Advanced Range Estimation for Electric Busses with Physics Informed Machine Learning

Thesis

Presented in Partial Fulfillment of the Requirements for the Degree Bachelor's with Honors Research Distinction in the Graduate School of The Ohio State University

By

Radu Pavel Jr.

Graduate Program in Mechanical Engineering

The Ohio State University

2023

Thesis Committee

Stephanie Stockar, Advisor

Marcello Canova

Abstract

Given the growing focus on environmentally sustainable practices and the desire for cost effective solutions, electric buses have caught the eye of many public transportation companies. To make electric buses an ideal addition to a fleet, they must complete required routes in all conditions, making accurate range finding of these buses an invaluable tool. A current approach for range estimation is to develop energy-based models of components and integrate them in a larger model that predicts the overall battery power draw, estimating the remaining range available. Such an analytical model is limited by the variety of extraneous variables affecting the system (traffic, temperature, passenger count), individual components which are difficult to model accurately, as well as finite access to required data and parameters for calibration and verification. In this context, the proposed research aims to improve the state of the art of range estimation for electric vehicles by combining data driven machine learning techniques with physics-based analysis (PBA). This combined model is applied to a case study of the regenerative braking in electric buses. First, a feed forward neural network model was trained to estimate regenerative braking based on available experimental data, then this network was integrated into a physics-based bus model. This implementation was then used to assess the capabilities of the combined model to account for various lapses in data quality, and how the overall accuracy can be improved from using a strictly analytical model. The combined model resulted in a clear improvement of the regenerative braking modeling, and therefore an improvement in the analytical modeling of the electric bus.

Acknowledgments

Thank you to Professor Stockar for her incredible patience and guidance throughout this entire process. Thank you to the members of CAR lab, and to the members of the COTA bus project. Special thank you to the members of my committee, Dr. Stockar and Dr. Canova.

Field of Study

Major Field: Mechanical Engineering

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

For the successful deployment of an electric bus for public transit, it must be able to complete its routes in a variety of conditions. Weather, passenger load, number of stops, traffic, and more all affect the range of the vehicle, which makes predicting the range difficult [1]. Precisely estimating the range of an electric bus is vital to determining if it will be able to complete all required routes for deployment. Furthermore, with inaccurate range estimation a driver is forced to conserve more battery than is required, further reducing the effective range of the vehicle [2]. A common approach for range estimation is to use physics-based models (PBM) that predict the energy demand of each component in the electric powertrain. This approach is especially well suited to buses, due to the scheduled and recurrent nature of their operating goals [3]. However, there are distinct limitations to be overcome for this approach to be successful. Some components of the bus are unable to be described analytically, there are a number of assumptions introduced in the modeling phase [4-10], and there is often poor availability of high-definition data to calibrate and verify such a model [3].

A possible solution for improving the accuracy of PBMs is to combine them with data driven models. There is a significant body work that leverages the flexibility of Machine Learning (ML) with the physical consistency of PBM [2,11-14]. For example, Figure 1 outlines a variety of different structures used to fuse ML and PBM techniques [11]. These structures range in complexity of implementation and applications.

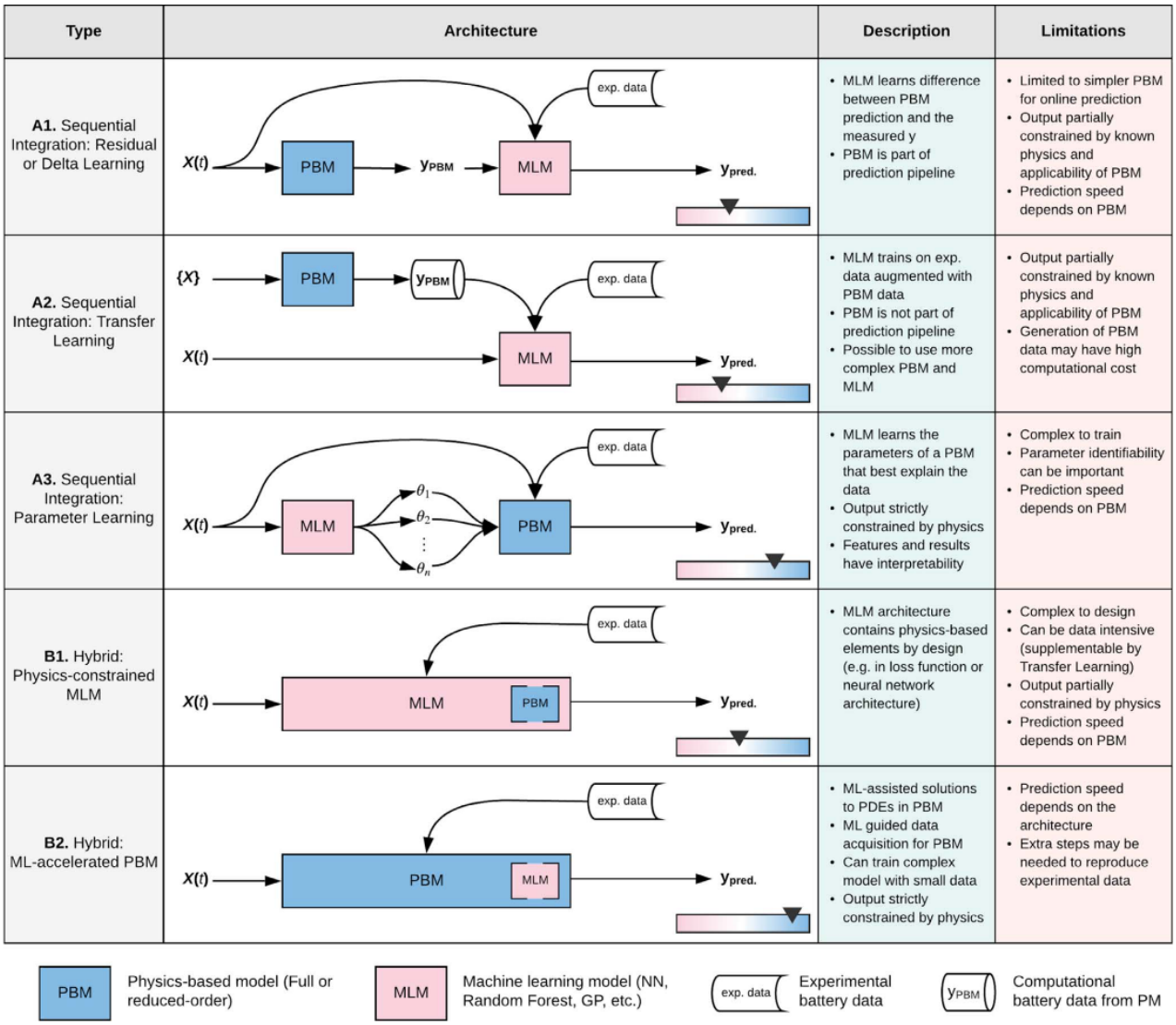

Figure 1: Examples of ML and PBM Integration

Integration types A1 and A3 are of particular relevance for the application in this thesis, as they provide a clearer boundary between the prediction obtained from the data driven model and the analytical model.

While machine learning is a powerful tool, it is limited by both quality and quantity of data available [12]. In many cases, high-fidelity data with consistent sampling time and proper synchronization is not available to take full advantage of a purely data

driven approach. Moreover, collecting a large enough sample of data for the training that is representative of the whole operating range of the system is often time and cost prohibitive. In the context of the electric bus range prediction, this process would require pulling a bus from service for several weeks, which would result in a disruption of service. For this reason, testing on a dynamometer system with the objective of collecting data for model calibration and verification is limited to only a day.

Alternatively, lower-fidelity data can be collected while the vehicle is deployed. While this option does not disrupt the service, there are several limitations related to the sampling rate of the data collected and test repeatability.

Due to the practical limitations of performing exhaustive tests as well as collecting reliable and consistent data, a modeling approach that achieves the desired range prediction accuracy when only limited information is available is crucial.

This thesis is structured as follows. Chapter 2 presents the analytical model used to simulate an electric bus servicing a Central Ohio Transit Authority (COTA) route that was developed as part of prior work. Then, Chapter 3 will outline how a data driven approach was combined with the analytical model to apply a Physics-Informed Machine Learning (PIML) approach to simulating bus behavior. Once the implementation and integration of the model was completed, a set of metrics including terminal SoC are assessed in Chapter 4. The model comparison shows that, while data driven modeling is known to be sensitive to quality and quantity of data, by combining a data driven approach to analytical modeling, this limitation can be overcome. Finally, conclusion and future work will be summarized in Chapter 5.

Chapter 2: Analytical Cota Bus Model

Current Approach to Bus Range Modeling

The work in this thesis leverages an energy-based electric bus model that was developed at The Ohio State University (OSU) – Center for Automotive Research (CAR) as part of a CAR Consortium project directed by COTA. The bus model is based on the Newflyer Xcelsior CHARGE NG 40 pictured in Figure 2. The main specifications of the bus are outlined in Table 1.

Figure 2: Newflyer Xcelsior Electric Bus

Table 1: Xcelsior Main Specifications

Length	41 0 (12.50m)
Width	102 (2.6m)
Max. Passengers	40 Seated, 44 Standing
Curb Weight	28,850 lb (13,086 kg)

Motor	160 kW, 1,033 lb-ft
Battery Pack	440 kWh
Range (Nominal)	213 mi

The components of the bus are each isolated in the model, and a zeroth order approximation is used for the battery. The block diagram of the model is shown in Figure 3.

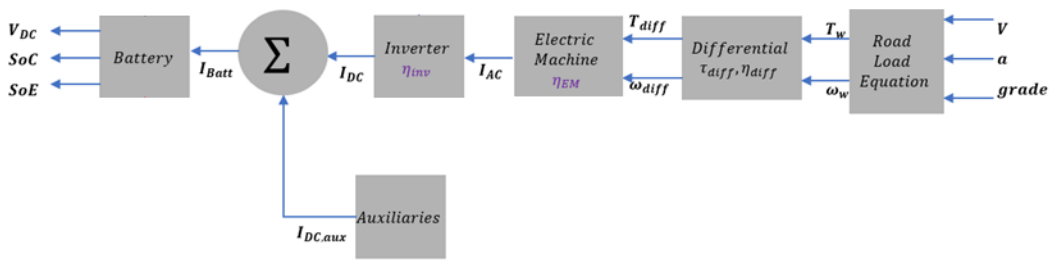

Figure 3: Block Diagram of Simulink Bus Model

Governing Equations

The foundation of the model is the Road Load equation:

$$\left(M + M_{eq}\right)\frac{dV}{dt} = F_w - \frac{1}{2}\rho_a C_d A_f V_{eff}^2 - C_r M g cos(\alpha) - M g sin(\alpha) \qquad (1)$$

where F_w is the force at the wheels, M is the mass of the vehicle, M_{eq} is the equilibrium inertial mass of the vehicle, ρ_a is the density of the air, C_d is the drag coefficient of the vehicle, A_f is the frontal area of the vehicle,V and V_{eff} are the velocity and effective velocity of the vehicle respectively, C_r is the coefficient of rolling resistance, and α is the grade coefficient [19]. The effective velocity is calculated as:

