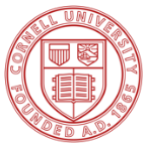


Cornell University
Advisor: Dr. Martin Wells

National Aeronautics Space Administration
Sponsor: Dr. Chetan Kulkarni

Battery Health Management for Small Satellites

Jingyi Chen (jc2498), Shuo Feng (sf587), Bingjie Huang (bh572),
Lina Takemaru (llt45)

May 14, 2021

Executive Summary

Lithium-ion battery health management and prognostics are essential in ensuring the success of devices such as electric cars, aircrafts, power grids, and satellites. However, Lithium-ion batteries begin to degrade under certain extreme conditions with consequential outcomes. To avoid these dangerous situations and ensure the success of certain satellite missions, the International Space Station (ISS) and the National Aeronautics and Space Administration (NASA) aim to gain more insight into battery health and performance over time. More specifically, we hope to provide valuable information regarding remaining system operations that can be supported under battery power.

We have divided the discussion of the statistical models we implemented into two main sections — State of Health (SoH) analysis and forecasting. Within each of these sections, we present multiple models with the goal of predicting when and why Lithium-ion batteries fail. The findings that we discuss in this manuscript have the potential to provide significant information about Lithium-battery performance to researchers and scientists at NASA. The health management of these power sources can greatly contribute to transformative technologies to ensure safety, optimize cost, facilitate operational decision making, and extend the life of satellite missions. As Lithium-ion batteries become more and more commonplace in dynamic systems, estimation of battery functionality is critical.

1. Introduction

Recent developments and improvements in Lithium-ion battery health, cost, and efficiency have emphasized their crucial role in a variety of consequential and pivotal innovations, such

as electric cars, aircrafts, and smartphones. The International Space Station (ISS), in part run by the National Aeronautics and Space Administration (NASA), is particularly interested in advancing satellite battery health. In the foreseeable future, the ISS plans to launch a handful of small satellite missions largely powered by Lithium-ion battery packs. In order to ensure these satellites' success, NASA's researchers and scientists hope to extend the lifetime of these space missions.

In order to achieve this goal, dependable, energy-dense, and durable power sources are essential. Lithium-ion batteries begin to deteriorate under certain conditions such as high temperature and overcharge, which may lead to serious accidents when a device is in operation (Kaneko et al., 2017). Thus, it is important to develop approaches that produce accurate predictions of age-dependent changes in battery performance. As such, in this report, we aim to build statistical models capable of detecting battery failure. The dataset provided to us from NASA Ames Researcher, Dr. Chetan Kulkarni, consists of time-series data from a Lithium-ion battery aging experiment from 2015. This data can be found in and downloaded from the NASA prognostic data repository (Bole, 2015).

It is widely accepted that as batteries age, charge storage capacity decreases while internal resistance increases (Kaneko et al., 2017). Because voltage and resistance are subject to these time-dependent fluctuations, we chose to quantify these measurements to provide insight into battery failure (Broussely, et. al., 2005). Using this information, we chose to focus on two indicators of battery performance and health — voltage and State of Health (SoH), which we further explain in their respective sections.

2. Data Description

Here, we provide a brief description of each of the features used when training and testing our models. Note that these descriptions are based on the explanations contained in the data file provided to us by Dr. Chetan Kulkarni (Bole, 2015).

i. **type**: identifies the period of the corresponding observation: "C" = Charging, "D" = Discharging, "R" = Resting
ii. **relativeTime**: vector of the sample time in seconds; reference is the start of each cycle
iii. **time**: vector of the time in seconds; reference is the start of the experiment
iv. **voltage**: vector of sample voltages in Volts (V)
v. **current**: vector of sample currents in Amps (A)
vi. **temperature**: vector of sample temperatures in degrees Celsius (C)
vii. **date**: date and time of the start of each cycle in dd-Mon-yyyy HH:MM:SS format

To explain this data further, we categorized our data into three types of profiles based on how the experimental data was collected — reference, pulsed load, and random walk. Note that the following descriptions are based on the same explanations in the data file mentioned above (Bole, 2015). Before we proceed, we would like to clarify some battery terminology used heavily in our analysis. The term discharge refers to the phenomenon in which electricity is *taken from* the battery, whereas the term charge refers to the setting in which electricity is *sent to* batteries fur use. During the charging process, the chemical processes used during discharge are reversed and electricity is stored for future use ("How Lithium-ion Batteries Work", 2021).

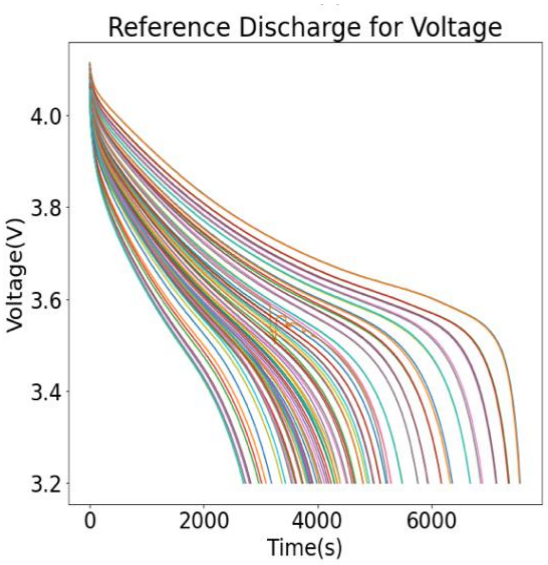


Figure 1a: *Voltage of reference discharge profiles over time*

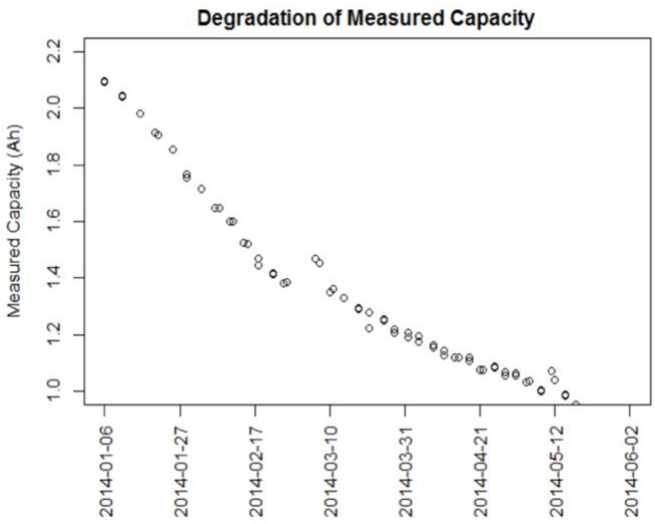


Figure 1b: *Degradation of measured capacity over time*
Note. From "Description of Room Temperature Random Walk Charging and Discharging Data Sets" by Dr. Bole, B., 2015.

2.1 Reference

In this stage, the data collected serves as a baseline for future comparisons of battery degradation. For our models, we focused on two stages in this period — reference charge and reference discharge. One reference charge cycle and one reference discharge cycle were performed after every 1500 Random Walk (RW) cycles, which are explained in detail in section 2.3 below.

i. **Reference charge:** In these cycles, the batteries are charged using a constant current of 2A until they reach 4.2V. The batteries are then charged using a constant voltage until the current falls below 0.01A.

ii. **Reference discharge:** In these cycles, the batteries are discharged at 2A until the voltage exceeds 3.2V.

To visualize trends in the reference period more clearly, we plotted a subset of the profiles over time. As depicted in figure 1a, as the batteries age, the voltage decreases.

This implies that the differences in voltage across these reference discharge periods are a result of battery health degradation (Bole, 2015). Further, figure 1b depicts the decrease in measured capacity over time — as batteries age, their charge storage capacity decreases due to an increase in internal resistance (Bole, 2015).

2.2 Pulsed Load

One major type of current utilized in electrical experiments is pulsed current, in which the application of current pulsates for a certain period of time, followed by a period of rest. For our models, we focused on two stages in this period — pulsed load discharge and pulsed charge. One pulsed load discharge cycle and one pulsed charge cycle were performed after every 3000 RW steps. The purpose of these steps is to gauge transient changes in battery dynamics.

i. **Pulsed load discharge:** In these cycles, the pulsed current discharge was applied at 1A for 10 minutes, followed by a 20-minute rest period.
ii. **Pulsed charge:** In these cycles, the pulsed current charge was applied at 1A for 10 minutes, following by a 20-minute rest period.

To visualize the effects of pulsed load discharge and charge, we plotted changes in current and voltage during these periods. Figure 2a illustrates a 1A pulsating current. The height of the vertical lines represents the pulse strength. The horizontal lines at 1A occur when the pulse is activated (charging), while those at 0A occur when the pulse is deactivated (discharging). Figure 2b shows a similar trend as figure 1a; as the battery ages, its ability to hold charge decreases. At the beginning of the experiment, the battery

is able to reach 4.2V during the pulsed charge stage but decreases to around 3.7V as the experiment progresses. This large voltage drop is predominantly due to an increase in the internal resistance of the battery circuits (Farrier & Bucknall, 2019).

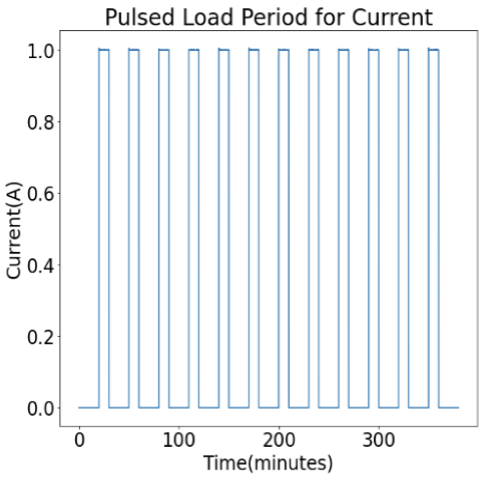


Figure 2a: *Current in pulsed load discharge and charge periods*

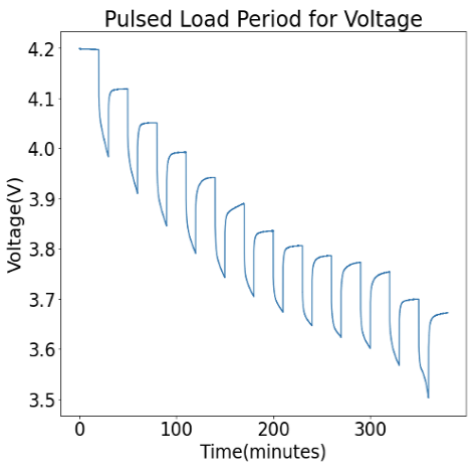


Figure 2b: *Voltage in pulsed load discharge and charge periods*

2.3 Random Walk (RW)

In the RW cycles, randomness is introduced to better represent the effect of different currents on battery degradation. Initially, a charging or discharging current is randomly select from the following set of values: { -4.5A, -3.75A, -3A, -2.25A, -1.5A, -0.75A, 0.75A, 1.5A, 2.25A, 3A, 3.75A, 4.5A}. The negative values represent **charge (random walk)**, and the positive values represent **discharge (random walk)**. The battery then undergoes charging or discharging at the selected current until its voltage surpasses 4.2V or its duration exceeds five minutes. Next, there is less than one second of rest, and a new value is selected. One RW cycle consists of one discharge/charge stage and one rest stage.

As expected, in figures 3a and 3b, there is a lot more randomness in the measurements of current and voltage. During the second half of the RW cycles, the battery reaches the

upper (4.2V) and lower (3.2V) voltage bounds more frequently, especially in the stages that demand a current greater than 2A. This is again due to the degradation of battery performance and health over the course of the experiment.

