Battery Second-Use Reconditioning Based on Sparse Data and Models

Cooperative Research and Development Final Report

CRADA Number: CRD-21-17528

NREL Technical Contact: Kandler Smith

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated under Contract No. DE-AC36-08GO28308

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Technical Report
NREL/TP-5700-93187
June 2025

Battery Second-Use Reconditioning Based on Sparse Data and Models

Cooperative Research and Development Final Report

CRADA Number: CRD-21-17528

NREL Technical Contact: Kandler Smith

Suggested Citation
Smith, Kandler. 2025. *Battery Second-Use Reconditioning Based on Sparse Data and Models: Cooperative Research and Development Final Report, CRADA Number CRD-21-17528.* Golden, CO: National Renewable Energy Laboratory. NREL/TP-5700-93187. https://www.nrel.gov/docs/fy25osti/93187.pdf.

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated under Contract No. DE-AC36-08GO28308

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Technical Report
NREL/TP-5700-93187
June 2025

National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

NOTICE

This work was authored by the National Renewable Energy Laboratory for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Vehicle Technologies Office. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government.

Cover photos (clockwise from left): Josh Bauer, NREL 61725; Visualization from the NREL Insight Center; Getty-181828180; Agata Bogucka, NREL 91683; Dennis Schroeder, NREL 51331; Werner Slocum, NREL 67842.

NREL prints on paper that contains recycled content.

Cooperative Research and Development Final Report

Report Date: February 5, 2025

In accordance with requirements set forth in the terms of the CRADA agreement, this document is the CRADA final report, including a list of subject inventions, to be forwarded to the DOE Office of Scientific and Technical Information as part of the commitment to the public to demonstrate results of federally funded research.

Parties to the Agreement: Smartville, Inc.

CRADA Number: CRD-21-17528

CRADA Title: Battery Second-Use Reconditioning Based on Sparse Data and Models

Responsible Technical Contact at Alliance/National Renewable Energy Laboratory (NREL):

Kandler Smith | kandler.smith@nrel.gov (Primary PI, for Sandra Loi | sandra.loi@nrel.gov)

Andrew Schiek | andrew.schiek@nrel.gov (Co-Author)

Name and Email Address of POC at Company:

Michael Ferry | mferry@smartville.io

Sponsoring DOE Program Office(s):

Office of Energy Efficiency and Renewable Energy (EERE), Vehicle Technologies Office

Joint Work Statement Funding Table showing DOE commitment:

Estimated Costs	NREL Shared Resources a/k/a Government In-Kind
Year 1	$100,000.00
TOTALS	$100,000.00

Executive Summary of CRADA Work:

NREL will aid Smartville, Inc. to rapidly characterize electric vehicle (EV) battery state-of-health (SOH), characterize lifetime models, and forecast lifetime under various second-use reconditioning strategies.

CRADA benefit to DOE, Participant, and US Taxpayer:

- Assists laboratory in achieving programmatic scope, and/or
- Uses the laboratory's core competencies.

<u>Summary of Research Results</u>:

Scope of Work:

NREL and Smartville will work together under the U.S. DOE's Recycling Prize Phase II Voucher program. Smartville would like to deploy their battery reconditioning/second-use business model in geographically diverse locations, for multiple models of used EV battery packs for which there is little or no state-of-health (SOH) or lifetime data. Based on limited data, NREL will help Smartville propose simplified life models, develop rapid SOH characterization approaches involving machine learning, and apply those diagnostic/modeling-based methods to inform Smartville's "life balancing" technical solution.

Task Descriptions:

Based on the CRADA, NREL's objectives in its work with Smartville can be broken down into three parts:

1. Meet regularly for status and planning.
2. Propose a simplified battery life model.
3. Development of SOH and life estimation algorithms.
4. Inform life-balancing reconditioning strategy.

We approached these objectives in two ways:

A. A physics-based life-prediction model
B. Machine learning (ML) heath estimation models.

Objective 1 was addressed primarily with method A. Using data from a publicly sourced vehicle database called Plug-In America we developed a physics-informed simplified life model. Objectives 2 and 3 were influenced by both Method A and Method B. Method A continued to use data from Plug-In America, whereas Method B utilized a publicly sourced database from MIT to develop a ML pipeline and tested this on data collected from Smartville on Nissan and Tesla batteries. In the following sections we'll address how we completed the objectives.

