EXTRAPOLATING THE EXPERIMENTAL DATA TO PREDICT THE LONGEVITY OF LI-BATTERY

JiaGuo[1], HeshengGong[2] and Bo Gao[3]

[1]Taizi Education and Technology, Co. Ltd., Beijing, China
[2]Eindhoven University of Technology, Netherland
[3]Beijing Jiaotong University, Beijing China

ABSTRACT

In search of a particular lithium battery with reliable safety and high energy, quantities of research have been focused on the chemical substances for the Anode and Cathode, respectively. In Cui's laboratory, an efficiency of 98.54% for more than 600 cycles as well as long lifespan beyond 900h in a LiCu-Ag@Li cell can be realized. A high cyclability of 98% capacity can be achieved after 1000 cycles along with a long lifespan of 1500h in a SiOxCy@Li cell, which both prevents electrons from piercing through a separator, and leverages the efficacy of the lithium-ions via a binder. Thanks to Cui et al. and Severson et al., we either have got approved for or searched for the published data regarding the lithium-ion battery's lifespan and chart a series of diagrams that reveal the curve-shaped trend line and unexpected surges in the first, middle and last few cycles of a cell's life. The more a shocking cusp (outliers) surfaces, the more a decline steepens. We compare the data from the laboratory to on-board batteries and build a polynomial regression in order to predict the life end of those cells. While the non-linear regression is unable to best fit every moment of a cell's decrepitude, our team create a regression model to increase the accuracy of predication to an average of 97.693% in the primary test according to the first 30-225 cycles, then seek the optimization for longevity forecast by programming solver and hyperparameter, and finally find a(non-fixed) relationship between the speed and acceleration during the period of a cell's degradation. SVM model has also been created along with its corresponding 3D pattern with Temperature considered and so has the model Multiple Regression but the cost/benefit analysis will be continued in future study of relevant subject for prediction on newly-bonded cells or all-purpose commercial batteries.

KEYWORDS

prediction accuracy, non-linear regression, speed, acceleration, optimization

David C. Wyld et al. (Eds): SEAU, NLMLT, ITCS, CST, DMS -2023
pp. 93-104, 2023. CS & IT - CSCP 2023 DOI: 10.5121/csit.2023.132007

1. INTRODUCTION

In order to combine an on-board battery with higher energy, higher safety and lower cost, this paper will present a non-linear regression and an optimized model for best-fitting and forecasting the end of the life of Li- batteries. In a laboratory, Cui et al.[1]conjugate SiOxCy to the 3D collector as an implosion-proof anode, which indicates an ultra long cyclability in both half and full cells. According to an array of the on-board data of more than a hundred of cells with a range of 2150 cycles and a maximum of 2300 cycles, a model of non-linear regression is set up to predict the end of life for a cell (80% of State of Charge, SoC). Given the three aspects, the outliers in the very beginning of a cell's lifetime, the trend in the first 100 cycles (in consistency with the findings of Severson et al.) and the comparison among 3 cycle-groups (30-225, 30-400, 30-450), we classify, compare and choose in the train set the first 30-225 cycles during which period the state of capacity degradation is steady and sound and which can reflect a declining trend normally with a negative slope. Machine learning allows the evaluation of life prediction with references to the formulas in physics and mathematics. To be specific, the regression model designed cannot only be interpreted in the way of big data and statistics whereby we are 95% confident that life expectancy of a cell can be predicted with a convergence of accuracy between (97.693+0.007) % and (97.693-0.007) % in the test set but it can also be analyzed in calculus and physics as to a velocity-acceleration relationship.

