Predicting Degradation of Fuel Cell Membrane Electrode Assemblies from Buses

Harsh Srivastav
Alexandre Bayen

Electrical Engineering and Computer Sciences
University of California, Berkeley

Technical Report No. UCB/EECS-2025-107
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2025/EECS-2025-107.html

May 16, 2025

Predicting Degradation of Fuel Cell Membrane Electrode Assemblies from Buses

by Harsh Srivastav

Research Project

Submitted to the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, in partial satisfaction of the requirements for the degree of **Master of Science, Plan II**.

Approval for the Report and Comprehensive Examination:

Committee:

Professor Alexandre Bayen
Research Advisor

May 15, 2025

(Date)

* * * * * * *

Professor Clayton J. Radke
Second Reader

May 16, 2025

(Date)

Predicting Degradation of Fuel Cell Membrane Electrode Assemblies from Buses

by

Harsh Srivastav

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Masters of Science

in

Electrical Engineering and Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Alexandre Bayen, Chair

Spring 2025

Predicting Degradation of Fuel Cell Membrane Electrode Assemblies from Buses

Abstract

Predicting Degradation of Fuel Cell Membrane Electrode Assemblies from Buses

by

Harsh Srivastav

Masters of Science in Electrical Engineering and Computer Science

University of California, Berkeley

Professor Alexandre Bayen, Chair

The Alameda Contra Costa Transit District (AC Transit) has provided lifetime data for a limited fleet of their buses that were run on hydrogen fuel cells. Opportunities to study such longitudinal data in clean energy transportation systems are uncommon, particularly over the complete operational lifespan of vehicles. This thesis leverages that dataset to investigate whether early-life performance indicators can be used to accurately predict long-term degradation behavior of fuel cell systems. We have run machine learning models on the predictions of the data to study whether later life degradation behavior can be accurately estimated from earlier time performance. To do so, we have explored using recurrent neural network type models, with variations in both the loss function and the data being trained upon. Long term predictions (over the course of years) are also presented for multiple buses with trends analyzed. These findings have implications for predictive maintenance, fleet management, and the broader deployment of hydrogen-powered transportation systems. Additionally, we discuss the limitations of the modeling approach and suggest future directions for improving predictive accuracy with hybrid modeling strategies and additional contextual data.

Contents

List of Figures

Acknowledgments

I will be the first to admit how surprised I am this thesis was written. I want to express my sincere gratitude to several people in fact, without whom this thesis would not exist. First, Professor Ali Mesbah, who admitted me to the PhD program at Berkeley for Chemical and Biomolecular Engineering without having published in any journals simply because he believed in me from my application essay. In the beginning of my PhD, I started earlier than most in taking courses outside my field, initially in Mathematics. Prof. Bryan McCloskey is to be heavily thanked in this regard for allowing me to take math classes in my first year, pushing back Chemical Engineering requirements, and helping me realize early that math courses were not ones I wanted to continue taking.

Instead, I began taking computer science courses towards the end of my second year. My experience taking CS 61B that Fall with Prof. Josh Hug confirmed I made the correct decision. Around this point, the CS department at Berkeley became (or perhaps had been for some time?) drastically underfunded, leading the department to make the difficult decision to prevent graduate students from other departments from taking the undergraduate classes, including me. The graduate classes were still open, however, many of them in machine learning, so I began taking those instead. Several such classes later and I had unintentionally completed the requirements for the Master's in EECS except for the thesis...and applying.

Many of the details of how I fell in this situation are worth mentioning. My roommate, Kushal Nimkar, was the one who found this opportunity to apply for the Master's program as a PhD student in another department. Someday, perhaps, I will be able to repay him for how his short remark one evening significantly changed the trajectory of my life. In the search for advisors, I initially approached Prof. Gireeja Ranade, who pointed me towards my current advisor, Prof. Alex Bayen. Through his generosity, he was willing to advise me after a five minute conversation where I waited outside his office and walked him from one meeting to another. Initially, when I applied for the Master's program, I was mistakenly rejected and then later accepted after the error was corrected — Patrick Hernan was invaluable in noticing and fixing the mistake.

My PhD advisors, Prof. Clayton J. Radke and Adam Z. Weber, were generous enough to allow me to devote a not insignificant portion of my time towards a field neither of them were particularly familiar with. The timing was quite fortunate as we had been sent a large amount of data on fuel cell buses with no prior planned projects within the group. To this end, I would like to thank Jose Vega, James Souza, and Cecil O. Blandon from the Alameda Contra Costa Transit District for their help in providing the data taken over the lifetime of the fuel cell buses. Rangachary Mukundan was extremely patient and helpful in guiding me through interpretations of the data and their implications.

It should go without saying, but I deeply appreciate the friends and collaborators I have made in Berkeley, in research and beyond, regardless of whether they are in Chemical Engineering, Computer Science, or neither. This thesis represents not simply my efforts, but the tiny little efforts of many of them.

Chapter 1

Introduction

Neural networks have revolutionized predictive modeling capabilities for machine learning, offering unparalleled methods of mimicking complex relationships contained in data. Their power includes the ability to match large ranges of functions accurately based on inputs through layers of abstraction to make accurate predictions for new targets. Indeed, sufficiently dense multilayer neural networks can capture any Borel measurable function to any desired degree of accuracy, making them universal function approximators.[1] This theoretical guarantee has been a cornerstone in justifying the application of neural networks to a wide variety of predictive tasks, from time-series forecasting to classification and regression problems. The back-propagation algorithm paved the way for making the training of such networks more efficient and feasible.[2] Although the underlying function being used for training can be resembled closely, this does not mean the overall function captures the real world trend as intended. This effect is usefully captured by the Bias-Variance Tradeoff, which details limitations in the predictiveness of the model due to simplifying assumptions (bias) and inaccuracies due to the data trained upon (variance).[3] Many techniques were invented to reduce the difference between the predicted function and reality, such as Lasso, Ridge Regression, Batch Renormalization, and more.[4, 5, 6, 7]

In practice, deep neural networks, which allow the approximation of complex behavior arising from large datasets, yield incredible performance in predictive analytics. Residual neural networks (ResNets) became especially helpful for image processing/recognition and denoising where the output looked similar to the input.[8] Convolutional Neural Networks (CNNs) became a standard for predicting features obtained from image processing.[9] Moving beyond simple predictions, modeling time implicitly was introduced through the development of Recurrent Neural Networks (RNNs), which demonstrated how recurrent connections can capture temporal structures in sequential data.[10] RNNs were trained using the same gradient descent/back-propagation algorithm for original fully-connected (dense) neural networks, which led to their own set of problems in certain applications. Exploding gradients still posed an issue (somewhat alleviated by normalization), but dying gradients were largely prevented using Long-Short Term Memory Networks (LSTMs) which introduced forget gates to determine the extent of remembering long-term dependencies. In this way, LSTMs combined

ideas incorporated by both RNNs and ResNets. In the world of natural language processing (NLP), models using transformer ideas (based on principles of deep neural networks) have set new records for machine translation, image generation, classification, summarization, and more.[11, 12, 13, 14, 15, 16]

Neural networks have also exceled in domains where traditional statistical modeling has struggled, particularly because of their robustness to noisy and unreliable subsets of data using the techniques mentioned above. In particular, sequential data (where the output depends on a past series of inputs) is often particularly noisy and requires careful handling. In stock forecasting, RNNs and LSTMs have been demonstrated to have improved accuracy over memory-free models, including random forests and logistic regression.[17] Taking in extensive weather and storm data, neural network models are increasingly used for weather predictions and mapping storm trajectories, providing accurate 10-day forecasts rapidly.[18, 19] Researchers have also developed models for predicting clinical events and mortality risk via RNNs, outperforming several baselines.[20, 21]

Hydrogen-based conversion and utilization hold promise for mitigating and severely reducing significant issues that arise from traditional non-renewable energy sources, including harmful emissions.[22] Hydrogen-based conversion technologies (including fuel cells and electrolyzers) offer numerous benefits, such as high energy efficiency, environmental sustainability, operational flexibility, energy security, high energy density, and grid independence.[23, 24] To promote advances in the hydrogen economy, the Department of Energy has developed consortia related to specific efforts involved, from hydrogen fuel cells (Million Miles Fuel Cell Truck (M2FCT)), hydrogen production (HydroGen and H2NEW), and efficient larger-scale manufacturing processes (Roll-to-Roll (R2R)).[25, 26, 27, 28] This multi-faceted approach is considered essential due to the significant challenges hydrogen faces in becoming a viable alternative energy source.

In this dissertation, we focus on the efforts of the M2FCT consortium, which in part focuses on the lifetime durability of Proton-Exchange Membrane Fuel Cells (PEMFCs). The M2FCT consortium has set a target of achieving a hydrogen fuel cell system durability equivalent to powering a heavy-duty vehicle for one million miles. A simple schematic of the cross section of a hydrogen fuel cell is depicted in Figure 1.1.