As shown in figures 1-3, quantifying and describing fluctuations in current and voltage would provide a good indication of battery failure. In the following sections, we introduce and thoroughly explain the models we implemented in hopes of achieving this goal.

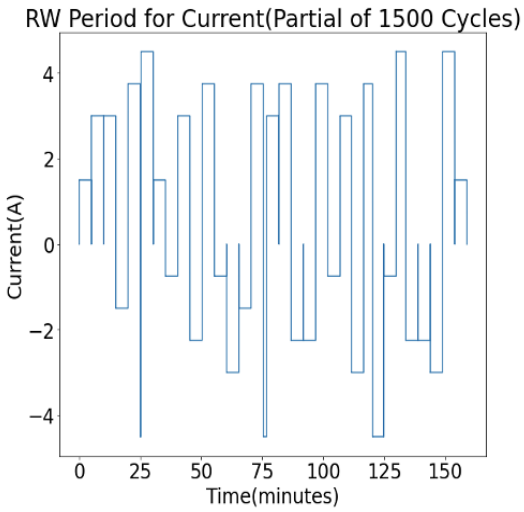


Figure 3a: *Current in 1500 RW cycles*

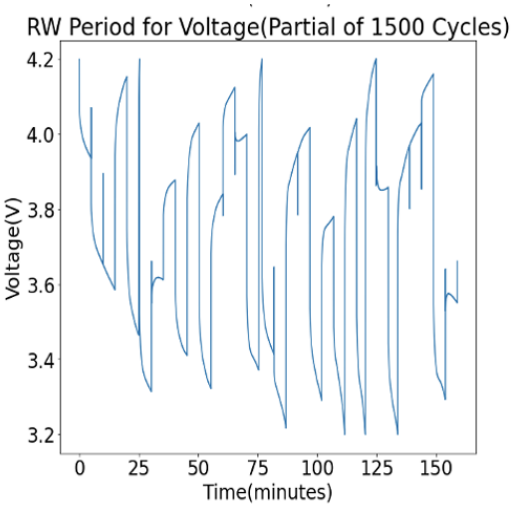


Figure 3b: *Voltage in 1500 RW cycles*

3. Battery Health and Performance Analysis

Here, we explain all the methods, techniques, and algorithms we implemented. We begin with the definitions and purposes of a few key terms we utilized when analyzing battery health and performance for characterization, such as internal capacity and resistance. We then define a handful of formulas in detail to provide more insight into the reasoning behind our models. In our explanations of State of Health, we lay out the purposes and importance of SoH Analysis, and how it assists us in detecting battery failure. Further, we explain our forecasting models and compare the results.

3.1 Feature Measurements

Battery Voltage, Current, Temperature, and Time

There are four basic readings in a battery system — voltage, current, temperature and time. Here, the time feature is treated as the cycle number of the battery's current state. Using these four measurements, we visualized the dynamic changes of battery performance as the number of cycles increased, or equivalently, as the battery aged.

Discharging Time

Discharging time is the amount of time a battery took to complete one discharging stage for each cycle. As the battery ages, the duration to complete one discharging cycle decreases due to a lower storage capacity, as shown in figure 4a. This indicates that the battery is losing voltage at a faster rate than its initial ideal condition.

Discharging time is measured as the difference between the last entry and the first entry in relative time readings for each discharging stage provided in the dataset. As the number of cycles increases, discharging time is expected to decrease linearly. However, in the pulsed load period, the discharging time is recorded as either the amount of time the battery takes to reach 3.2V, or 10 minutes (600 seconds) if the voltage never reaches 3.2V. As a result, the discharging time is not a strictly linearly decreasing trend for the pulsed load period, as shown in figure 4b.

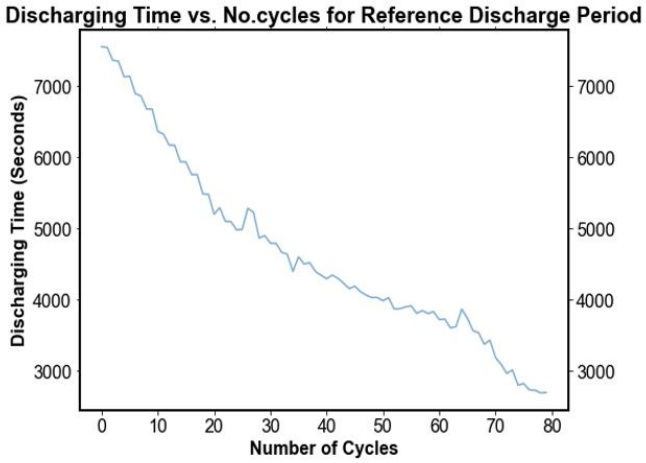


Figure 4a: *Discharging time in the reference discharge period*

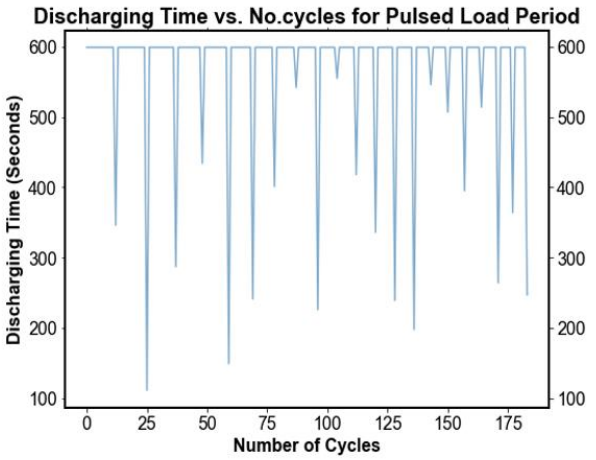


Figure 4b: *Discharging time in the pulsed load period*

Internal Resistance

Internal resistance provides an indication of battery impedance, which we refer to as an ideal voltage source, in conjunction with a resistance series. Here, we use the term ideal voltage source to represent the initial voltage capability of the Lithium-ion battery at the start of the experiment. In addition, for simplicity, we treat resistance and impedance the same conceptually. At this stage, the battery is at its peak performance. The resistance or impedance measurement is dependent on the size, state of charge, chemical properties, age, temperature, and the discharge current of the battery.

In a battery circuit, the battery is assumed to exhibit the characteristics of an ideal voltage source at the start of the experiment. Despite how the battery is loaded, the voltage at the source terminals, also referred to as open circuit voltage, will always stay the same. We use the term open circuit voltage to refer to the maximum voltage a battery can reach at its peak performance (Qi & Wang, 2012). However, a battery's ability to act as an ideal voltage source

changes as it ages. As a result, a battery system can be created using an open circuit voltage and a load voltage (resistor) in series. Internal Resistance, measured in Ohms, can be calculated from the open circuit voltage, V_{ocv}, load voltage, V_{bat}, and current, I in the following formula:

$$R_b(SOC, T) = \frac{V_{ocv}(SOC, T) - V_{bat}(SOC, T)}{I_{pulse}}$$

As the battery ages, internal resistance will increase. Figures 5a and 5b show the internal resistance against the number of cycles for the reference discharge period and the pulsed load period. We can observe that internal resistance increases overall, consistent with what we expect due to increasing battery impedance. However, in the pulsed load period, instead of linearly increasing, the internal resistance fluctuates up and down recursively. This is because the open circuit voltage in the pulsed load period refers to the voltage in each pulsed stage cycle, not for the ideal source voltage.

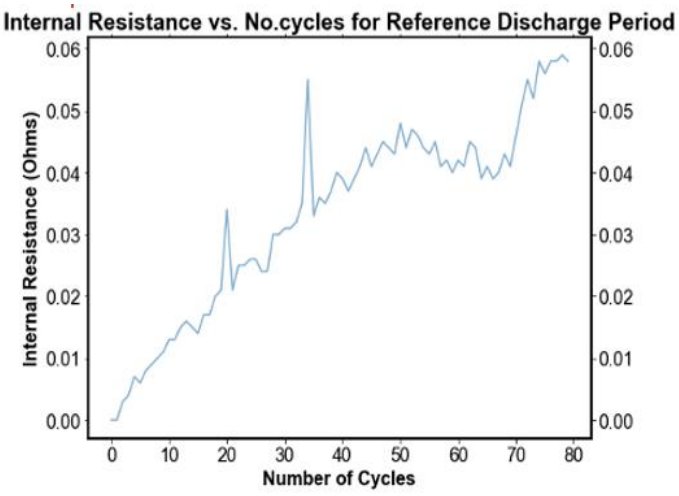


Figure 5a: *Internal resistance in the reference discharge period*

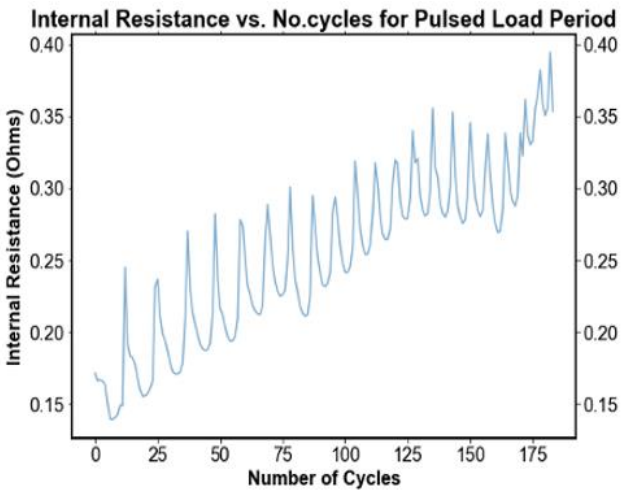


Figure 5b: *Internal resistance in the pulsed load period*

Capacity

Battery capacity is a measure of the charge stored. Capacity (measured in Amp hours) represents the maximum amount of stored energy of a battery under certain specified conditions. As a battery ages, the capacity is expected to decrease. The measurement of capacity is proposed as the integral of current on discharging time as follows:

$$C_d = \int_0^{t_{cutoff}} Idt$$

As mentioned above, discharging time in the pulsed load period does not follow a linear pattern. Therefore, the capacity against the number of cycles in the pulsed load period does not exhibit a clear pattern. As observed in figure 6a, the overall trend of capacity for reference discharge and random walk is decreasing, while in figure 6b, there is a pulsating increase and decrease in capacity. Again, this is what we expect.