Task 1 (NREL & Participant): Schedule and attend virtual progress meeting every 2 weeks

Progress meetings between the project partners was held regularly.

Task 2: Simplified Battery Life-Model

NREL will propose a simple, yet general life model that can be rapidly parameterized for batteries from new EV types/models that have not been previously tested by Smartville or its partners. The simplified life model may take into account first-life use history, e.g., calendar age of pack, vehicle mileage or total Amp-hour-throughput associated with pack, and geographic location(s) where vehicle was registered, allowing estimation of ambient temperature exposure history From previous work we have developed a generalized life model of the form: $q = 1 - k_{cal} \cdot t^{p_{cal}} - k_{cyc} \cdot EFC^{p_{cyc}}$. From that generalized equation we fit for the coefficients k_{cal} and k_{cyc} and the exponents p_{cal} and p_{cyc}. Some examples of our previous work are shown in the figure below with their fitted equations in the Figure.

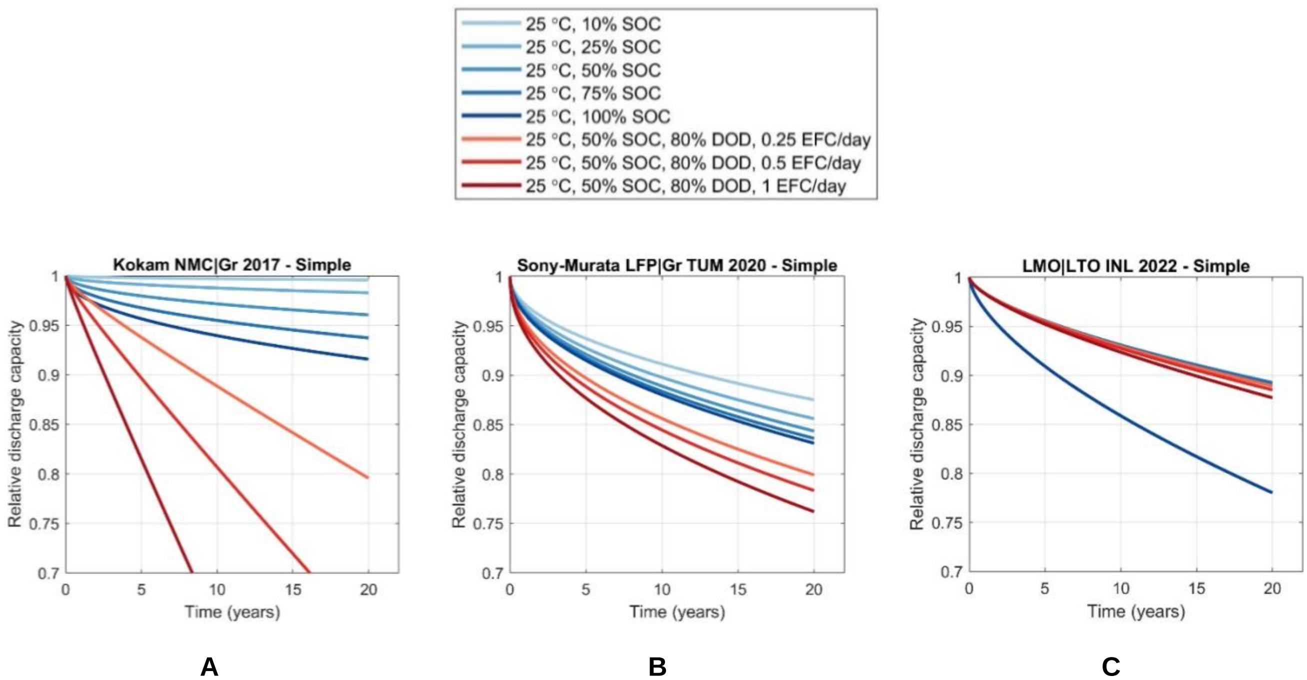


Figure 1: Previous Models

A) NMC|Gr Chemistry, Fitted Parameters: $k_{cal} = p_1 \cdot \exp\left(p_2/T\right) \cdot \exp(p_3 \cdot U_a)$; $k_{cyc} = (p_4 + p_5 \cdot DOD + p_6 \cdot C_{rate})$; $p_{1,4,5,6} > 0$