$$V_{eff} = V_{veh} - V_{wind} \qquad (2)$$

where V_{veh} is the velocity of the vehicle, and V_{wind} is the wind velocity. The grade coefficient is calculated using the road grade:

$$\alpha = \arctan\left(\frac{grade}{100}\right) \tag{3}$$

The wheel torque is:

$$T_w = F_w R_{w,loaded} \tag{4}$$

where $R_{w,loaded}$ is the loaded radius of the wheel which is 98% of wheel radius R_w:

$$R_{w,loaded} = 0.98 R_w \tag{5}$$

Finally, the relationship between wheel velocity and the wheel angular velocity is:

$$V_w = \omega_w R_{w,eff} \tag{6}$$

where V_w is the wheel velocity which is assumed here to be the same as the vehicle speed V_{veh}, ω_w is the wheel angular velocity, and $R_{w,eff}$ is the effective radius at the wheel:

$$R_{w,eff} = 0.95 R_w \tag{7}$$

The differential torque is then obtained with:

$$T_{diff} = \frac{T_w}{\tau_{diff} \cdot \eta_{diff}} \tag{8}$$

where η_{diff} is the efficiency of the differential, assumed constant, and τ_{diff} is the constant final drive ratio ($\tau_{diff} = 0.176$). Similarly, the differential output speed is:

$$\omega_{diff} = \omega_w \cdot \tau_{diff} \tag{9}$$

Then, the power of the electric machine is then calculated by converting the mechanical power at the shaft into electrical power:

$$P_{EM} = \begin{cases} \frac{1}{\eta_{EM}} T_{diff}\omega_{diff}, & T_{diff}\omega_{diff} \geq 0 \\ \eta_{EM} T_{diff}\omega_{diff}, & T_{diff}\omega_{diff} < 0 \end{cases} \tag{10}$$

where η_{EM} is the electric motor efficiency:

$$\eta_{EM} = \begin{cases} \eta_{EM,gen}, T_{diff} > 0 \\ \eta_{EM,regen}, T_{diff} < 0 \end{cases} \tag{11}$$

$\eta_{EM,gen}$ is the efficiency of the machine while in traction, which is determined by a Willans line approximation [20]. This approximation is shown in Figure 4 as an efficiency map.

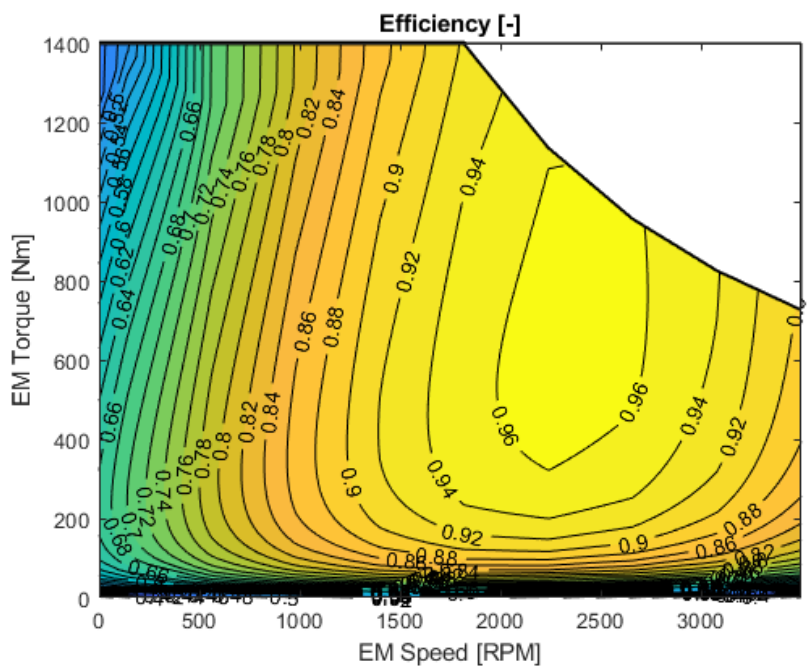

Figure 4: Willans Line Model Estimated Efficiency Map

Conversely, $\eta_{EM,regen}$ is the regen efficiency, which is currently represented using a lookup table that is function of bus SoC, derived experimentally. The maximum torque as a function of EM Speed can be found in figure 5.

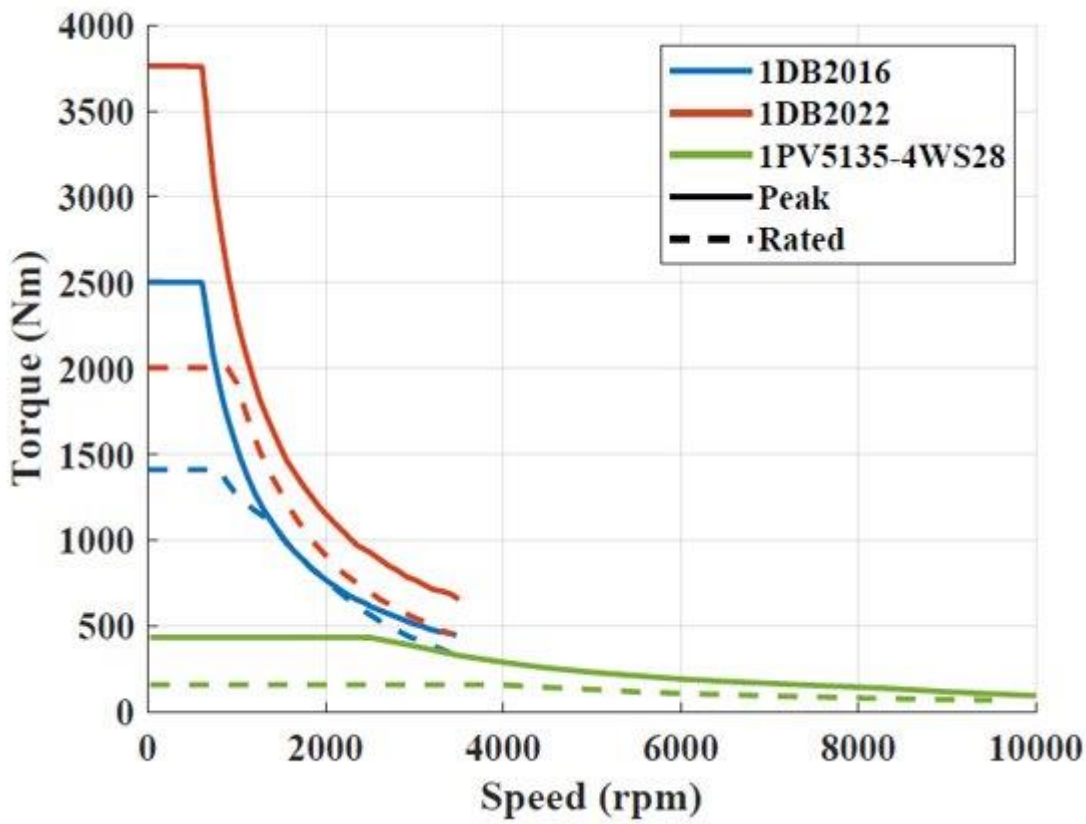

Figure 5: Max Torque as a Function of EM speed

The inverter is modeled considering a static approximation with a constant efficiency that depends on whether the vehicle is operating in traction or regenerative mode:

$$P_{EM-Inv} = \begin{cases} \frac{1}{\eta_{M-Inv}} P_{EM}, & P_{EM} \geq 0 \\ \eta_{M-Inv} P_{EM}, & P_{EM} < 0 \end{cases} \tag{12}$$

where $\eta_{M-Inv} = 0.95$.

To predict the battery state of charge (SoC) and voltage, a zero-order equivalent circuit model, show in Figure 6 is used [19].

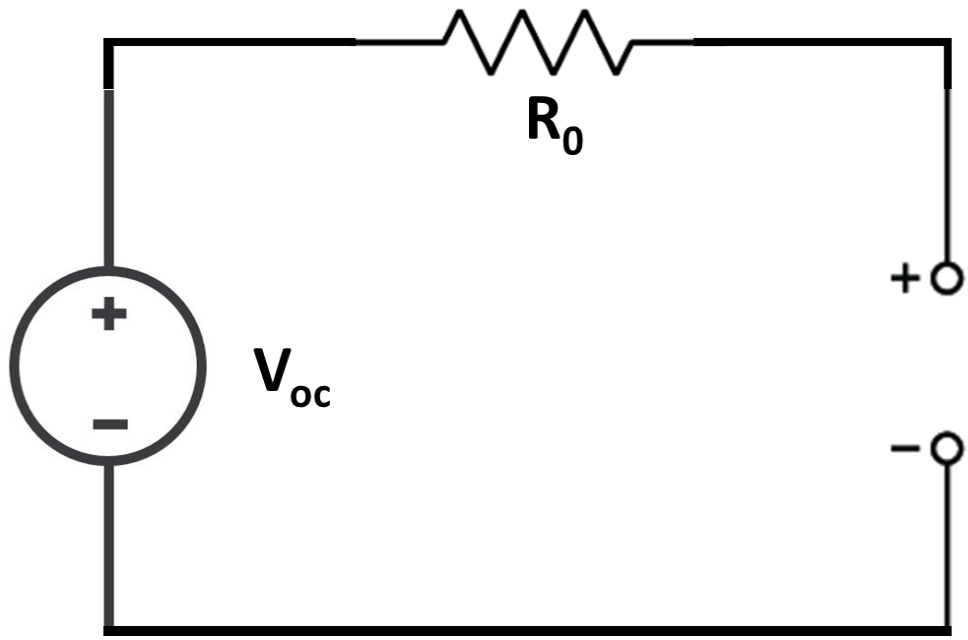

Figure 6: Zero Order Battery Model

The Power transferred to the cells of the battery can be found from the power transmitted to the battery

$$P_{cell} = \frac{P_{batt}}{N_p N_s} \tag{13}$$

where N_p is the number of battery packs, and N_s is the number of battery cells. From this, the current to each cell can be found

$$I_{cell}(t) = \frac{V_{OC} - \sqrt{V_{OC}^2 - 4P_{cell}R_0}}{2R_0} \tag{14}$$

where R_0 is the internal resistance is defined in Equation 16, and V_{OC} is the open circuit voltage of the battery.

The batttery voltage is calculated using Kirchoff's Voltage law

$$V_{DC}(t) = V_{OC}(SoC) - R_0 I_{batt}(t) \tag{15}$$

where I is the battery current. For this application, the open circuit voltage is assumed to be function of the SoC while the temperature is assumed constant. Conversely, the internal resistance is assumed constant during either charging or discharging events:

$$R_0 = \begin{cases} R_{discharge}, & I > 0 \; or \; P_{batt} > 0 \\ R_{charge}, & I < 0 \; or \; P_{batt} < 0 \end{cases} \tag{16}$$

The battery current is determined by considering both the motor current I_{M-Inv} and the auxiliary current $I_{aux-Inv}$, both at the inverter:

$$I_{batt} = I_{M-Inv} + I_{aux-Inv} \tag{17}$$

To predict the SoC of the battery, the model relies on Coulomb counting:

$$\frac{d}{dt} SoC = -\frac{I}{C_{nom}} \tag{18}$$

where C_{nom} is the nominal battery capacity. The parameters of the bus and bus components are summarized in Table 2.