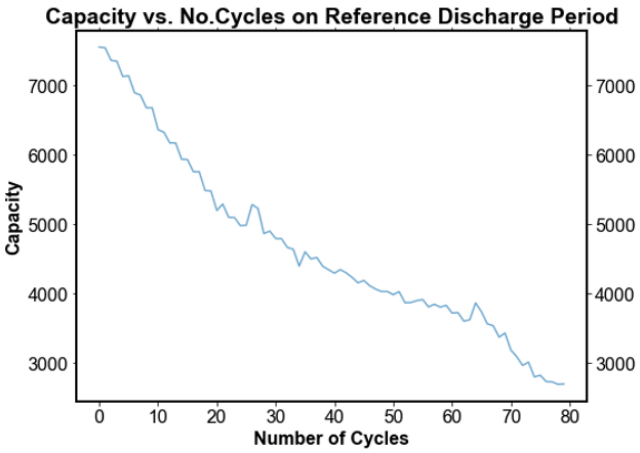


Figure 6a: *Capacity in the reference discharge period*

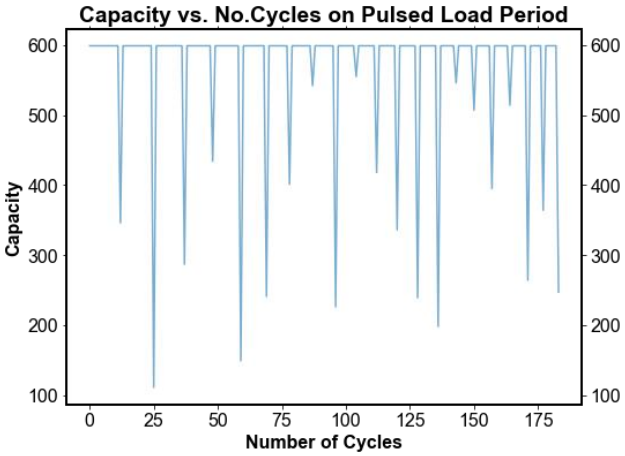


Figure 6b: *Capacity in the pulsed load period*

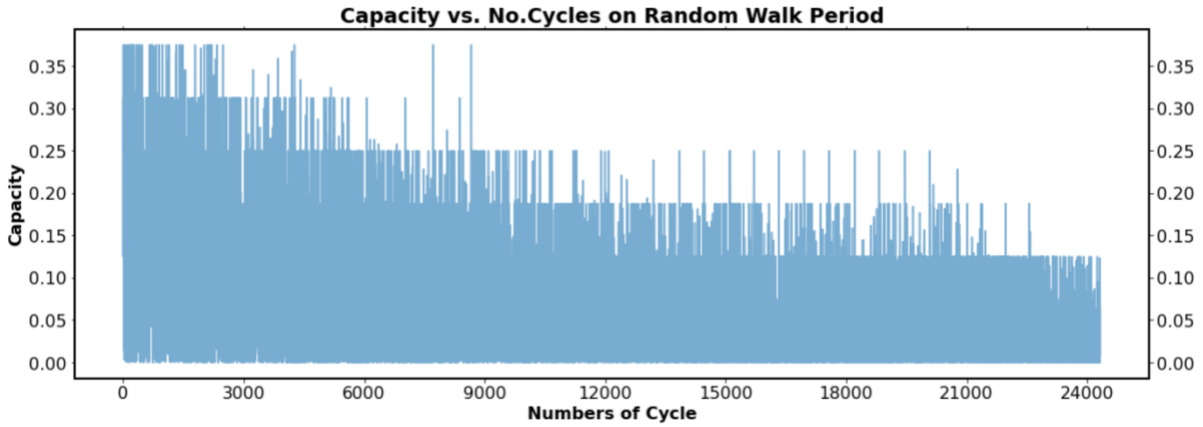


Figure 7: *Capacity in the random walk period*

State of Charge (SoC) Measurement

The SoC of a battery is its remaining capacity, measured using a voltage-based method.

The SoC of a battery is used to describe its remaining capacity and is a very important parameter for controlling battery system management. As the SoC is an integral parameter that reflects battery performance, an accurate measurement can provide protection for and improvements in battery life. There are various methods to measure and estimate SoC.

SoC is calculated using the voltage-based method mentioned above. There is an approximately linear relationship between the discharging voltage and the number of cycles. As such, SoC is calculated using the following formula, where V is the average discharging voltage for each discharging cycle. SoC decreases as the number of cycles increases. We show these calculations visually in figures 8a and 8b, in which we can see the linearly decreasing trend and pulsating trend consistent with all other calculations and measurements.

$$SOC = \frac{V - V_{min}}{V_{max} - V_{min}} \times 100\%$$

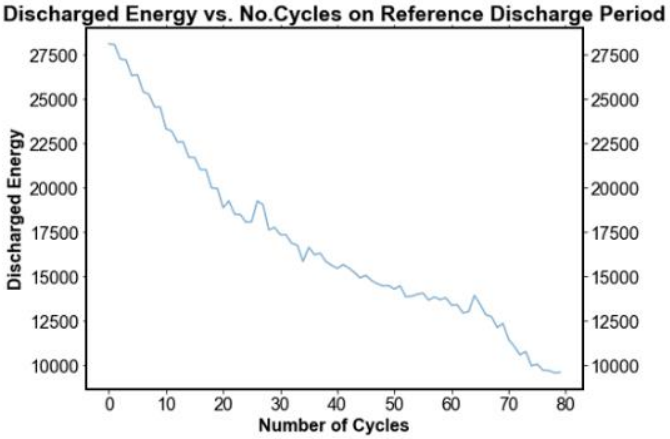


Figure 8a: *State of Charge in the reference discharge period*

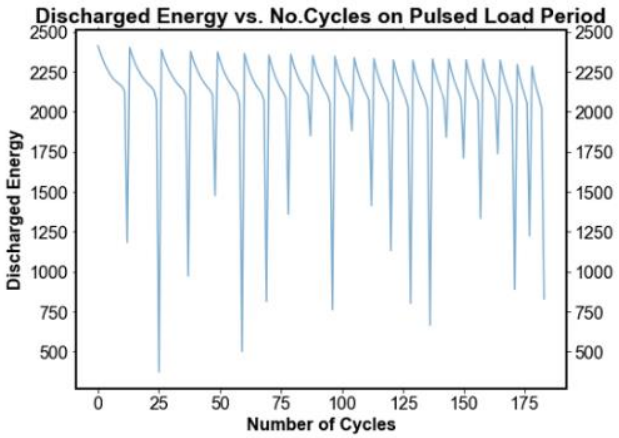


Figure 8b: *State of Charge in the pulsed load period*

Discharged Energy

Discharged energy is the amount of stored energy used during discharging stages. It is calculated as the product of capacity and voltage, using the average discharging voltage for each cycle. The formula is as follows: $Q = C(t) \times V(t)$. As discussed in previous sections, the discharged energy decreases as the battery ages due to a decrease in capacity. As a result, the amount of energy transferred decreases as well.

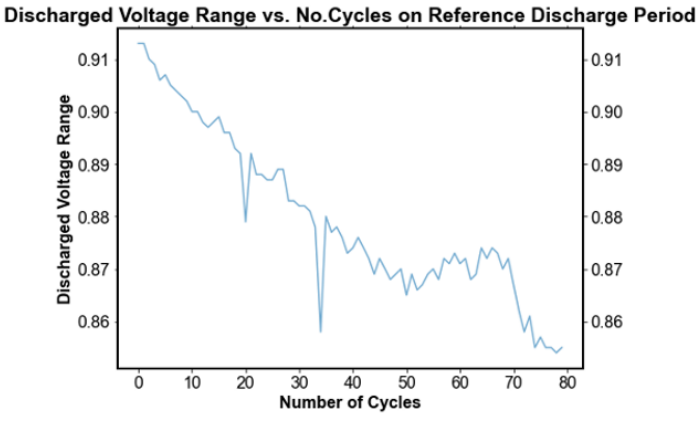


Figure 9a: *Discharged energy in the reference discharge period*

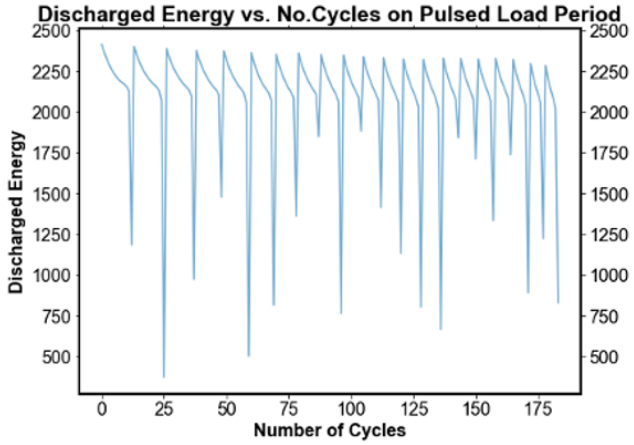


Figure 9b: *Discharged energy in the pulsed load period*

Discharged Voltage Range

Discharged voltage range is the voltage difference for each discharge cycle from beginning to end. This feature measures how much voltage changes during discharge, which is also a reflection of a battery's ability to transfer energy. The calculation for discharged voltage range is straightforward; it is simply the maximum voltage minus the minimum voltage for each discharging cycle. Figures 10a and 10b illustrate the behavior of these calculations in the reference discharge period and in the pulsed load period. For lack of repetition, we simply acknowledge that the trends are largely the same as in previous sections.

$$\text{Discharged Voltage Range} = V_{max} - V_{min}$$

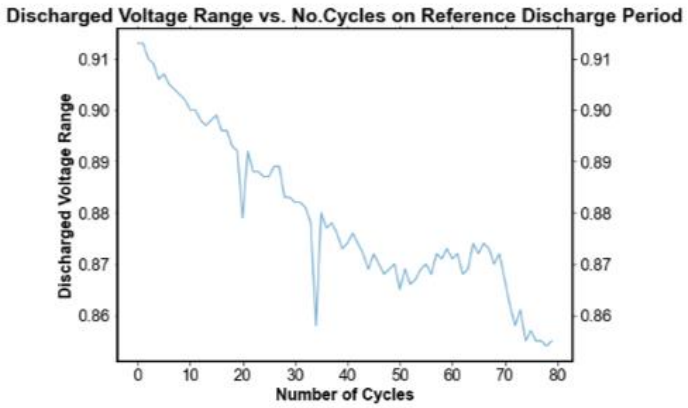


Figure 10a: *Discharged voltage range in the reference discharge period*

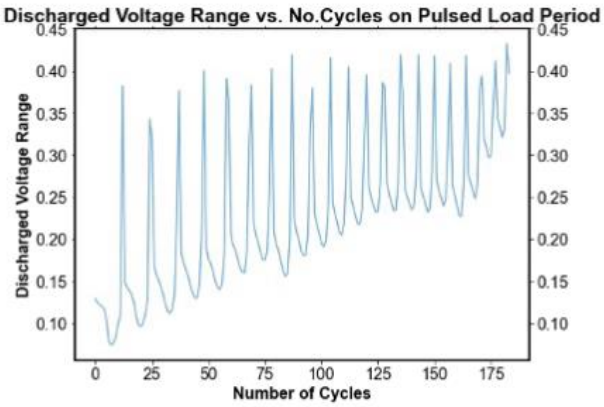


Figure 10b: *Discharged voltage range in the pulsed load period*

3.2 State of Health (SoH) Analysis

3.2.1 Overview of SoH

Lithium-ion battery cells are used in many applications, such as portable electronics, electric vehicles, and aerospace manufacturing. Lithium-ion batteries are one of the most popular rechargeable batteries. It is crucial to measure and estimate its rechargeable capabilities and its useful lifetime. Therefore, an indication of battery health condition should be determined to reflect the current health situation. In applications, the SoH of a battery is introduced to

represent battery system health, a measurement of the battery's ability to store and deliver electrical energy, compared to the initial ideal condition. The SoH parameter should be a combination of a variety of physical feature measurements.

This section deals with the algorithms utilized for SoH estimation, based on the reference discharge period and pulsed load period. Since the situation of the experiment implemented in the random walk period is significantly more complex and casual than that in the reference discharge period and pulsed period, the SoH analysis on the random walk period is harder to perform. Important features in addition to the raw measurements in the data set will be introduced and the technical environment specifications for SoH parameter determination will be explained. An overview of the estimation methods of SoH in cycle prediction will also be presented. In addition, this section will present model selection for SoH estimation algorithms based on several proposed and implemented techniques. Finally, the evaluation procedures are outlined based on real-time data measurements. This procedure, combined with the result output, delivers a readily comprehensible interpretation for all readers.

