenarios 2 and 4, the pre-stop cycle already features a stop, so a new deceleration maneuver does not need to be created.

The acceleration is formed in the same way in scenarios 1 and 2. In scenarios 3 and 4, the acceleration in the post-stop cycle is modified if the segment after the stop needs to be shortened in order to correct the distance, which is demonstrated in Fig. 6. Otherwise the acceleration in the post-stop cycle is used without modifications.

![Fig. 6. Modifying the acceleration in scenarios 3 and 4. F and G are the points where acceleration begins in the post-stop and synthetic cycles, and H is where the synthetic cycle is connected to the post-stop cycle.](image)

When shortening the segment after the stop as shown Fig. 6, the distance covered during the new acceleration is calculated as:

\[
s_{GH} = \int_{t_G}^{t_H} v_s \, dt = \int_{t_{pre}}^{t_f} v_{pst} \, dt + s_{cor} \tag{3}
\]

where \( s_{cor} \leq 0 \). In case the distance correction requires the distance to be increased instead, constant speed driving will then be added at the first speed peak after the stop.

The algorithm determines the durations of the synthetic stops based on the scenario:

- In scenario 1, the stop lasts 12.27 s, which was found to be the average duration of a stop in the measured cycles.
- In scenario 2, the stop duration in the pre-stop cycle determines the duration of the stop.
- In scenario 3, the stop duration in the post-stop cycle determines the duration.
- In scenario 4, the stop duration is the average of the durations of the stops in the pre-stop and post-stop cycles.

With the chosen definitions, the percentage of time spent overall at bus stops in a large number of synthesized cycles should be approximately equal to that of the measured cycles.

2) SKIPPING A STOP

In scenarios 5 to 8, the bus stop is passed without stopping. An example of scenario 5 is shown in Fig. 7. The pre-stop and post-stop cycles are connected by connecting the speed peaks nearest to the stop. The connection is formed in such a way that the distance covered in the synthesized cycle between points I and L is:

\[
s_{IL} = \int_{t_1}^{t_L} v_s \, dt = \int_{t_{pre}}^{t_f} v_{pre} \, dt + \int_{t_f}^{t_{pst}} v_{pst} \, dt + s_{cor} \tag{4}
\]

By connecting the cycles in this manner, the acceleration between points I and L can vary. The only limitation set for the acceleration is that its absolute value cannot exceed a set threshold value. If it does exceed that, the point where the synthesized cycle joins the post-stop cycle is moved forward until the acceleration is within the limits. The acceleration threshold was set to 1.1 m/s² based on the findings in [35] that the absolute value of longitudinal acceleration of a ground transport vehicle should not exceed this value in order for the ride to feel comfortable for the passengers. Furthermore, the measurements showed that this threshold was almost never exceeded outside of brief periods at low speeds during stop or start maneuvers. Hence, it was assumed that bus drivers generally tend to avoid exceeding this boundary. The acceleration threshold value is used in scenarios 5-8.
In scenario 6, depicted in Fig. 8, the synthesized cycle diverges from the pre-stop cycle at the last speed peak before the bus stop and connects to the post-stop cycle in such a way that the distance covered during the connection phase between points M and N is:

$$s_{MN} = \int_{t_M}^{t_N} v_s \, dt = \int_{t_M}^{t_P} v_{pre} \, dt,$$

and the distance between points N and O is equal to $s_{cor}$. If the absolute value of the acceleration between points M and N exceeds the aforementioned threshold value, point N is then moved further along the post-stop cycle.

In scenario 7, the synthesized cycle follows the pre-stop cycle until the point where the bus stop was passed in the pre-stop cycle. That point is then connected to the first speed peak after the stop in the post-stop cycle. The distance covered when connecting the two cycles is:

$$s_{RS} = \int_{t_R}^{t_S} v_s \, dt = \int_{t_Q}^{t_S} v_{ps} \, dt + s_{cor}. \quad (6)$$

Similar to the previous two scenarios, the connection point S is moved further along the post-stop cycle if the acceleration between points R and S is not below the threshold value.

3) CYCLE SYNTHETIZATION METHOD VALIDATION

The validity of the cycle synthetization method was examined by generating a large number of synthetic cycles and comparing their statistical properties to those of the measured cycles. Ten thousand synthetic cycles were generated. The acceleration and speed distributions of the measured and synthesized cycles are presented in Fig. 11. Acceleration values of zero are included in the (-0.2, 0.0] bin.