B) LFP|GR Chemistry, Fitted Parameters: $k_{cal} = p_1 \cdot \exp\left(p_2/T\right) \cdot (\exp(p_3 \cdot SOC) + p_4)$; $k_{cyc} = (p_5 + p_6 \cdot DOD + p_7 \cdot C_{rate}) \cdot \exp\left(p_8/T\right) \cdot \exp\left(-p_9/T\right)$; $p_{1,4,5,6,7,8,9} > 0$

C) LMO|LTO Chemistry, Fitted Parameters: $k_{cal} = p_1 \cdot \exp\left(p_2/T\right) \cdot \exp(p_3 \cdot U_a)$; $k_{cyc} = \left(p_4 + p_5 \cdot \exp\,(p_6 \cdot DOD)\right) \cdot \exp\left(p_7/T\right) \cdot \exp\left(-p_8/T\right)$; $p_{1,4,5,6,7,8} > 0$

In the work with Smartville we used the Plug-In America data to fit a model to the generalized equation of the format: $q = 1 – b_1\ (SOC) \exp(-E_{a1} / T)\ t^{1/2} – b_2\ (DOD) \exp(-E_{a2} / T)\ N_{100}$. An example fit is shown below.

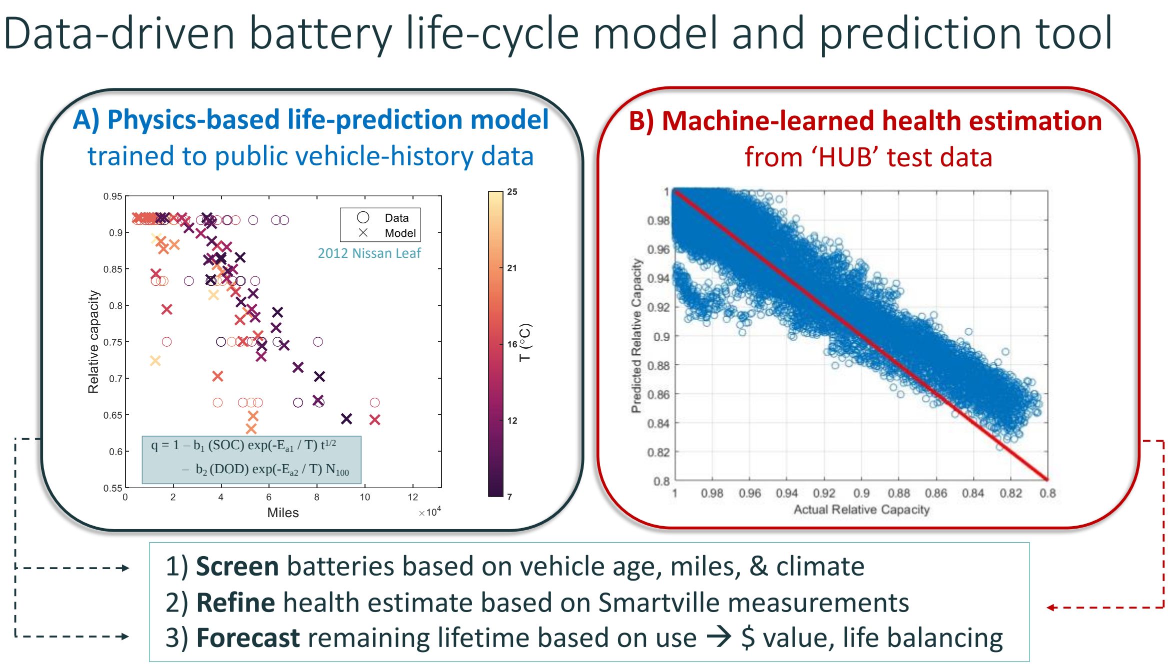


Figure 2*: Data-driven battery life-cycle model and prediction tool

The goal of this first objective was to create a simplified battery life-model, which we achieved from our domain knowledge and previous work. This provides a useful comparison for models that more rigorous, can give us an idea if the extra computational time produces significantly better results than that of a simple model. We will discuss some of this comparison after presenting our more complicated models.