Table 2: Bus Parameters [21]

PARAMETER	**VALUE**	**SOURCE/REASON**
Mass (Curb Weight), M	$13086\ kg$	Specification
Effective Mass, M_{eff}	1.1×M	Approximation. Ignores - λ ,J_{EM},J_w,η_t
Fully Loaded, M_{Load}	$18813\ kg$	Maximum Passenger capacity – 84; 150lbs per adult
Air Density	$1.225\,\frac{kg}{m^3}$	Air density at sea level, Temperature variation: [-15°C,35°C]

Drag Coefficient	0.65	Literature - Airflow studies –Base model & 2010 Autonomie report
Frontal Area	$7.72\ m^2$	$90\ \%\ A_f$
Rolling Resistance	0.008	Approximate maximum values from Continental
Tire Radius	$0.4929\ m$	Specification - 305/70R22.5
Nominal Battery Capacity, C_{nom}	841.13 Ah	Calculated from Dyno Data
Internal Battery Resistance R0	$1.5\ m\Omega$	Calculated from Dyno Data

Data Collection

A Newflyer electric bus was tested on the heavy-duty chassis dyno at CAR with the objective of collecting calibration and validation data for the physics-based model described in this Chapter. During the single day of testing, several coast-down tests from various initial vehicle velocities were performed, for a total of 1100 seconds of data. The summary of the data collected is provided in Table 3. It is worth noting that the dyno data provides the highest fidelity data, due to the repeatability of the tests and high precision measurements that are typically not available when the bus is deployed. For this reason, dyno data will be used in this thesis for the training and verification of the PIML.

Table 3: Data collected from Dyno Testing

Data acquisition device	Measured variable	Resolution/Sample Rate
	Auxiliary inverter Current	0.01s
	Motor Inverter Current	0.01s

Oscilloscope	DC-DC Voltage	0.01s
	Battery Current	0.01s
Fluke (connected to Battery Pack)	Absolute Current	1s
	Voltage	1s
Dynamometer	Speed	0.1s

A second dataset, referred to as the DRIVE dataset, was collected from a bus in service, via DAQ instruments on board. The same measured quantities as the one in Table 3 were collected, but at inconsistent sample rates due to both connectivity issues and on-board instrumentation. While the on-road dataset is larger than the one generated using the dyno testing, the quality of the signals is affected by noise, gaps, and inconsistent sampling rate. Hence, the Dyno dataset will be used for validation, and the DRIVE data will be used for verification of the model.

Accuracy of Analytical Model

A sample simulation result obtained on the Dyno data is shown in Figure 7, where the predicted battery current, state of charge and battery voltage are compared against the experimental data. The model predicts the final state of charge within 0.16%. While the model shows good overall agreement, the battery current from the model underpredicts the current from experiments, particularly during negative current events. This is verified by plotting the negative current and the corresponding error distribution, Figure 8.

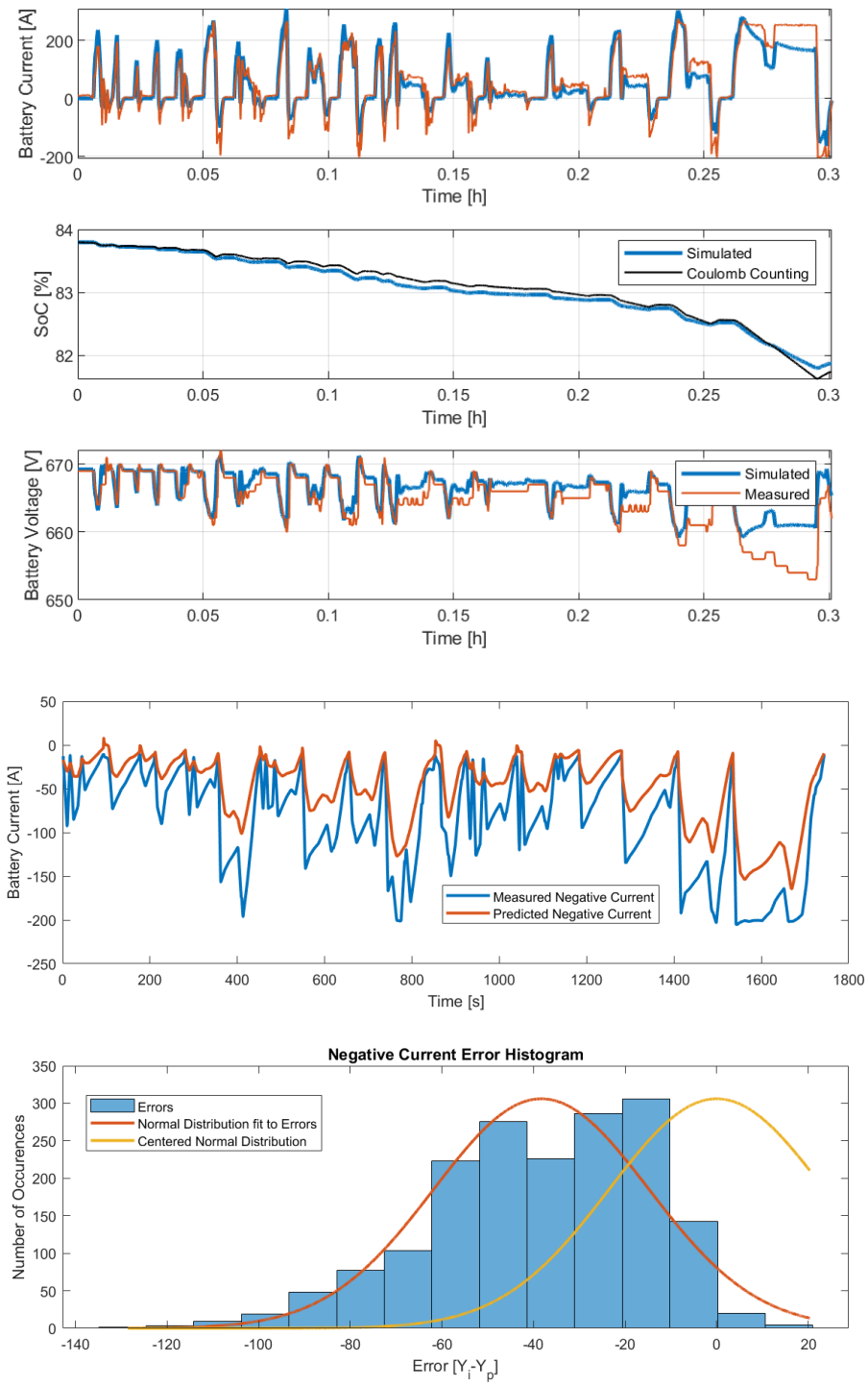

Figure 8

Figure 7: Analytical Model Simulation Results

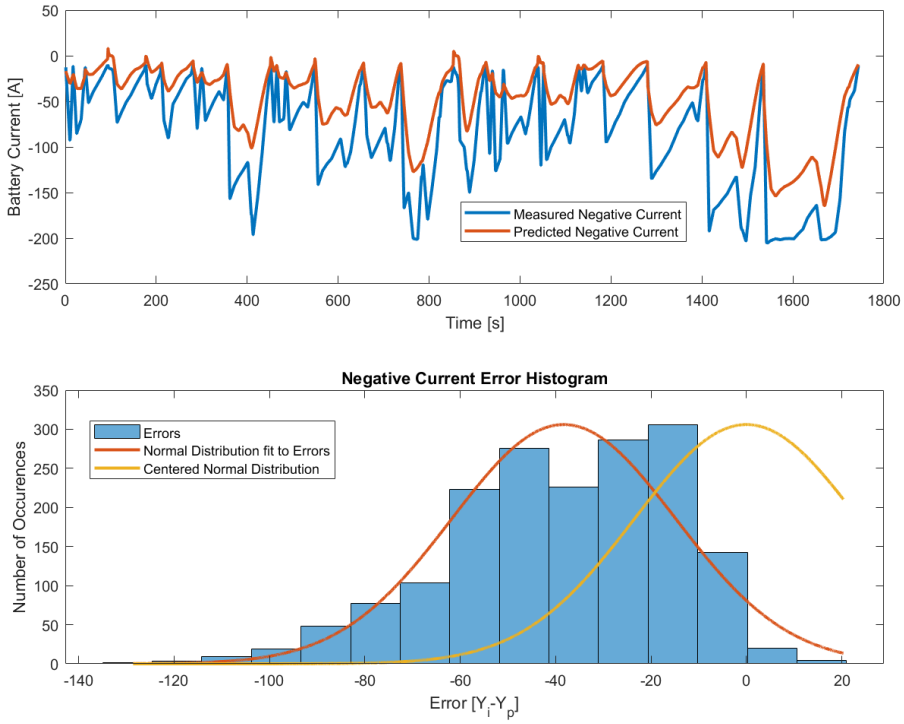

Figure 8: Analytical Model Negative Current Errors

Figure 8 clearly shows that the model fails to capture the regenerative behavior of the bus during coast down phases and, as a result, there is a significant error in predicting the battery charging. Moreover, the error distribution of the negative current prediction is far from the ideal normal distribution with a mean of zero, as shown in Figure 8. This systematic error compounds and gets worse in longer datasets, eventually significantly underrepresenting the change in SoC due to regenerative braking. Being able to improve the modeled regenerative braking through physics informed machine learning will improve the ability of the model to estimate the remaining range of a bus while on route.

Chapter 3: Combined Physics-based and Data Driven Modeling

Regenerative Braking as a Case Study

The current bus model has a significant limitation in its ability to accurately predict SoC due to challenges in modeling regenerative braking. Regenerative braking is a complex process that involves several factors that are difficult to analyze analytically. These factors include the system's efficiencies, which are not well defined, the battery management system's limitations on power transfer to and from the battery to ensure safe operation, and external inputs such as temperature and vehicle mass that are uncertain and can affect the vehicle power demand. For these reasons, a data-driven approach for estimating the regenerative braking is proposed in this thesis.

Due to the modularity of the energy-based model, the integration of a data-driven component is straightforward. Specifically, the regen efficiency coefficient approximating the power loss between the differential and battery can be replaced with a neural network which is trained on relevant data.

Structure and implementation

As shown in Figure 1, there are many possible integrations for physics informed machine learning models. The strengths of each should be considered in a case-by-case basis when being applied to modeling problems. In the case of regenerative braking, there is a clear and distinct set of inputs to the subsection, as well as a clear division from the rest of the model, which allows for isolated training and validation of a neural network. This leads to the implementation of structure B2, as shown in Figure 9, which can be interpreted as combination of structures A1 and A2 of Figure 1.

