Titre: Title:	A Hybrid Machine Learning and Physics-based Approach for Accurate Energy Consumption Modeling of Electric Buses in Public Transport
Auteurs: Authors:	Lucas Franck Frederic Adam, Robert Pellerin, & Bruno Agard
Date:	2025
Type:	Article de revue / Article
Référence: Citation:	Adam, L. F. F., Pellerin, R., & Agard, B. (2025). A Hybrid Machine Learning and Physics-based Approach for Accurate Energy Consumption Modeling of Electric Buses in Public Transport. Sensors & Transducers, 268(1), 45-58. https://www.proquest.com/scholarly-journals/hybrid-machine-learning-physics-based-approach/docview/3212840245/se-2?accountid=40695

Titre de la revue: Journal Title:	Sensors & Transducers (vol. 268, no. 1)
Maison d'édition: Publisher:	International Frequency Sensor Association
URL officiel: Official URL:	https://www.proquest.com/scholarly-journals/hybrid-machine-learning-physics-based-approach/docview/3212840245/se-2?accountid=40695

Sensors & Transducers, Vol. 268, Issue 1, April 2025, pp. 45-58

Sensors & Transducers
Published by IFSA Publishing, S. L., 2025
http://www.sensorsportal.com

A Hybrid Machine Learning and Physics-based Approach for Accurate Energy Consumption Modeling of Electric Buses in Public Transport

[1,2,3] * Lucas ADAM, [1,3,4] Robert PELLERIN, and [1,2,3] Bruno AGARD
[1] Polytechnique Montréal - Department of Mathematical and Industrial Engineering, Canada
[2] LID, Data Intelligence Laboratory, Polytechnique Montréal, Canada
[3] CIRRELT - Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation, Canada
[4] Jarislowsky/AtkinsRéalis Research Chair in the Management of International Projects, Canada
[*] E-mail: lucas.adam@polymtl.ca

Received: 12 Nov. 2025 /Revised:31March 2025 /Accepted:15 April 2025 /Published:30 April 2025

Abstract: Public transport organizations are increasingly concerned about reducing air pollution, leading many to transition their fleets into electric vehicles (EVs). In this context, limited battery range and charging times remain significant hurdles. Precise modeling of electric bus energy consumption is crucial. Still, existing methods often face difficulties due to the complexities of real-world conditions, such as diverse driving patterns and external factors. To tackle this, the study proposes a hybrid model combining physical principles and machine learning using real-world data from 30 buses across 130 routes over one year. Key variables like passenger load, weather, and route characteristics are incorporated. Several machine learning models, including MLP, KAN, and XGBoost, are compared using Mean Absolute Percentage Error (MAPE). The hybrid model outperforms others, achieving a low MAPE of 5.59 % on test data and 5.79 % on validation data with a low Standard Deviation. Additionally, models incorporating operational factors, such as bus lines and time of day, enhance prediction accuracy. The study concludes that integrating physical laws with machine learning offers a more accurate and stable approach to energy consumption modeling, providing a promising framework for fleet management and energy efficiency in public transport systems.

Keywords: Public transport, Electric bus, Telemetry, Energy consumption, Big data, Artificial intelligence.

1. Introduction

As consciousness arises about the role of road transport in air pollution, the global trend points towards the use of more energy-efficient vehicles and the reduction of greenhouse gas emissions. According to Krawiec et al. [2], this affects not only personal vehicles but also public transport vehicles.

The share of electric-powered buses operated by public transport organizations is expected to reach 100 % by 2050. Despite the significant advantages associated with electric vehicles, such as zero tailpipe emissions [3] and reduced noise pollution, which contributes to improved overall life quality in urban areas [4], these innovative solutions come with their challenges [3]. Some of the most critical issues are their limited battery range and longer charging times [5]. This highlights the crucial need for accurately modeling discharge patterns in electric vehicles [1].

While many researchers have proposed approaches based on simulated data [5], these methods often fall short in real-world applications, where "energy consumption prediction is challenging due to complicated driving cycles and a wide range of influential factors" [6]. Modeling electric energy consumption remains a complex challenge [7].

Numerous approaches have been proposed in the scientific literature, each with distinct advantages and limitations. Rule-based models use fundamental physical principles to simulate vehicle-environment interactions, incorporating factors such as rolling resistance, aerodynamic drag, road gradient, and acceleration forces. These models offer a mechanistic understanding of energy consumption. In contrast, data-driven models uncover hidden relationships between variables using historical data, offering a more empirical approach [6].

This article proposes a different approach by combining physical and machine learning approaches to predict the required energy consumption of electric buses per trip. As such, the paper's main objective is to provide a method to incorporate telemetry and planning data into prediction models that can make energy predictions at fleet scale, considering the specifics of multiple bus lines and vehicles.

The remainder of the paper is structured as follows. First, the following section presents the literature review. The proposed models and the data used in this study are then presented. The performance of the proposed models is then evaluated, followed by an analysis of variable influences. Finally, limitations and future works are discussed in the discussion and conclusion sections.

2. Literature Review

The literature identifies a broad range of parameters influencing energy consumption in electric vehicles. Physical factors such as speed, acceleration, weather conditions, ambient temperature, vehicle mass, and road slope are frequently cited as key determinants [7]. Data-driven models, however, expand upon these by incorporating additional variables, including stop frequency, traffic conditions, trip timing, and in-vehicle climate control usage, offering a more nuanced understanding of real-world energy consumption [7]. These models can be categorized into three main types [8].

The first category includes statistical models, such as linear regression and ordinary least squares. While effective for small datasets, they assume linear relationships that may not fully capture complex dependencies. The second category consists of machine learning models, including support vector regression, XGBoost, and Random Forest. These models handle intricate, non-linear relationships and automatically identify relevant features from large datasets. However, they require extensive feature engineering and struggle with highly unstructured data. The third category comprises neural network models, such as deep neural networks, Long Short-Term Memory (LSTM) networks, and Transformers. These excel at modeling time-series data and capturing long-term dependencies, but they demand significant computational resources and large datasets for effective training.

Several studies advocate for hybrid models combining physical and data-driven approaches to leverage the strengths of both paradigms [8]. By integrating physical laws with machine learning insights, such models enhance accuracy and generalization. Despite the growing interest in electric vehicle energy modeling, relatively few studies focus specifically on electric buses [2]. Most recent work, particularly those utilizing real-world datasets, centers on passenger cars [9]. While valuable, these approaches often fail to account for the unique challenges of public transport modeling, such as cyclical usage patterns and significant variations in vehicle mass due to passenger loading [10].

Beckers et al. [11] proposed a physical modeling approach to quantify the impact of friction forces due to cornering on energy consumption. They found that cornering losses account for approximately 3 % of propulsion system energy consumption. While this study offers valuable insights into specific aspects of energy loss, such as friction during turns, it does not address broader operational variables like passenger load and weather conditions, which can also significantly affect energy consumption.

In contrast, Fiori et al. [12] adopt a more comprehensive decomposition of power consumption, separating it into two components: motion-related and auxiliary consumption. The motion-related component, modeled using physical principles and calibrated with data from 435 trips, is validated with an impressive RMSPE of 2.56 %. In addition, the authors also consider auxiliary consumption, such as HVAC systems, by modeling it as a polynomial function of outdoor temperature. While this approach is more holistic, accounting for factors like HVAC energy consumption, it is limited by the need for separate calibration per vehicle and possibly per line, which makes it impractical for large-scale, real-time operational use.