3.2.2 Battery Failure Threshold

The battery ability's fading is caused by the loss of reagents in the battery electrodes through irreversible uses in the battery. This loss of reagents will result in reduced Capability, SOC and increased Internal Resistance. As the above feature analysis and dynamic figures shows, these features, separately or combined, can reflect the health situation of a battery to some extent. To determine whether the battery in current time is healthy or not, and how much useful lifetime left, a failure threshold point of SOH based on indication parameters should

be pre-determined so that the current battery can be directly classified as healthy if its health indicator parameter does not reach failure threshold, as unhealthy if its health indicator reaches failure threshold. In addition, it will be better if the useful lifetime left of a battery can be calculated through some way.

Battery Failure Threshold for Reference Discharge Period

For reference discharge period, SOC can be used as a health performance indication, and it decreased from 1 to 0 in battery whole life as aging. Therefore, health condition can be reflected in the value of SOC, battery is very healthy at its initial ideal stage when SOC equals 1, very unhealthy at its end of life when SOC reaches 0. From experience a general replacement threshold for SOH performance is suggested at 25 % of initial SOC.

$$threshold = 25\% \times SOC_{initial}$$

Therefore, the failure threshold would SOC = 25%, in other words, when SOC drops below 25%, the battery is indicated as unhealthy. Since SOC is not a directly measurement from basic readings, the goal is to convert this failure threshold, based on SOC, to another indication which can be easily detected and understood. As a result, cycle number can be the corresponding failure threshold to represent to health condition of battery. In the case, when SOC reaches 25%, the corresponding cycle number is No.65, therefore, cycles after No.65 are classified as unhealthy.

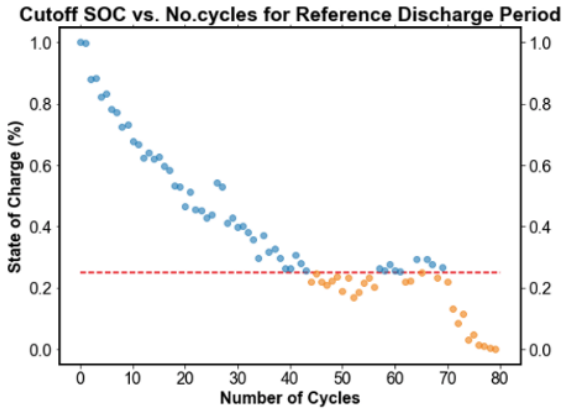


Figure 11a: *Cutoff State of Charge % in the reference discharge period*

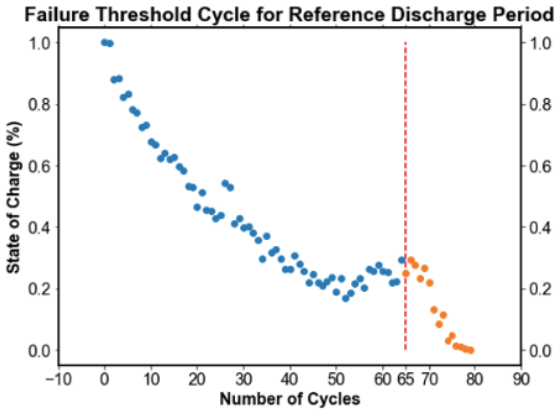


Figure 11b: *Failure State of Charge % in the reference discharge period*

Battery Failure Threshold for Pulsed Load Period

For pulsed load period, SOH can be evaluated through internal resistance, where a maximum R_{in} value may be established as a limit parameter for the SOH. A general replacement threshold for SOH performance is suggested at 200 % of initial R_{in}.

$$threshold = 200\% \times R_{initial}$$

In this case, the failure threshold for pulsed load period is R_{in} = 3.4 Ohms, when R_{in} reaches 3.4, the battery is unhealthy, and the corresponding cycle number No.150. Therefore, all cycles after No.150 are classified as unhealthy.

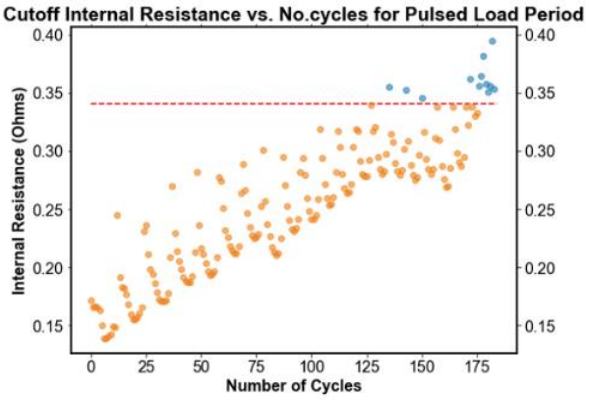


Figure 12a: *Cutoff for internal resistance in the pulsed load period*

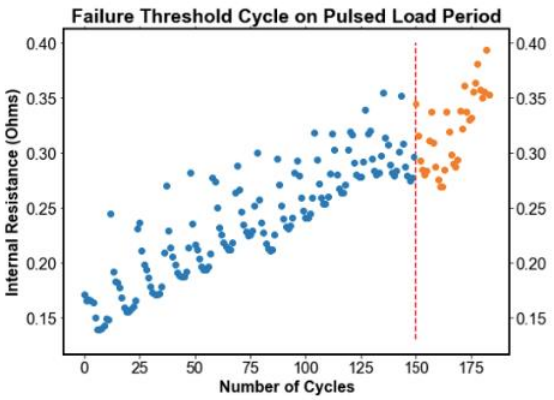


Figure 12b: *Failure internal resistance in the pulsed load period*

3.2.3 Algorithm for Estimating SOH

Since the cycle number is a straightforward representation of lifetime and the failure threshold is pre-determined, the goal is to estimate the real-time cycle with real-time feature measurements. The input includes basic data: temperature, voltage, current, discharging time, and additional features such as capability, internal resistance etc. Then, we performed variable selection starting with the full variable model. After selecting the best subset of variables, we fit models using different algorithms on these selected variables and compared the cycle estimation accuracies. Finally, the model with highest estimation accuracy was selected as the final model, which can estimate the current cycle number of the battery and report its health status and remaining useful lifetime.

Data

- Predictor Variables: Current, Voltage (Open Circuit), Voltage (Battery), Temperature, Capacity, Internal Resistance, Discharging Time, SOC, Discharged Energy, Discharged Voltage Range
- Response Variable: Cycle

Methods

- **Feature Selection**

  Starting will the full model with all 11 variables, use Ordinary Least Squares Regression and AIC as criteria to perform variable selection.

  - Selected Variables for Reference Discharge Period: Current, Voltage (Open Circuit), Voltage (Battery), Temperature, Capacity, Internal Resistance, Discharging Time, Discharged Energy, Discharged Voltage Range

- Selected Variables for Pulsed Load Period: Current, Voltage (Open Circuit), , Temperature, Internal Resistance, SOC, Discharged Voltage Range

- **OLS Regression**

  Fit OLS regression model with selected variables for reference discharge period and pulsed load period separately. Then, use the model to estimate cycles for each period, round the result to an integer value.

- **Random Forest Regression: iterate regression for depth = [5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15]**

  Random Forest Regression is a supervised learning algorithm that uses ensemble learning method for regression. A Random Forest operates by constructing several decision trees on selected variables. Iterate random forest regression on depth = 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, respectively, and output estimated coefficient parameters.

Model Selection Criteria

For all fitted models above, calculate the following criteria parameters by each model for reference discharge period and pulsed load period, separately. Then select the best model by the following priority, for each period.

- **Estimation Accuracy (1st)**

  Since the response variable is number of cycles, which is an integer, while the regression model might generate decimal numbers, then the normal model selection criteria MSE might not be appropriate here. As the goal is to evaluate model's ability to estimate cycle correctly, therefore, instead of MSE, calculate the portion of correct

estimation among all cycles is more logical. Define estimation accuracy as the ratio of number of correct estimations to the number of total cycles.

$$\text{Accuracy} = \frac{\#of\ correct\ \ estimation}{\#of\ total\ \ cycles} \times 100\%$$

Since models produce decimal number, need to first round results to integer, then compare with actual number of cycles.

- **Modified Accuracy (2nd): if estimation error = 1 cycle, count as correct**

  If the result of estimation accuracy defined above is not desired, too low for an estimation model, then a reasonable slight change of accuracy evaluation will be needed. If treat estimations, which has estimation error only 1 cycle with actual cycle, as correct estimation, then the modified estimation accuracy would be better than original estimation accuracy. The reason behind this modification is that even though the model is good enough, the estimated result would come out a near decimal number which might be rounded to the cycle next to the actual one, and since battery lifetime experiment usually consists of many cycles, estimation with 1cycle error could be ignored.

- **R Squared (3rd)**

  Ordinary R Square could also be used to test model performance.

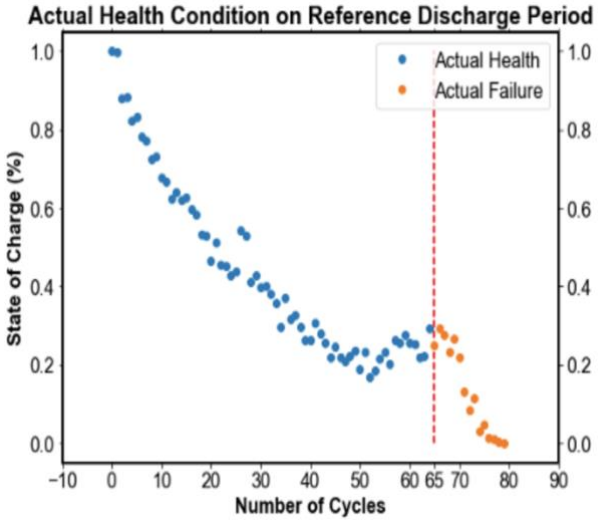
- **Test MSE (4th)**

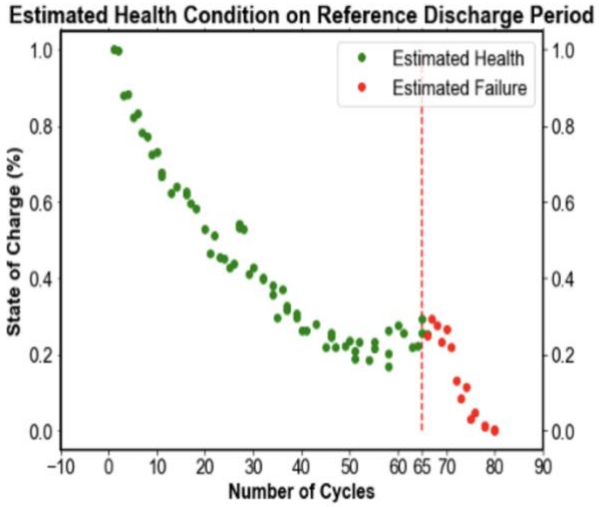
  Ordinary test MSE can also be used to test model performance, and the whole period is used to calculate test MSE since sometimes the feature parameters fluctuate drastically due to some chemical properties, so only use a subset for test MSE might generate large abnormal MSE.

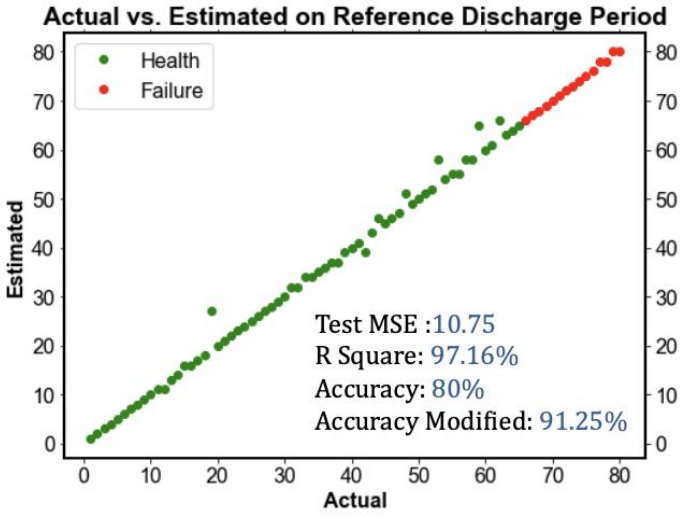
3.2.4 Model Analysis

Reference Discharge Period

As figures 13a and 13b show, the random forest regression model with a depth of 9 generates well-performing estimations based on cycle number. The estimated results have a similar distribution to the actual data for both the healthy and unhealthy stages. The estimation accuracy is 80% and the modified accuracy is 91.25%, as shown in figure 13c.