This ambitious goal aims to significantly enhance the commercial viability of hydrogen fuel cells in long-haul trucking applications. Obtaining this target would represent a five-fold increase in the longevity of current combustion-engine driven trucks, making hydrogen fuel cells a more practical and cost-effective alternative that would take over the trucking industry. Additionally, while heavy duty vehicles such as trucks and buses make up less than 10% of overall vehicles present on roads, they comprise up to 25% of overall emissions and fuel consumption, delineating an important mitigation target.[29] With almost all of the major automakers investing in FCV development as well Toyota, Hyundai, and Honda as offering commercial FCV automobiles, along with governments worldwide promoting hydrogen energy through subsidies and initiatives, the industry is at a pivotal moment.[30, 31, 32]

In addition to partnerships with trucking manufacturing companies, M2FCT also maintains a tie to the Alameda Contra Costa Transit District (AC Transit) with its fuel cell

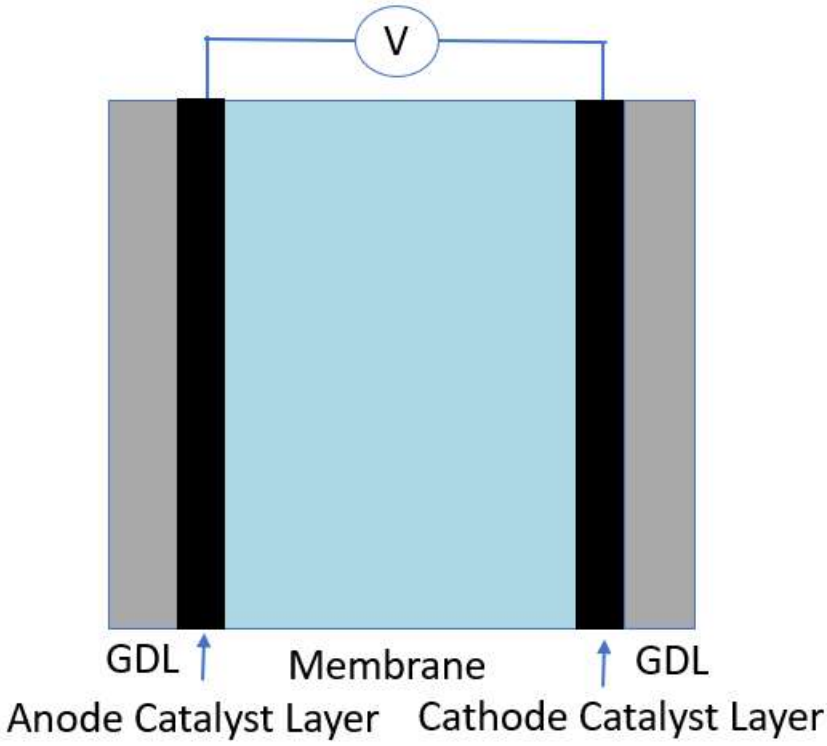

Figure 1.1: A simple schematic of a proton-exchange hydrogen fuel cell (not drawn to scale). The gas diffusion layers (GDLs) represent porous components where the reacting gases enter while the catalyst layers constitute electrodes where the gases react. The membrane is constructed of a special material that forces only positively charged groups through, forcing the electrons through the external circuit drawn overhead.

buses. AC Transit operates a fleet of 626 buses to transport residents in the Alameda and Contra Costa counties of California. Under the Legacy Fuel Cell Bus Study, 13 of these buses (labeled FC4 – FC16, "FC" for fuel cell) were hybrid models, running on both battery technology and hydrogen fuel cell stacks.[33, 34] This included an initial deployment of three Van Hool A330 fuel cell buses in 2006, followed by the addition of 10 more Van Hool fuel cell buses between 2010 and 2011.[35] The average operational lifespan of the Legacy fleet was 25,763 hours, with three buses exceeding 30,000 hours of operation without the need for major maintenance. Performance metrics for these buses were recorded during operational maintenance visits over the courses of their lifetimes (until the buses were decommissioned). A visual of a hydrogen fuel cell stack is present in Figure 1.2.

Hydrogen-fueled FCVs offer a pathway toward a zero-emission future, but their success will depend on overcoming key technical and economic barriers including the lack of data and understanding of the durability of the cells and components over time in real-world

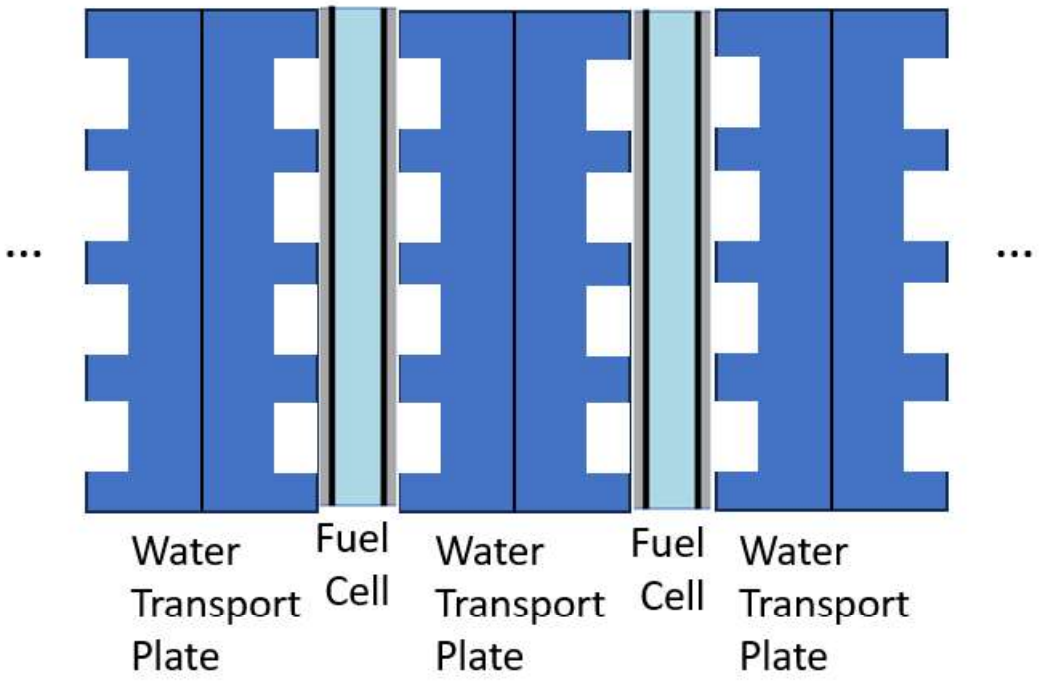

Figure 1.2: The layout of a hydrogen fuel cell stack (layers not drawn to scale). The fuel cells present are (with minor modifications) the same as in Figure 1-1 while the water transport plates (WTPs) build up voltage through providing conductive pathways while also allowing for removal of water generated within the cells. The inset cuts out of the WTPs represent channels for water to exit the system and reactant species to enter.

operation. This work aims to bridge the gap in mastery over the comprehension of predicting the performance of fuel cell systems over their lifetimes. The dataset is quite unique and has not been studied in this context prior. Multiple neural network architectures, including LSTMs and transformer-based networks, have been utilized for comparison in this task. These networks forecast the state of the fuel cell stacks after a certain period of time, not necessarily limited by the training data. Various enhancements are introduced to improve prediction accuracy, and their impacts are systematically assessed. The findings of this work have the potential to inform the design, operation, and maintenance strategies for the next generation of zero-emission vehicles.

Chapter 2

Background

The performance of fuel cell systems is often characterized in terms of polarization curves, which provide a graphical representation of the applied voltage and output current of a fuel cell under its operation conditions. These curves are invaluable in identifying key power metrics and delineating potential losses associated with various electrochemical phenomenon within the cells. An example of a polarization curve with the different loss regimes is depicted in Figure 2.1.In a typical polarization curve, the cell voltage, in volts (typically normalized by the number of cells present in the system), is plotted on the y-axis against the current density $[A/cm^2]$, where the relevant area is the cross-sectional area of the fuel cells or stack (assumed constant). The curve displays three prominent features, indicative of the different dominant potential loss mechanism.

At low current densities, the curve drops steeply due to activation losses. These activation losses arise largely from two different sources: hydrogen crossover from the anode through the membrane to the cathode and Butler-Volmer activation kinetics. Hydrogen crossover occurs due to slow reaction kinetics in the anode, resulting in an excess in the reactant gas, that permeates through the membrane and reacts on the cathode instead. Over the course of the lifetime of a fuel cell, this permeation eventually leads to further degradation of the membrane.[36] The Butler-Volmer equation details how the current density through a reaction at an electrode is impacted by the voltage bias in the solid phase:

$$i = i_0 \left[\exp\left(\frac{\alpha_a F(V - U_r)}{RT}\right) - \exp\left(-\frac{\alpha_c F(V - U_r)}{RT}\right)\right] \tag{2.1}$$

where i_0 is the exchange current density, α_a,α_c are the anodic and cathodic transfer coefficients (respectively), U_r is the equilibrium cell voltage (1.23 V for ORR), $R = 8.314$ J/(mol-K) is the ideal gas constant, $F = 96,485$ C/mol e^- is Faraday's constant, and T is the temperature. Thus, for current to flow as a result of the HOR and ORR, logarithmic drop in the voltage must occur.