Furthermore, Beckers et al. [13] take a similar physics-based approach to model energy consumption, focusing on factors like motor efficiency and rolling resistance based on road surface conditions. This study also emphasizes calibration through real-world testing campaigns. However, it struggles with incorporating regenerative braking effects, a critical aspect of modern electric vehicles that could influence energy recovery and overall efficiency. This limitation further highlights a common challenge across physics-based models. While they provide valuable insights into specific components of energy use, they often fail to capture the full complexity of vehicle operation, particularly in terms of energy recovery mechanisms like regenerative braking.

Blades et al. [14] addressed the challenge of generating realistic drive cycles, which is often limited by the cost and complexity of real-world data collection. By developing a generic Battery Electric Bus (BEB) model and applying machine learning techniques such as linear regression, LASSO, ridge regression, Random Forest, and Extra Trees, they identified the average road gradient and speed as the most influential factors affecting energy consumption. The study reported MAPE errors ranging from 2.1 % to 10.67 % and recommended including additional factors such as passenger load and weather conditions – factors that are explicitly considered in our own approach. This aligns with the growing recognition that contextual factors beyond the road and vehicle characteristics play a crucial role in predicting energy consumption.

Similarly, Basso et al. [15] also utilized machine learning techniques to predict energy consumption, focusing on multiple buses and a range of time aggregation levels (5, 10, 15, and 30 minutes). Their use of 30 variables, including a physics-based consumption estimate, and their application of models like linear regression, Random Forest, SVR, and neural networks resulted in a best-performing model (linear kernel SVR) with a MAPE of 22 %, a significant improvement over the physics-based model's MAPE of 43\%. The study underscored the importance of long-term telemetry data and multi-route contextual variables, which are central to our research. While their approach was more data-intensive, it highlights the necessity of incorporating diverse contextual factors to improve prediction accuracy.

Meanwhile, Pamula and Pamula [16] took a deeper dive into deep learning approaches, specifically comparing LSTM, Deep Learning with Autoencoder (DLNA), and Multilayer Perceptron (MLP) models on two bus routes. They found that DLNA outperformed the other models with a MAPE of 6.2 %, followed closely by MLP (6.7 %) and LSTM (7.2 %). Despite deep learning's promise in sequential energy prediction, their study was constrained by a limited sample size and a narrow range of routes.

Xu et al. [17] partly tackled this issue by analyzing a larger dataset of a year's worth of data from ten buses operating on a single route. They compared Elastic Net, GPR, and LightGBM models, with the best-performing models achieving MAPEs of 9.17 %, 7.54 %, and 7.71 %, respectively. Elastic Net was noted for its generalizability, and the study emphasized the significant impact of external temperature, driver behavior, vehicle characteristics, and road attributes on energy consumption. These findings are consistent with the notion that operational factors beyond the vehicle itself – such as weather conditions and driver behavior – are essential to understanding and predicting energy use. However, the study also pointed out the frequent omission of passenger load variation.

Finally, Sennefelder et al. [18] analyzed driver behavior across 30 diesel buses and extrapolated the findings to electric buses. After filtering the data, they were left with 149 trips from 24 vehicles, highlighting the inherent challenges of real-world data collection. Using the Neighborhood Component Feature Selection (NCFS) algorithm, they classified energy consumption into three levels and found that Random Forest achieved the highest accuracy at 83 %. This study reinforced the complexities of real-world data acquisition and the importance of robust filtering techniques when dealing with large and noisy datasets.

By comparing these studies, we can see the evolution of data-driven approaches in predicting energy consumption for electric buses. Early studies, such as Pamula and Pamula [16], focused more on identifying key influencing factors. As the field progressed, studies like Basso et al. [15] and Blades et al. [14] incorporated more complex models and larger datasets, addressing some of the limitations of earlier approaches.

Table 1 summarizes the key variables used in different studies related to bus energy consumption prediction. It highlights the specific factors considered by the key studies we have identified. By examining these variables, we can better understand the scope and limitations of previous research and refine our approach to predicting bus energy consumption. Table 2 summarizes the key characteristics of the highlighted studies on energy consumption prediction for electric buses. This table highlights the context variables such as the number of vehicles, model types, temperature ranges, data types, and study durations. By comparing these attributes, we gain insight into the diverse approaches and setups used in previous work and identify potential gaps or areas for further exploration in our research.

The limited focus on electric buses is often attributed to challenges in data availability, as proprietary datasets or small sample sizes (such as [12, 13, 15]) hinder generalizable conclusions [19]. Many studies analyze a restricted number of bus routes [12, 15, 16] or small fleets [11-15], limiting the applicability of findings. This limitation in route diversity is a challenge we also aim to address in our work, as broader route diversity can significantly affect model generalization and robustness.

To the best of our knowledge, existing models in the literature do not directly incorporate variables such as vehicle number, route number, and direction. We believe that including these variables could significantly enhance machine learning models by helping to identify cyclical patterns or consumption behaviors that vary by bus, ultimately improving predictions. Furthermore, these pieces of information are easily accessible from scheduling software, allowing for seamless integration by transport companies.

3. Proposed Model

To overcome the limitations mentioned, we seek to leverage an extensive real-world dataset from 30 buses operating on over 130 routes over an entire year, enabling us to identify seasonal influences on energy consumption predictions. Our approach incorporates a diverse range of influential variables – including, speed, acceleration, passenger load, temperature, vehicle identification, and route attributes – offering a robust and practical framework for energy consumption prediction in electric bus fleets. Additionally, our industrial partnership provides access to buses equipped with advanced telemetry systems, ensuring reliable data acquisition. This collaboration allows us to address key gaps in the literature and refine our understanding of electric bus energy consumption.

Table 1. Summary of Variables Used in Previous Studies.

	Physical Model			**Hybrid**	**Machine Learning**			
Article	[13]	[11]	[12]	[15]	[14]	[16]	[17]	[18]
Speed	×	×	×	×	×		×	×
Acceleration	×	×	×	×	×		×	×
Passenger Load		×	×	×	×			
Temperature			×	×			×	×
Slope	×			×	×			
Weather Conditions						×		
Distance			×	×	×	×		
Travel Time					×	×		
Number of Stops					×			
Elevation					×	×	×	
Time of the Day					×		×	
Trip Number			×		×			
Day of Week							×	
Cornering Angle	×							

Table 2. Summary of Study General Context. Phy.: Physical model; ML: Machine Learning model; Aux.: Auxiliary model; Sim.: Simulated.

Article	[13]	[11]	[12]	[15]	[14]	[16]	[17]	[18]
Number of vehicles	1	1	1	6	1		10	30
Number of lines			1	1		2	1	
Model Type	Phy.	Phy.	Phy. and Aux.	Hybrid	ML.	ML.	ML.	ML.
Max Temperature (°C)			17	17	20	24		
Min Temperature (°C)			-15	11	1	-5		
Data Type	Real	Real	Real	Real	Sim.	Real	Real	Sim.
Study Duration	1 trip	62 hours	2 months	1 day		9 months	1 year	11 days
City			Howald	Santiago		Jaworzno	Beijing	Seville
Country			Luxembourg	Chile		Poland	China	Spain

We first introduce the physical model found in the literature. Then we detail how some forces will be grouped to create consumption variables for the hybrid model. Finally, we detail which machine learning approaches will be used to create the hybrid models that will be compared in the next sections.