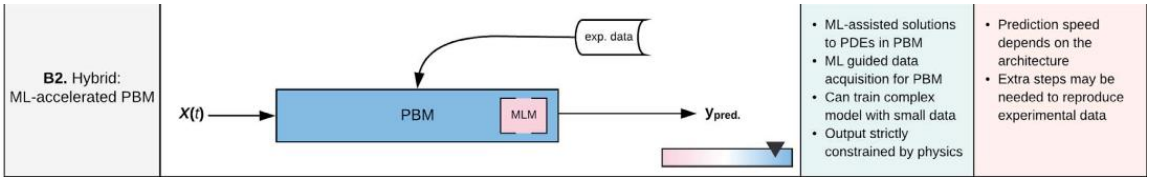

Figure 9: Selected Model Structure [11]

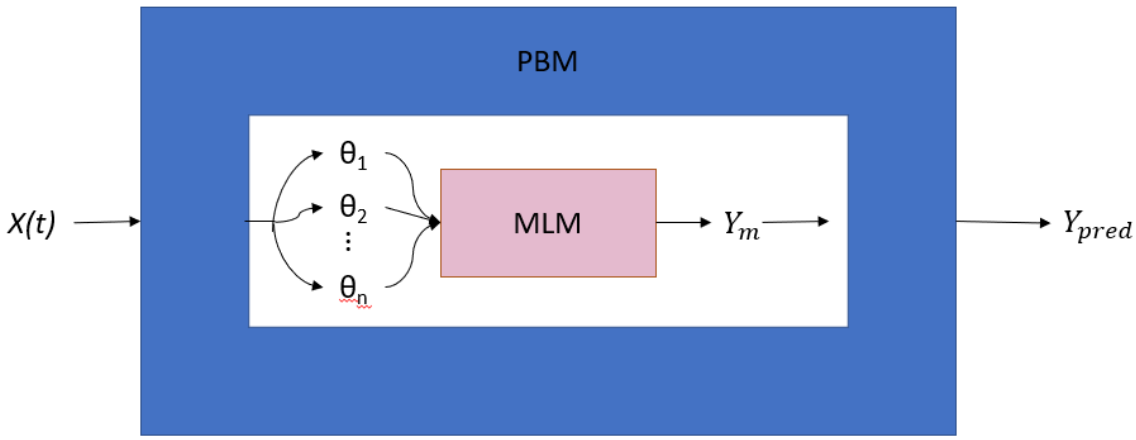

Figure 10: Model Structure (detailed)

In Figure 10, $X(t)$ represents the inputs to the combined model, in this case velocity and route data. Then, θ_n represents internal model variables solved for by the PBM, which is then inputted to the machine learning model that provides an output Y_m. This internal output is then used by the PBM to calculate the model output Y_{pred} which in the case of the electric bus is the battery state of charge at any instance of time.

The MLM implementation replaces the regen coefficient described in Equation 11, and it is integrated in the model as shown in Figure 11. Here, the data driven regenerative braking model is applied directly to the Electric Machine block in the

analytical model, this allows access to $\theta = [T_{diff} \quad \omega_{diff} \quad \eta_{EM}]^T$. These inputs were selected to inform a MLM model of regenerative braking as a result of their availability within the model, and their clear and direct relationship to regenerative braking power generated by the motor. Similarly, the desired network output was determined to be motor power (P_{EM}) due to its required availability in the model.

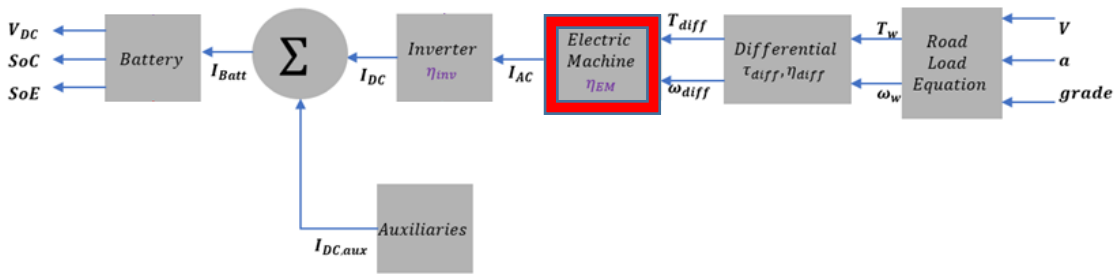

Figure 11: Location of NN Implementation to PBM

Neural network Data Processing

To perform the training, a dataset of training variables must be generated which informs the neural network (NN) of the power regeneration behavior. This step of the process is particularly challenging because none of the NN inputs θ are directly measured. However, the single component models in the PBM can be exercised so that the inputs to the NN are determined from the experiments. Specifically, the Electric Machine Efficiency, Differential Torque, and Differential angular velocity can be related to the network output, Electric Machine Power. The calculation is performed using Equation 1 to derive the Differential torque, Equation 9 to derive the Differential angular velocity, and the Willans line model is used to calculate the approximate electric machine

efficiency η_{EM}. To calculate a training output for the network, Equation 17 is used to derive the Electric Machine Power.

$$P_{EM} = I_{batt} * V_{batt} * \eta_{inv} \tag{19}$$

Once these values are calculated and the inputs to the NN are made available, the next step consists of parsing the data so that only segments relevant to regenerative braking operations are used. Because a data driven model is based in pattern recognition, the training data must be as relevant to the desired pattern as possible. In this case, the desired pattern is regenerative braking, and therefore the data driven model should be trained with the measurement associated with a negative bus acceleration. To impose this requirement, the bus acceleration is approximated by computing a discretized derivative of its velocity. Then, the indexes k associated with

$$a_{veh}(k) \leq -0.2m/s^2 \tag{20}$$

are stored. The boundary was selected to avoid noise in the data close to zero acceleration. To guarantee that the data was the result of a sustained deceleration, only negative accelerations that met the condition in Eq. 18 for 4 consecutive timesteps are selected as part of the training dataset.

Once the intermediate variables θ are calculated using the component models and experimental data and then parsed to only consider negative accelerations, a comparison between θ and the complete model prediction is performed. The results of this comparison are shown in Figure 10. Overall, the two approaches show very similar trends, but are not an exact match. This is expected due to the compounding of modeling

errors when the complete PBM is used, together with some desynchronization due to the parsing.

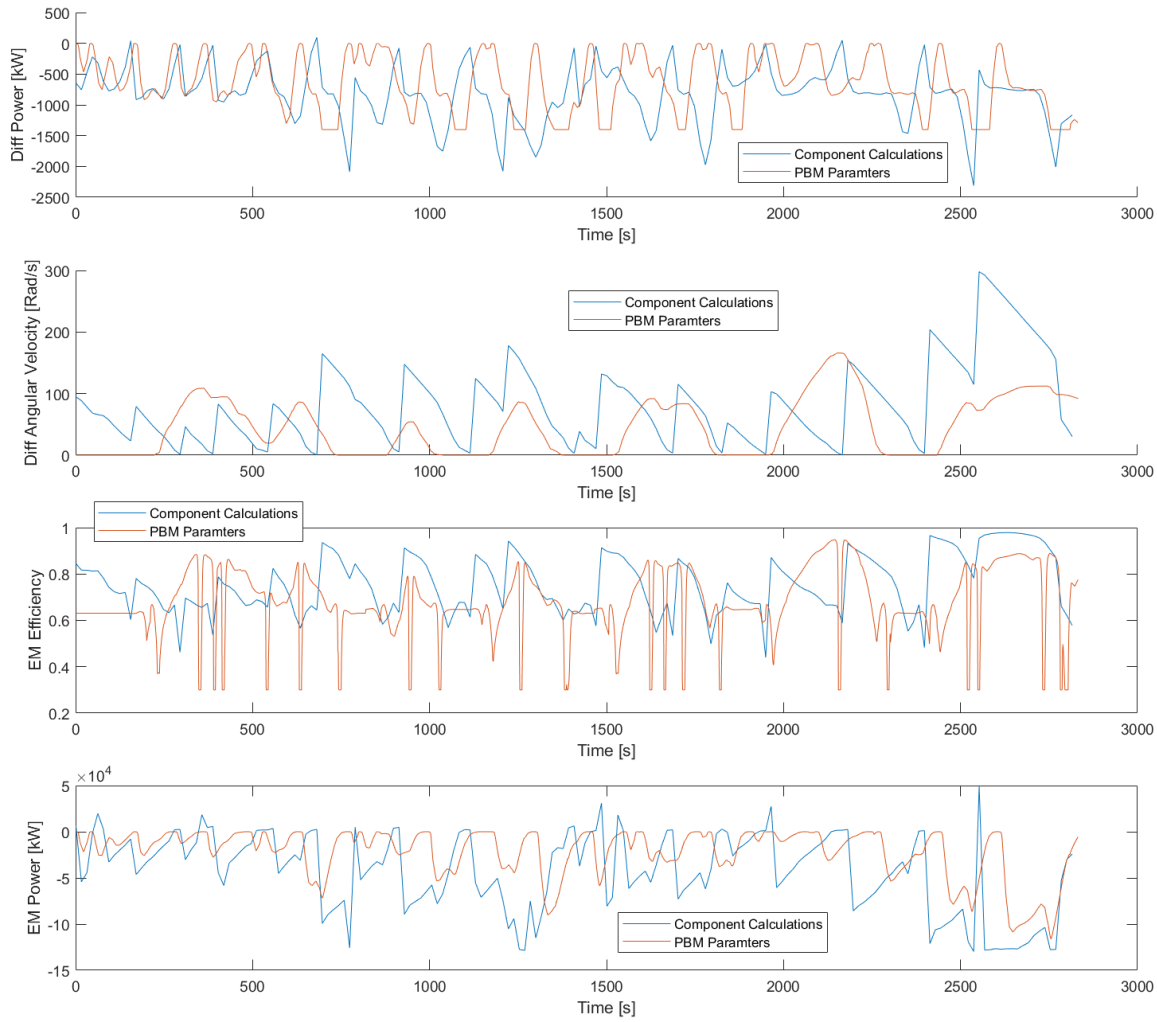

Figure 12: Component Calculated vs PBM Calculated Variables

Normalization approach

Good practice for training neural networks is to normalize the input and output data between 0 and 1:

$$\theta_n = \frac{\theta}{\bar{\theta}} \tag{21}$$

$$Y_{m,n} = \frac{Y_m}{\overline{Y_m}} \tag{22}$$

where $\bar{\theta}$ is the normalization factor for the inputs and $\overline{Y_m}$ is the factor for the outputs. Initially, the data was normalized based on the calculated training inputs for the Neural Network. This yielded good results in the network verification, but it becomes impractical for the integration with the model due to the difference in input ranges, as seen in Figure 10. For this reason, an alternate normalization of the inputs and output is used, which is based on the averages of the variables predicted by analytical model.

Network Training

The structure chosen for the neural network is a feed forward neural network (FNN) since the prediction is a regression problem [22]. For training, then testing against the validation set, the data was randomly split into 10 sections using a MATLAB function for random permutation. These 10 sections were then randomly divided into validation and verification pools, with 8 sections going to validation and 2 going to verification. The network tunable parameters and corresponding values are summarized in Table 4.

Table 4: NN tunable Parameters

Network Parameters	**Value**
Number of hidden layers	6 (1 Batch normalization, 1 dropout layer)

Size of fully connected layers 1-4	[100, 250, 200, 100]
Dropout rate	0.1
Minibatch size	4
Activation type	Leaky Relu
Max Epochs	60

The network parameters were selected by trial and error, retraining the network with various combinations of different parameters. The metric used to evaluate the success of a training was the RMSE of the validation and verification predictions as compared to the real values.