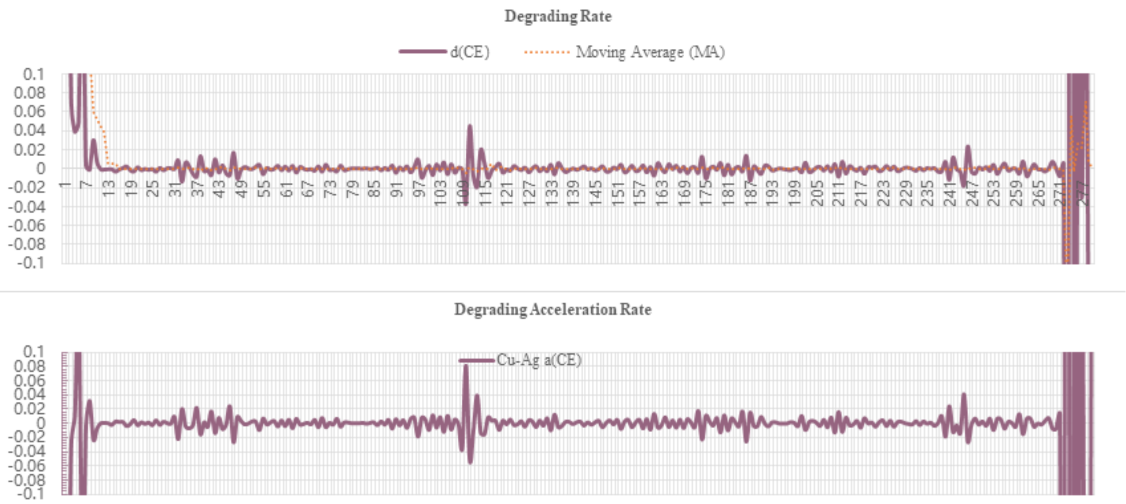


Fig. 1. first derivative and second derivative of discharge capacity in Graph

As the figure shows, a laboratory-used cell runs from Cycle 1 to Cycle 270. The first derivative or the degrading speed of CE beats no more than a rate of 0.02 except for the out branches in the head, middle (indicative of a critical point), and tail of the two wavelengths, for Cu-Ag Li-battery. So does the second derivative or the acceleration rate.

2. LIMITATION AND RESTRICTIONS

Admittedly, the better the chemical property is, the longer life cycle will be and the higher coulomb effect becomes. The fact is, however, that it is unattainable to perfect the performance of each parameter in a battery. There is no doubt that researchers in this field have long eluded to reach a consensus upon measurements as to what kind of lithium-ion battery has verdict as poor, good, or excellent since many factors such as safety, energy effect, costs etc. are being offset or contradicted to each other. Indicators also have defects: for example, decay rate is conventionally calculated by dividing discharge-current-difference $\Delta Q_{n-2}^{discharge}$ into the number of cycles. Yet the equal weight 1/n allocated to each variable is unreasonable since the decreasing rate cannot be averaged. Discharge capacity not only ejects sometimes a curve line, sometimes a straight line from the initial stage to the end, but it also produces different decay rates, decrease-rejected or decrease-accelerated. To trace down its aging by simply creation of a parabola is nearly a mission impossible, because the trajectory freely goes beyond the limit of 2500 cycles, as is shown in the graphs right below.

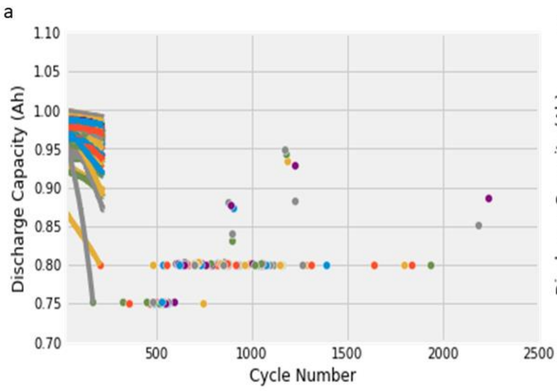
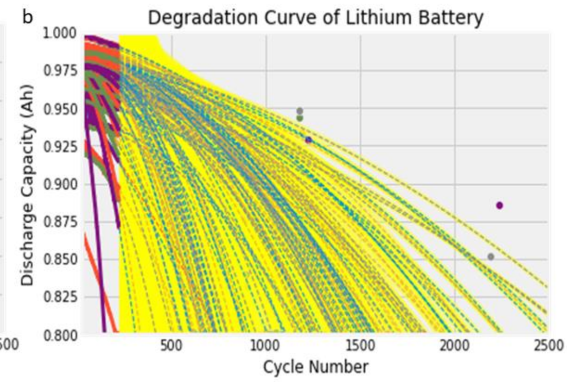


Fig.2.a. normalized capacity good fitness line of capacity in Cycle 30-225



B. Simulated Parabolas of all the Cells

This paper develops the prediction with simply a polynomial regression model and interprets the relationship between the energy efficiency and life span of a battery by employing theoretical and experimental results, regardless of temperature fluctuations. One part of data is derived from Laboratory of Photoelectric Control on Surface and Interface where the performance of Cu and Cu-Ag batteries is registered. Another part is sourced from the paper on the prediction of Li-battery's end cycle, where in Sevens on et al. published the data of 139 on-board cells. In the real world, data may vary with seasonality and timing, electric resistance (ER), etc. and that is why discharge capacity cannot fully explain the variation of degradation in forecast and visualization.