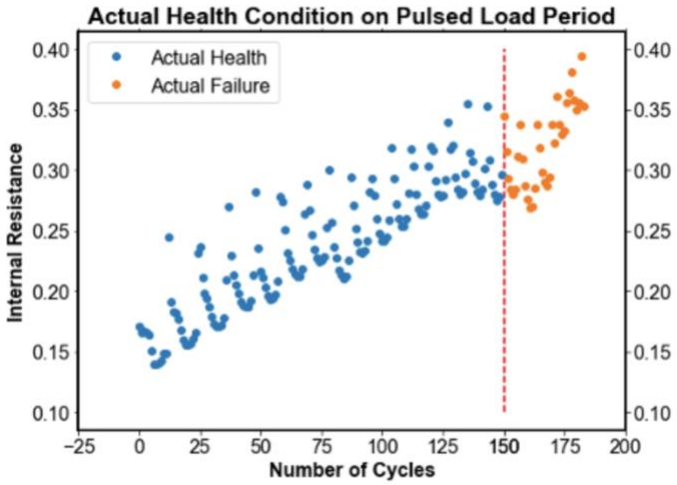


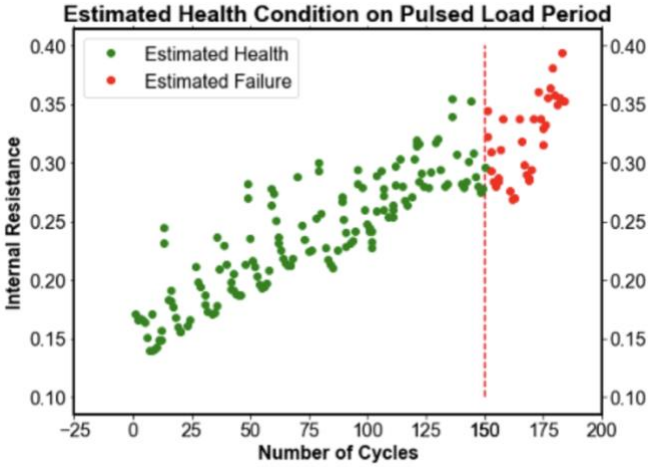


Pulsed Load Period

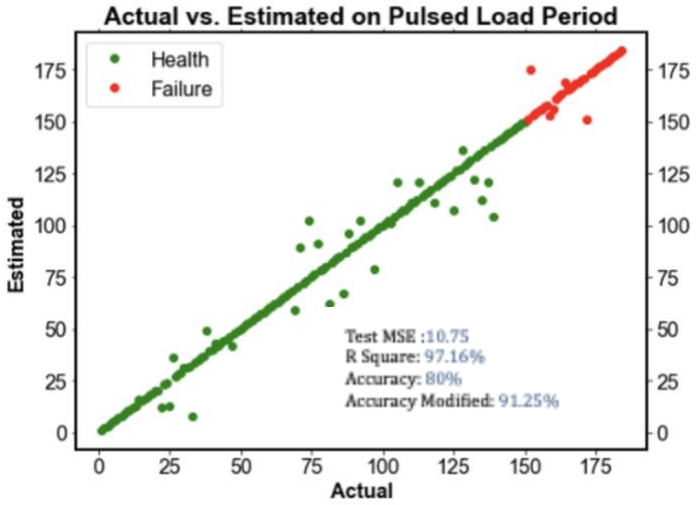
Best model: Random forest regression with depth = 11

Compare actual cycle vs. predicted cycle





The random forest regression model with depth = 11 generates good estimation on cycle number, the estimated results are distributed quite similar to the actual data distributed, both for healthy and unhealthy part. Following plot and parameters of estimated value vs. actual value shows a linear relationship, indicate good fit. Estimation accuracy is 80.43%, and modified accuracy is 82.61%, good fit.



3.2.5 Results

Testing these models with input of measurement generates the following output.

Reference Discharge Period

```
------------------------------------------------------------------------
For the follwing input data on reference discharge period:
current                  0.999901
voltage_ocv              3.724370
voltage_bat              3.598367
internal_resistance      0.040004
duration              4345.000000
temperature             29.676644
soc                      0.262072
discharge_energy      15633.362521
discharge_range          0.873000
capacity              4344.571479
The estimated current cycle is at No. 40
There are 25 cycles left to reach failure threshold
The actual current cycle is at No. 40
There are actual 25 cycles left to reach failure threshold
Correct Estimation
```

```
------------------------------------------------------------------------
For the follwing input data on reference discharge period:
current                  0.999895
voltage_ocv              3.724370
voltage_bat              3.595677
internal_resistance      0.041004
duration              4191.000000
temperature             30.225704
soc                      0.246318
discharge_energy      15067.907180
discharge_range          0.872000
capacity              4190.561986
The estimated current cycle is at No. 46
There are 19 cycles left to reach failure threshold
The actual current cycle is at No. 46
There are actual 19 cycles left to reach failure threshold
Correct Estimation
------------------------------------------------------------------------
```

Pulsed Load Period

```
------------------------------------------------------------------------
For the follwing input data on pulsed load period:
current                0.999900
voltage_ocv            4.029693
internal_resistance    0.357614
temperature           33.418471
soc                    0.380808
discharge_range        0.333000
The estimated current cycle is at No. 180
There are estimated 30 cycles overused beyond failure threshold
The actual No. cycle is:  180.0
The actual current cycle is at No. 180
There are actual 30 cycles overused beyond failure threshold
Correct Estimation
------------------------------------------------------------------------
```

```
------------------------------------------------------------------------
For the follwing input data on pulsed load period:
current                0.999900
voltage_ocv            4.029693
internal_resistance    0.357614
temperature           33.418471
soc                    0.380808
discharge_range        0.333000
The estimated current cycle is at No. 180
There are estimated 30 cycles overused beyond failure threshold
The actual No. cycle is:  180.0
The actual current cycle is at No. 180
There are actual 30 cycles overused beyond failure threshold
Correct Estimation
------------------------------------------------------------------------
```

The result output is easy to understand. When we input the full set of features, which contains 11 variables, into the model, the output will show the significance of the features. The model helps to know which features are important to the SOH indicator in cycle number estimation, therefore, in later experiment, design and measurement of these important features will obtain much consideration. Selected features for reference discharge period

and pulsed load period are different due to different circuit design in these two periods. With input data, model will estimate which cycle number is the battery currently at, how many cycles left to reach the failure threshold, actual cycle number, actual useful lifetime left and correct estimation or error estimation. If the current cycle number is beyond the failure threshold cycle number, model will print out 'overused', also, the estimation error can be easily compared in the output. This result output is comprehensive to all people since cycle number is a direct reflection of useful lifetime.

3.3 Battery Voltage Forecast

3.3.1 Overview of voltage forecasting

After a long-term use on battery, the internal resistance within battery changes dramatically and battery voltage also drops with change of internal resistance.

Therefore, it's important to analyze the underlying principles and methods to predict for the battery voltage.

In the dataset, each observation selected is defined to be one cycle. For each cycle, we took the average value for the columns 'current', 'time', 'voltage', 'temperature' and created new variables: 'mean_current' , 'mean_time', 'mean_'voltage' and 'mean_temperature'. Here, we predicted the battery voltage for each cycle on the random walk period and the reference discharge period separately.

3.3.2 Forecasting battery voltage on the random walk period

Description and Data Cleaning:

Given 112682 data observations for the random walk period (discharge, charge and rest types), we randomly selected the data in the random walk period with observations' index $100*i$ where $i \in [0,1126]$. For convenience, we threw away the Rest type ("R") and kept the discharge and charge types' data on the random walk period. Therefore, without data of Rest type, we have 584 observations which are corresponding to 584 cycles on the random walk period.

Here, given those 584 cycles, we predicted the variable 'voltage' ('mean_voltage') for each cycle using the Ordinary Least Squares (OLS) and predicted the residuals of the 'voltage' using Exponential Smoothing. Specifically, we utilized multivariate variables to fit OLS models. To select statistically significant variables, we used the best OLS model with the least Akaike Information Criterion (AIC) to determine what features to keep in our model. Then, we used residuals returned by the best OLS model as a predictor. Later on, we forecasted the residuals in Exponential Smoothing with trend and seasonality. During the whole modeling process, we split our data into a training set (before May 2014, 374 cycles) and validation set (after May 2014, 210 cycles) for model evaluation.

Exploratory Data Analysis:

The plot of the timeseries is the Figure 3.3.2 a. For the column 'mean_temperature', we found that those high temperatures are more likely to occur in later periods (especially after May), see Figure 3.3.2 b. This is caused by battery degradation, similar to the decreasing on battery voltage.

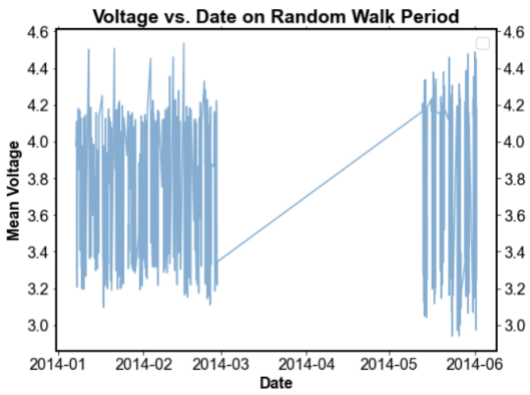


Figure 3.3.2 a

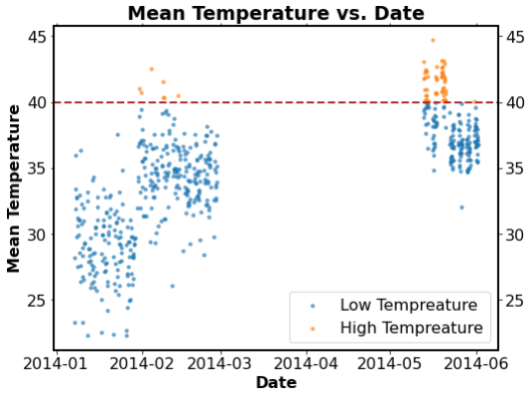


Figure 3.3.2 b

Modeling and Prediction:

Dep. Variable:	y	R-squared:	0.898
Model:	OLS	Adj. R-squared:	0.897
Method:	Least Squares	F-statistic:	2550.
Date:	Fri, 14 May 2021	Prob (F-statistic):	2.18e-288
Time:	11:03:59	Log-Likelihood:	367.84
No. Observations:	584	AIC:	-729.7
Df Residuals:	581	BIC:	-716.6
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	3.9154	0.043	91.144	0.000	3.831	4.000
mean_current	-0.1344	0.002	-71.306	0.000	-0.138	-0.131
mean_temperature	-0.0053	0.001	-4.287	0.000	-0.008	-0.003

Omnibus:	0.233	Durbin-Watson:	1.894
Prob(Omnibus):	0.890	Jarque-Bera (JB):	0.304
Skew:	0.043	Prob(JB):	0.859
Kurtosis:	2.928	Cond. No.	280.