Results of Training

Figures 13 and 14 show the predictions of the final trained FNN.

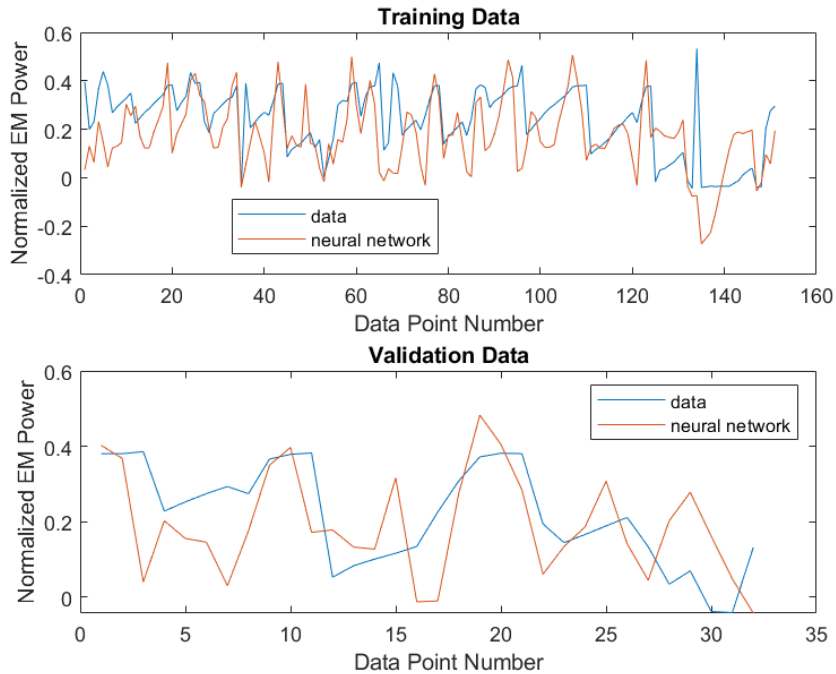

Figure 13: Network Results Compared to Data

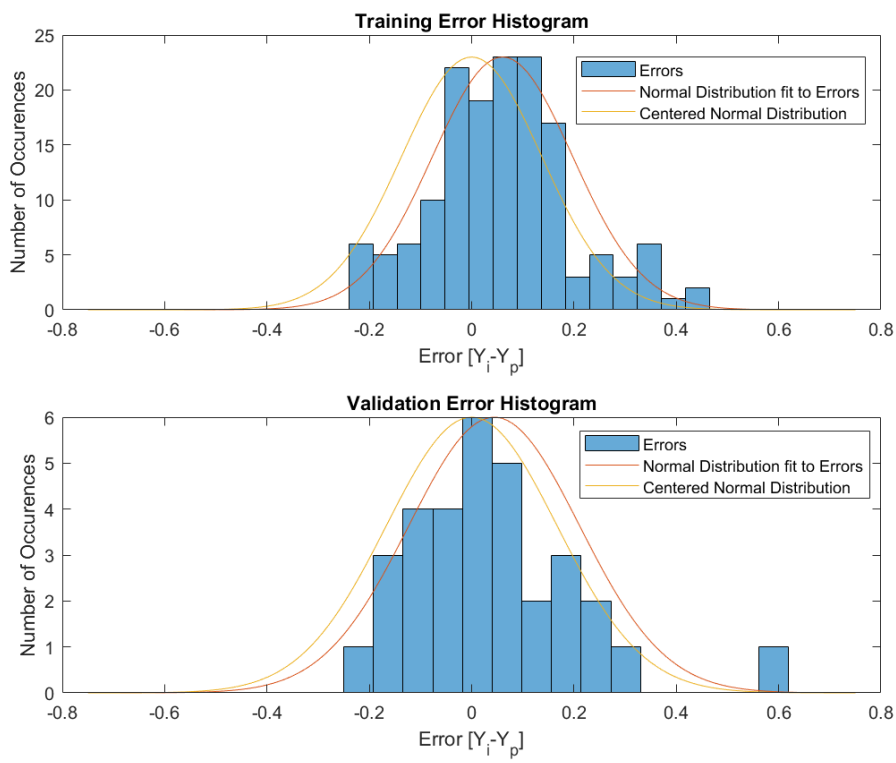

Figure 14: Trained network error histograms

Figure 14 shows that the network likely has some small systematic error, overestimating the regeneration of power from braking, but overall, the preliminary results for the trained network are promising. Next, this network will need to be integrated into the analytical model to assess its ability to improve the prediction of the regenerative braking in the electric bus. Assessing the training of the network by itself only shows that it can correctly predict the experimentally calculated values earlier and does not necessarily correlate to an improvement in the analytical model's accuracy.

Network Implementation in Simulink

Sample Model

To implement the network in the model, first the network must have an equivalent Simulink implementation, to match the PBM. To test the Simulink NN implementation, a sample model was created in Simulink that takes the network inputs and predicts the output variable just as the "Predict" function does. This also allows for testing of the normalization of the inputs and denormalization of the output.

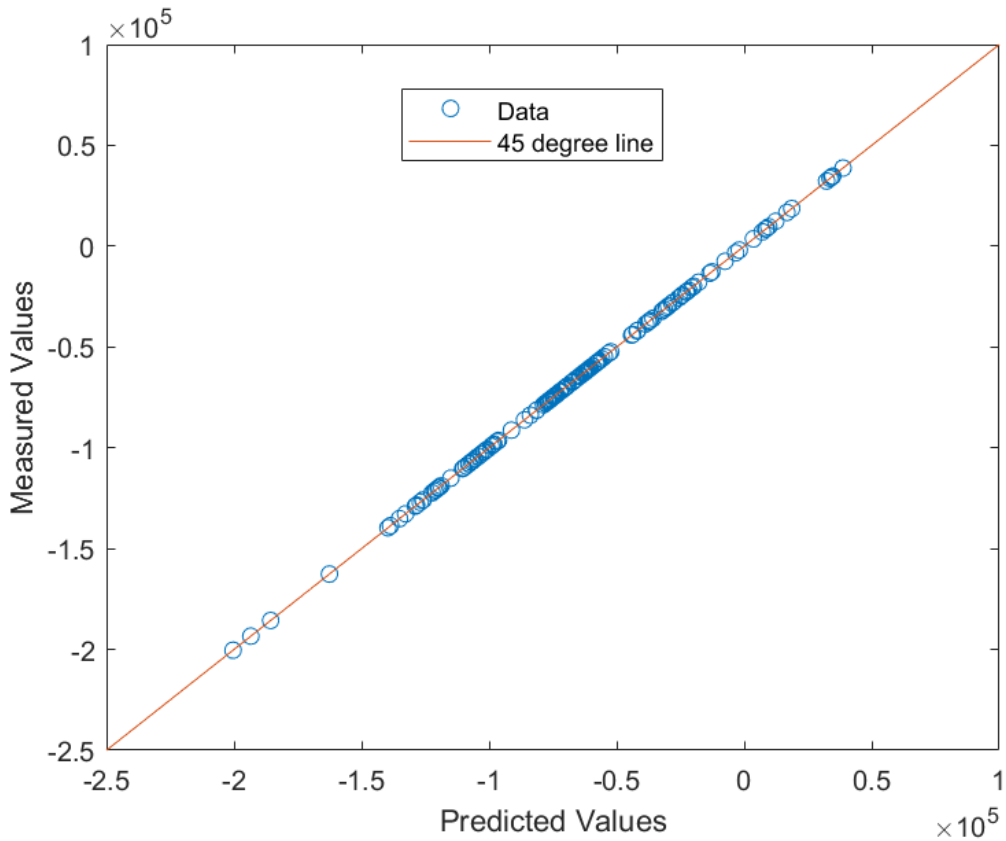

Figure 15: Sample Model Verification

The results of this sample model testing are shown in figure 15, where the Simulink implementation of the network matches the original MATLAB code.

Experimental Runs and Testing Metrics

The integration of the trained NN in the PBM requires the removal and replacement of the simple regen coefficient, as shown in Figure 11. The validation of the enhanced model is performed on the first set of dyno data. First the comparison is performed on the training set, then the model is used to predict the SoC in a driving cycle that was not part of the training. To determine the improvements to the model, three main metrics are considered:

- The error distribution of the battery current during regenerative braking.
    - This metric highlights the improvements achieved through the integration of the data driven method.
- The terminal SoC error
    - This metric quantifies the improvement of the model in the ability of predicting the bus driving range.
- The error distribution of the battery current during traction.
    - This metric ensures that the data driven model only impacts the regenerative braking portion of the vehicle operation.

Results

First, the model is assessed by its ability to predict validation data, the same dataset which was used during the training of the NN. A simulation run performed on

this data is shown in figure 16, comparing the analytical and combined model. The final SoC error in the original model is 0.16%, and the final SoC error in the Combined model is 0.55%.

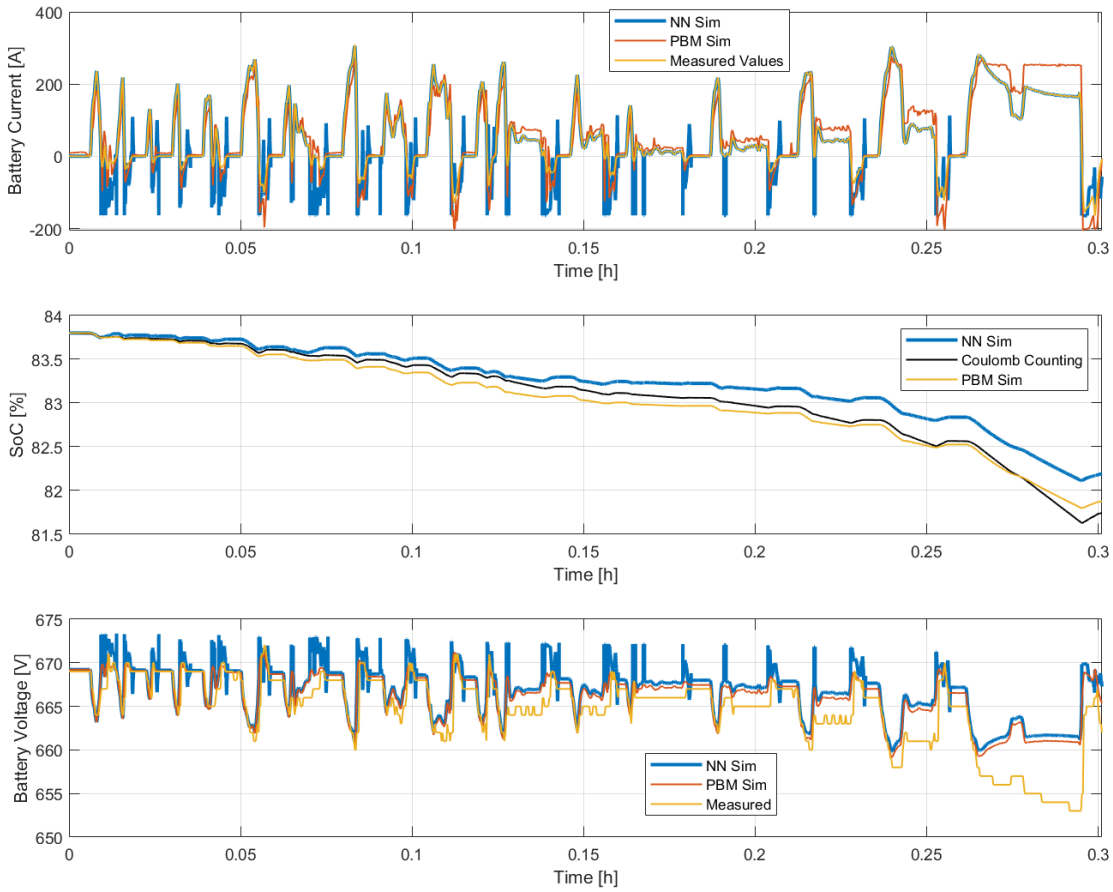

Figure 16: Model Verification Data Results

From this output, it is clear that both models provide a strong prediction of the bus behavior during the test on the dyno. Moreover, the error on the positive current prediction is shown in Figures XYZ, where the two models show the same behavior. This is expected as the objective of the NN is only to correct the negative current.