The test results in the laboratory under perform due to outliers and insufficient short-lived batteries (for new material trials) despite the featured data that often reads room temperature and minuscule internal resistance.

Table 1.Measurementof Sample Type

Sample Type	Temperature(K)	Internal Resistance(ohm/cm)
Cu	300	8.31E-07
Cu-Ag	300	6.99E-07

3. SOLUTIONS

In our experiment, regular or irregular, the outliers frequent a significant deviation from the trendline. On one hand, the eerie situation is invariably followed by a phenomenon of acceleration of the capacity degradation, i.e., the aberrated values are not meaningless in predicting the trend of CE; on the contrary, they herald a faster decline instead, as is illustrated in Fig. 3.b. One the other hand, other outliers can be regarded as noisy to be replaced by the corresponding data in the previous cycle, such as the outliers in red circle in Fig. 3. a charted by Severson et al., while redundant data in the first and last cycles are deleted as the experimental errors. Robust data propel both statistical accuracy and practical application. Thus, the cleaned data in a cell's aging process display a non-linear degradation behavior, thereby helping optimize the prediction for further diagnosis and prognosis.

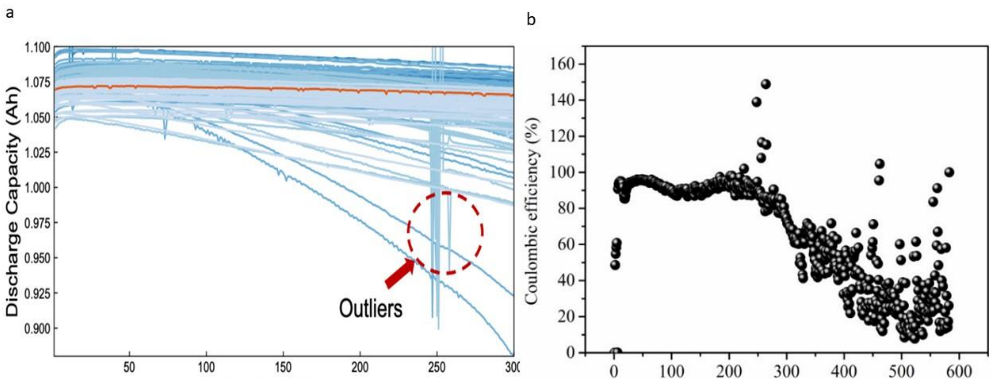


Fig.3.a. outliers both in on-board cells (Severson) b. CE in laboratory (Cui)cells

4. RELATED WORKS

Previous research is dedicated to the prediction for the cycle life of lithium-ion battery (e.g., lithium iron phosphate) by virtue of$\Delta Q(V)$, the difference of discharge capacity voltage, using the first 300, 100 or even merely 5 cycles. Published in 2019, the paper Data-Driven Prediction of Battery Cycle Life before Capacity Degradation illustrates how the very best model designed by Severson et al.[2] can project the cycle life based on only first 100 cycles (from 2 to 100) of 124 lithium iron phosphate (LFP) cells whose life cycles range from 150 to 2300. Surprisingly, her team found that at least 75% of the cells have higher capacity at cycle 100 relative to cycle 2 (with a median increase of 0.2%); discharge capacity in three quarters of the cells ticks up in the first 100 cycles. Another eminent contribution is that Severson et al. reduce a test error to 9.1%

with the first 100 cycles and achieve a classification accuracy of 97.5% within the first 5 cycles. Furthermore, another team configure and transform ML models to decrease the error through skewness or kurtosis of ΔQ100−10(V). From other research perspective, coulomb efficiency (CE) is one of the key indicators to study the degradation of batteries on a cycle-to-cycle basis. In this way, Yang Fangfang et al.[3] predict the longevity of the Li-battery cell with an accuracy of above 95% based on the initial 100-200 cycles. The fitness level of the actual and predicted values against Normalized Capacity is later enhanced to such a high level as R-square at 0.9971-0.9979 with a two-term logarithmic model, as a batch of BAK 18650 LFP batteries are sampled. In 2021, Liu WM [4]uses the regression model (ΔQ(V) versus the number of cycle) based on the first 300 cycles (from 2 to 300) of LFP batteries, abating the test error from 9.1% to 7.3%.