Task 3: SOH and Life Estimation Algorithms

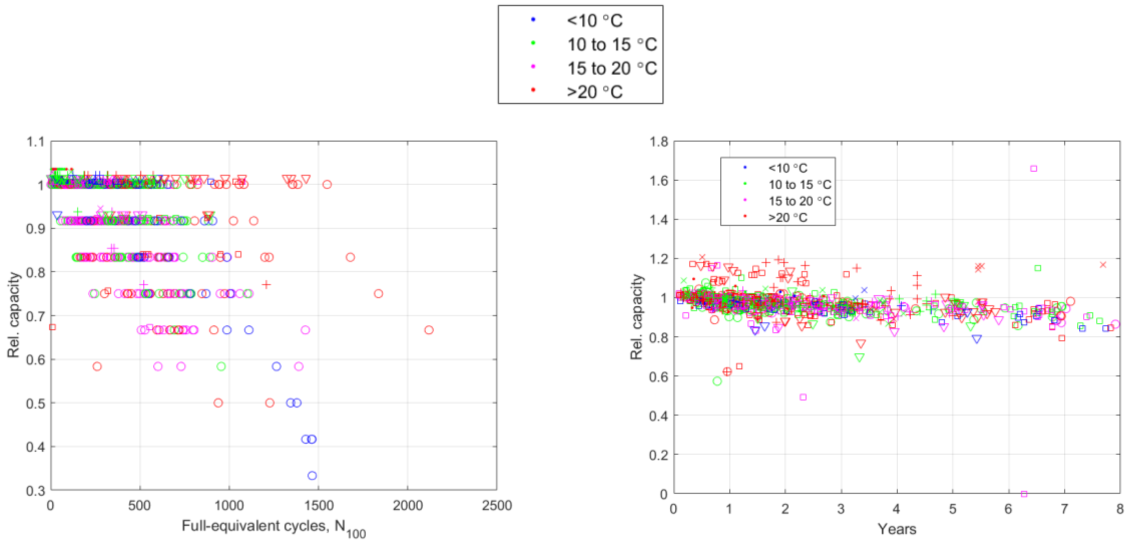
NREL will propose an algorithm topology that fuses multiple disparate data streams to forecast SOH and life. Those data streams include:

1. Battery first-life use history
2. Smartville's characterization data when the battery is first received.
3. Periodic diagnostic pulses injected into the batteries during operation under grid service applications (e.g., energy shifting, demand charge, arbitrage)
4. Machine learning correlation of the diagnostic pulse responses to full capacity.
5. Real-time operational data under grid service applications
6. Smartville's characterization data after the battery has completed its life-balancing and reconditioning.

NREL will document the algorithm using flowcharts and write and deliver prototype software code to accomplish the SOH/life estimation. Rigorous validation is beyond the scope of this present effort.

Physic-Based Life Prediction

This objective can be broken down into the two approaches: physics-based life prediction and ML SOH estimation. For the physics-based life prediction we used the simplified life model to influence the model derived from the Nissan Leaf and Tesla Model S data. As can be seen from the figure below we have a wide range of SOH values even if some of them may be outliers.



A **B**

Figure 3: A) Nissan Data Set B) Tesla Data Set

Using these data sets and the aforementioned simplified life model we fit the equation: $q = 1 - b_1 (SOC) \exp(-E_{a1} / T)\, t^{1/2} - b_2 (DOD) \exp(-E_{a2} / T)\, N_{100}$. What we're looking for with this fit is that the model coefficients match the relationships we expect to see (based on physics) with the given variables. For both the Nissan and Tesla data sets we saw this and example of the fit is shown in Figure 4. From Figure 4 we do see some higher error at the beginning of life, but otherwise have residual error less 0.1, and there are no trends between error and temperature. This gives us confidence in saying that our model is fit well for the given temperatures and SOH range.