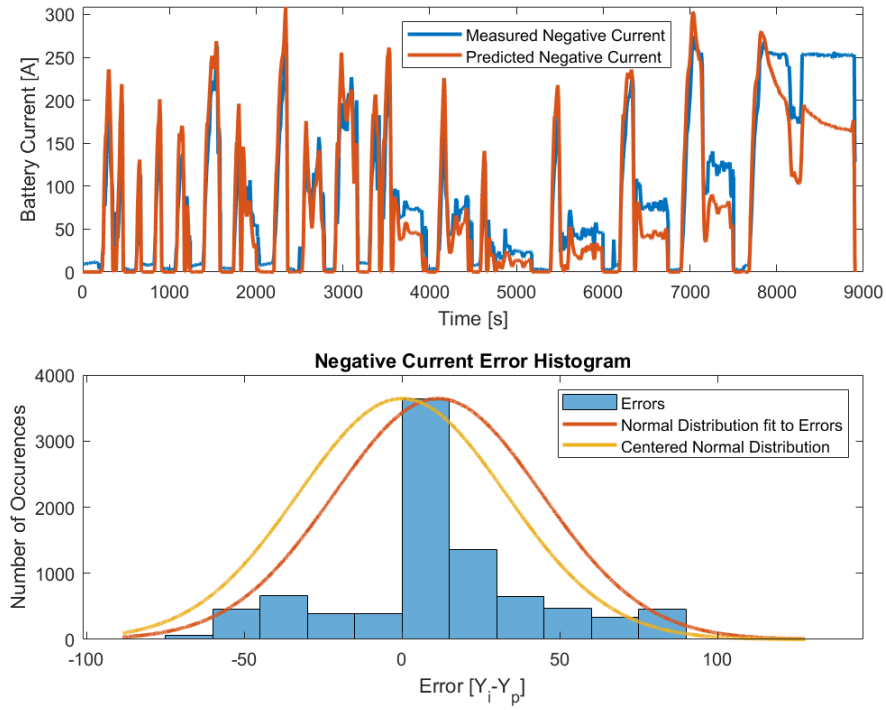

Figure 17: Physics Based Model Positive Current Error

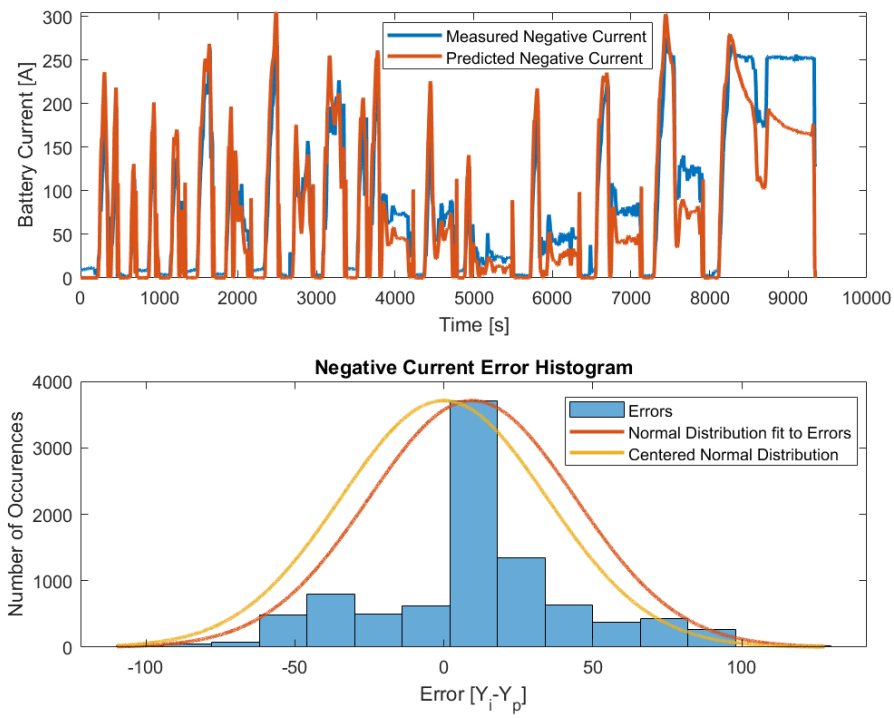

Figure 18: Combined Model Positive Current Errors

To assess the combined models' improvement to the regenerative braking prediction, the plots the negative portion of the battery are shown in Figures XYZ. As observed for the validation of the Neural Network alone, the error distribution of the battery current obtained from the NN-enhanced model follows approximately a normal distribution centered about zero. Some gaussian noise will always be expected in even the best predictions, therefore this visualization more clearly shows the performance of the model as opposed to the mean of the error.

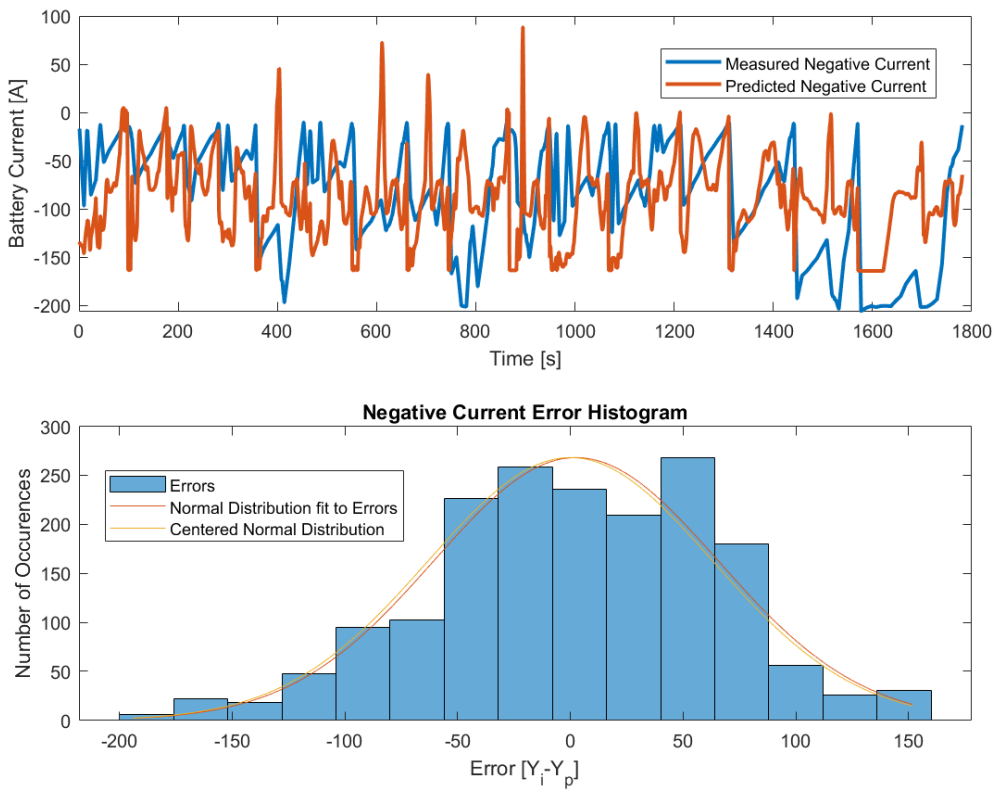

Figure 19: Combined Model Negative Current Errors

Moreover, these results show a clear improvement in the negative current prediction from the entirely physics-based model to the combined model. Though the magnitudes of

errors are slightly larger in the combined model, the mean of the distribution is closer to zero compared to the original implementation. Although the negative current predictions are better in the combined model, the SoC prediction is slightly worse. This is probably because, by correcting the errors in the negative part of the current predictions, there are no longer any errors to offset some of the errors in the positive predictions. In fact, when analyzing the results from the original model, both positive and negative error distributions showed a consistent shift indicating a tendency of the model to overestimate both negative and positive power. When the model was coupled, this effect was balanced. This feature is then lost in the combined mode, where the negative current portion is corrected while the traction is unmodified.

The model validation is done using the DRIVE data, these missions were not part of the training and therefore this provides an evaluation of the accuracy beyond the dataset. A simulation performed with the DRIVE data yielded the results shown in figure 16. The terminal SoC error for the analytical model was 1.57%, and the terminal SoC error for the combined model was 0.65%.

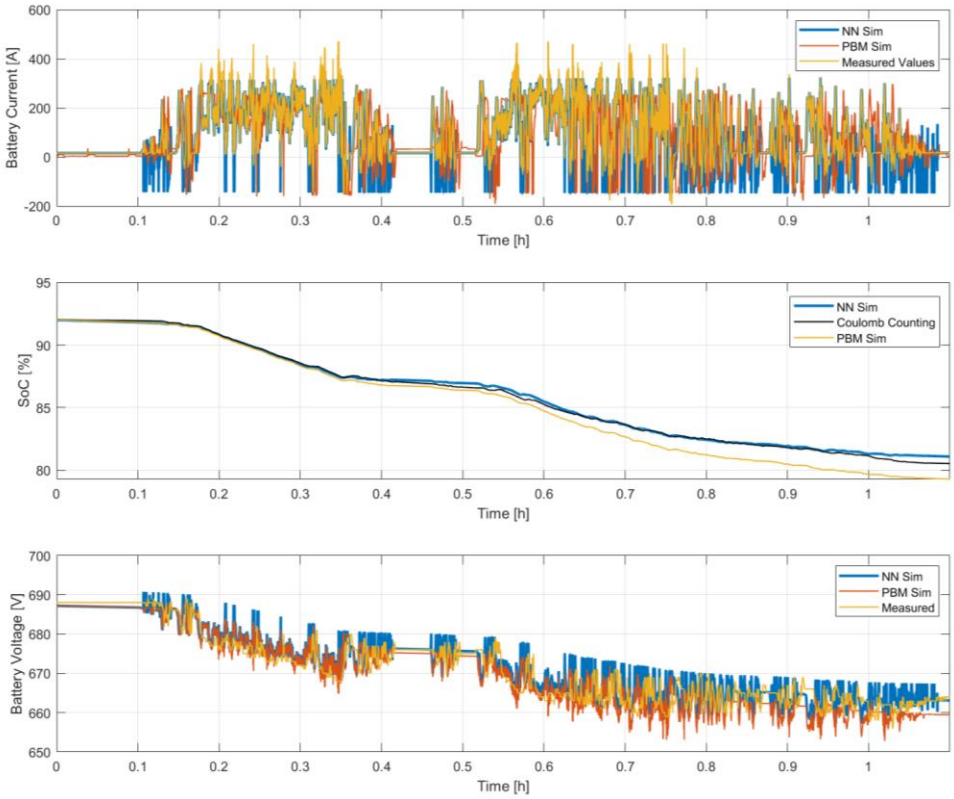

Figure 20: Model Validation Data Results

This shows a clear improvement in the modelling of the bus's range with the combined model. To verify that this improvement is due to the NN modeling the regenerative braking, the negative battery current predictions are analyzed.