5. Physical and Mathematical Formulations

Coulombic Efficiency Sequence

$CE_1=\frac{C_d}{C_c}$,

where Coulombic Efficient is denoted as CE, Capacity of Discharge as C_d, and Capacity of Charge as C_c.

$CE_2 = (CE_1)'+CE_1+\varepsilon$
$CE_3 = (CE_2)' + CE_2 + \varepsilon$
$CE_4 = (CE_3)'+CE_3+\varepsilon$
…
$CE_{n-1} = (CE_{n-2})' + CE_{n-2} + \varepsilon$

$$CE_n = (CE_{n-1})' + CE_{n-1} + \varepsilon$$

$CE_n = (CE_{n-1})' + (CE_{n-2})' + (CE_{n-3})' + \cdots + (CE_1)' + CE_1 + \varepsilon$ (1)
$CE_n - CE_1 = (CE_{n-1})' + (CE_{n-2})' + (CE_{n-3})' + \cdots + (CE_1)' + \varepsilon$

Assume that n is the number of the life cycle of a cell, then

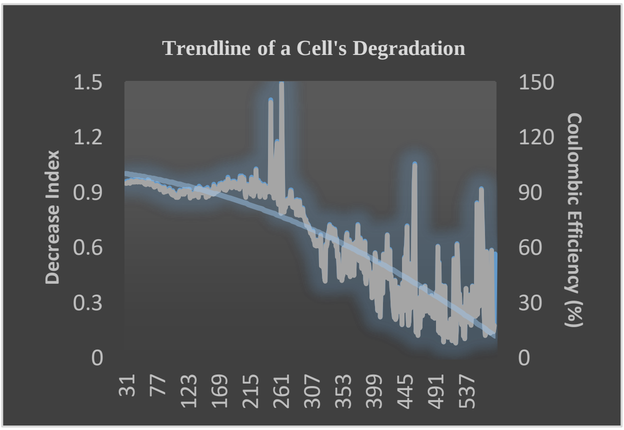


Fig. 4. the quasi-parabola-shaped trend line

6. NON-LINEAR REGRESSION MODEL

Since the graph of CE takes on a parabola-like curve, a polynomial (degree=2) regression model is set up.

$C_d = ax^2 + bx + c,$

where x is the cycle of the battery, a is the acceleration of the degradation and b is the speed of degradation and c is the initial value of the capacity.

When the critical point does not occur, i.e., a is arbitrarily getting close to 0, the battery degrades in a quasi-straight line with the slope, b starting from the initial value, c but the capacity is not stable in the first few cycles and therefore our calculation starts from the 30st cycle. We assume that b is the initial falling speed as it is the indicator of degradation rate in Cycle30st. Those cycles are more likely to compose a positive slope or decreasing line. That is the reason why we choose the cycle interval [30, 225).

When the critical point does occur, i.e.,$a \neq 0$, the capacity decreases at an increasing rate and bends more sharply, a process that helps shape a curve similar to a parabola.

$C_d = ax^2 + bx + c$ （a<0,indicating that the parabola faces down; b<0, when the initial speed goes down, b>0, when the initial speed goes upward）.

To reduce the error, we set up D(x).

Let D(x)=a*x^2+b*x+c+ ε,

where a, b, c are constants; a is related to acceleration, b initial falling speed, c initial value of discharge capacity, and ε is the error.

$D'(x)=2 * a * x + b$,

where $D'(x)$ is the first derivative of discharge capacity, b is the initial value of the falling speed of the discharge capacity.

$D''(x)=2a$,

where D’’(x) is the second derivative of discharge capacity, 2a is A, the falling acceleration of the discharge capacity.

$D_k - D_{k-1} = 2 * a * x_k + b$ (1)
$D_{k-1} - D_{k-2} = 2 * a * x_{k-1} + b$(2)
…
$D_3 - D_2 = 2 * a * x_3 + b$ (k-2)

In order to derive the value of $D_k - D_2$, add up all the expressions from (1) to (k-2), respectively and yield

$D_k - D_2$ =2*a*($x_k + x_{k-1} + \cdots + x_3$) + b*(k-2)

According to Calculus (first derivative)
$D_k - D_{k-1} = D_k' + \ \varepsilon$ (1)
$D_{k-1} - D_{k-2} = D_{k-1}' + \ \varepsilon$(2)
…
$D_3 - D_2 = D_3' + \ \varepsilon$(k-2)