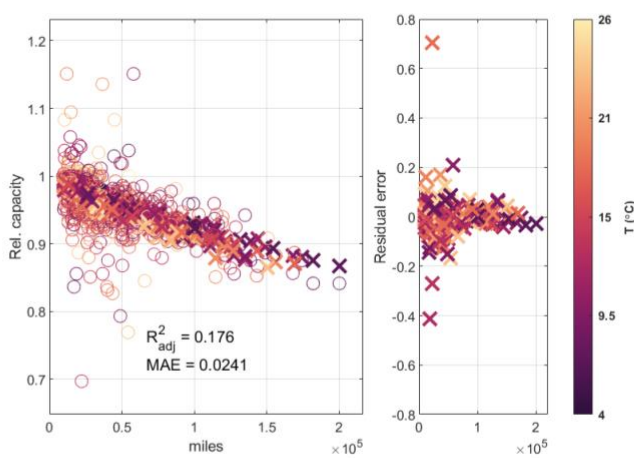
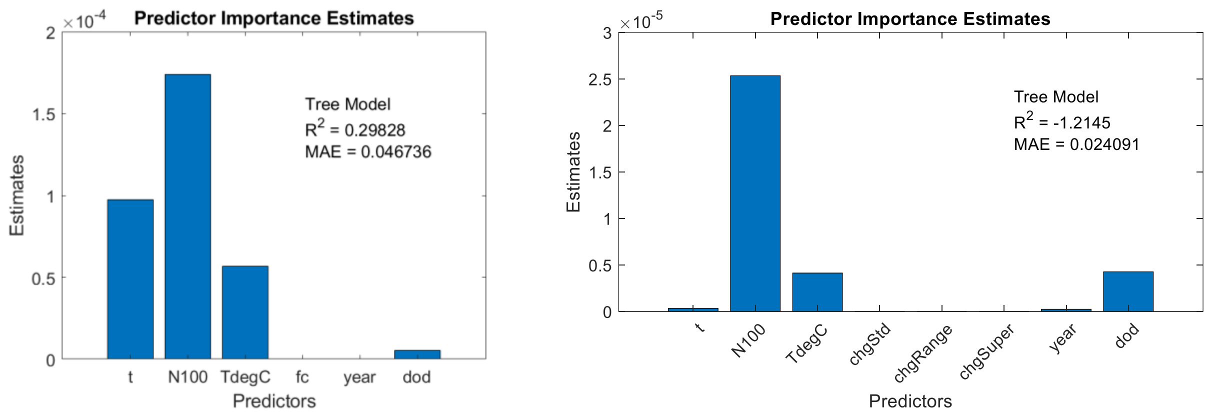


Figure 4: Tesla Model S Fit

One shortcoming of this model is that we could improve inline prediction due to metrics such as the R^2 adjusted being low (0.176). One attempt to improve this was to develop a black box ML model on the same data streams. An additional advantage to this outside of improved inline accuracy is the ability to look at how important different features are for predicting SOH. From Figure 5 we see that Temperature and Cycles were very important to both the Nissan and Tesla models, but time was only important for the Nissan model and depth-of-discharge (DoD) was important for the Tesla. By looking at the predictor importance it can give us an idea what data we need to collect in future testing/monitoring and may help us explore relationships we had not considered in our physics model.



A **B**

Figure 5: Predictor Importance Scores for A) Nissan Leaf Tree Model B) Tesla Model S Tree Model

Although this model helps improve our inline prediction compared to the strictly physics-based model, it has issues predicting outside of the given SOH range. In Figure 6 we see that once we get outside of our given range the model starts to predict an increase in SOH. This is incorrect, and we know based on theory that many ML models do not predict well outside of their given range. Given that our physics-based model produced the expected dependence on temperature, time, and cycling, we feel safe using this for prognostics and forecasting, and can use the ML model to help improve inline estimates.

Machine Learning State-of-Health Estimation

We have three data sources for this part of the estimation. The publicly sourced MIT data set and the Nissan and Tesla data sets from Smartville. The MIT data set has 38,765 data points while the Nissan and Tesla datasets consist of less than 500 data points each. Having the larger data set makes it more reliable for training our model. From each data set there was summary data (temperature, internal resistance, SOH) and cycling data (voltage and current curves).

For predicting SOH from these data streams, we developed a ML pipeline to make the estimation and evaluation process more streamlined. The pipeline can be broken into three sections: feature engineering, feature selection, and model fitting. The process is outlined in Figure 6.