In the acceleration histograms, the average difference in the size of the bins between the measured cycles and generated cycles is 0.47 %-points. In the speed histograms, that value is 0.21 %-points. Other statistical properties of the measured and synthetic cycles are compared in Table IV. The statistical
The first phase is conducted once the bus stops where the passengers in the bus during a cycle in the measurements, process. Distorted to any significant extent in the cycle synthetization acceleration or deceleration mode. The comparison indicates creeping, or idling, it is then considered to be in either acceleration or deceleration mode. The comparison indicates that the statistical properties of the original data set are not distorted to any significant extent in the cycle synthetization process.

![Acceleration and speed histograms of the measured cycles and 10 000 synthetic cycles.](image)

**FIGURE 11.** Acceleration and speed histograms of the measured cycles and 10 000 synthetic cycles.

**TABLE IV**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Measured cycles</th>
<th>Synthetic cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average speed (km/h)</td>
<td>20.582</td>
<td>20.419</td>
</tr>
<tr>
<td>Avg. acceleration (m/s²)</td>
<td>0.547</td>
<td>0.541</td>
</tr>
<tr>
<td>Avg. deceleration (m/s²)</td>
<td>0.506</td>
<td>0.508</td>
</tr>
<tr>
<td>Avg. cycle duration (s)</td>
<td>1823</td>
<td>1837</td>
</tr>
<tr>
<td>Acceleration percentage</td>
<td>31.66%</td>
<td>32.39%</td>
</tr>
<tr>
<td>Deceleration percentage</td>
<td>34.15%</td>
<td>34.37%</td>
</tr>
<tr>
<td>Cruise percentage</td>
<td>15.63%</td>
<td>15.11%</td>
</tr>
<tr>
<td>Creep percentage</td>
<td>2.01%</td>
<td>1.49%</td>
</tr>
<tr>
<td>Idle percentage</td>
<td>16.55%</td>
<td>16.65%</td>
</tr>
</tbody>
</table>

### C. PASSENGER VARIATIONS

The passenger numbers are randomized in two phases:

1. Determine the overall average number of passengers in the bus during the cycle.
2. Determine the number of passengers in the bus during each segment.

The first phase is conducted once the bus stops where the vehicle will stop at during the new cycle have been randomized. The average number of passengers in the bus during the cycle can then be sampled from a conditional (bivariate) normal distribution as [38]:

\[
X_1 | X_2 = x_2 \sim N \left( \mu_1 + \frac{\sigma_1}{\sigma_2} \rho (x_2 - \mu_2), (1 - \rho^2)\sigma_1^2 \right)
\]  

where \( x_2 \) is the randomized total number of stops at bus stops during the new cycle, \( \mu_1 \) is the overall mean number of passengers in the bus during a cycle in the measurements, \( \mu_2 \) is the measured mean number of stops at bus stops during a cycle, \( \sigma_1 \) is the standard deviation of the measured average passenger numbers, \( \sigma_2 \) is the measured standard deviation of the total number of stops at bus stops during a cycle, and \( \rho \) is the Pearson correlation coefficient between the measured average number of passengers and stops performed at bus stops during a cycle.

After the average number of passengers during the whole cycle has been determined, the passenger numbers on the individual segments are then sampled from a separate multivariate normal distribution. The mean number of passengers in the bus on each segment and a covariance matrix were calculated based on the measurements. The covariance matrix was established as:

\[
\Sigma = \text{cov} \left( \begin{bmatrix} n_{1,1} & \cdots & n_{1,M-1} \\ \vdots & \ddots & \vdots \\ n_{N,1} & \cdots & n_{N,M-1} \end{bmatrix} \right)
\]  

where \( N \) is the number of cycles, \( M \) is the number of segments, and \( n_{i,j} \) is the number of passengers in the bus on the \( j \)-th segment in the \( i \)-th cycle. The last segment was omitted from the matrix, as there are always zero passengers in the bus on the drive from the terminus to the fast-charging station.

Hence, passenger numbers can be sampled from a distribution defined by the vector of mean values and the covariance matrix. 100000 passenger number sequences would be sampled. Then, the sequence with the overall average value closest to the value previously determined in the first phase with the bivariate distribution would be chosen. By modelling the passenger numbers in these two phases, the correlation between the number of stops and the average passenger load could be modelled as well as the correlations of the passenger numbers on the different segments.

The use of multivariate normal distribution in the second phase required the sampled passenger numbers to be rounded to the nearest integer. Limitations had to be set for the sampled passenger numbers; namely, the number of passengers in the bus could not:

- be negative,
- exceed a defined limit, or
- change between segments in case the bus would not stop at the bus stop between the segments.