The literature categorizes as "Rule-based" [19] the model used to assess energy consumption defined by equations (1) to (6):

$$P_{\text{total}} = \frac{k v_t}{\eta_m \eta_t}\left(F_{rr} + F_d + F_g + F_a\right), \tag{1}$$

$$F_{rr} = C_r m g \cos \alpha_t, \tag{2}$$

$$F_d = \frac{\rho_a}{2} C_d A_f v_t^2, \tag{3}$$

$$F_g = m g \sin \alpha_t, \tag{4}$$

$$F_a = m \delta a_t, \tag{5}$$

$$E_{\text{battery}} = \int_t^T P_{\text{total}}, \tag{6}$$

where P_{total} is the power provided by the battery at time t. F_{rr}, F_d, F_g and F_a are respectively forces due to rolling resistance, drag, road gradient, and acceleration. v_t, a_t and α_t are the velocity, acceleration, and road slope at time t. The constants used are C_r and C_d, rolling resistance and drag coefficients, ρ_a air density, A_f the bus front surface, m vehicle mass, g gravitational acceleration, rotational inertia, η_m and η_t motor and transmission efficiencies, whose values are derived from [15]. The regeneration factor k is varying from 0 to values close to 1 based on the technology used. Integrating those elements for a trip leads to $E_{battery}$, the total energy consumption during the considered trip.

This model has several possible variations, depending on the available data. The details of these physical model variants are summarized in Table 3, while the architecture of each model is illustrated in Fig. 1. Model A is used when the data of speed, load, slope, Heating, Ventilation, and Air Conditioning (HVAC) power are available and also when the driver utilizes regenerative braking. Model B applies when

HVAC data is unavailable. Model C is used when energy generation is deactivated due to unsafe driving conditions, such as slippery road surfaces during freezing or snowy weather. Model D is employed when only the speed profile is available.

Table 3. Summary of the different models and their associated parameters.

Model	Available Data	Load	Slope
A	Speed, Acceleration, Slope, Load, HVAC	Variable	Variable
B	Speed, Acceleration, Slope, Load	Variable	Variable
C	Speed, Acceleration, Slope, Load	Variable	Variable
D	Speed, Acceleration	Empty Weight	0

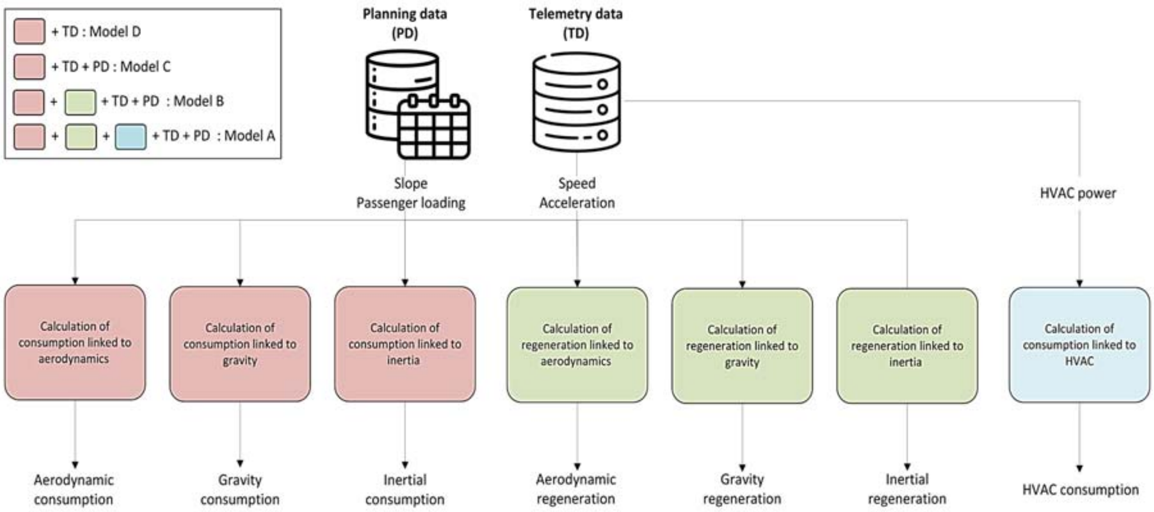

Fig. 1. Architecture of physical models (A, B, C, D).

We propose addressing the limitation of conventional physical models with a physical model coupled with a machine learning model. The physical model provides the consumption linked to aerodynamics, gravity, and inertia as well as their regeneration contributions. Instead of adding the energy of every force, we split them into three power groups described by equations 7 to 9. To calculate the energy consumption of a contribution (gravity, aerodynamic or inertia) during the trip, equation 10 is used. If the battery power is positive, then k = 1 and consumption contributions are calculated, if the battery power is negative, k = 0.3 and regenerative contributions are calculated.

$$P_{gravity} = \frac{kv_t}{\eta_m \eta_t}(C_r mg \cos\alpha_t + \\ +mg \sin\alpha_t), \tag{7}$$

$$P_{aerodynamic} = \frac{kv_t}{\eta_m \eta_t}\left(\frac{\rho_a}{2} C_d A_f v_t^2\right), \tag{8}$$

$$P_{inertia} = \frac{kv_t}{\eta_m \eta_t}(m\delta a_t), \tag{9}$$

$$E_{\text{contribution}} = \int_t^T P_{\text{contribution}} \tag{10}$$

From a physical point of view, the rolling resistance and aerodynamic drag cannot regenerate energy. In our case, "aerodynamic regeneration" or "gravity regeneration" is their consumption during the regeneration phase. After multiple tests, constructing these variables allows the following machine learning to better predict the energy consumption than if each force and its related consumption or regeneration is considered independently.

Based on the literature review, we identified that the Multi-Layer Perceptron (MLP) is a common machine learning approach, so we chose to compare its performance. Since its release in April 2024, the article by Liu et al. [21] has sparked the interest of many machine learning researchers due to its potential to replace architectures using dense layers from MLPs. Also, ensemble methods like XGBoost appeared promising for their ability to represent complex relationships between data [8].

In the first model, M1, the energy contributions are used directly as inputs to a Multilayer Perceptron (MLP). In the second model M2, the vehicle identification number, bus line number, direction, day of the week, time of the day, and trip distance are added to provide for each prediction and adjust the forecast accordingly. The M3 model uses the variables used by M2 with two new variables, average energy and average speed. The M4 Model uses a Kolmogorov-Arnold Network (KAN) while the M5 model uses an XGBoost model with the same variables as M3. The details of those hybrid models are given in Table 4. The overall process is illustrated in Fig. 2. We will investigate the influence of the data available on the accuracy of prediction using these model variants. Less data available is expected to lead to worse performance of the associated predictions.

Table 4. Summary of the different models and their associated parameters.