At intermediate current densities instead, the voltage decline becomes linear, given by Ohm's fundamental relation,

$$V = iAr \tag{2.2}$$

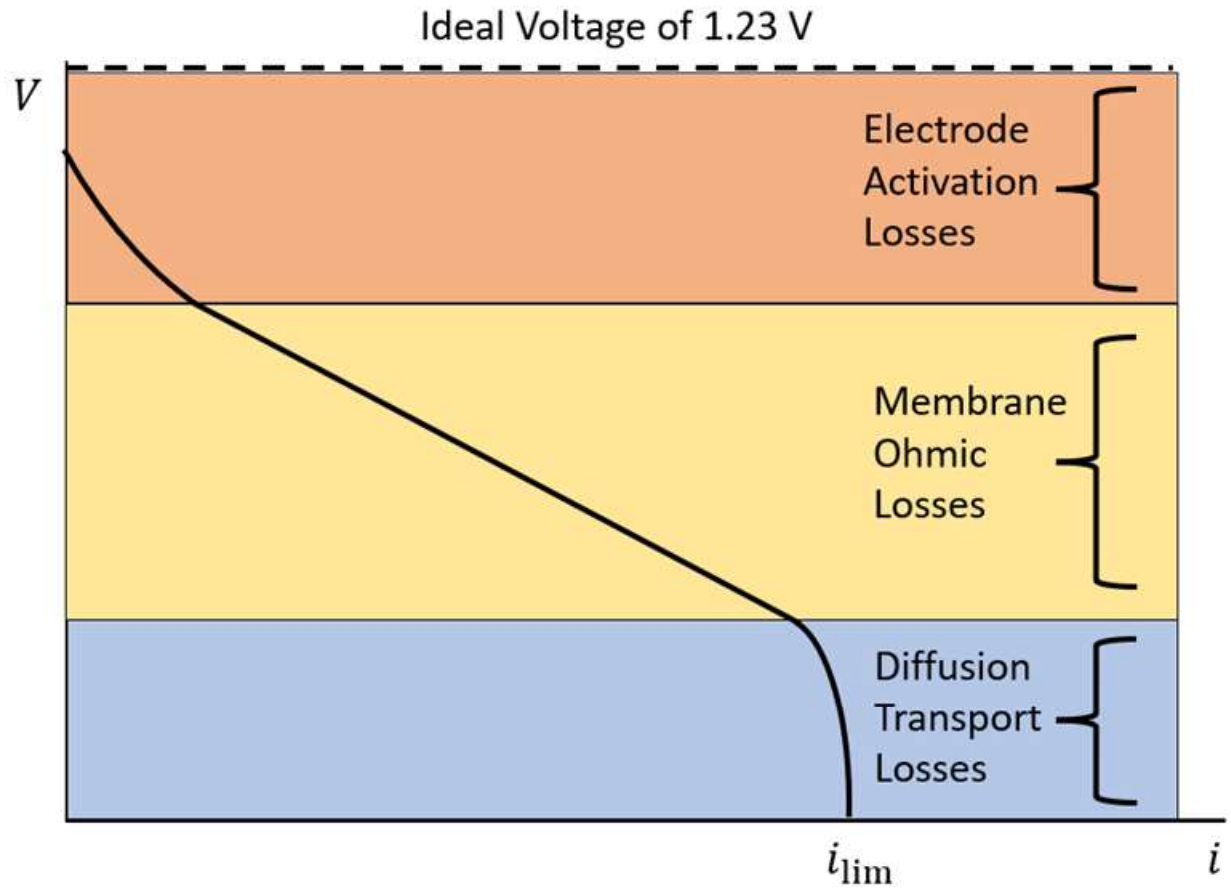

Figure 2.1: The polarization curve shows the average voltage of a fuel cell, component, or associated system, V, against the current density of the cross-sectional area, i. The activation, ohmic, and transport loss regimes all dominate in the respective bars depicted in the colors shown, though all three are present in all operating conditions. The ideal voltage represents the thermodynamic limit at 25℃ while the limiting current, i_{lim} is the maximum producible current.

where A is the cross-sectional area and r is the resistance of the component. This regime is dominated by the resistance to ionic conduction through the membrane and electronic resistance through the catalyst layers (usually minor). The slope of the ohmic region thus directly relates to the internal resistance of the cell.

Higher current densities include a further rapid potential drop due to mass transport limitations in the cathode. These losses occur when the oxygen supplied through the cathode GDL cannot diffuse to the catalyst sites at a sufficient rate or become impeded through water generation from the ORR, restricting access. Although this failure manifests in the cathode, this region represents failures in the GDL architecture in removing product water and supplying the needed oxygen to the reaction sites.

The physical stacks from two different buses (named FC4 and FC12) were provided by AC Transit for further experimental testing. Each bus contained two stacks of 290 proton-exchange-membrane fuel cells (PEMFCs). Each PEMFC has an active area of 420 cm^2 and is comprised of a Gore 18 µm membrane [37] with symmetric anode and cathode catalyst layers with a nominal loadings of 0.4 $mgPt/cm^2$ and 12 mm thickness. The stacks employed WTPs at ambient pressure with hydrogen and air, as have been described elsewhere. [38, 39] Figure 2.2 shows a picture of a stack on a pallet for scale.

The stacks are then integrated into the hardware of the bus through streams for the inlet gas streams and electrical lines connecting the stacks to the motor. The stacks built up voltage in series to supply power for the bus, with fresh hydrogen and air feeds entering into each stack. The cooling water entered cross-current into the stacks, mixing with the outlet water product stream. A simple visualization of the system is present in Figure 2.3.

The voltage data provided by AC Transit covers a narrow range of the degradation mechanisms covered here, in most cases only documenting ohmic regime information. Readings often captured only fragments of polarization curves, or in many cases, only a single point along the polarization curve. In many cases, although other aspects of the fuel cell were tested, either the voltage or the current were not measured, resulting in a timepoint that cannot be utilized in this study. Still further difficulties arose from artifacts in the measurements due to either the devices used or the bus system surrounding the fuel cell stacks. Additionally, operating conditions when recording the data varied and were not held consistent, sometimes even during the same timepoint. Further, the routes of the buses were not recorded, hinting that past performance may not be fully representative of future outcomes. Examples of polarization curves provided by AC Transit along with ones taken by in-house fuel cell test stands are depicted in Figure 2.4.

Despite the irregularities and gaps in the dataset, the presence of a full lifetime time-series of even partial polarization data offers the opportunity to explore the capabilities of deep neural networks without using traditional physics-based techniques. The hope through this modelling is the data-driven approach of using deep neural networks, we can capture degradation trends found not simply in the fuel cells themselves, but also in the surrounding system used to integrate them into the bus' hardware.

Figure 2.2: A photograph of a stack as extracted from the system of bus FC4. The horizontal lines through the stack represent the individual fuel cells while the pipe visible in the front of the stack shows the hydrogen input line. Pallet dimensions were approximately 5'x5'.

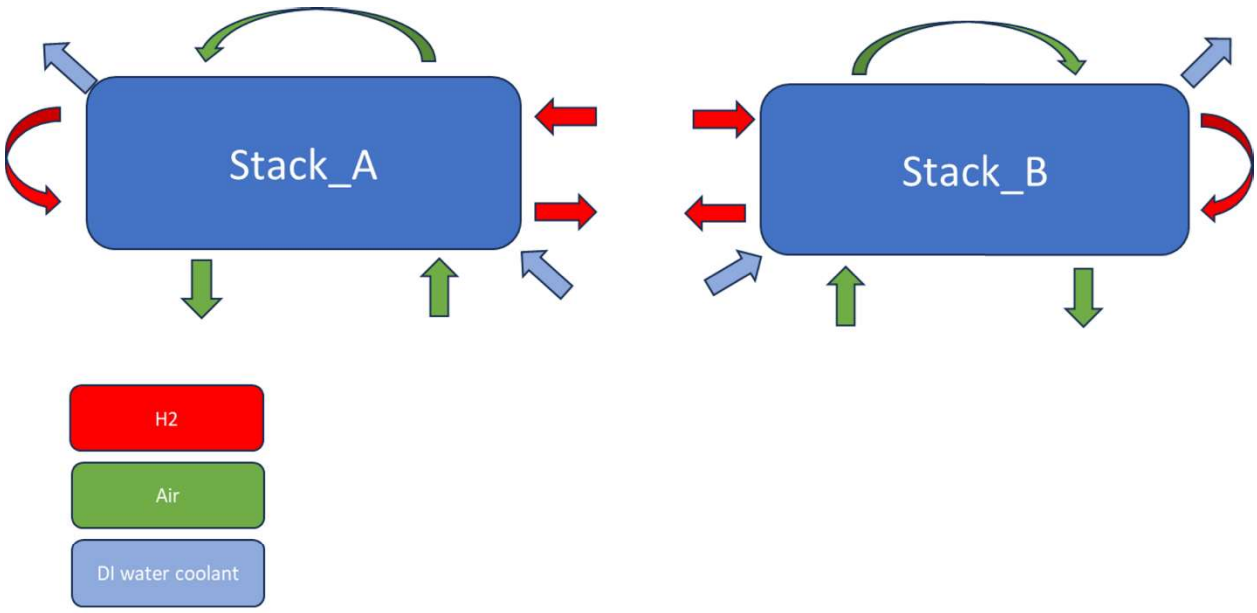

Figure 2.3: A diagram of the bus system, given by AC Transit. Each stack receives a fresh supply of hydrogen which is recycled on the opposite end of the stack after a single pass through the cells. The recycled streams are passed through the same cells, divided by indents on the WTPs. The air is similarly supplied to each cell while the cooling water undergoes a single pass and is mixed with the product water on outlet.