When 10000 cycles were synthesized, the mean number of passengers in the bus during the synthetic cycles was only 0.10 lower than in the measured cycles. In the synthetic cycles, the correlation coefficient between the number of stops at bus stops and the average number of passengers in the bus during a cycle was 0.56 while in the measurements it was 0.60. These results show that the developed passenger generation algorithm produces realistic passenger numbers that correlate with the number of stops at bus stops in a realistic manner. An
example of a synthetic cycle with passenger numbers is shown in Fig. 12.

The mass of a passenger was defined as 70 kg. As the maximum allowed payload of the bus was known to be 5500 kg [39], the maximum passenger limit was set to 78. The limit was not reached in the simulated runs. The highest number of passengers in the bus at any time during the 10000 synthesized cycles was 38. This is reflective of the fact that the examined bus line is not particularly crowded. In the 37 measured cycles, the highest number of passengers observed in the bus was 29. The significant difference between the maximum passenger numbers was simply due to the low number of measured samples and high number of synthetic samples.

![FIGURE 12. Example of a synthetic Espoo bus line 11 driving cycle.](image)

D. SIMULATION MODEL

The simulation model used in this study was presented in [40]. A flow chart of the model is shown in Fig. 13. The virtual driver model was altered so that it converts the driving cycle from the time vs speed format to distance vs speed. Using a distance-based driving cycle ensures that the bus always travels the correct distance during the cycle. Performing stop maneuvers correctly is ensured by switching to the equivalent point in the time-based cycle when the bus is close to stopping. Once the bus starts moving again, the driver model switches back to the distance-based cycle.

![FIGURE 13. Flow chart of the simulation model [40].](image)

The simulation model was validated to match the city bus equipped with the data acquisition system. As the model had originally been validated for a different bus, many parameters needed to be reconfigured. The nominal parameters of the simulation model are presented in Table V. The battery pack model was changed from the LiFePO4 model in the original version to an LTO battery model. The auxiliary device power demand was defined as the average demand measured during the validation cycle. However, it must be noted that the bus is heated with a separate diesel heater, and the power demand of the heater was not measured. Hence, the heater power was excluded from the simulation model.

The model was validated by comparing the simulated SOC, current, and voltage of the battery to those measured with the on-line data acquisition system during a cycle (Fig. 14). The average ambient temperature during the validation cycle was -6 °C. The SOC change during the actual cycle was 19.8 %. In the simulation, the change was 19.77 %. The root-mean-square (RMS) error of the simulated voltage was 3.497 V, and the RMS error of the simulated current was 21.89 A.

![FIGURE 14. Speed, SOC, battery voltage, and battery current during the simulated and actual validation cycle.](image)

<table>
<thead>
<tr>
<th>Table V: The Nominal Parameters of the Simulation Model.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>General</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Battery pack</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Electric motor</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
IV. RESULTS AND DISCUSSION

An energy consumption distribution was acquired for the route by running 10000 synthesized cycles with the simulation model and collecting the average consumption during each cycle into a histogram (Fig. 15). The statistical parameters of the acquired energy consumption distribution are listed in Table VI.

![Figure 15. Energy consumption distribution for 10000 simulated synthetic driving cycles.](image)

<table>
<thead>
<tr>
<th>Energy consumption parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.9139 kWh/km</td>
</tr>
<tr>
<td>Min. (difference to the mean in %)</td>
<td>0.7473 kWh/km (-18.3 %)</td>
</tr>
<tr>
<td>Max. (difference to the mean in %)</td>
<td>1.0783 kWh/km (+17.9 %)</td>
</tr>
<tr>
<td>Difference between min. and max.</td>
<td>0.3311 kWh/km</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.0431 kWh/km</td>
</tr>
</tbody>
</table>

The mean energy consumption during a cycle was similar to consumption values that have been presented for BEBs in other recent papers [27], [29]. The results show that the confidence interval defined by the standard deviation was equal to 9.4 % of the mean consumption. The total range of the energy consumption was equal to 36.2 % of the mean.

In order to assess the amount of simulated cycles needed to acquire a reliable energy consumption distribution with the developed method, the overall average energy consumption per cycle was plotted as a function of the number of simulated cycles (Fig. 16). The mean and standard deviation with varying numbers of cycles can be found in Table VII.