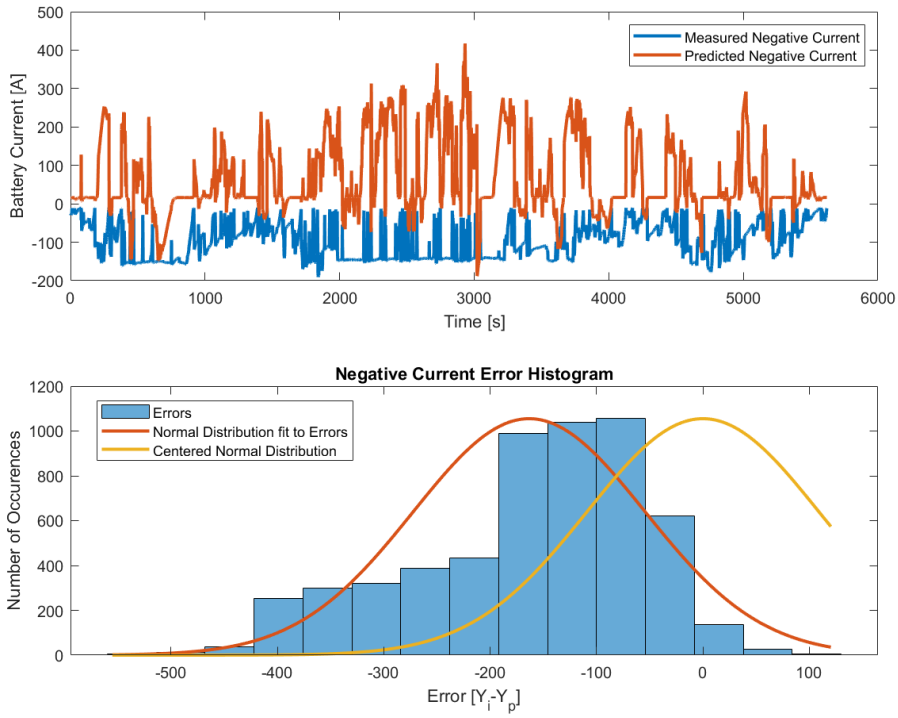

Figure 21: Analytical Model Negative Current Errors (Validation)

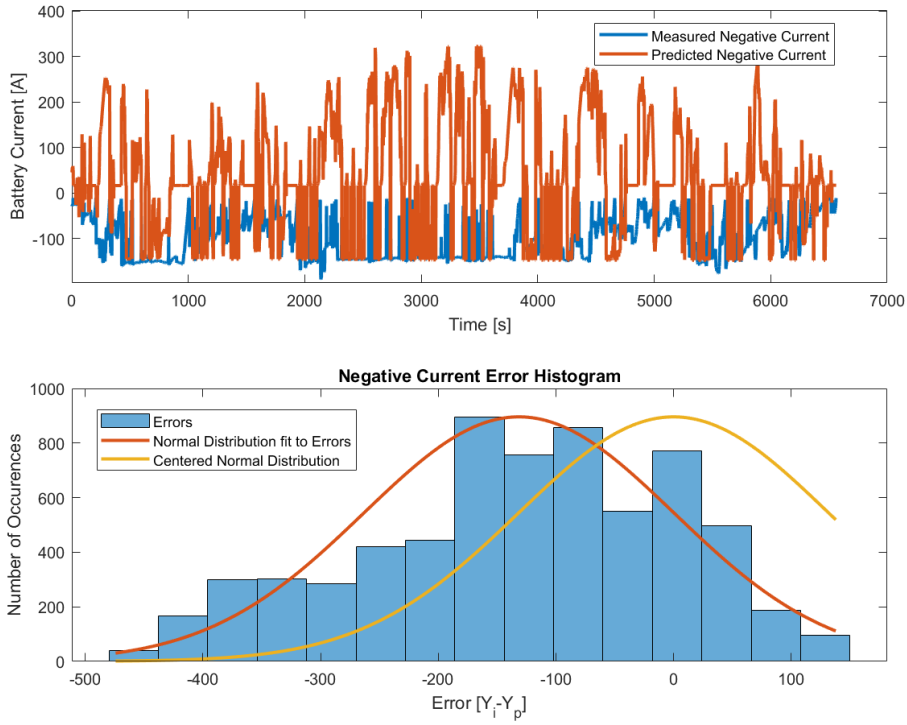

Figure 22: NN Model Negative Current Errors

Though the negative current errors are not as clearly improved as the verification set, there is still an improvement in the error distribution. It is clear from the battery current plots in figures 17 and 18 that the NN is better capturing the negative portion of the battery current. There is also a significant positive current prediction in the sections of the route data where a negative current is measured. This is possibly due to the much lower fidelity data quality in the drive data samples, as well as some possible desynchronization.

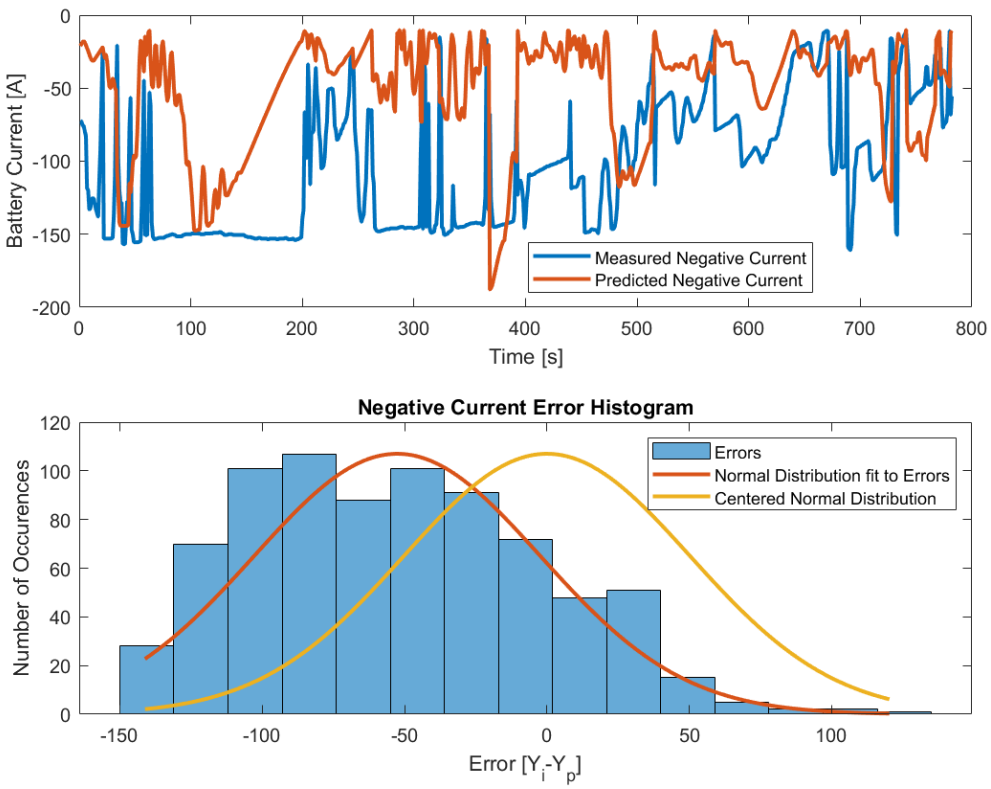

Figure 23: Analytical Model Negative Current Errors (Validation) Filtered for negative measured current and negative predicted current

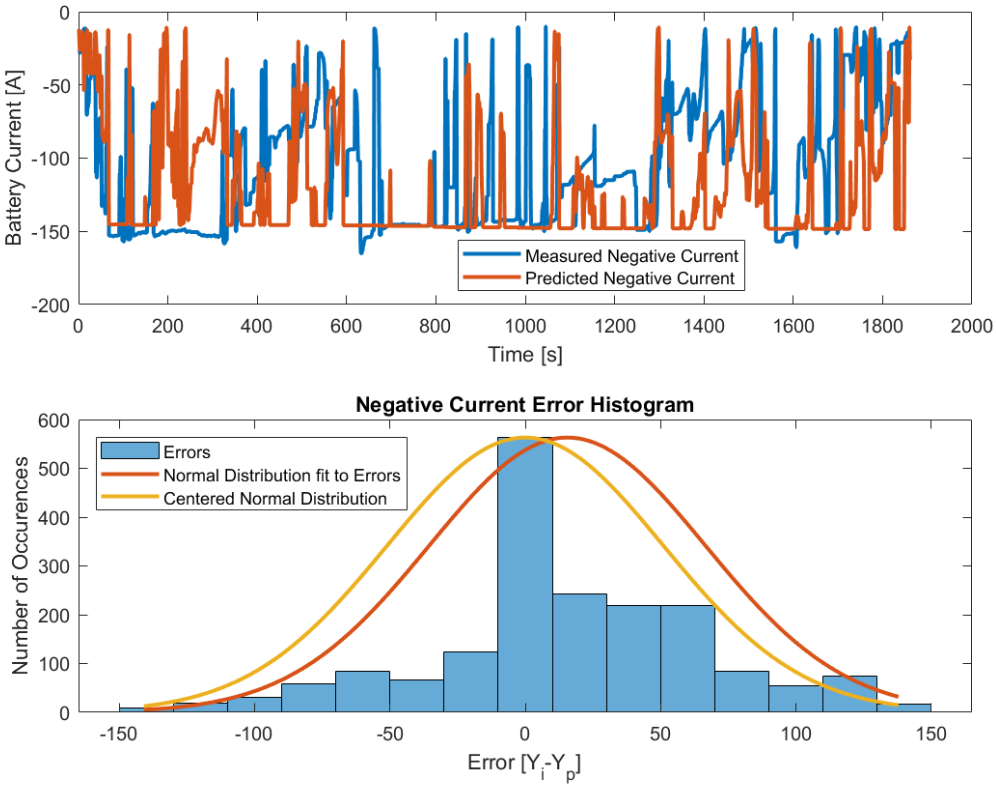

Figure 24: NN Model Negative Current Errors Filtered for negative measured current and negative predicted current

The combined model predicts better when the battery current would be negative in the bus, as it has over twice as many data points. Second, it shows that the error distribution is much better in its predictions and results in a much better prediction of the negative current in the bus for this validation set. The results data is summarized in Tables 5 and 6, with an energy based approach shown in tables 7-10.