Therefore,

$D_k - D_2 = D_k' + D_{k-1}'$+…+$D_3'$+ ε

=$\int_3^k D_x' dx$

=$\frac{1}{2} * \left(D_3' + D_k'\right) * x_{k-2}$ (Area of Trapezoid)

$D_k' = D_3'$ +A*x_{k-2}=b+2a*x_{k-2}

where A is a Constant as acceleration and b is the initial speed of the degradation.
Therefore, $D_k - D_2$:

2*a*($x_k + x_{k-1} + \cdots + x_3$) + b*(k-2) = $\frac{1}{2} * \left(D_3' + D_k'\right) * x_{k-2}$

=$\frac{1}{2}$*(b+b+2a*x_{k-2})* x_{k-2}

= (b + a*x_{k-2})* x_{k-2}

x_{k-2} is a continuous instead of discrete number in life-cycle prediction while k is,however, an integer (the last cycle number). We are confident that a and b have multiple relationship, though, with minor residuals, unsure about the concrete and certain times between the two variables.

2*a*$\sum_{i=2}^{k} x_i$+b*(k-2)=(b+a*x_{k-2})*x_{k-2}

Let's further do an example and assume that k=813,$x_k = 818.29$ as a predicted end of life cycle, then

2*a*$\frac{(2+818.29)}{2}$*817+b*811=(b+a*816.29)*816.29

By test in python, we prove that a and b have multiple relationship (rounded a minor residue up to a whole number).

7. Machine-Learning Approaches

We clean the battery data by standardization and normalization and then classify them by train and test sets in which some of the cells are further grouped attributive to the negative slope of the initial speed and others of the cells to the positive slope, respectively and the classified groups are put into two independent loops. By iteration and batch processing, coefficients of the polynomial regression, a, b and c are derived, which attests the value of c to be near the initial capacity and b to be near the initial speed. More surprisingly, it is noticeable that a and b have a numerical relationship. Therefore, we deprive part of a and b values as hyper-parameters and retain the remaining for optimization. We have found a non-fixable but multiple relationship between a and b. For the value of b is the initial speed of the degradation, the value of a can be assessed and reasoned based on the calculated value of b. However, the value of b can be positive or negative, which may impact the hypothesis, so we build another consistent algorithmic model between a and b in the train set before running data in the test set and finally prove our assumption.

According to the data from the randomly-selected cells, we have tested the accuracy from the classified Cycle 30-225, 30-400, 30-450. It turns out that the farther the cycles are chosen, the closer the values between the evaluated and actual can be as per the graph. However, when our model is applied, the regression line, though unable to fit the actual degradation, can attain an accuracy ranging from 93.253% to 99.725% in terms of the prediction based on the data in Cycle 30-225 better than its counterparts. As is shown below, the data are neither coincident nor unexpected in light of the above-shown working on physics and mathematics. It may be observed that the battery life ends up with 0.9 instead of 0.85 or 0.80, but barely will the missing data affect the outcomes, because the model is built on those data of a typical cell of 919,051 cycle index and 139 such cells. No matter how far the experimental data records, the predicted end point is expected to fall in proximity to a cell’s actual longevity as follows,