![Figure 16. Average energy consumption per cycle versus number of simulated cycles.](image)

<table>
<thead>
<tr>
<th>Num. of cycles</th>
<th>Mean (kWh/km)</th>
<th>Std. dev. (kWh/km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.9239</td>
<td>0.0457</td>
</tr>
<tr>
<td>500</td>
<td>0.9194</td>
<td>0.0427</td>
</tr>
<tr>
<td>1000</td>
<td>0.9159</td>
<td>0.0423</td>
</tr>
<tr>
<td>2000</td>
<td>0.9142</td>
<td>0.0429</td>
</tr>
<tr>
<td>5000</td>
<td>0.9146</td>
<td>0.0425</td>
</tr>
</tbody>
</table>

Table VII shows that both the mean and standard deviation of the consumption were within 1 % of the final values after only 500 cycles. After 2000 cycles, the mean consumption differed from the final value by only 0.03 %.

In order to evaluate which driving cycle properties had the most influence on the energy consumption, correlations between various cycle parameters and energy consumption were assessed with the Pearson correlation coefficient (Table VIII).

![Table VI. Energy consumption results of the simulation studies.](image)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Pearson correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of stops</td>
<td>0.8053</td>
</tr>
<tr>
<td>Average speed</td>
<td>-0.7525</td>
</tr>
<tr>
<td>Number of stops at bus stops</td>
<td>0.7059</td>
</tr>
<tr>
<td>Cruise percentage</td>
<td>-0.6417</td>
</tr>
<tr>
<td>Average driving speed</td>
<td>-0.6271</td>
</tr>
<tr>
<td>Average passengers</td>
<td>0.6120</td>
</tr>
<tr>
<td>Average deceleration</td>
<td>-0.5954</td>
</tr>
<tr>
<td>Idle percentage</td>
<td>0.5658</td>
</tr>
<tr>
<td>Average acceleration</td>
<td>0.5574</td>
</tr>
<tr>
<td>Creep percentage</td>
<td>0.1720</td>
</tr>
<tr>
<td>Deceleration percentage</td>
<td>-0.1541</td>
</tr>
<tr>
<td>Acceleration percentage</td>
<td>-0.0560</td>
</tr>
</tbody>
</table>

The Pearson coefficients show that the number of stops made during a cycle had the strongest linear correlation with the energy consumption. The average speed also had a high degree of correlation with the consumption. It should be noted that the number of stops strongly influenced most of the other parameters: the absolute values of the correlation coefficients between the number of stops and the other cycle parameters were 0.4 or higher except for average passengers, average deceleration, as well as the percentages of acceleration, deceleration, and creep. Hence, the number of stops can be considered clearly the most influential cycle parameter on the energy consumption of the ones listed in the table. This result is in line with the findings of previous studies that have shown the frequency of stops to be one of the most significant factors in the energy consumption of buses [29], [41], [42]. The energy consumption influence of the number of stops and the average number of passengers was further examined.

The energy consumption as a function of stops per kilometer is presented in Fig. 17. The lowest number of stops made during a cycle was 9 and the highest was 38. A linear regression model was fitted to the data using the least squares method.
The linear regression model shows that an increase of one stop per kilometer increased the energy consumption by approximately 0.104 kWh/km. Stop maneuvers increase consumption because energy is wasted during the accelerations and the auxiliary devices consume energy while the bus is idling. Stops performed at bus stops and traffic stops were observed to differ in terms of impact on energy consumption. Linear regression analysis showed that one stop at a bus stop increased consumption by 0.14 kWh and one traffic-related stop caused an increase of 0.08 kWh. The difference in the influence of the different types of stops was likely caused by the correlation between the number of stops at bus stops and the average passenger load.

The energy consumption as a function of the average number of passengers in the bus during the cycle is presented in Fig. 18. A linear regression model was again fitted to the result data with the least squares method. Linear regression was used so that the influence of passengers and stops on the consumption could be assessed using the slope of the linear model.

According to the regression model, a one passenger increase in the average number of passengers increased the consumption by approximately 0.013 kWh/km.

The energy consumption effects of the number of stops and the average number of passengers were compared using the Global Sensitivity Analysis (GSA) method based on the Sobol approach, which was implemented in MATLAB in [43]. The GSA results indicated that the variation in the number of stops contributed to the variation in the energy consumption approximately 3.26 times as much as the variation in the average number of passengers did. One reason why the passenger load had a lesser effect on the energy consumption variations is that the passenger load was relatively low. The highest average number of passengers in the bus on any cycle was 14.33, and the mean value was 5.78. On a more crowded bus line, the passenger load would likely contribute more to the energy consumption variations. Although the results presented here are case-specific, they do demonstrate clearly that even with BEBs, which are capable of kinetic energy recovery, the energy consumption is significantly influenced by the stop maneuvers.