Model	Model Architecture	Context
A, B, C, D	Physical	None
M1	Physical followed by MLP	None
M2	Physical followed by MLP	Vehicle identification number, Bus line number, Direction, Day of the week, Time of the day, Trip distance
M3	Physical followed by MLP	Vehicle identification number, Bus line number, Direction, Day of the week, Time of the day, Trip distance, Mean Temperature, Mean Speed
M4	Physical followed by KAN	Vehicle identification number, Bus line number, Direction, Day of the week, Time of the day, Trip distance, Mean Temperature, Mean Speed
M5	Physical followed by XGBoost	Vehicle identification number, Bus line number, Direction, Day of the week, Time of the day, Trip distance, Mean Temperature, Mean Speed

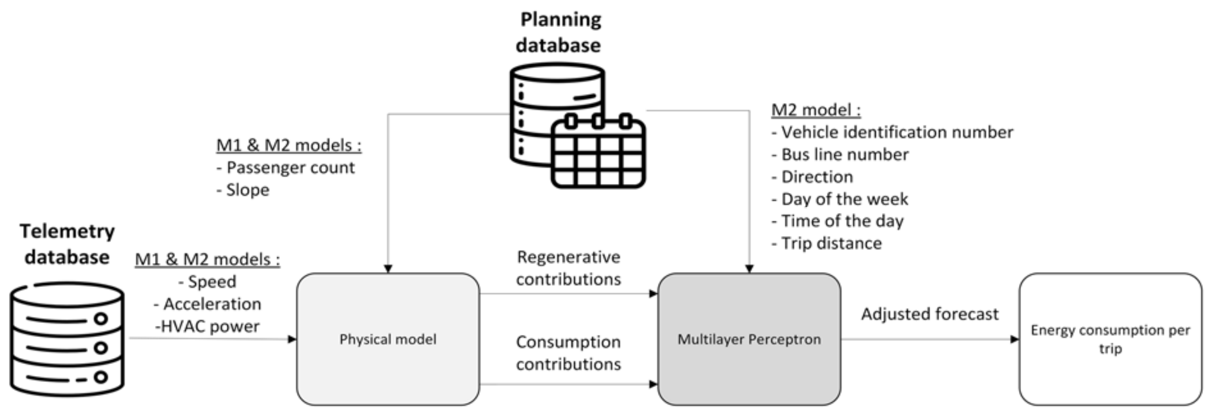

Fig. 2. Architecture of the proposed models (M1 and M2).

4. Data Description

As previously mentioned, the data used for this study comes from the operational databases of a transport company. These data are not derived from simulations or controlled test campaigns but from the real-world use of these vehicles. This case study encompasses nearly 100000 trips across 30 electric buses, covering over 25000 work cycles on nearly 130 different bus routes over one year. Each work cycle consists of multiple trips on various lines.

The telemetry system implemented is designed to monitor vital signs such as speed, heating power, and the number of active battery packs. It is primarily used to send alerts and error messages for maintenance purposes. Additionally, variables such as acceleration, slope, or passenger count are not directly measured in vehicles but must be reconstructed by merging data from different sources. A detailed list of the variables considered is provided in Table 5.

We have chosen to conduct an initial exploratory study to highlight the specific characteristics of this application case. The driving speed varies from 0 to 90 km/h, which is quite common in a large city, with the higher end occurring when the bus drives on interurban portions. The mean acceleration of 0.19 m/s² corresponds to a very gentle acceleration, which maximizes passenger comfort. The higher value of 0.77 m/s² is more noticeable for passengers but remains safe for carrying standing passengers. The slope mean is approximately 0, highlighting the cyclical behaviors of most trips. When a bus faces a positive slope during its trip, it also faces a downward slope later. The maximum and mean values of 1.5 and -1.47 rad, respectively, are unrealistic, representing slopes close to 90 degrees. After further investigation of the values, we found that the altitude gathered from GPS positioning is inaccurate for public uses, as it does not have high spatial resolution. This is the kind of noise our models must correct to provide the best possible forecasts.

This distance varies from 1 km to 39 km, with a mean value of 9.7 km. This highlights the significant variability of the routes we are trying to forecast energy consumption. It shows that when real operational data is used, it is often far from the simulations and idealized cases used in previous literature studies, justifying further studies on real-world datasets. The temperature varies from -14 °C to 44 °C, with a mean of 15 °C, highlighting the various environmental conditions these buses must face. It also covers a broader range of temperatures than most literature studies. It is worth noting that these values are subject to inconsistencies as they are measured directly by bus sensors. These measurements might be higher in winter than the actual exterior temperature because the vehicle temperature heats the sensor due to heat propagation. In the summer, if exposed directly to sunlight, the rays might heat the sensor, again disrupting its measurements. We have chosen to use bus temperature sensors rather than those of meteorological stations because they could better represent micro-climates. The average duration of a trip is 32 minutes, which is relatively short for public transport. However, the maximum value of 240 minutes (4 hours) indicates the presence of exceptional situations, such as breakdowns. During these trips, it is likely that the vehicle's energy consumption is not truly representative of nominal usage.

Table 5. Description of Variables, Sources, Types, and Sampling Methods.

Variable	Source	Type	Sampling
Speed	Telemetry system	Float	Triggered
Acceleration	Calculated from speed	Float	Per second
Slope	Calculated from GPS altitudes and distance between stops	Float	Mean between planned stops
Distance	Calculated from speed	Float	Per second
Temperature	Telemetry system	Float	Triggered
Duration	Planning database	Float	Trip
Passenger count	Planning database	Integer	Between planned stops
Energy consumption	Telemetry system	Float	Triggered
Vehicle identification number	Planning database	Categorical	Trip
Bus line number	Planning database	Categorical	Trip
Direction	Planning database	Categorical	Trip
Day of the week	Planning database	Categorical	Trip
Time of the day	Planning database	Categorical	Trip

The average number of passengers is 21, which is reasonable for an urban bus. The maximum value of 190 passengers describes exceptionally crowded trips, such as in high-density areas or during city events attracting large crowds.

The average energy consumption is 13.7 kWh, which seems reasonable for an urban bus. However, the maximum value of 73 kWh indicates long or particularly energy-consuming trips during strong accelerations, significant fluctuations in the number of passengers, or prolonged use of air conditioning and heating systems. The negative value indicates that the buses do not consume energy on specific trips but rather generate it. While this observation may initially seem impossible, it is caused by a particular route covering only a short trip with a steep slope. When the bus goes downhill with passengers, it generates energy. When it goes uphill, it consumes a significant amount of energy to overcome gravity. This particular route is not cyclical. To observe the downhill and uphill slopes, two different directions of the route need to be considered. Therefore, in our scheduling database, we will have two different trips.

We have chosen not to work on full working days but rather on individual trips so that we can later optimize the scheduling of working days based on forecasts of individual trips. The results of this preliminary study are presented in Table 6.

Table 6. Descriptive Statistics of Variables.

Variable	Min	Max	Mean
Speed (km/h)	0	90	18
Acceleration (m/s²)	0	0.77	0.19
Slope (rad)	-1.47	1.5	0
Distance (km)	1	39	9.7
Temperature (°C)	-14	44	15
Duration (minutes)	0	240	32
Passenger count	0	190	21
Energy consumption (kWh)	-0.8	73	13.7

5. Preprocessing and Model Evaluation

To use these models, the data from the telemetry system must first be preprocessed (see Fig. 3) Only trips with a quality index higher than the threshold required by our partner are considered. The quality index, expressed as a percentage, measures the proportion of scheduled stops not served by the bus. The lower the index, the fewer planned stops were adhered to.