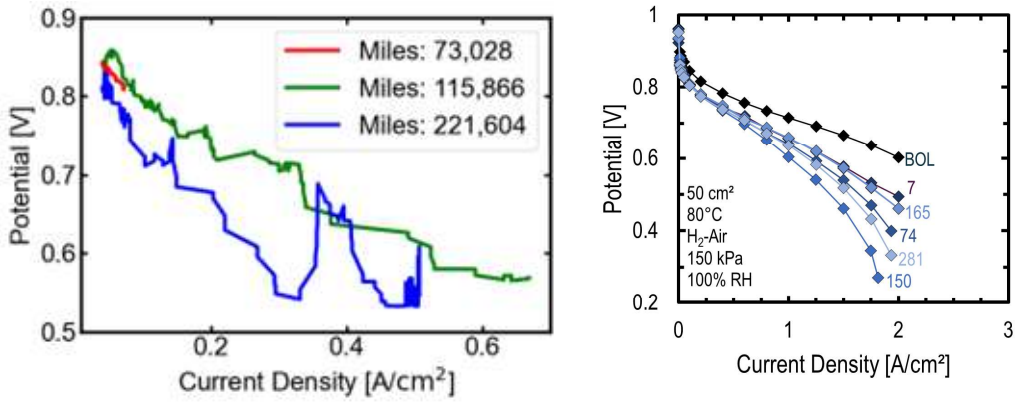

Figure 2.4: (Left) Stack-averaged polarization data given by AC Transit with mileages of the bus recorded. (Right) Data taken from individual cells from the decommissioned stack A from FC4, except for a beginning-of-life (BOL) cell, which was included for reference (courtesy of Arthur Dizon at Lawrence Berkeley National Laboratory).

Chapter 3

Design of Predictive Models

We begin by preprocessing the data for anticipated irregularities and gaps. First, the ranges of current densities over which the voltages were recorded were varied and limited to the ohmic regime. Due to the noisy nature of the data, the physical data was passed through a convolution filter that averages over every 10 points. The physical data in the subsequent plots are all from post-processing through this filter. Thus, we have fit each polarization curve to a linear relationship that the model tries to predict, characterized by its slope and intercept. An example fit line is given in Figure 3.1.

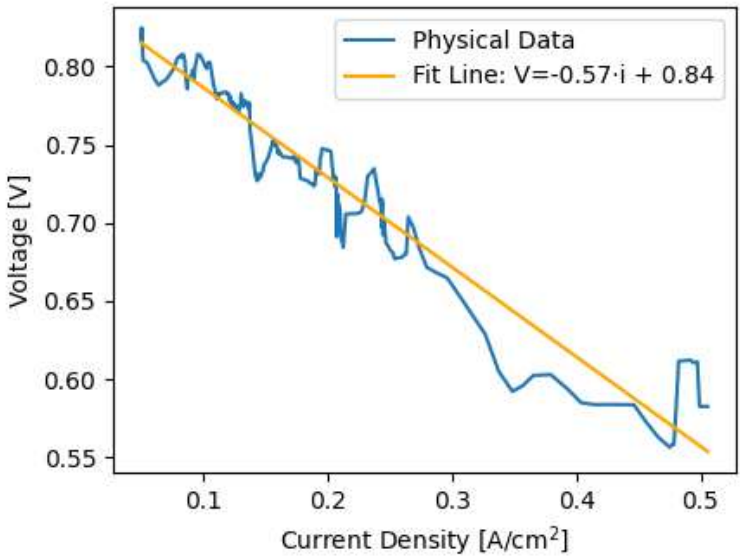

Figure 3.1: The fit line for the polarization data for FC4 on February 24, 2022 with 253,242 miles.

In many cases, the polarization data was not tested for the second stack, stack B. How-

ever, as demonstrated in Figure 3.2, stack B rarely yielded differences from stack A, especially as both stacks were arranged in series. Therefore, the data for stack A was uniformly assumed to be representative of stack B as well. In feature extraction, however, the role of missing stack B data was left as a form of dropout on the training data and passed through without modification.

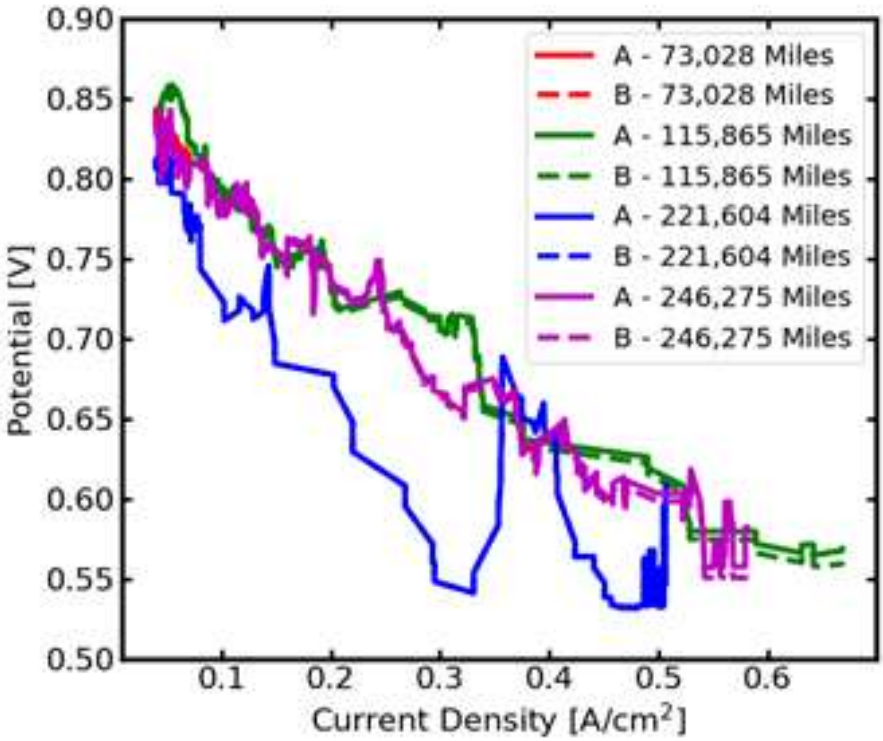

Figure 3.2: The polarization curves taken for four different time points over the lifetime of FC4. The performance is very similar for both stacks (solid and dotted lines), coinciding indistinguishably in all except the 115k miles case at higher current densities.

As the slope and intercept of the fit line of the target date and mileage were the targets of the neural networks, the target shape was simply a vector of 2 components. In contrast, when performing feature extraction, the operating conditions when performing the polarization curve, along with the previous 4 dates of data were compressed as a single feature vector. These operating conditions included the ambient air temperature of the inlet streams, the air mass flow rate, the fuel (hydrogen) mass flow rate, the air exhaust temperatures for both stacks, the recycled hydrogen pressure, the anode inlet pressures, and the ambient air pressure. In addition to these conditions which can be directly controlled, the date and

mileage of the bus was passed as an input. The mileage is already a floating number while the date was taken as a difference of the number of days since January 1st, 2010. The mileage directly captures a measure of age for the fuel cell stacks and although the materials involved in the fuel cell stacks themselves are unlikely to degrade over the course of 10 years, they may create significantly more degradation for the external system.

A sequence of four such feature vectors was then passed into the LSTM with the hope of performing few shot prompting. When predicting forward into the future, the polarization data for every 10 cells was also passed in as inputs for the previous 4 days. A summary of the data being broken into features and targets is present in Figure 3.3.