Table 5: Terminal SoC Values of Simulation Outputs

	EXP SoC	PBM SoC	PBM Error	PIML SoC	PIML Error
Validation	0.8174	0.8188	0.16%	0.8219	0.55%
Verification	0.808	0.7928	1.57%	0.8107	0.65%

Table 6: Negative Current Errors of Simulation Outputs

	Model Type	Average Error [A]	Error STD [A]	Cumulative Abs Error
Validation	PBM	-40.04	22.933	6,487
	PIML	5.228	61.47	8,623
Verification	PBM	-52.579	51.033	4,829
	PIML	15.54	50.336	7,351

Table 7: Validation Energy Metrics

Metrics	Tot. Energy [kWh/mi]	Consumed [kWh/mi]	Regenerated [kWh/mi]	Terminal SoC
Test Data	2.1671	2.6936	0.5265	0.8174
PBM	2.0367	2.3442	0.3075	0.8188 (0.16% error)
PIML	1.7002	2.3994	0.6993	0.8219 (0.55% error)

Table 8: Validation Regen and Consumption Power Errors

Metrics	Consumed [kWh/mi]	Consumed Error	Regenerated [kWh/mi]	Regen Error
Test Data	2.6936		0.5265	
PBM	2.3442	-14.57%	0.3075	-41.61%
PIML	2.3994	-12.7%	0.6993	32.79%

Table 9: Verification Energy Metrics

Metrics	Tot. Energy [kWh/mi]	Consumed [kWh/mi]	Regenerated [kWh/mi]	Terminal SoC
Test Data	2.1154	2.3963	0.2810	0.8080
PBM	2.3434	2.4382	0.0948	0.7928 (1.57% error)
PIML	2.0154	2.4250	0.4097	0.8107 (0.65% error)

Table 10: Verification Regen and Consumption Power Errors

Metrics	Consumed [kWh/mi]	Consumed Error	Regenerated [kWh/mi]	Regen Error
Test Data	2.3963		0.2810	
PBM	2.4382	1.17%	0.0948	-66.26%
PIML	2.4250	1.18%	0.4097	45.8%

Chapter 5: Conclusions and future work

Accuracy Improvements

By implementing a physics informed machine learning approach to modeling the range of a COTA electric bus, a clear improvement in modeling the regenerative braking in the system was found, as compared to a traditional analytical approach. In the validation set, there was a significant improvement in the terminal SoC error, which directly correlates to an improvement in range prediction of the electric bus. This was also shown to come from the better current predictions in the regenerative braking part of the predictions.

Success Despite Low Fidelity Data

This approach to PIML was done under a scarce data conditions. The training set had only 183 elements for the neural network to learn from. Despite this, the trained network was able to be integrated into the analytical model and provide a clear improvement to regenerative braking modeling in the bus. Due to the nature of data driven modeling, having more data to train and test with would only further improve the modeling capabilities of such a combined model.

By combining a data driven model with an analytical framework, the sparse dataset did not have a prohibitive effect on the model. Despite some point errors, and inconsistent behavior of the neural network, the analytical framework which it was informing resulted in these errors being insignificant to the overall range prediction.

Future Work

Incremental improvements to current model

Though the results yielded by the combined model are very good, there are several clear opportunities for incremental improvement to the model:

The current model was trained with parsed data which was at a lower acceleration than -0.2m/s^2 to avoid noise close to 0 measured acceleration. With the current dataset, more extensive filtering could be done to avoid some of the noise close to 0 acceleration, rather than avoiding the problem. This would allow the network to train on more data close to 0 acceleration and result in a more accurate regen estimate at low changes in acceleration.

Another possible improvement, specifically in the data driven part of the combined model, would be to build some analytical constraints into the network itself, essentially adjusting the loss function based on physical parameters. One such example would be placing a high emphasis on outputting a negative P_{EM} in cases with negative torque (a physically logical result).

Future model development

If additional data were collected for the purpose of tuning a combined model, there are some distinct properties which should be a focus for data collection:

- Wide variation in data conditions collected
    - Allows better training of MLM, and results in less extrapolation for PIML implementation

- Consistent properties of collected data, including sample time, data synchronization, and post collection processing
- Identification of desired MLM inputs and outputs prior to data collection
    - Allows testing to be designed in such a way to maximize the correlation between inputs and outputs for the desired pattern to be recognized by the MLM.

Should these properties be applied to a regenerative braking case study, additional dyno testing would include a focus on various braking patterns applied during coast down tests. It would also include a measurement of the pedal position synchronized with other data collection. The specific design of these tests could be better determined by a study of the current combined model weaknesses when faced with subsections of current verification data. This would allow a data collector to determine what range of inputs would be best suited to improving MLM training.

Bibliography

[1] Pan, Chaofeng, et al. "Driving range estimation for electric vehicles based on driving condition identification and forecast." *AIP Advances* 7.10 (2017): 105206.

[2] Hong, Joonki, Sangjun Park, and Naehyuck Chang. "Accurate remaining range estimation for electric vehicles." *2016 21st Asia and South Pacific design automation conference (ASP-DAC)*. IEEE, 2016.

[3] Papa, Gregor, Marina Santo Zarnik, and Vida Vukašinović. "Electric-bus routes in hilly urban areas: Overview and challenges." *Renewable and Sustainable Energy Reviews* 165 (2022): 112555.

[4] Eisel, M., Nastjuk, I., Kolbe, L.M., 2016. Understanding the influence of in-vehicle information systems on range stress – Insights from an electric vehicle field experiment. Transp. Res. Part F Traffic Psychol. Behav. 43, 199–211.

[5] Franke, T., Krems, JF, 2013. What drives range preferences in electric vehicle users? Transp. Policy 30, 56–62

[6] Kubička, M., Sciarretta, A.S., Cela, A., Mounier, H., Thibault, L., Niculescu, S.I., 2016. About prediction of vehicle energy consumption for eco-routing, in: IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC. Institute of Electrical and Electronics Engineers Inc., Rio de Janeiro, Brazil, pp. 1096–1101. https://doi.org/10.1109/ITSC.2016.7795693

[7] Lamantia, M., Su, Z., Chen, P., 2021. Remaining Driving Range Estimation

Framework for Electric Vehicles in Platooning Applications, in: Proceedings of the American Control Conference. Institute of Electrical and Electronics Engineers Inc., pp. 424–429.

[8] Scheubner, S., Thorgeirsson, A.T., Vaillant, M., Gauterin, F., 2019. A Stochastic Range Estimation Algorithm for Electric Vehicles Using Traffic Phase Classification. IEEE Trans. Veh. Technol. 68, 6414–6428

[9] Shen, H., Wang, Z., Yang, K., Lamantia, M., Chen, P., Wang, J., 2021. Comparison of Different Variable Combinations for Electric Vehicle Power Prediction Using Kernel Adaptive Filter, in: Modeling, Estimation and Control Conference. pp. 858–863.

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

[11] Aykol, M., Gopal, C. B., Anapolsky, A., Herring, P. K., van Vlijmen, B., Berliner, M.D., Bazant, M. Z., Braatz, R. D., Chueh, W. C., & Storey, B. D. (2021). Perspective—Combining Physics and Machine Learning to Predict Battery Lifetime. *Journal of The Electrochemical SoCiety*, *168*(3), 030525. https://doi.org/10.1149/1945-7111/abec55

[12] Zhu, Q., Sun, Z., Liang, X., Xiong, Y., & Zhang, L. (2020). OCoR: Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems. *Proceedings - 2020 35th IEEE/ACM International Conference on*

Automated Software Engineering, ASE 2020, 883–894. https://doi.org/10.1145/3324884.3416530

[13] Chao, M. A., Kulkarni, C., Goebel, K., & Fink, O. (2020). *Fusing Physics-based and Deep Learning Models for Prognostics*. http://arxiv.org/abs/2003.00732

[14] West, Brian H., et al. "Development of data-based light-duty modal emissions and fuel consumption models." *SAE transactions* (1997): 1274-1280.

[15] Zhang, Yuhe, et al. "Remaining driving range estimation of electric vehicle." *2012 IEEE International Electric Vehicle Conference*. IEEE, 2012.

[16] Vepsäläinen, Jari, et al. "Development and validation of energy demand uncertainty model for electric city buses." *Transportation Research Part D: Transport and Environment* 63 (2018): 347-361.

[17] Pevec, Dario, et al. "Electric vehicle range anxiety: an obstacle for the personal transportation (r) evolution?." *2019 4th international conference on smart and sustainable technologies (splitech)*. IEEE, 2019.

[18] Namburu, Setu Madhavi, et al. "Systematic data-driven approach to real-time fault detection and diagnosis in automotive engines." *2006 IEEE Autotestcon*. IEEE, 2006.

[19] Guzzella, Lino, and Antonio Sciarretta. *Vehicle propulsion systems*. Vol. 1. Springer-Verlag Berlin Heidelberg, 2007.

[20 Rizzoni, Giorgio, Lino Guzzella, and Bernd M. Baumann. "Unified modeling of hybrid

electric vehicle drivetrains." *IEEE/ASME transactions on mechatronics* 4.3 (1999): 246-257.

[21] Grunditz, E.. “Design and Assessment of Battery Electric Vehicle Powertrain, with Respect to Performance, Energy Consumption and Electric Motor Thermal Capability.” (2016).

[22] AGGARWAL, CHARU C. *Neural Networks and Deep Learning: A Textbook*. SPRINGER INTERNATIONAL PU, 2023.

[23] Kaiser E, Kutz JN, BruntonSL. 2018 Sparse identification of nonlineardynamics for model predictive control in the low-data limit.Proc.R.SoC.A474: 20180335.http://dx.doi.org/10.1098/rspa.2018.0335

[24] Brunton, Steven L., et al. “Discovering Governing Equations from Data by Sparse Identification of Nonlinear Dynamical Systems.” *Proceedings of the National Academy of Sciences*, vol. 113, no. 15, 2016, pp. 3932–3937., https://doi.org/10.1073/pnas.1517384113.

[25] “Simscape Electrical.” MATLAB & Simulink, MathWorks, https://www.mathworks.com/products/simscape-electrical.html.

Unified modeling of hybrid electric vehicle drivetrains." *IEEE/ASME transactions on mechatronics* 4.3 (1999): 246-257.

[26] K. Manohar, B. W. Brunton, J. N. Kutz and S. L. Brunton, "Data-Driven Sparse Sensor Placement for Reconstruction: Demonstrating the Benefits of Exploiting Known Patterns," in *IEEE Control Systems Magazine*, vol. 38, no. 3, pp. 63-86, June 2018, doi: 10.1109/MCS.2018.2810460.

[27] Zhang, Ruifeng, Bizhong Xia, Baohua Li, Yongzhi Lai, Weiwei Zheng, Huawen Wang, Wei Wang, and Mingwang Wang. 2018. "Study on the Characteristics of a High-Capacity Nickel Manganese Cobalt Oxide (NMC) Lithium-Ion Battery—An Experimental Investigation" Energies 11, no. 9: 2275.https://doi.org/10.3390/en11092275