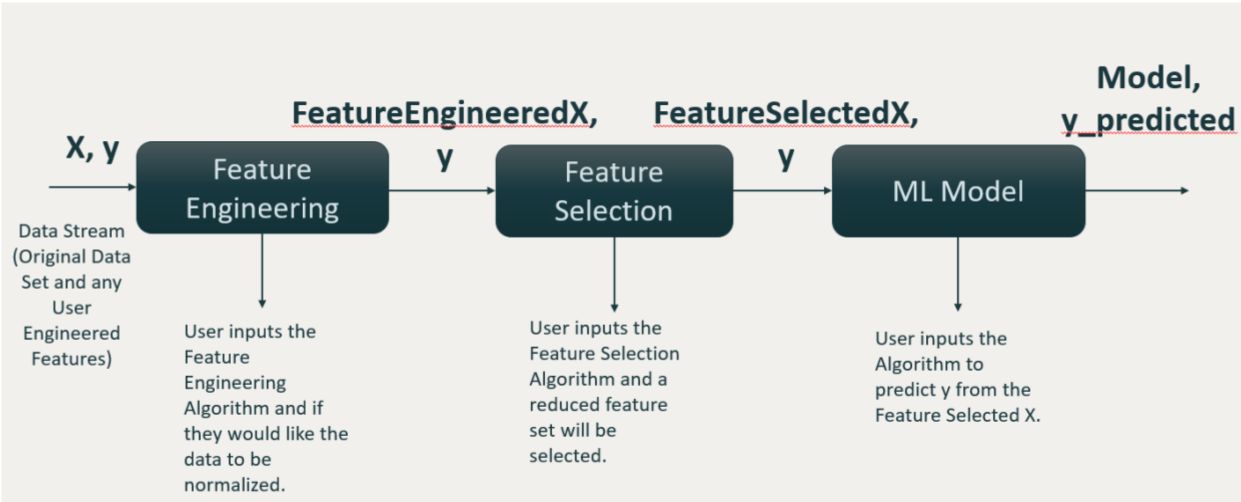


Figure 6: Machine Learning Pipeline Workflow

Feature engineering consists of a process of engineering or developing features from existing features. This can come from either an algorithm or manually. The algorithm we used was Principal Component Analysis (PCA). PCA reduces the dimensionality of the data set and the potential for correlated predictors by creating components that consist of all the predictors with varying coefficients. In Figure 7 we see an example of this applied to the MIT data set. Here it's clear that the predictors vary in what their contributions are. In Component 1, most of our contributions come from a range of current and temperature variables, whereas Component 5 is composed almost entirely from the minimum current. After selecting the components that explain 95% of the variance, we added these predictors to the complete feature set prior to feature selection. In addition to this automation, we generated features from the summary statistics of the voltage and current curves (minimum, maximum, range, variance). These features were also added to the available features before the feature selection step.

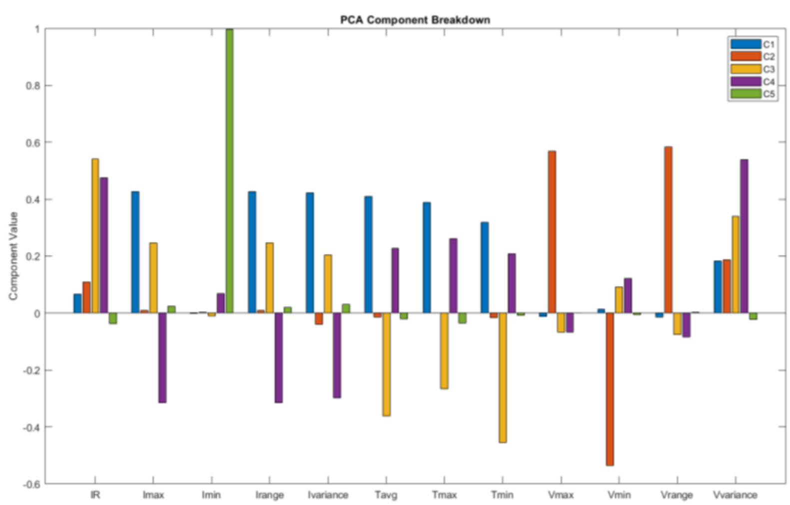


Figure 7: PCA Example on MIT data set

Feature selection is the process of reducing our predictors to only the most important. We used forms of penalized regression, penalized k-nearest neighbors, and fit scores to select the most descriptive features. This process allows us to reduce the possibility of over-fitting while gaining a better understanding of what data really correlates to SOH. The pipeline developed has a manual that describes all the feature selection algorithms in more detail. In general feature selection reduces the likelihood of creating an overfit or overcomplicated model by only using the important features. It also gave us an idea of what those important features are for our model. Similar to the feature importance discussed in the previous section this gives us an idea of what features have the largest contributions to variation in SOH.