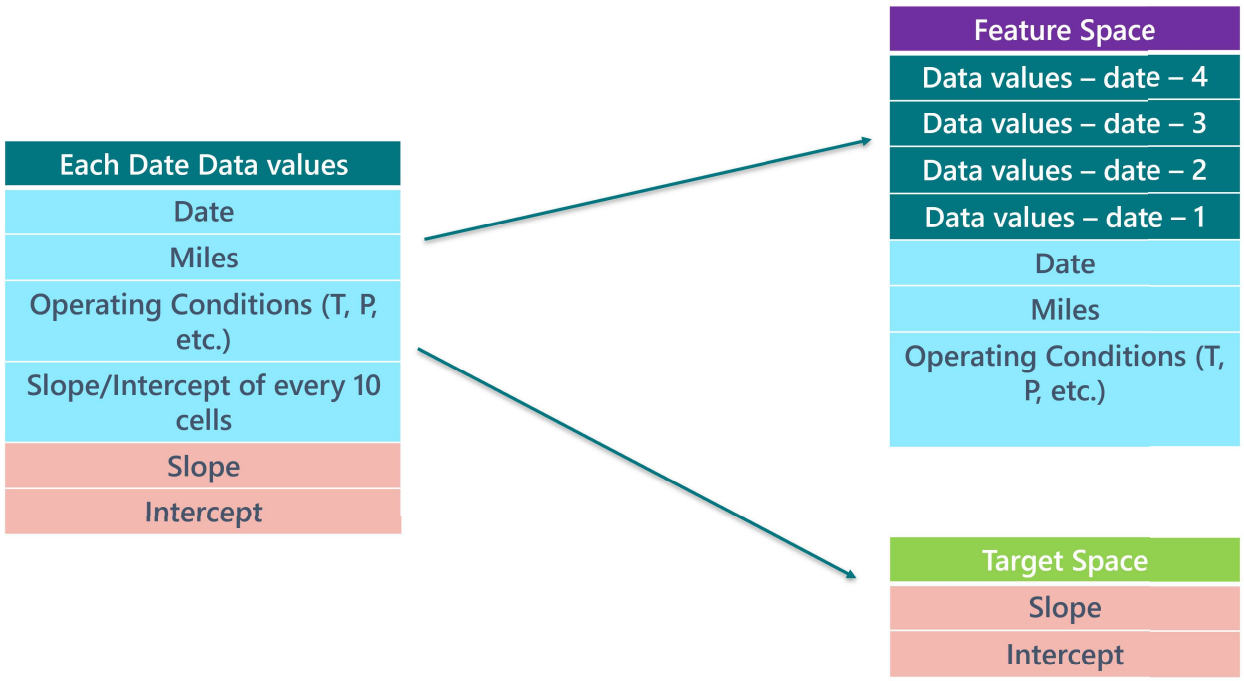

Figure 3.3: The data for each timepoint consists of the date, mileage, operating conditions (such as ambient temperature and pressure), and the polarization data for every 10 cells. The slope and intercept of the average polarization curve for stack A is maintained as a separate variable. Several dates worth of data make up the feature space while the target remains simply the overall slope and intecept.

Using the PyTorch library for Python 3.13.2, the LSTM is implemented with a hidden size of 32. The default zeros vector was utilized as the initial hidden state. To match the size of the outputs of the model (two for the slope and intercept of the overall fuel cell stack polarization curve), a fully connected layer is added after the LSTM layer. A simple mean squared error loss function, implemented using PyTorch's MSELoss, is initially used

for minimization and compared against the case when either the slope or the intercept target loss is weighted more heavily. More concretely, the unmodified loss function is

$$l(m_{pred}, b_{pred}, m_{target}, b_{target}) = (m_{pred} - m_{target})^2 + (b_{pred} - b_{target})^2 \tag{3.1}$$

where m and b stand for polarization slope and intercept, respectively. The optimizer utilized was Adam [40] with a learning rate of 0.001 and otherwise default parameters.

After arranging the data in sequences of four datapoints for the feature set and one for target, the sequences were divided into an 80:20 split between training data and validation data. As the most common usage for the model would be to predict towards the end of the lifetime of the bus using earlier time behavior, this split was not randomized such that the training data consisted of the first 80% of sequences for the bus and validation for the last 20%. A choice can be made in this case to train the models only on the particular bus being used to predict forward in time for or to train on multiple buses at a time and validate using the bus desired for prediction. This comparison is made explicit with a few examples depicting how close each case comes for representation the polarization data.

Chapter 4

Results

Initially, using only a single bus (FC4) to train upon, the LSTM-based model created using the above approach was trained on the first 80% of sequences and validated on the last 20% of sequences. The training was performed for 100 epochs, resulting in the following training and validation loss curves present in Figure 4.1. In future trainings of the models with different characteristics, the training and loss curves look similar and thus have not been replotted. To maintain a fair comparison, the loss has been normalized by the size of each dataset. For reproducible results, a seed of 1 has been passed into the PyTorch random number generator.

Just using the base model, predicting out forward towards the next datapoint, we obtain the following results for the two different days and mileages of the predictions, as displayed in Figure 4.2. Overall, the difference between predictions and targets is visually small, validating the assumptions and architecture of the model. Somewhat importantly, although the slope and intercept from the left figure have a higher error than the right, visually the left prediction line intersects the right, seemingly providing more accurate results for certain current densities where the bus might be more likely to operate.

As we can see, even using the unmodified loss function (PyTorch's MSELoss), we still yield fairly accurate predictions. We can demonstrate the affect of changing the importance of both the slope and intercept though on the predictions by increasing the weighting in the loss function:

$$l(m_{pred}, b_{pred}, m_{target}, b_{target}) = w_1(m_{pred} - m_{target})^2 + w_2(b_{pred} - b_{target})^2 \tag{4.1}$$

Previously, $w_1 = w_2 = 1$, but here, we raise each independently to 10. First, starting with the intercept, we have the results in Figure 4.3. Although the predicted curve for FC4, which was already more accurate, did not change significantly, the outputted predictions for FC7 improved somewhat in the region where the data was taken (again, despite only knowing the slope and intercept, so the model has no intuition about which current density range was in fact utilized).

For the slope predictions, we have (remarkably similar) the results in Figure 4.4. The results are better for FC7, but slightly worse for FC4. Requiring accuracy in general, either

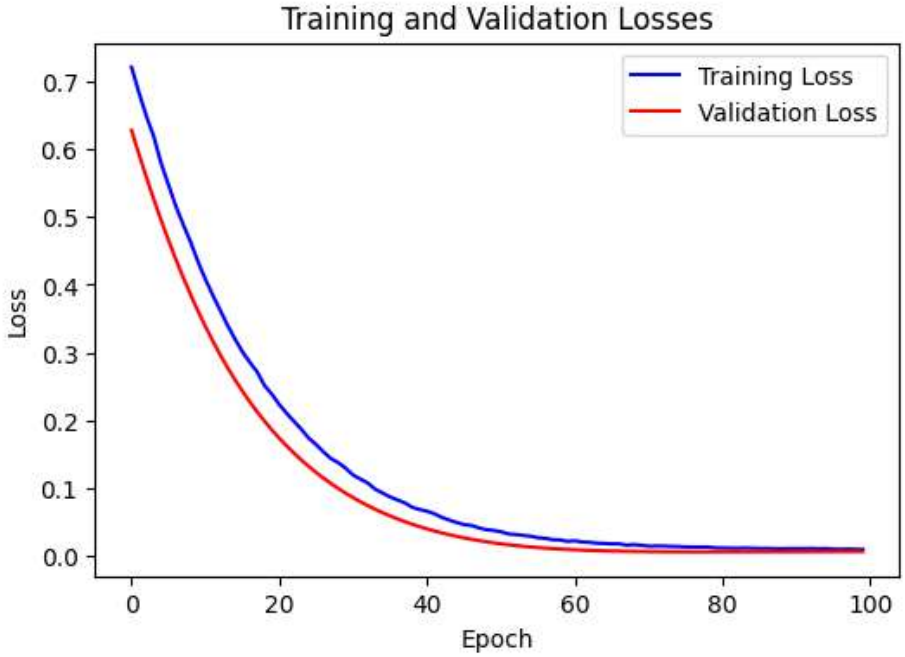

Figure 4.1: The training and validation losses for training the base model on only the FC4 data. Each loss is normalized by the size of its respective dataset.