Figure 3.3.2 c

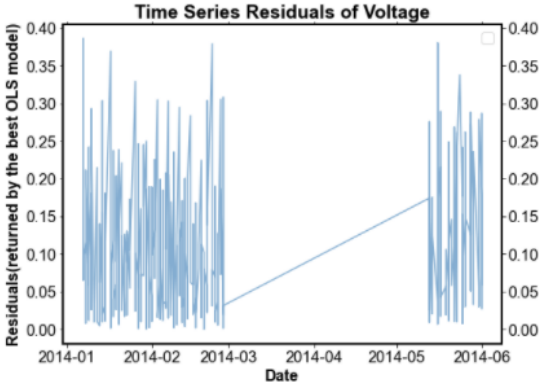


Figure 3.3.2 d

Using Ordinary Least Squares, we built a model that predicts 'mean_voltage', using predictors selected by the OLS model with the minimum AIC, so we got the predictors ('mean_current' and 'mean_temperature' , see Figure 3.3.2 c). We fit the OLS model again with the predictors and created the plot of the residuals returned by the model (see Figure 3.3.2 d) and checked its ACF and PACF (See Figure 3.3.2 e). Then we forecasted the residuals using Exponential Smoothing.

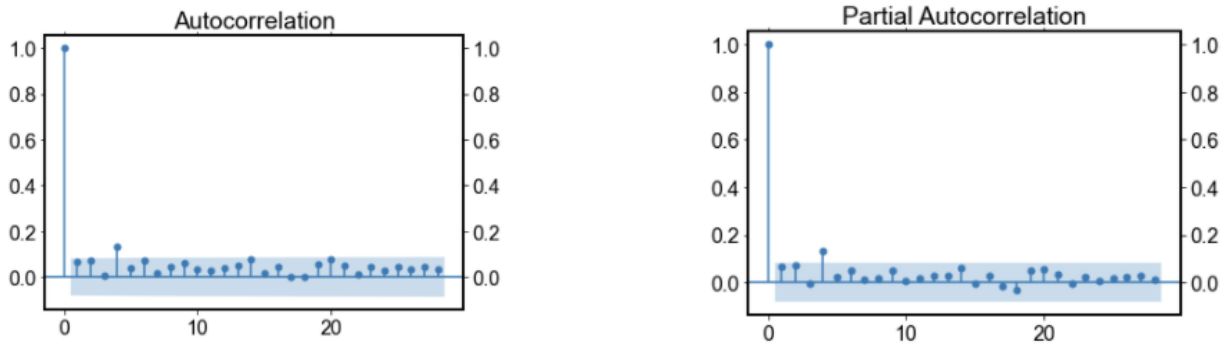


Figure 3.3.2 e

When fitting the model in Exponential Smoothing, we performed hyperparameter tuning on the following parameters: trend type = 'add', seasonal type = 'mul', smoothing level = 0.5, smoothing trend = 0, damping trend = 0, seasonal periods = 10 and smoothing seasonal = 0. With the parameters above, sum of the square error on the training set (374 cycles) is around 2.465 and mean square error on the validation set (210 cycles) is around 0.0063, which is pretty low (See Figure 3.3.2 f). Lastly, we also forecasted for voltage residuals for the next 100 cycles (index from 585 to 685, see Figure 3.3.2 g).

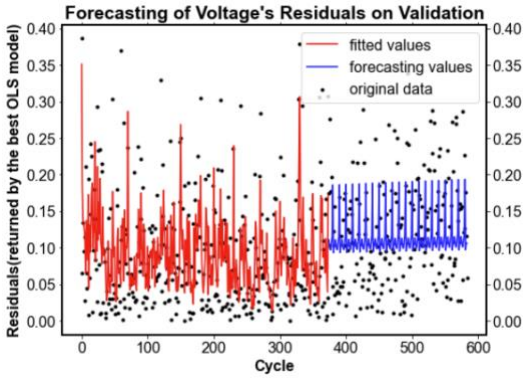


Figure 3.3.2 f

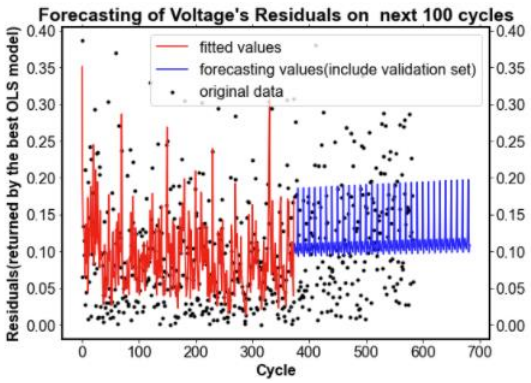


Figure 3.3.2 g

Summary:

To make forecasting on the battery voltage on the random walk period, we calculate that the Fitted voltage value = β0 + β1 * mean_current + β2 * mean_temperature (where β0 = 3.9154,

$\beta 1$ =-0.1344, β2 = -0.0053 calculated by the previous best OLS model). The prediction on battery voltage becomes: Predicted Voltage's value = Fitted Voltage's value +- Residual (for cycle n and predicted by the Exponential Smoothing model)

3.3.3 Forecasting battery voltage on the reference discharge period

Description:

On this period, we have 80 cycles(observations) on the reference discharge period. In addition to using the method similar to voltage prediction on the random walk period (Ordinary Least Squares (OLS) & Exponential Smoothing), we also introduced the following methods: Holt's Linear Trend / Holt's Exponential Trend / Holt's Damped trend models and Prophet Forecasting. During the whole modeling process, we split our data into a training set (before May 2014, 58 cycles) and a validation set (after May 2014, 22 cycles) for model evaluation. There is a decreased trend on voltage's data (Figure 3.3.3 a).

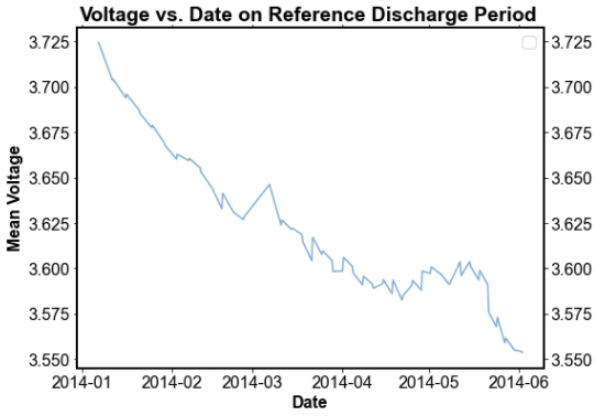


Figure 3.3.3 a

Dep. Variable:	y	R-squared:	0.893
Model:	OLS	Adj. R-squared:	0.888
Method:	Least Squares	F-statistic:	210.4
Date:	Fri, 14 May 2021	Prob (F-statistic):	1.03e-36
Time:	12:46:35	Log-Likelihood:	231.73
No. Observations:	80	AIC:	-455.5
Df Residuals:	76	BIC:	-445.9
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-52.6604	32.621	-1.614	0.111	-117.631	12.310
mean_current	56.3349	32.622	1.727	0.088	-8.638	121.307
mean_time(s)	-1.098e-08	7.49e-10	-14.668	0.000	-1.25e-08	-9.49e-09
mean_temperature	0.0008	0.001	1.038	0.303	-0.001	0.002

Omnibus:	7.227	Durbin-Watson:	0.218
Prob(Omnibus):	0.027	Jarque-Bera (JB):	7.551
Skew:	0.746	Prob(JB):	0.0229
Kurtosis:	2.795	Cond. No.	2.35e+11

Figure 3.3.3 b

Ordinary Least Squares (OLS) & Exponential Smoothing's result:

We utilized the features (mean_current, mean_time(s) and mean_temperature) to predict the target variable(mean_voltage) in OLS model and then we found that the best OLS model with the minimum AIC kept all features (See Figure 3.3.3 b) and the R^2 in the best OLS model is around 0.9 which shows the OLS model fits the data well. From the best OLS model's result, we created the plot of the residuals returned by the model (See Figure 3.3.3 c). Then we fit the model in Exponential Smoothing with additive trend and additive seasonality. Then, we forecasted voltage's residual on the validation set (See Figure 3.3.3 d).

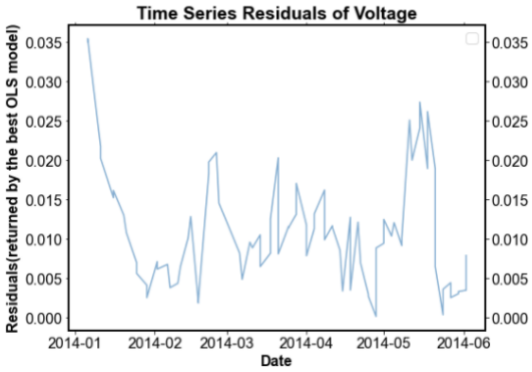


Figure 3.3.3 c

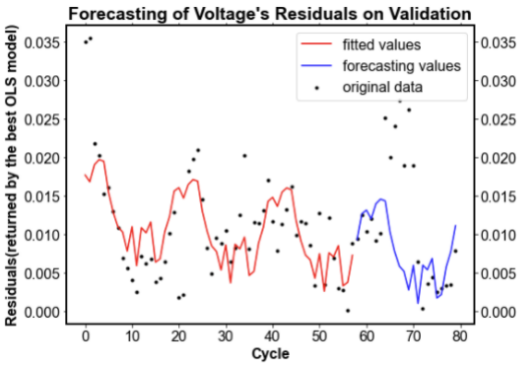


Figure 3.3.3 d

The sum of the square error on the training set (58 cycles) is around 0.0017 and the mean square error on the validation set (22 cycles) is around 0.00009, which shows that the errors generated by our models is almost negligible. Then, we forecasted voltage residuals for the next 80 cycles (cycle's index: [81, 160]. See Figure 3.3.3 e).

To make forecasting on the battery voltage on the reference discharage period, we calculate that the Fitted voltage value = β0 + β1 * mean_current + β2 * mean_time(s) + β3 * mean_temperature (where β0 = -52.6604, β1 = 56.3349, β2 = -1.098e-08 and β3 = 0.0008 calculated by the previous best OLS model). The prediction on battery voltage becomes:

Predicted Voltage's value = Fitted Voltage's value +- Residual (for cycle n and predicted by the Exponential Smoothing model).

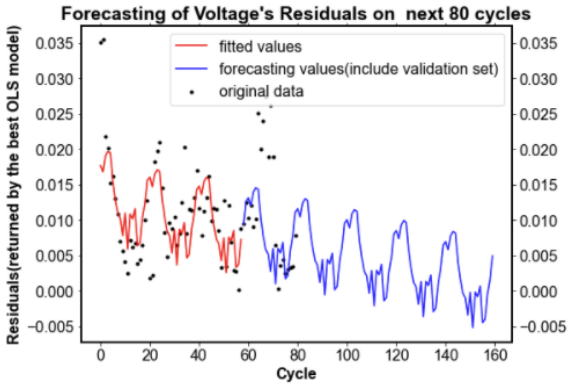


Figure 3.3.3 e

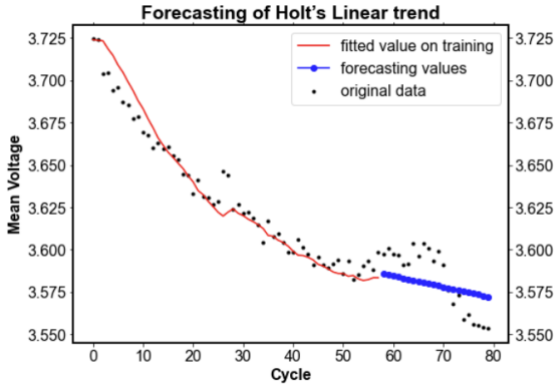


Figure 3.3.3 f (1)

Holt's Linear Trend / Exponential Trend / Damped Trend models' result:

Under different models, we found the best smoothing level and smoothing slope by considering the smallest mean square error on the validation set (After May 2014, 22 cycles). Then for each model, we forecasted the cycles on the validation set and plotted the forecast, fitted values, and original data. (See Figure 3.3.3 f (1) - (3)).