Model fitting is the final stage of the pipeline, here we fit a model to predict SOH. Two algorithms were used here: Gaussian Process Regression (GPR) and Random Forest (RF). Here we would train one of these algorithms with the down selected features. Due to the differing chemistries, we didn't pass the model fit from the MIT data to the Nissan or Tesla data sets, rather we only retrained the pipeline with the features engineered and selected by the MIT training. We were able to achieve 0.25% Mean Absolute Percentage Error (MAPE) with a relatively simple model, meaning no PCA and using LASSO for feature selection, which selected the fewest number of features, 3. With our best model that included PCA and more rigorous feature selection we achieved a MAPE of 0.11%. As addressed in the previous section ML models have limitations extrapolating. We have confidence in our model's ability to predict based on the results of the MIT data set and that it predicts well for Nissan and Tesla (seen in Figure 8), but the range of SOH values is less there than MIT's data set. To predict outside the given SOH range we either need a model to predict across chemistries or to increase our range of data on the Nissan and Tesla batteries.

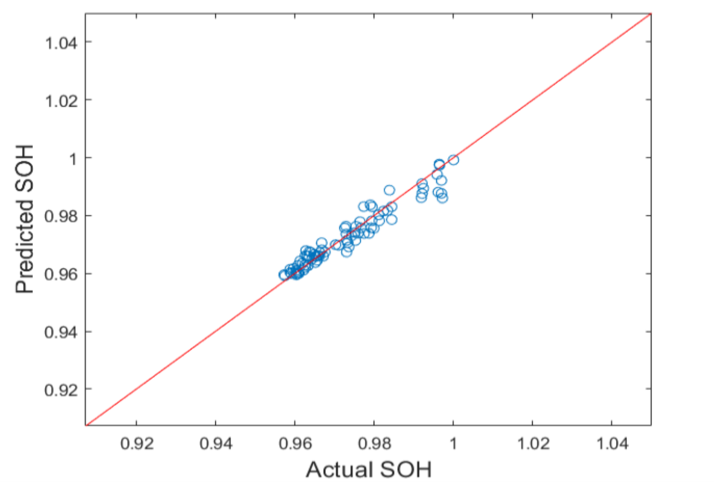


Figure 8: Example Nissan Fit

Task 4: Inform Life Balancing Reconditioning Strategy

Smartville's power electronics and life-balancing control strategy enable a first-use battery with heterogeneous cell SOHs to be reconditioned while it is performing grid services. The strategy cycles strong cells more and weak cells less to drive heterogeneous cell SOHs to become more homogeneous over several months. NREL will propose how the SOH and life estimation algorithm information from Task 2 can be used to inform Smartville's life-balancing control strategy.

NREL used the information in Objective 2 to help influence this strategy. In particular, our machine learning models (tree and ML pipeline) had elements of identifying predictor importance or selecting the most important features. As we expected time and cycles played important roles, but we have little control over those for future use. Identifying predictors like temperature and DoD give us ideas of how to condition the batteries for homogeneity or how to use them in the future. In addition to these generating features from the current and voltage curves give us the chance to understand how to more efficiently collect data in the future. In future work we can expand on this milestone through more creative feature engineering, more data streams, and larger SOH windows.

Task 5: CRADA Final Report – Preparation and submission in accordance with Article X.

This report serves to meet the requirement for the CRADA Final Report in accordance with Article X.

Participant Task Descriptions:

Task 1: Schedule and attend virtual progress meeting every 2 weeks

Task 1 above is inclusive for both partners.

Task 2: Testing Protocols

The Participant will describe to NREL its characterization test protocol. The Participant shall share example pack and cell-level current/voltage/temperature data for several battery packs undergoing that protocol. The battery pack data should be both pre- and post-life-balancing reconditioning. Those battery packs should consist of a mix of different SOHs (e.g. some barely aged and some severely aged. Explained in Task 3 above.

Task 3: Collect Data

The Participant will collect new diagnostic pulse data for NREL to use in machine learning.

Explained in Tasks 3 & 4 above.

Subject Inventions Listing:

None.

ROI #:

None.