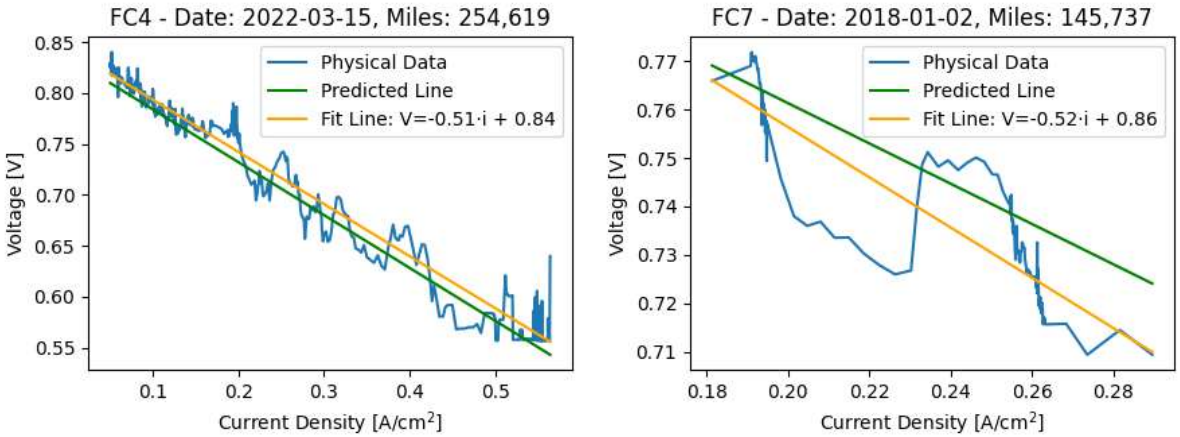

Figure 4.2: The predicted lines against the fit lines (and physical data that they represent) for training the base model on only each respective bus's data.

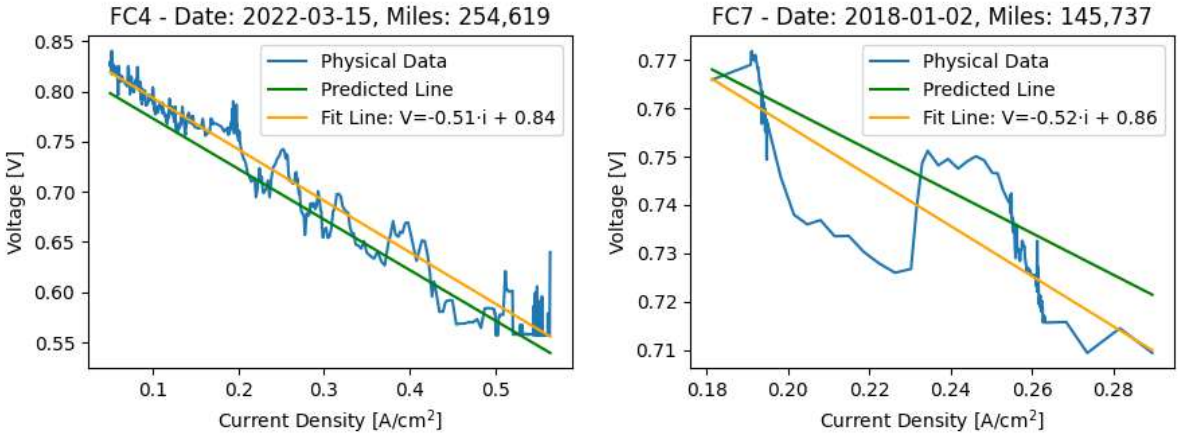

Figure 4.3: The predicted lines against the fit lines (and physical data that they represent) for training the base model on only each respective bus's data when weighting the intercept as 10x.

approach seems advantageous over unmodified MSELoss however.

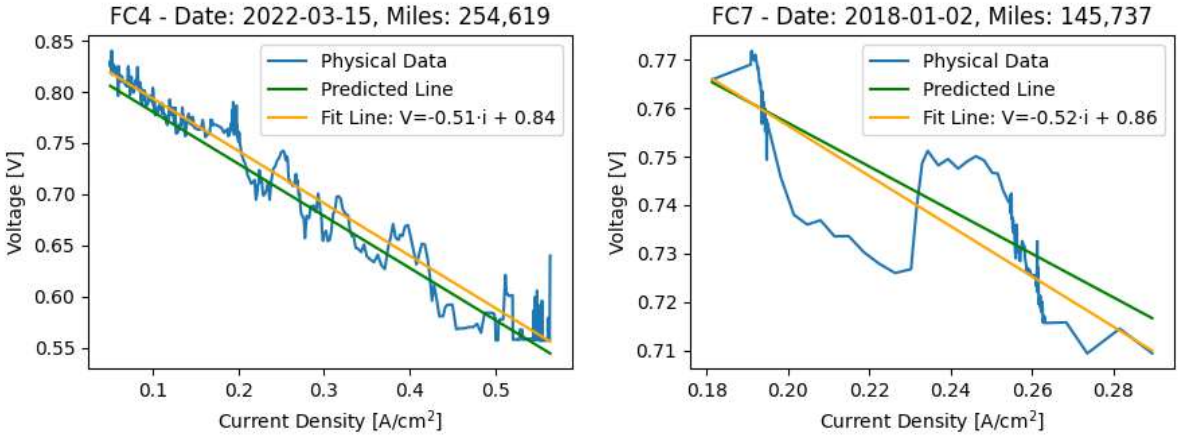

Figure 4.4: The predicted lines against the fit lines (and physical data that they represent) for training the base model on only each respective bus's data when weighting the slope as 10x.

From this point on, we modify the loss function to be the one where the intercept is weighted by 10x. We can additionally test the utilization of the model as being fixed when encountering new buses; essentially considered a portion of the buses as training data and fixing the parameters of the model for the future. This is less necessary with our smaller

dataset, but may become more helpful as we accumulate more data. In Figure 4.5, we predict the polarization data for FC15, only training on the information given from 8 other buses (FC4 - FC11). Despite having no further information about FC15 beyond the 4 previous dates of data, the model performs remarkably well, leading to the notion that the degradation mechanism of the buses is likely similar. Additionally, FC15 employs a much healthier fit line than the other buses above (FC4 and FC7) with a more gradual downslope and the model captures this trend, with an expected voltage at 0.5 A/cm^2 of 0.63 V, matching similar voltages present for 115k miles in Figure 3.2.

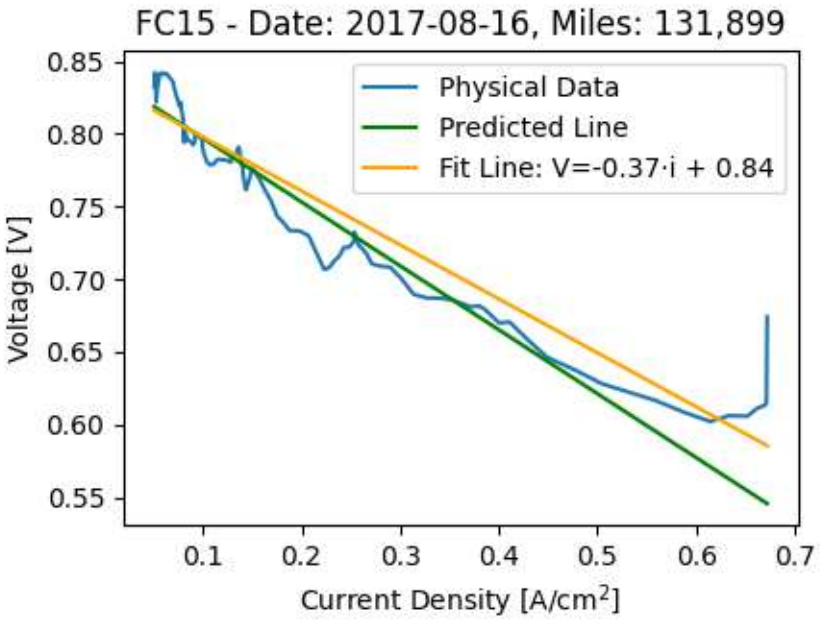

Figure 4.5: An example of a predicted polarization curve for FC15 using the version of the model that was trained only on buses FC4-FC11.

Finally, another utilization of the model is to predict out towards the end of a bus's lifetime to visualize when power output might become an issue. We can accomplish this by maintaining the same sequence prediction model designed above, but changing the date and mileage of the final sequence to be that of the last timepoint for each bus. Plotting the loss functions in Figure 4.6, we find that no obvious trend appears except the loss does move towards being minimized at the end of life. However, only based on the current dataset, we cannot recommend guidelines after which forecasting become accurate.

For visualization, we can additionally map the predictions given for a few of these dates for both buses as in Figure 4.7.