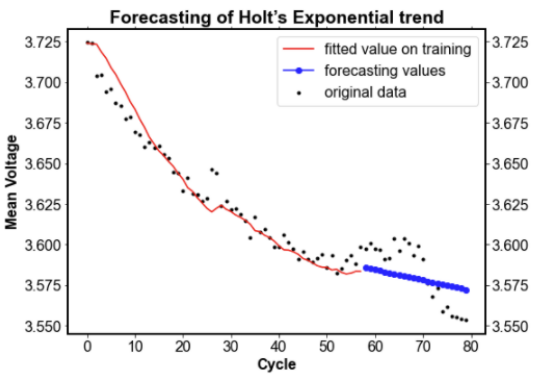


Figure 3.3.3 f (2)

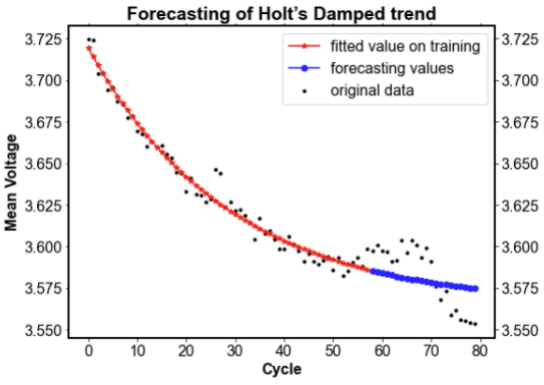


Figure 3.3.3 f (3)

Here, when evaluating models on the validation set, we found that the Holt's Damped Trend model has the mean square error = 0.01 and it performs the best when comparing AIC, Second-order AIC, and Bayesian Information Criterion (BIC) in different models (Figure 3.3.3 g).

Model	smoothing _level	smoothing _slope	AIC	BIC	BICC
Holt's Linear Trend	0.2	0.1	-536.9	-528.7	-535.3
Holt's Exponential Trend	0.2	0.1	-537.2	-529.0	-535.6
Holt's Damped trend	0.0	0.0	-589.5	-579.2	-587.2

Figure 3.3.3 g

Therefore, we forecasted the next 80 cycles using the Holt's Damped Trend model:

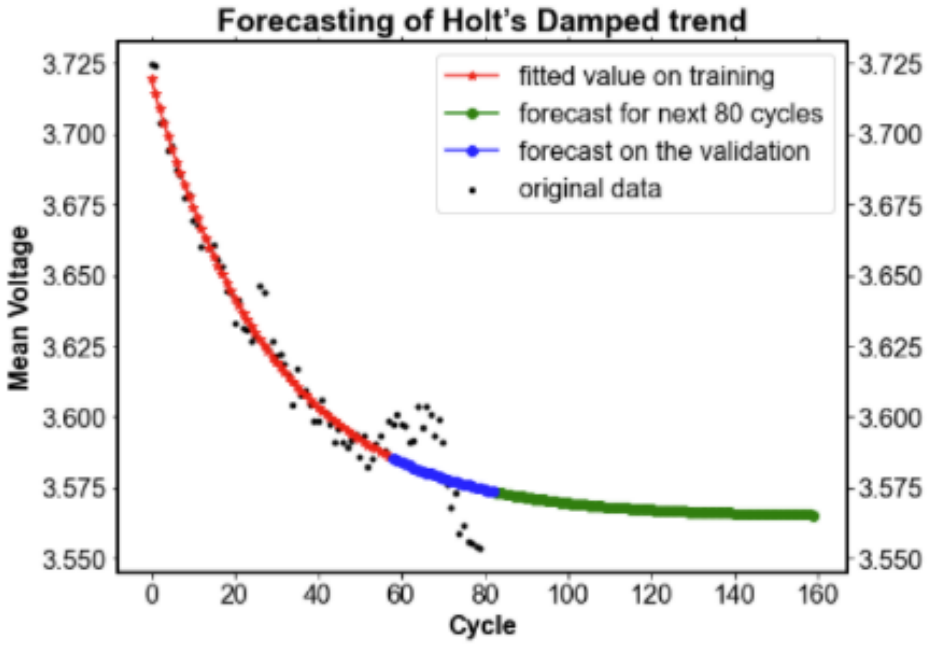


Figure 3.3.3 h

Prophet Forecasting's result:

Prophet is a popular forecasting method for time series data, which can control the various trend components and influence of multiple seasonality or seasonal patterns in the dataset. We fitted the Prophet model on the training set and predicted on the validation. Then, we did hyperparameter tuning on the seasonality mode = 'additive', changepoint prior scale = 0.05, holidays prior scale = 0.05 and n changepoints= 10 with the mean square error (MSE) on validation set = 0.0003 (See Figure 3.3.3 i).

Model Comparison:

We compared the Prophet model and the Holt's Damped Trend model here. Prophet's MSE = 0.0003 is much smaller than the MSE = 0.01 in the Holt's Damped Trend model. Also, from the plotting of Figure 3.3.3 j, the Prophet model performs better than the Damped trend.

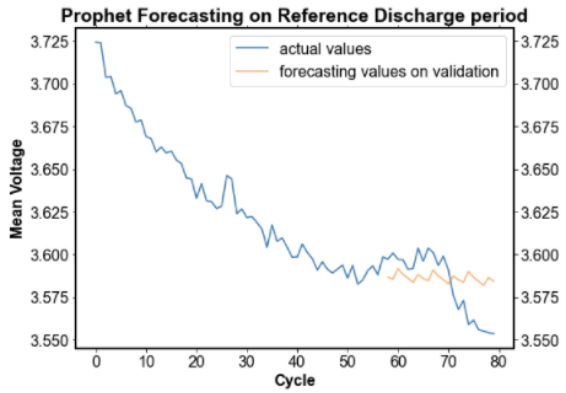


Figure 3.3.3 i

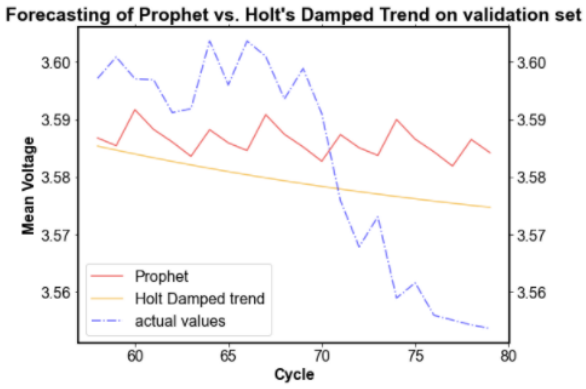


Figure 3.3.3 j

3.4 Battery Capacity Forecast

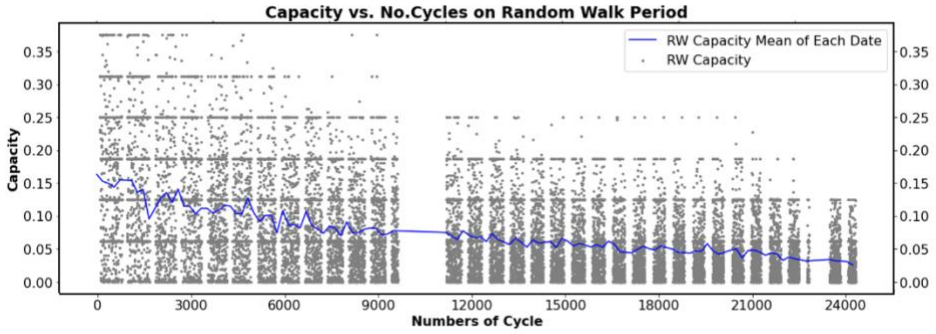
3.4.1 Overview of Capacity Forecast

Most measurements have indicated that Li-ion battery capacity decreases as a result of cycling, and the magnitude of this loss is dependent on both the number of cycles and the

depth of discharge that the battery is subjected to during these cycles (A. G. Ritchie, et. al., 2004). To approximate how long the battery will last before needing to be changed to a new battery, we examined the capacity during the cycling experiment and conduct a forecasting analysis based on the cycle for reference discharge period and date for random walk period. By constructing two time series models fitted for reference discharge and random walk periods, the models capture the pattern of capacity degradation of the future cycles by extracting meaningful information from preceding cycles.

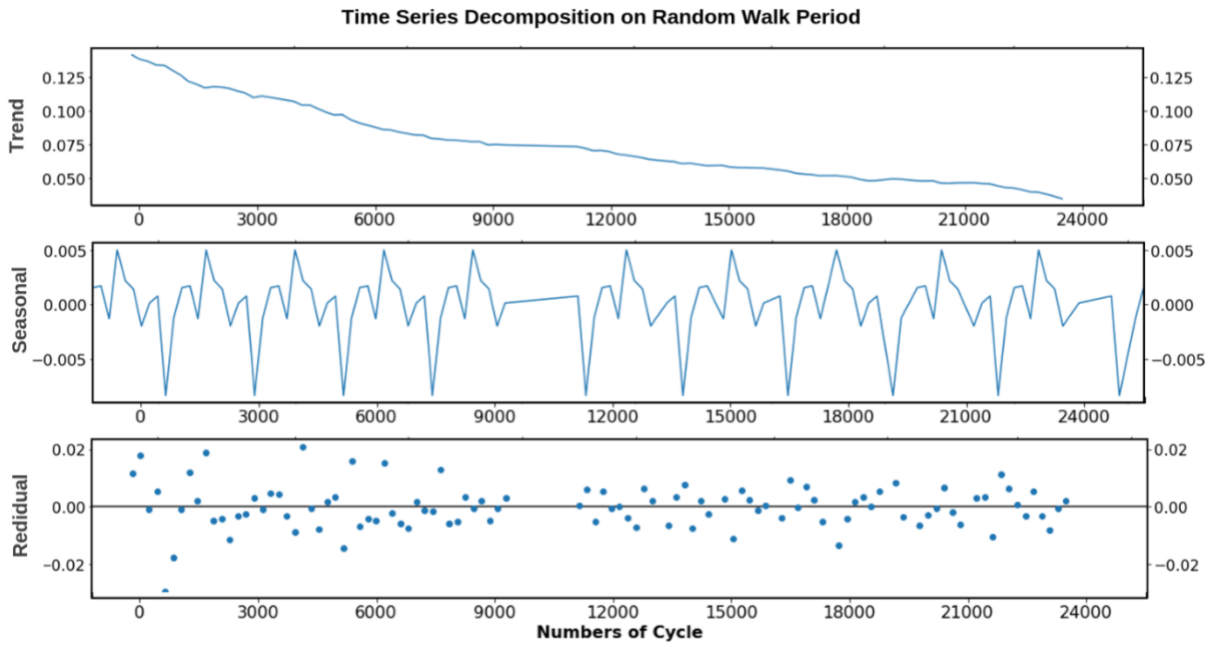
3.4.2 Data Resampling & Decomposition for Random Walk

Visualizing the overall trend of battery capacity for reference discharge period, it remains linearity over the cycles. While for the random walk period, a large number of cycles indicates the capacity decreases as a step function over the cycles. To fit a time series model effectively, we use the data resampling approach by evaluating the mean capacity of the same date as the reference capacity to avoid the uncertainty of the estimation.