We then use each trip's start and end times, provided by the planning data, to extract the corresponding speed profile, battery power, and HVAC cooling power from the telemetry database.

The vehicle's empty weight and the number of passengers determine the variable load. As mentioned earlier, the slope is not directly measured; it is calculated based on the altitude obtained from GPS coordinates and the distance between bus stops. To do this, we used GPS data tiles available from the Consortium for Spatial Information (CGIAR-CSI) [20]. These data tiles have a spatial resolution of approximately 30 meters. Since the stop position data comes from consumer-grade GPS services, they have a resolution on the order of ten meters. The altitude returned is accurate to the meter. Given that we work with horizontal precision on the order of a meter and distances between stops on the order of a hundred meters, the relative error on the slope is considered acceptable in relation to the precision of the telemetry system's measurements.

Regarding the vehicle's mass, we calculate it as the sum of the empty weight and the number of passengers on board, with an average body mass of 80 kg per passenger. The number of passengers is calculated from the boarding and disembarking at each stop. The data available indicates the number of passengers on the bus on arrival and departure from each planned stop. If passengers board or alight outside these planned stops, an algorithm assigns them to the nearest stop.

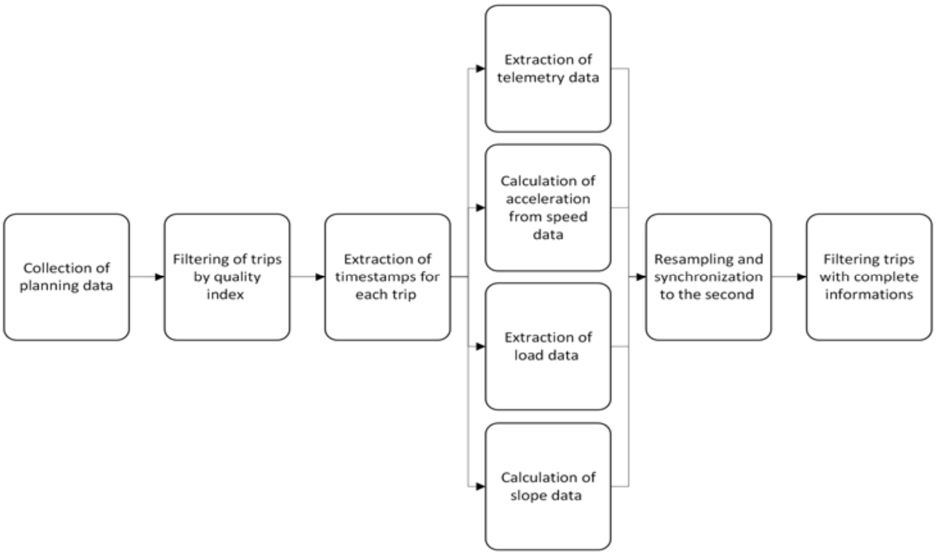

Fig. 3. Data processing from the telemetry system.

The data is then resampled to a per-second basis using forward filling. The timestamps for each measurement are not natively synchronized because each vehicle sends its telemetry data based on triggered sampling to save bandwidth. Even with these considerations, the data stored is on the order of terabytes per year for a fleet of a thousand vehicles.

During the preprocessing phase, less than 5 % of the measured trips do not meet the required quality index and are therefore removed from the analysis. This means the trips do not follow the planned path closely. For the remaining trips, 52 % allow for the complete calculation of the model (Model A) because of missing variable values. To further process data and make it useable by the hybrid models, categorical data such as vehicle identification number, line number, direction, bus number, day of the week, and time of the day have been processed using One Hot Encoding, whereas numerical values such as HVAC consumption, aerodynamic, gravity, inertia-linked consumptions, and regeneration have been standardized.

The performance of the models is evaluated by calculating the Mean Absolute Percentage Error (MAPE) between the actual and predicted energy consumption for each trip.

$$MAPE = 100 * \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \widehat{y_i}|}{y_i},$$

where n is the number of observations, y_i the true value, and $\hat{y}_i$ the forecasted value. We have chosen this evaluation metric because it does not depend on the observation scale, is easy to understand as a percentage, and because the absolute calculation prevents positive and negative values from canceling each other.

6. Results

This section will expose the results of our study after describing how we set up the training and evaluation framework.

6.1. Experimental Setup

This case study data is used with the models described in Table 1. To set the value of k for the physical model, we compared the total energy consumed with and without regeneration, ultimately concluding that $k = 0.3$ for our case study. We used the Optuna framework to tune the hyperparameters of the MLP, KAN, and XGBoost models. Their hyperparameters are essential for optimizing their performance in predicting energy consumption. The error is minimized over the training set, and the best model is the one that minimizes the error over the test set. Predictions of energy consumption are finally done on the validation set to validate the generalization of the model. Our complete dataset is split between three sub-datasets: training set 70 %, test set 20 %, and validation set 10 %.

For the MLP model, the first and second hidden layers have 46 and 32 neurons, respectively, with a learning rate of 0.001. The optimizer used is Adam, and the model trains for 2500 epochs with L1Loss as the loss function.

For the KAN model, the table lists the values for the number of neurons in the first and second hidden layers (hidden size1 and hidden size2), set at 29 and 20, respectively. Other hyperparameters include a grid search parameter (grid) of 3, a k-fold cross-validation value (k) of 2, a learning rate of 0.010, the use of the Adam optimizer, 250 epochs for training, and the use of L1Loss as the loss function.

The XGBoost model parameters include a learning rate of 0.068, 210 estimators (trees), and a maximum depth of 12 for the decision trees. The model also uses a subsample rate of 0.635 and a colsample bytree value of 0.679 to control the features considered in each tree. The gamma parameter is set to 0.001 to control the regularization of the model, and the minimum child weight is set to 8 to prevent overfitting.

These model architectures are shown in Table 11 from the Appendix. The output layer of each neural network model contains a single neuron. Multiple loss functions were compared to train the model, including L1 Loss, MSE, and MAPE. These results are presented in Table 7.

Table 7. Performance of Multiple Loss Functions during Training.

Loss Function	MAPE (%)	Standard Deviation
MAPE	9.6	76.4
MSE	5.6	28.7
L1 Loss	4.8	21.5

Depending on the loss function used during the training phase, the MAPE of the training phase varies greatly. One could think that the best way to minimize the MAPE is to use it during the training phase. However, this error introduces noise, leading to convergence issues. The Mean Squared Error (MSE) penalizes the prediction errors far from the mean more significantly, proving less effective than L1 Loss during our training phase. It scores a MAPE of 4.8 %, followed by MSE and MAPE. Their standard deviations reinforce our observation with the same ranking. For the rest of our neural network models, L1 Loss will be used.

Adam is the optimizer used to train the neural network models because it allows fast convergence due to its adaptive learning rate and low memory requirements [22]. The maximum number of epochs has been set to 2500 for the MLP and 250 for KAN because they allow the models to converge based on our dataset. For physical models, it is not necessary to separate the presentation of the results since there is no training phase, but we have chosen to evaluate the performance on the same data as the machine learning models.

6.2. Models Performances

The results from the different models are presented in Table 8 for the training dataset, Table 9 for the test dataset, and Table 10 for the validation dataset.