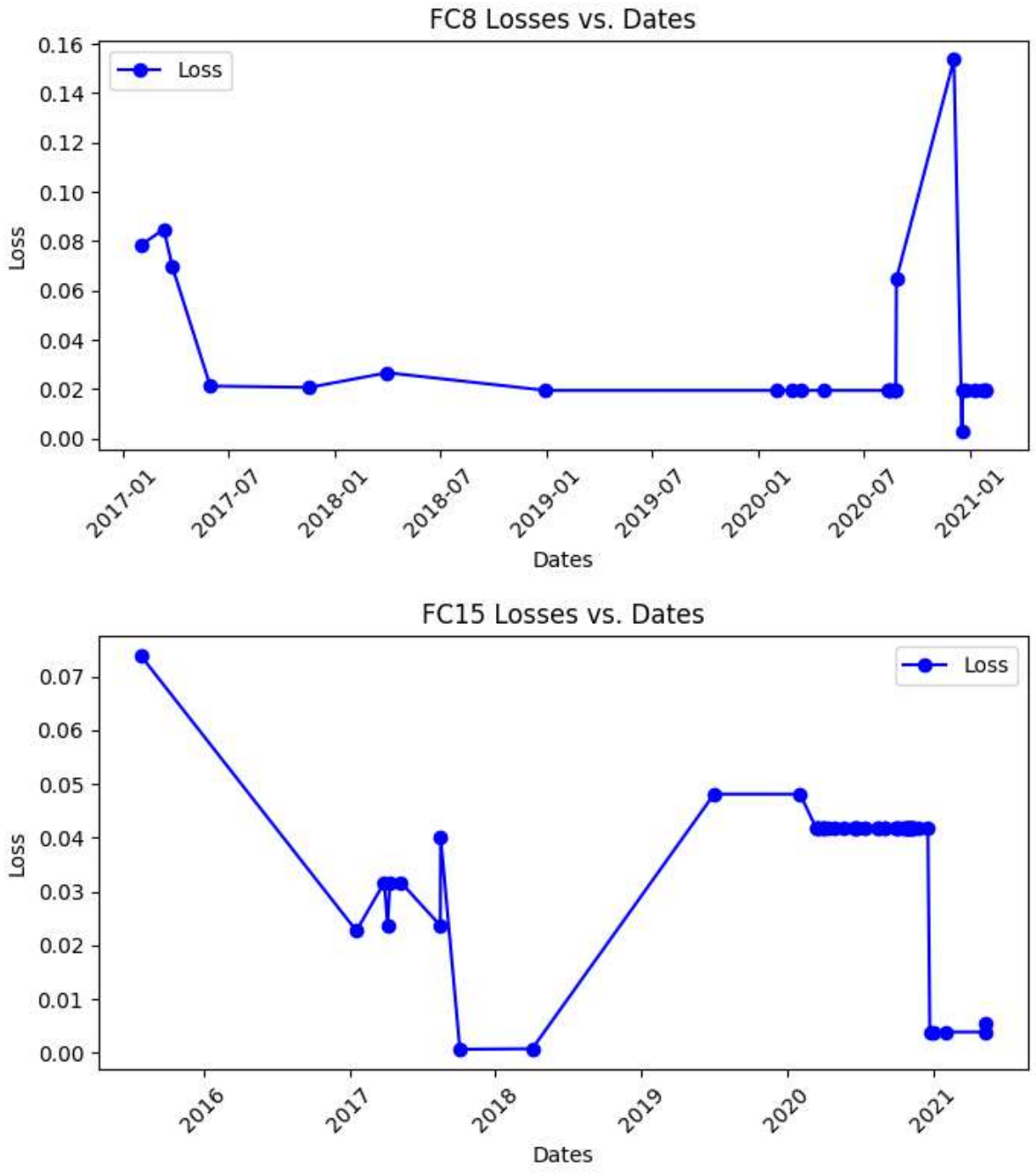

Figure 4.6: (Intercept-weighted) loss functions when predicting out towards the last time-point for each bus. For FC8, the two points towards the right with highest loss are due to outlier physical data, but has been included here as such data might be encountered in practice.

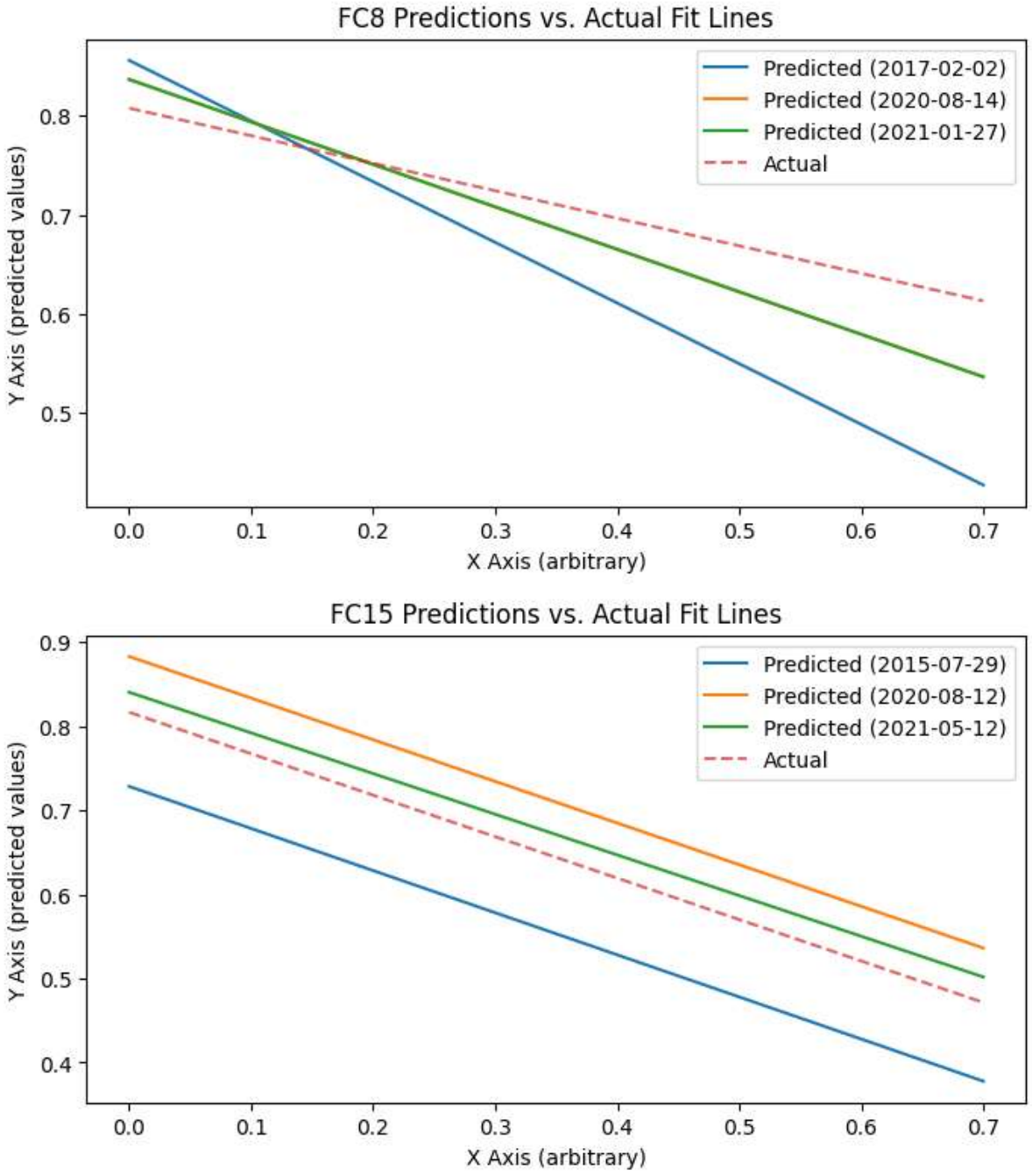

Figure 4.7: Predictions from the model training on each bus utilizing the (Intercept weighted) loss functions from the text. For FC8, the orange line coincides indistinguishably with the green line.

Conclusion

In this work, we have explored LSTM-based models of predicting fuel cell stack behavior given lifetime behavior of hydrogen fuel cell buses, with data supplied from AC Transit. These models utilize sequences of timepoints when the buses were brought in for maintenance for polarization curve predictions. We find the models to be remarkably accurate in many cases, often predicting forward towards the next timepoint when polarization data taken with ease. Long-term predictions suffer greater error here, however, and several mitigation strategies such as reducing noise in experimental measurements, polarizing the stacks to both the kinetic and mass transport limited regime, and operating said stacks at universal conditions during testing may generally improve not just the model reliability, but replicability. Future work on model architecture as well as more reliable and larger datasets will improve the robustness of these networks, allowing for empirical optimization of the performance of such technologies.

Bibliography

[1] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. "Multilayer feedforward networks are universal approximators". In: *Neural Networks* 2.5 (Jan. 1989), pp. 359–366. ISSN: 0893-6080. DOI: 10.1016/0893-6080(89)90020-8.

[2] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. "Learning representations by back-propagating errors". In: *Nature* 323.6088 (Oct. 1986). publisher: Nature Publishing Group, pp. 533–536. ISSN: 1476-4687. DOI: 10.1038/323533a0.

[3] Stuart Geman, Elie Bienenstock, and René Doursat. "Neural Networks and the Bias/Variance Dilemma". In: *Neural Computation* 4.1 (Jan. 1992), pp. 1–58. ISSN: 0899-7667. DOI: 10.1162/neco.1992.4.1.1.