After we have the reference capacity of random walk period, it shows some distinguishable patterns appear. We apply the time series decomposition to visualize the data into three distinct components which are trend, seasonality, and noise. As we can see, it shows a

downward trend over cycles, and variations specific to around 3000 cycles' frame which related to an obvious seasonality. The reason of this seasonal effect appears in the random walk period since after 3000 cycles, the experiment performs a pulsed current discharge and charge as a benchmark changes to battery transient dynamics. The residuals show high variability in the early stage and small noise at the end of the experiment indicates the more cycles for battery, the capacity degradation decreases steadily.



3.4.3 Model analysis and Parameters Selection

Long Short-Term Memory (LSTM) Time Series Model on Reference Discharge

LSTM as the variants of recurrent network, the data signals travel in backward directions as well as these networks have the feedback connections in its architecture. The flexible of LSTM architecture results from it can tune the hyperparameters like regularization term or learning rate easily. Since The capacity of the reference discharge period contains only 80

cycles lead us to choose a single hidden dense with 64 neurons to avoid the complexity of the model structure led to overfitting problem.

Autoregressive Integrated Moving Average (ARIMA) Time Series Model on Random Walk

ARIMA model drive its own lags and the lagged forecast errors to predict the feature values by characterized three parameters (p, d, q).

- p denotes the auto-regressive part which involving the effect of past values into the model.
- d is the integrated part which incorporating the amount of differencing in the model.
- q is the moving average part which setting the error of our model as a linear combination of the error values observed at previous time points in the past.

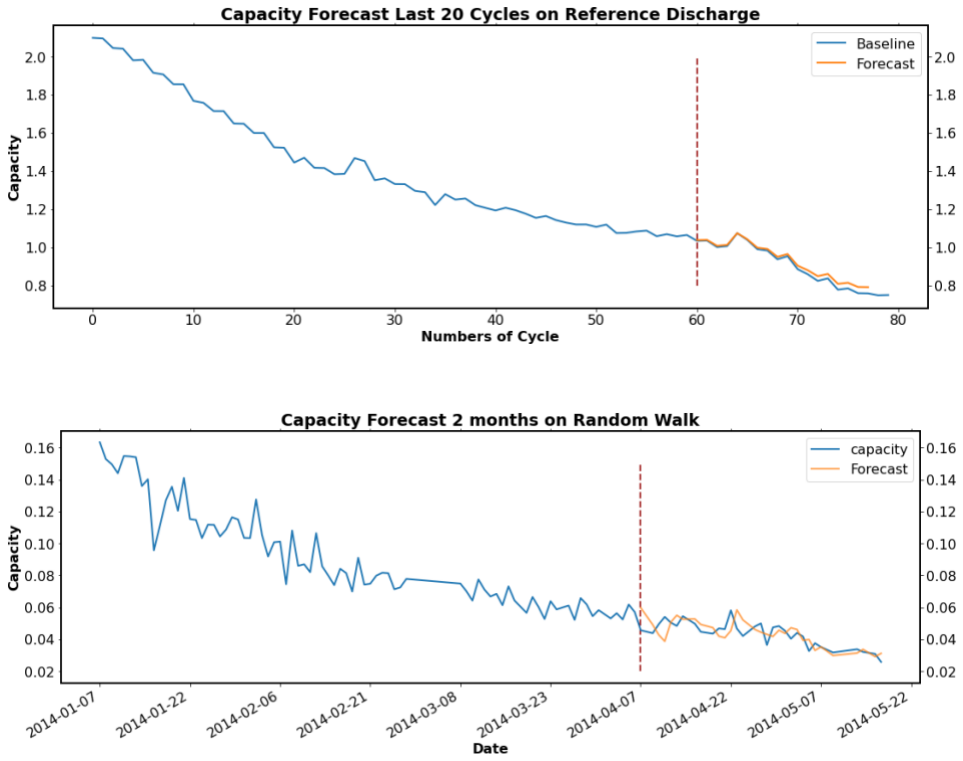
To capture seasonal effects, we make use of the seasonal ARIMA. And the s denotes the periodicity of the time series model.

By evaluating and making comparison of the models fitted with different combinations of these four parameters, we use AIC (Akaike Information Criterion) value as the reference to measure how well the models fit the data. Here's the result of parameters selection and the final parameter of the best AIC is (p=1, d=1, q=1).

```
Parameter Selection with AIC value:
ARIMA(0, 0, 0) - AIC:-231.89574880749822
ARIMA(0, 0, 1) - AIC:-347.2331374408138
ARIMA(0, 1, 0) - AIC:-669.9705725386666
ARIMA(0, 1, 1) - AIC:-694.5451393232695
ARIMA(1, 0, 0) - AIC:-678.3155592903652
ARIMA(1, 0, 1) - AIC:-727.1060886467964
ARIMA(1, 1, 0) - AIC:-690.7997081250879
ARIMA(1, 1, 1) - AIC:-692.7601189764337
```

3.4.4 Results

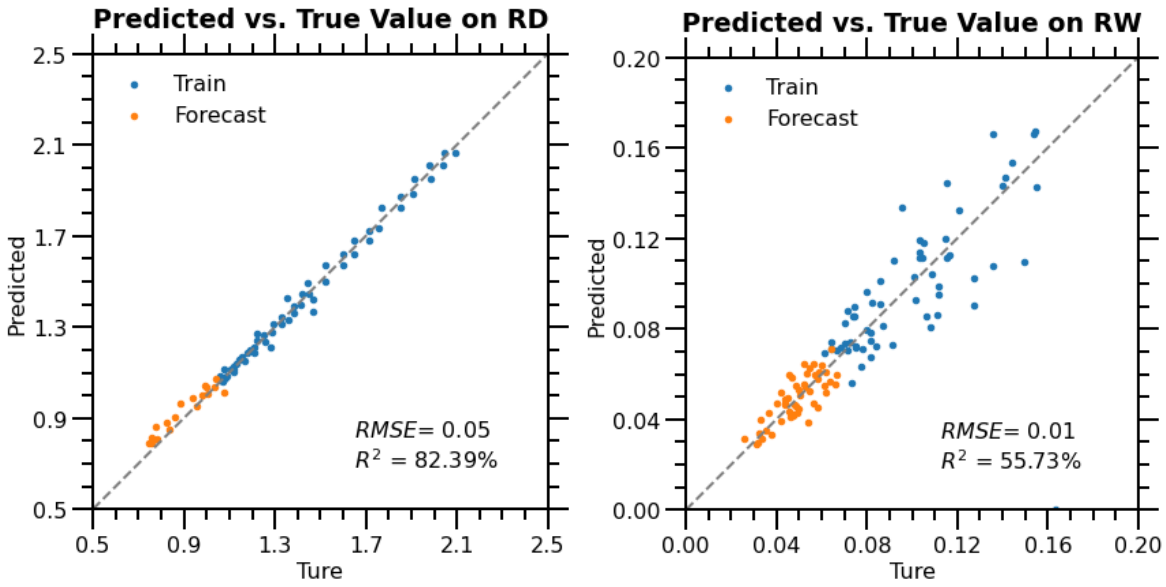
The results of forecast capacity on reference discharge and random walk period as follows:



For capacity forecast on reference discharge period, we use first 60 cycles as train set to train the LSTM time series model. The forecast part fit the baseline very well for last 20 cycles. Accuracy reaches to 82.39% and RMSE is 0.05.

The experiment baseline around 6 months range. Since we extract the mean of each date capacity for random walk, we use date as time stamp for ARIMA time series model and set first 4 months as train set to forecast the following 2 months. As we can see in the fitted plot, due to the unstable patterns of the baseline, the forecast part does not fit very well and there exists the fluctuation. Overall, the trend of forecast indicates the decrease pattern and becomes more and more flat because of the speed of capacity degradation becomes slow.

Accuracy reaches to 55.73% and RMSE is 0.01. In the plot of prediction and true value comparison, we also can observe for reference discharge period, the forecasting result indicated a better result than random walk period since the inherent linearity of reference discharge period and fluctuation patterns of random walk.



4. Conclusion

As discussed extensively in the above sections, the importance and potential of Lithium-ion batteries in technological advancements is enormous. Throughout this report, we have carefully outlined a number of statistical models that aim to predict battery failure. We have accomplished this using two main techniques/strategies — State of Health and forecasting. As discussed, random forest regression with feature selection, autoregressive integrative moving averages, exponential smoothing, ordinary least squares, damped trend models, and prophet forecasting were performed across all three periods.

Although it is difficult to determine what models before the best due to different evaluation criterion, we presented these results to the reader to use their own informed judgement. In the future, we recommend building more models that take into account more advanced calculations to represent battery health that were not mentioned here. Further, in this report, we discussed models and algorithms that represent the three periods separately. To make a model that encompasses all data, we must think of more complex methods. As the usage and importance of Lithium-ion batteries increases tremendously over the next few years, we hope that our analyses provide valuable insight to researchers at NASA to improve the longevity of their satellite missions.

References

Bole, Brian. (2015, March). Description of Room Temperature Random Walk Charging and Discharging Data Sets. Retrieved from https://ti.arc.nasa.gov/tech/dash/groups/pcoe /prognostic-data-repository/

Bole, B., Kulkarni, C., Daigle, M. (2014). Adaptation of an Electrochemistry-based Li-Ion Battery Model to Account for Deterioration Observed Under Randomized Use. *The proceedings of the Annual Conference of the Prognostics and Health Management Society*.

Broussely, M., Biensan, P., Bonhommeb, F., *et. al.* (2005). Main aging mechanisms in li ion batteries. *Journal of Power Sources*, *146*(1-2), 90-96.

Farrier, L., Bucknall, R. (2020). Investigating the Performance Capability of a Lithium-ion Battery System When Powering Future Pulsed Loads. *Energies*. 13(6):1357. https://doi.org/10.3390/en13061357

How Lithium-ion Batteries Work. (2021, May 12). Retrieved from https://www.energy.gov/science/doe-explainsbatteries.

Kaneko, F., Kawakami, T., Shinbo, K. *et al.* (2017). Judging Degradation of Lithium-Ion Batteries Using a Real-Time Capacitance Measurement. *Electrical Engineering Japan*, 199: 10-16. https://doi.org/10.1002/eej.22962

Manh-Kien Tran, Michael Fowler. (2020). A Review of Lithium-Ion Battery Fault Diagnostic Algorithms: Current Progress and Future Challenges.

Murnane, M., Ghazel, A. (2017). A Closer Look at State of Charge (SOC) and State of Health (SOH) Estimation Techniques for Batteries. *Analog Devices.*

Naha, A., Khandelwal, A., Agarwal, S. *et al.* (2020). Internal short circuit detection in Li-ion batteries using supervised machine learning. *Sci Rep*, 10, Article 1301. https://doi.org/10.1038/s41598-020-58021-7

Qi, B., Wang, J. (2012). Open-circuit voltage in organic solar cells. *Royal Society of Chemistry,* 22: 24315-24325. 10.1039/C2JM33719C

Ritchie A. G., Recent developments and likely advances in lithium rechargeable batteries, *J. Power Sources,* 136 (2004) 285-289.

Tessier, O.A., Dubois, M.R., and Trovão, J. P. (2016, June 19-22). Real-Time Estimator Li-ion Cells Internal Resistance for Electric Vehicle Application*. EVS29 Symposium.*

Woongchul Choi. (2020). A Study on State of Charge and State of Health Estimation in Consideration of Lithium-Ion Battery Aging.