Table 8. Model Performance Comparison on Training Dataset.

Variable	MAPE (%)	Standard Deviation
A	21.15	42.25
B	17.47	32.00
C	28.02	90.68
D	31.89	83.84
M1	12.72	48.14
M2	5.99	21.63
M3	4.78	18.13
M4	7.05	33.37
M5	3.00	17.67

The results of the models tested on the training, test, and validation datasets show a clear difference in terms of performance and stability, with M5 (Physical followed by XGBoost) dominating. This model achieves the best results across all datasets, with a very low MAPE of 3.00 % on the training set, 5.59 % on the test set, and 5.79 % on the validation set, accompanied by low standard deviations, indicating an excellent ability to generalize and provide stable predictions. This remarkable performance can be attributed to the approach combining physical data and using XGBoost. This gradient-boosting decision tree model is particularly effective for this type of task. M3 (Physical followed by MLP with additional features, including temperature and mean speed) closely follows, with a MAPE of 4.78 % on training, 6.05 % on test, and 6.30 % on validation, but its standard deviation is higher for both training and validation datasets, indicating some instability in its predictions. The M4 model, built with a KAN instead of MLP, scores a MAPE of 7.05 % on training, 7.58 % on the test, and 7.79 % on validation. It is, therefore, not more performant than a conventional MLP and requires substantially more training time than the MLP. M2 (Physical followed by MLP with basic features) presents similar results to M3, with a MAPE of 5.99 % on training and 7.78 % on test, but its standard deviation remains reasonably controlled, suggesting good stability. However, it indicates that using the Mean speed and Mean Temperature as predictors helps improve the prediction. In contrast, A and B models show poor performance, with high MAPEs (21.15 % and 17.47 % on training, respectively) and significant standard deviations, reflecting their inability to capture complex relationships between different variables and provide reliable predictions. It is worth noting that Model B (Without HVAC consumption) provides better results than Model A (With HVAC consumption). This seems counterintuitive since one would think a more complete model would yield better results. This can be explained by the overall tendency of physical models to over-predict the power consumption to put the vehicle in motion. When HVAC consumption is added, the prediction quality increases, thus worsening the error. Model C is the second least performing model, with exceptionally high MAPEs (up to 28.02 % on training and 27.11 % on test), and very high standard deviations. Finally, the worst-performing model is D, with a MAPE of up to 33.34 % on the validation dataset and a standard deviation of 137.57. This indicates that for public transport vehicles and, more widely, vehicles with high variations in mass, it is essential to consider the energy that can be gained from regenerative braking and its loading. In conclusion, M5 clearly stands out as the model to prioritize for predicting the energy consumption of electric buses, thanks to its combination of physical data and the XGBoost algorithm. In contrast, models based solely on physical features or simpler architectures such as A, B, and C should be considered less suitable for this task and be used only when few variables are available.

6.3. Variables Contributions

To assess the influence of the various variables used in our study on the prediction output, we used SHAP (SHapley Additive exPlanations) [23]. SHAP is a widely used method that provides interpretable machine learning models by attributing the output of a model to each feature based on cooperative game theory. It calculates Shapley values, which represent the contribution of each feature to the prediction, allowing us to understand the impact of each variable on the model’s output. This technique is valuable for identifying which factors most influence our predictions and for providing better transparency in the model's decision-making process. This method will be used to explain the performance of the best performing model, M5. As it uses a XGBoost model, TreeExplainer proposed by Lundberg et al. [24] is used to perform the SHAP analysis. The results, presented in a summary plot based on the validation dataset (Fig. 4) highlight the feature importance for the model prediction. Features are ranked by their impact on the prediction, from top to bottom. The color gradient, from blue to red, indicates the range of feature values from Low to High in the dataset. Each point represents a single observation, and the horizontal positioning of these points reflects their contribution to the model output.

Table 9. Model Performance Comparison on Test Dataset.

Variable	MAPE (%)	Standard Deviation
A	20.64	22.87
B	16.91	17.38
C	27.11	40.46
D	32.99	88.38
M1	12.28	16.88
M2	7.78	17.38
M3	6.05	11.84
M4	7.58	19.49
M5	5.59	8.38

Table 10. Model Performance Comparison on Validation Dataset.

Variable	MAPE (%)	Standard Deviation
A	20.98	24.52
B	13.13	16.73
C	26.65	31.77
D	33.34	137.57
M1	12.36	19.62
M2	8.06	20.27
M3	6.30	15.69
M4	7.79	16.82
M5	5.79	9.01

The cumulative distance is the most important feature. It represents a clear trend of consuming more energy as the distance increases. This relationship is intuitive as longer distances usually mean more time spent driving and consequently using energy. Inertia represents the energy used to put the vehicle in motion, the more the bus must stop, the more energy it uses to restart. Aerodynamic and Gravitational energy show similar patterns as increased values improve the overall consumption. Counterintuitively, the model associates higher inertia and gravitational regeneration values with increased consumption. This contradicts physical principles, as these types of energy should reduce total consumption. The explanation for this might lie in their strong correlation with consumption variables. When a vehicle regenerates energy during braking, it still requires additional energy to get back in motion. Additionally, the vehicle typically experiences uphill and downhill transitions, as bus routes are cyclic, with few exceptions. The large variations in regeneration values are partly due to the method used to differentiate between motor-driven and regenerative power, as engine power data is not available. Since the power is negative during regeneration, an offset related to constant vehicle consumption (e.g., lighting, battery management) distorts the interpretation. Even during stops, we could not clearly identify the "base consumption" of the vehicle. This means that consumption or regeneration could be attributed to the wrong variable during transition phases, which creates strong noises the model must deal with. The XGBoost might then misunderstand the contribution of inertia, gravitational, and aerodynamic regeneration to the overall consumption.

An increased Heating energy is also strongly linked with higher consumption, with extreme SHAP values that rival those of distance. The average speed seems to be inversely proportional to total energy consumption. This might appear counterintuitive as well, but it is in line with the expert observations from the company: lower speeds result in longer travel times, leading to higher heating energy use.

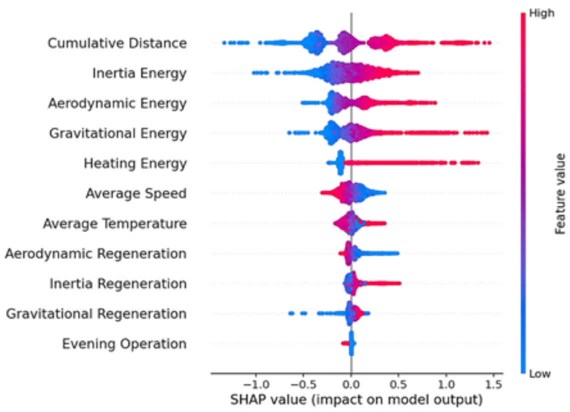

Fig. 4. Main SHAP values on the validation dataset.