[4] Sergey Ioffe and Christian Szegedy. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". In: (Mar. 2015). arXiv:1502.03167 [cs]. DOI: 10.48550/arXiv.1502.03167. URL: http://arxiv.org/abs/1502.03167.

[5] Robert Tibshirani. "Regression Shrinkage and Selection via the Lasso". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 58.1 (1996). publisher: [Royal Statistical Society, Oxford University Press], pp. 267–288. ISSN: 0035-9246.

[6] A. E. Hoerl and R. W. Kennard. "Ridge Regression". In: *Encyclopedia of Statistical Sciences*. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/0471667196.ess2280.pub2 DOI: 10.1002/0471667196.ess2280.pub2. John Wiley & Sons, Ltd, 2006. ISBN: 978-0-471-66719-3. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/0471667196.ess2280.pub2.

[7] Arthur E. Hoerl and Robert W. Kennard. "Ridge Regression: Biased Estimation for Nonorthogonal Problems". In: *Technometrics* (Feb. 1970). publisher: Taylor & Francis Group. ISSN: 1048-8634. URL: https://www.tandfonline.com/doi/abs/10.1080/00401706.1970.10488634.

[8] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: (Dec. 2015). arXiv:1512.03385 [cs]. DOI: 10.48550/arXiv.1512.03385. URL: http://arxiv.org/abs/1512.03385.

[9] Kunihiko Fukushima. "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position". In: *Biological Cybernetics* 36.4 (Apr. 1980), pp. 193–202. ISSN: 1432-0770. DOI: 10.1007/BF00344251.

[10] Jeffrey L. Elman. "Finding structure in time". In: *Cognitive Science* 14.2 (Apr. 1990), pp. 179–211. ISSN: 0364-0213. DOI: 10.1016/0364-0213(90)90002-E.

[11] Sishun Liu et al. "Fine-Tuned Transformer Model for Sentiment Analysis". In: *Knowledge Science, Engineering and Management*. Ed. by Gang Li et al. Springer International Publishing. Cham, 2020, pp. 336–343. ISBN: 978-3-030-55393-7. DOI: 10.1007/978-3-030-55393-7_30.

[12] Jacob Devlin et al. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. https://arxiv.org/abs/1810.04805v2. [Online; accessed 2025-04-07]. Oct. 2018. URL: https://arxiv.org/abs/1810.04805v2.

[13] Jacob Devlin et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In: (May 2019). arXiv:1810.04805 [cs]. DOI: 10.48550/arXiv.1810.04805. URL: http://arxiv.org/abs/1810.04805.

[14] Huiwen Chang et al. "Muse: Text-To-Image Generation via Masked Generative Transformers". In: (Jan. 2023). arXiv:2301.00704 [cs]. DOI: 10.48550/arXiv.2301.00704. URL: http://arxiv.org/abs/2301.00704.

[15] Ali Hatamizadeh et al. "DiffiT: Diffusion Vision Transformers for Image Generation". In: (Aug. 2024). arXiv:2312.02139 [cs]. DOI: 10.48550/arXiv.2312.02139. URL: http://arxiv.org/abs/2312.02139.

[16] *PEGASUS: A State-of-the-Art Model for Abstractive Text Summarization*. [Online; accessed 2025-04-07]. URL: https://research.google/blog/pegasus-a-state-of-the-art-model-for-abstractive-text-summarization/.

[17] Thomas Fischer and Christopher Krauss. "Deep learning with long short-term memory networks for financial market predictions". In: *European Journal of Operational Research* 270.2 (Oct. 2018), pp. 654–669. ISSN: 0377-2217. DOI: 10.1016/j.ejor.2017.11.054.

[18] Andrew R. Chow. *How Meteorologists Are Using AI to Forecast Hurricane Milton*. [Online; accessed 2025-04-07]. Oct. 2024. URL: https://time.com/7081372/ai-hurricane-forecasting/.

[19] Clive Cookson. "AI outperforms conventional weather forecasting methods for first time". In: *Financial Times* (Nov. 2023). [Online; accessed 2025-04-07]. URL: https://www.ft.com/content/ca5d655f-d684-4dec-8daa-1c58b0674be1.

[20] M. Aczon et al. "Dynamic Mortality Risk Predictions in Pediatric Critical Care Using Recurrent Neural Networks". In: (Jan. 2017). arXiv:1701.06675 [stat]. DOI: 10.48550/arXiv.1701.06675. URL: http://arxiv.org/abs/1701.06675.

[21] Edward Choi et al. "Doctor AI: Predicting Clinical Events via Recurrent Neural Networks". In: (Sept. 2016). arXiv:1511.05942 [cs]. DOI: 10.48550/arXiv.1511.05942. URL: http://arxiv.org/abs/1511.05942.

[22] *Fuel Cells*. [Online; accessed 2025-03-17]. URL: https://www.energy.gov/eere/fuelcells/fuel-cells.

[23] Adam Rossi. *Electrolyzer and Hydrogen Fuel Cell Safety*. [Online; accessed 2025-03-24]. Mar. 2023. URL: https://totalshield.com/blog/electrolyzer-and-hydrogen-fuel-cell-safety/.

[24] Bloom Energy. *What are the Advantages of Hydrogen Fuel Cells?* [Online; accessed 2025-03-24]. Apr. 2024. URL: https://www.bloomenergy.com/blog/what-are-the-advantages-of-hydrogen-fuel-cells/.

[25] *Research — M2FCT*. [Online; accessed 2025-03-24]. URL: https://millionmilefuelcelltruck.org/research.

[26] *Home — Hydrogen Program*. [Online; accessed 2025-03-24]. URL: https://www.hydrogen.energy.gov/.

[27] *HyMARC: Hydrogen Materials Advanced Research Consortium*. [Online; accessed 2025-03-24]. URL: https://www.energy.gov/eere/fuelcells/hymarc-hydrogen-materials-advanced-research-consortium.

[28] *R2R: Roll-to-Roll Consortium*. [Online; accessed 2025-03-24]. URL: https://www.energy.gov/eere/fuelcells/r2r-roll-roll-consortium.

[29] jennynuss. *Six Ways Berkeley Lab is Bringing Clean Hydrogen to the World*. [Online; accessed 2025-04-07]. Oct. 2023. URL: https://newscenter.lbl.gov/2023/10/06/six-ways-berkeley-lab-is-bringing-clean-hydrogen-to-the-world/.

[30] *Toyota Fuel Cell Electric Vehicles — Toyota Europe*. [Online; accessed 2024-12-09]. URL: https://www.toyota-europe.com/electrification/fcev.

[31] *2023 Nexo Fuel Cell — Vehicle Overview — Hyundai USA*. [Online; accessed 2024-12-09]. URL: https://www.hyundaiusa.com/us/en/vehicles/nexo.

[32] *2025 Honda CR-V e:FCEV — Hydrogen Fuel Cell Vehicle*. [Online; accessed 2024-12-09]. URL: https://automobiles.honda.com/cr-v-fcev.

[33] *AC Transit Fuel Cell Bus Longevity Study Report Summary — FTA*. [Online; accessed 2025-03-30]. URL: https://www.transit.dot.gov/research-innovation/ac-transit-fuel-cell-bus-longevity-study-report-summary.

[34] *Zero Emission Bus Technology Analysis — Alameda-Contra Costa Transit District*. [Online; accessed 2025-03-30]. URL: https://www.actransit.org/zebta.

[35] *AC Transit Bus Roster — Alameda-Contra Costa Transit District*. [Online; accessed 2025-03-30]. URL: https://www.actransit.org/index.php/ac-transit-bus-roster.

[36] Qianwen Tang et al. "Review of hydrogen crossover through the polymer electrolyte membrane". In: *International Journal of Hydrogen Energy* 46.42 (June 2021), pp. 22040–22061. ISSN: 0360-3199. DOI: 10.1016/j.ijhydene.2021.04.050.

[37] *GORE-SELECT® Membranes(PEM) for Fuel Cells — Gore*. [Online; accessed 2024-12-09]. URL: https://www.gore.com/products/gore-select-membrane.

[38] Adam Z. Weber and Robert M. Darling. "Understanding porous water-transport plates in polymer-electrolyte fuel cells". In: *Journal of Power Sources*. 10th EUROPEAN LEAD BATTERY CONFERENCE 168.1 (May 2007), pp. 191–199. ISSN: 0378-7753. DOI: 10.1016/j.jpowsour.2007.02.078.

[39] Xiaofeng Wang et al. "Impact of In-Cell Water Management on the Endurance of Polymer Electrolyte Membrane Fuel Cells". In: *Journal of The Electrochemical Society* 161.6 (May 2014). publisher: IOP Publishing, F761. ISSN: 1945-7111. DOI: 10.1149/2.084406jes.

[40] Diederik P. Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: (Jan. 2017). arXiv:1412.6980 [cs]. DOI: 10.48550/arXiv.1412.6980. URL: http://arxiv.org/abs/1412.6980.