Certain limitations in this study limited the scope of the results. Firstly, driving cycles were only recorded during winter months. Consequently, the synthesized cycles only represented driving behavior during that time of year. It is unknown how much the driving cycles would differ during different times of year on this route. The simulation model was also validated only for a specific ambient temperature. Thus, the acquired energy consumption distribution can only be considered valid for winter months and for that specific ambient temperature. The results are likely representative for suburban routes with similar average speeds and stop intervals, but separate analyses should be conducted for highly congested urban routes as well as higher speed highway routes in order to examine how much the results differ.

Furthermore, the energy demand of the diesel heater was not included in the simulations due to the unavailability of measurement data from the heater. Using linear accelerations in the artificial driving portions of the synthetic cycles may also have had some influence on the results. However, the influence is assumed to be marginal, as the statistical properties of the measured cycles were shown to not become significantly distorted in the synthetization process. Additionally, traffic was not explicitly modeled in the synthetization process although the effects of different traffic conditions are inherently included in the measured cycles. Due to the random selection of the measured cycles for each segment in the synthesized cycles, segments recorded during different times of day could be mixed in the synthetization process. However, this was not considered an issue because higher traffic hours were found to make no significant difference to the driving maneuvers in the measured cycles.

This was assumed to be due to special bus lanes, buses having priority in traffic lights, and a general lack of significant congestion on the route even during the busiest hours. On routes where the rush hours have a significant impact on the driving maneuvers, cycles could be synthesized separately with rush hour and non-rush hour measurement data.

V. CONCLUSIONS

A novel segment-based method for generating a large number of varying synthetic driving cycles and passenger numbers for a specific bus route was developed in this study. The validity of the method was tested by comparing the statistical properties of generated cycles to those of the original measured cycles. The comparison showed that the method maintains the properties of the measured cycles with good
accuracy. The synthetization method was used to analyze the energy consumption variations inflicted by driving cycle variations and passenger load uncertainty on a suburban bus route in Espoo, Finland by running 10000 synthetic cycles with a validated BEB simulation model.

The energy consumption during each simulated cycle was collected into a histogram. The shape of the histogram closely resembled a normal distribution with a mean value of 0.914 kWh/km and a standard deviation of 0.043 kWh/km. The range of the energy consumption variation was equal to 36 % of the average consumption during a cycle. The results indicated that a reliable energy consumption distribution forecast could be acquired with only 2000 simulated synthetic cycles. Further examination of the results showed that the energy consumption correlated particularly strongly with the number of stops performed during the cycle. The passenger load variations had a lesser impact on the energy consumption variation, but the effect was still clearly recognizable. It seems likely that the passenger load would have a more significant impact on a busier bus line, as the number of passengers in the bus on the examined line tends to be low.

The type of energy consumption analysis presented in this paper could prove advantageous in the design process of BEBs. The powertrain design could be based upon a simulated energy consumption forecast if the route the bus will operate on is known, allowing the components to be scaled more appropriately. However, consideration should nonetheless be given to worst-case scenarios in order to avoid making inflexible designs. Additionally, this type of energy consumption forecasting could be used in the case of designing range-extender equipped electric buses. The forecast could be used such that a desired percentage of runs of a route could be driven with pure electric power, with the range-extender being deployed only in the highest-consuming cases. Being able to scale the battery more accurately for a given route would be particularly beneficial, as oversizing batteries contributes significantly to electric vehicle prices. Moreover, energy consumption forecasting could be used when designing timetables for bus lines operated with BEBs, as the energy consumption influences the amount of time required for charging. Hence, the new method could provide substantial benefits to the PTA, route operators, bus manufacturers, and ultimately to public transport users in the form of improved cost efficiency and more reliable schedules.

The methods presented in this paper could be enhanced by incorporating other uncertainties such as weather condition variations into the simulations. This would further improve the accuracy of the energy consumption distribution acquired from the simulated synthetic driving cycles. Additionally, analysis could be conducted for various different bus routes in order to examine how dependent the acquired results are on the bus route characteristics.

ACKNOWLEDGMENT

The authors are grateful to Linkker Ltd and Sami Ruotsalainen for providing support and data for this study.

REFERENCES