Lower average speed is also linked to increased stops and higher inertia consumption. The average temperature shows no clear trend in energy consumption, with blue and red points merging near zero. This might be caused by specific heating thresholds between full diesel, full electric, or both heating modes simultaneously when the weather is particularly cold. A previous study explained that cold weather reduces the battery capacity of vehicles. Our model does not consider the available energy from the battery but rather the expected consumption. In further work, we would like to use charging data to understand the impact of temperature and state of health on the real available energy. This also highlights that averaging temperature over a trip masks important nuances and is not a well-suited feature for the model. For instance, a bus starting its day from a heated or cooled depot requires time to adjust to ambient temperatures, introducing bias in the mean temperature. On the other hand, temperature measurements from meteorological stations do not reflect what the vehicle faces in its microclimate, such as when a bus is exposed to the sun on a summer day. Investigating the impact of both temperature measurements on the model would be a valuable future direction. Lastly, the time of day has a relatively low impact on energy consumption. Still, energy consumption decreases during evening operations when traffic is lighter and fewer passengers are on board.

Although their SHAP contributions are negligible compared to the previously analyzed features, examining the remaining SHAP values can provide valuable insights into specific operational patterns. Fig. 5 highlights that certain bus lines are associated with improved energy efficiency, while others exhibit the opposite trend. Similarly, some vehicles, such as Vehicle 4, consistently show higher energy consumption than others. This could be related to a vehicle in less good working condition or with low-pressure tires and could suggest maintenance.

While Evening Operation in Fig. 4 was linked to reduced energy consumption, Afternoon Peak Operations and, more notably, Midday Operation display increased consumption patterns. This is primarily influenced by passenger load and traffic density. Such insights could help transit planners adopt a more strategic approach when assigning vehicles to specific routes, adjusting for expected service conditions to optimize energy efficiency.

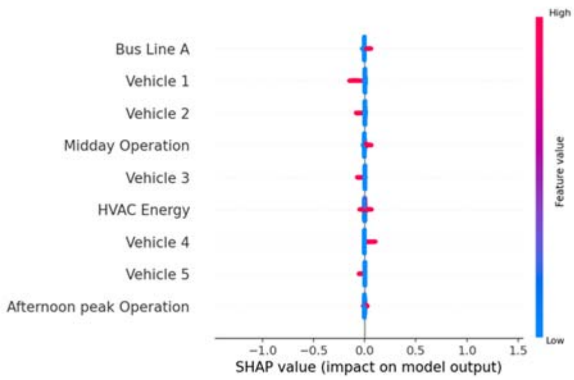

Fig. 5. Following SHAP values.

7. Discussion

In the long run, our models are designed with the ambition to support operational teams in optimizing the dimensioning of electric vehicle batteries for future bus fleets. By accurately predicting the energy needs of these vehicles, we can better plan battery sizes that will meet operational demands throughout the life cycle of the fleet, ensuring that they can handle various driving conditions and route profiles. This will improve fleet management efficiency and reduce unnecessary battery oversizing, contributing to both economic savings and sustainability.

In the short term, our consumption prediction model presents an exciting opportunity to refine the daily operational planning of electric buses. We have observed that the remaining charge at the end of the workday often remains above 50 %, suggesting that bus operations are managed conservatively to avoid service interruptions. This conservative approach, while ensuring reliability, leads to an under-utilization of the battery's capacity. Our forecasting tools allow for a more precise estimate of the energy required for each trip, enabling the operational teams to make more informed decisions on battery use. As a result, we can safely utilize a higher proportion of the battery's available energy without compromising service reliability, contributing to the overall cost-efficiency of electric bus operations.

Furthermore, while buses are often labeled as "100 % electric", they are still equipped with diesel-powered heating systems to supplement electric heating during extreme weather conditions. These thresholds for switching to diesel heating are often set arbitrarily, and using our model to assess their impact on energy consumption could provide important insights. By analyzing the effect of these heating systems on total energy consumption during different routes, we could determine whether it would be feasible to rely solely on electric heating for specific trips, thereby reducing the environmental impact of bus operations. This would directly contribute to the broader goal of decarbonizing public transport.

Additionally, our model could be integrated with real-time adjustments. For example, energy consumption forecasts would be used to plan daily routes, and adjustments could be made dynamically based on real-time telemetry data. If energy consumption up to a given point exceeds the forecast, diesel heating could be used to conserve energy. Conversely, if energy consumption is lower than anticipated, electric heating could be prioritized to reduce emissions. This flexible approach could also lead to a new strategy of opportunistic charging, where buses would remain parked and charge for a few minutes between successive trips, ensuring they have enough energy to complete their service without interruptions.

These real-time adjustments could have additional benefits beyond day-to-day operations. Simulations using this model could be invaluable for better dimensioning charging infrastructure at different stages of fleet deployment. For instance, as the fleet transitions from diesel to electric buses, it will be crucial to ensure that charging stations are appropriately scaled to meet the increased demand. This includes forecasting the energy needs of buses as the fleet grows and ensuring that charging infrastructure can keep pace with the new fleet composition. The ability to predict energy consumption at various phases of the project (from the introduction of electric buses to the full electrification of the fleet) will allow city planners and fleet operators to make data-driven decisions regarding where and when to deploy new charging stations.

Moreover, our model could help optimize the operational patterns of the fleet by minimizing the number of charging stations required while maintaining an equivalent level of service. By ensuring that buses are appropriately charged before they go on their assigned routes, we can reduce charging downtime, improve the fleet's overall efficiency, and thus, reduce costs.

Like other researchers in this field, our model's significant advantage lies in its adaptability to different datasets. We are currently gathering telemetry data from buses operating in different Canadian cities, each with distinct topologies, climates, and travel behaviors. By applying our model to this diverse range of data, we can evaluate its robustness and accuracy across different operational contexts.

Moreover, our model has potential beyond the public transportation sector. The same principles that drive the consumption patterns of electric buses can be applied to other industries where consumption cycles are well-defined. For example, sectors like aviation, where aircraft undergo similar energy usage patterns, and mining, where heavy vehicles operate in harsh conditions, could benefit from integrating our forecasting model. Testing the model in other industries would allow us to validate its generalizability and refine its capabilities, expanding its scope and impact.

Another promising avenue for future work would be the integration of our model within an optimization framework for vehicle assignment and charging scheduling. While traditional optimization models often rely on probabilistic demand forecasts, they tend to oversimplify the complexities of real-world operations. An approach combining forecasts with real-time adjustment could provide a more dynamic and accurate solution. For instance, rather than relying on static predictions, real-time adjustments based on energy consumption data could be made to adapt charging schedules and vehicle assignments on the fly. This would improve the charging infrastructure's efficiency and the fleet’s operational planning, ensuring that vehicles are always charged when needed without over-committing charging resources.

Another important consideration, often overlooked in existing literature, is the energy consumption associated with off-route movements and non-passenger operations. These activities consume significant energy, including repositioning buses between depots, pre-heating the vehicle for passenger comfort, or repositioning buses between different routes. Experts estimate that these activities could add up to 10\% to the energy consumption forecasts for each trip. Our future work will focus on including these "off-trip" consumption factors in the model, ensuring that we provide a more realistic estimate of energy consumption that reflects the full operational cycle of the buses.

The impact of operational practices on battery lifespan is a key consideration in the electrification of buses. For instance, while effective in maximizing autonomy, opportunity charging may accelerate battery degradation due to more frequent and rapid charge cycles. Similarly, using electric heating, although more environmentally friendly than diesel heating, can cause faster battery discharge, thus reducing its lifespan. Optimizing these practices based on battery charge levels and thermal impact would be crucial to minimize premature wear while maximizing energy efficiency and battery longevity.

Finally, we recognize the growing complexity of machine learning models in the field, particularly with the rise of large-scale artificial intelligence models and big data. While simpler models often require fewer computational resources, the recent advances in large language models (LLMs) and foundation models have introduced new challenges. These models, which offer great promise in terms of generalizability and flexibility, are often very large and computationally expensive. If these models are used for optimizing operations and supporting sustainability goals, their high training and inference costs could offset the energy savings from fleet electrification. It is crucial, therefore, to consider the energy cost of these models in the broader context of sustainability. We must balance the benefits of advanced AI models and the environmental costs associated with their deployment. Future work could focus on evaluating the energy consumption of these models and integrating that assessment into decisions about their use in optimizing electric vehicle fleets.

8. Conclusion

Our case study first highlighted the complexity of operational data from public transport companies compared to simulated data. We demonstrated the limitations of physical models in a real operational context. To address these challenges, we combined these physical models with various machine-learning approaches, allowing us to fine-tune predictions based on the operational context. The hybrid model combining physical principles with XGBoost delivered the best performance, with an error of 5.59 % on the test set and 5.79 % on the validation set, with respective standard deviations of 8.38 and 9.01.

These results demonstrate the model's ability to adapt to noisy data while providing effective energy consumption predictions across different routes.

References

[1]. L. Adam, R. Pellerin, B. Agard, Analysis of rule-based models for predicting energy consumption in electric bus fleets: a case study with telemetry data, in *Proceedings of the 5th Winter IFSA Conference on Automation, Robotics & Communications for Industry 4.0/5.0 (ARCI'25)*, 2025, pp. 68-73.

[2]. S. Krawiec, B. Łazarz, S. Markusik, G. Karoń, et al., Urban public transport with the use of electric buses – development tendencies, *Transport Problems*, Vol. 11, 2016, pp. 127-137.

[3]. S. Corbet, C. Larkin, J. McCluskey, The influence of inclement weather on electric bus efficiency: evidence from a developed European network, *Case Studies on Transport Policy*, Vol. 12, 2023, 100971.

[4]. B. Zhang, N. Niu, H. Li, Z. Wang, et al., Could fast battery charging effectively mitigate range anxiety in electric vehicle usage?, *Transportation Research Part D: Transport and Environment*, Vol. 95, 2021, 102840.

[5]. A. Jahic, M. Eskander, E. Avdevicius, D. Schulz, Energy consumption of battery-electric buses: review of influential parameters and modelling approaches, *B&H Electrical Engineering*, Vol. 17, Issue 1, 2023, pp. 7-17.

[6]. Y. Chen, Y. Zhang, R. Sun, Data-driven estimation of energy consumption for electric bus under real-world driving conditions, *Transportation Research Part D: Transport and Environment*, Vol. 98, 2021, 102969.

[7]. L. K. Lim, Z. A. Muis, W. S. Ho, H. Hashim, et al., Review of the energy forecasting and scheduling model for electric buses, *Energy*, Vol. 263, 2022, 125773.

[8]. X. Zhang, Z. Zhang, Y. Liu, Z. Xu, et al., A review of machine learning approaches for electric vehicle energy consumption modelling in urban transportation, *Renewable Energy*, Vol. 234, 2024, 121243.

[9]. P. Iora, L. Tribioli, Effect of ambient temperature on electric vehicles' energy consumption and range: model definition and sensitivity analysis based on Nissan Leaf data, *World Electric Vehicle Journal*, Vol. 10, Issue 2, 2019, 21.

[10]. Y. Chen, G. Wu, R. Sun, A. Dubey, et al., A review and outlook of energy consumption estimation models for electric vehicles, *arXiv preprint*, 2020, arXiv:2001.07429.

[11]. C. J. J. Beckers, I. J. M. Besselink, H. Nijmeijer, Assessing the impact of cornering losses on the energy consumption of electric city buses, *Transportation Research Part D: Transport and Environment*, Vol. 86, 2020, 102360.

[12]. C. Fiori, M. Montanino, S. Nielsen, M. Seredynski, et al., Microscopic energy consumption modelling of electric buses: model development, calibration, and validation, *Transportation Research Part D: Transport and Environment*, Vol. 98, 2021, 102978.

[13]. C. J. J. Beckers, I. J. M. Besselink, J. J. M. Frints, H. Nijmeijer, Energy consumption prediction for electric city buses, in *Proceedings of the 13th ITS European Congress*, 2019.

[14]. L. A. W. Blades, T. Matthews, T. E. McGrath, J. Early, et al., Predicting energy consumption of zero emission buses using route feature selection methods, *Transportation Research Part D: Transport and Environment*, Vol. 130, 2024, 104158.

[15]. F. Basso, F. Feijoo, R. Pezoa, M. Varas, et al., The impact of electromobility in public transport: an estimation of energy consumption using disaggregated data in Santiago, Chile, *Energy*, Vol. 286, 2023, 129550.

[16]. T. Pamuła, D. Pamuła, Prediction of electric buses energy consumption from trip parameters using deep learning, *Energies*, Vol. 15, Issue 5, 2022, 1747.

[17]. Z. Xu, J. Wang, P. D. Lund, Y. Zhang, Analysis of energy consumption for electric buses based on low-frequency real-world data, *Transportation Research Part D: Transport and Environment*, Vol. 122, 2023, 103857.

[18]. R. Sennefelder, R. Martín-Clemente, R. González-Carvajal, Data driven characterization and predictive classification of energy economy for public transport, in *Proceedings of the International Conference on Computational Science and Computational Intelligence (CSCI')*, 2021, pp. 1254-1262.

[19]. R. M. Sennefelder, P. Micek, R. Martín-Clemente, J. C. Risquez, et al., Driving cycle synthesis, aiming for realness, by extending real-world driving databases, *IEEE Access*, Vol. 10, 2022, pp. 54123-54135.

[20]. CGIAR-CSI SRTM, SRTM Data, https://srtm.csi.cgiar.org/srtmdata/

[21]. Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, et al., KAN: Kolmogorov-Arnold networks, *arXiv preprint*, 2024, arXiv:2404.19756.
[22]. D. P. Kingma, J. L. Ba, Adam: a method for stochastic optimization, *arXiv preprint*, 2014, arXiv:1412.6980.
[23]. S. M. Lundberg, S. Lee, A unified approach to interpreting model predictions, *arXiv preprint*, 2017, arXiv:1705.07874.
[24]. S. M. Lundberg, G. Erion, H. Chen, et al., From local explanations to global understanding with explainable AI for trees, *Nature Machine Intelligence*, Vol. 2, Issue 1, 2020, pp. 56-67.

Appendix

Table 11. Hyperparameters of the neural network models and XGBoost.

Architecture	Parameter	Value
KAN	hidden_size1	29
	hidden_size2	20
	grid	3
	k	2
	learning_rate	0.010
	optimizer	Adam
	epoch	250
	Loss	L1Loss
XGBoost	learning_rate	0.0299
	n_estimators	975
	max_depth	48
	subsample	0.642
	colsample_bytree	0.614
	gamma	1.476e-5
	min_child_weight	56
MLP	hidden_size1	46
	hidden_size2	32
	learning_rate	0.001
	optimizer	Adam
	epoch	2500
	Loss	L1Loss