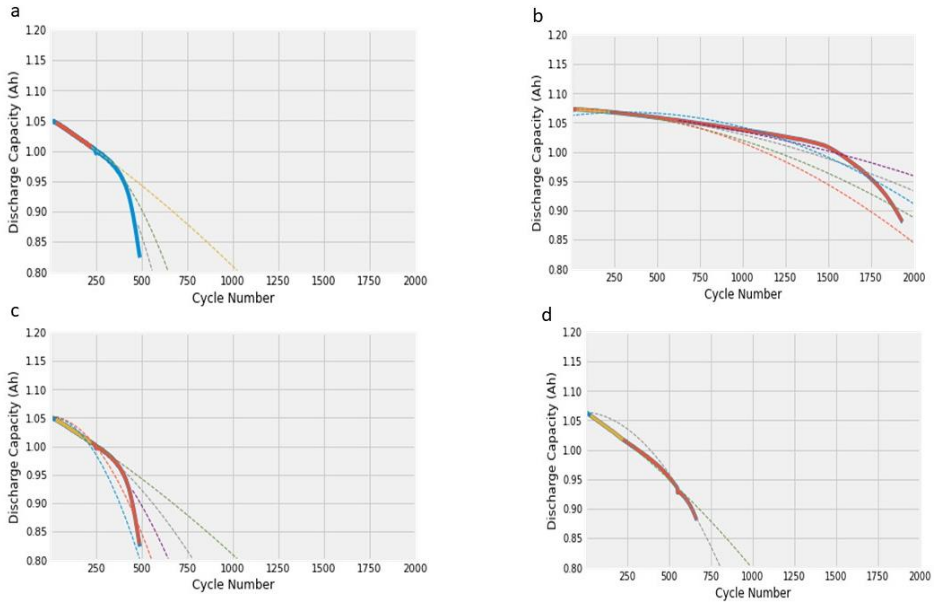


Fig. 5. Algorithmic Optimization of the four cells with b <0 represented by illustrations a, b, c and d, respectively

Table 2. Statistical Table

Battery index	Actual	Predicted	Accuracy
1	[813]	[818.29]	99.349%
2	[490]	[508.81]	96.161%
3	[666]	[668.83]	99.575%
4	[541]	[577.50]	93.253%
5	[1009]	[1011.78]	99.725%
6	[828]	[833.46]	99.341%

When the initial speed is greater than zero or b > 0, the model can also be agilely used for the train set, as the architecture is designed and developed for a classified result on prediction.

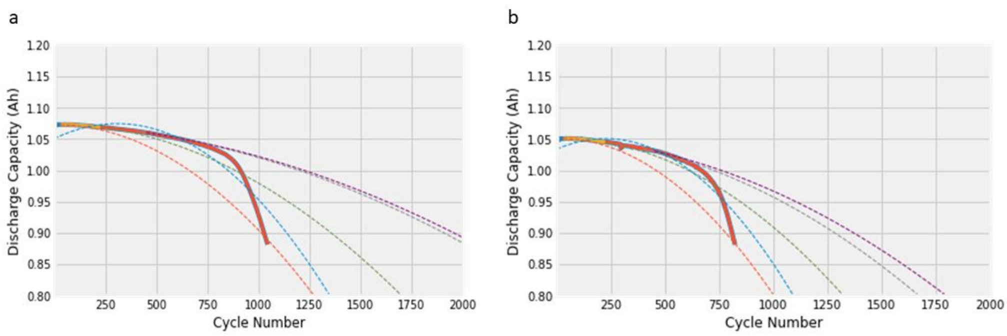


Fig. 6.AlgorithmicOptimization of the two cells with b > 0 represented by illustrations a and b, respectively

Table 3. Statistical Table

Battery index	Actual longevity	Predicted	Accuracy
34	[824]	[829.415]	99.343%
7	[1045]	[1068.448]	97.756%

After all, statistics include probability and it cannot escape from errors. In the first test, a majority of cells satisfy an accuracy above 90% out of 139 cells but some parameters still need to be algorithmically optimized to better the predicative results.

8. EXPERIMENTAL RESULTS

In the first stage, we take advantage of the model of polynomial regression to simulate and predict the life cycle of cells with an accuracy from 93.253% to 99.725% for the majority of cells, although the success rate is far from perfection. The reason is that the degradation of a cell is not a free fall without the control of electric resistance (ER) or fluctuation of temperature. Multivariate model should have been considered with ER included, for example, but theoretically, the relationship among currents, ER and Voltage is well known in physics $I=V/R$ and thus we can represent V and R with capacity (I). The variable discharge capacity can empirically reflect the lurking variables, voltage and ER in the non-linear regression model.

In the second stage, we run the model in the train set and test set, respectively and the accuracy can reach above 95% in average with RMS 14.531 in the train set and 74.081% with RMS 466.388 in the (before-classification) test set. The percentage accuracy for the cells can be averaged as high as 97.693% in the primary test (for 88 cells) and 97.313% in the secondary test (for 118 cells). Further experiment needs to be continued in order to explore the relationship between a and b or between the acceleration and speed of li-ion battery degradation.

```
In [3]:
from sklearn.metrics import mean_squared_error
from math import sqrt
#root_mean_sq_err
rms = sqrt(mean_squared_error(actual, predict))
print(rms)

14.531838400696193

In [79]:
from sklearn.metrics import mean_squared_error
from math import sqrt
#root_mean_sq_err
rms = sqrt(mean_squared_error(A, Predicted))
print(rms)

466.3880113944437
```

Fig. 7. python code in calculating RMS

In the third stage of optimization, programing solver and embedded loops are used to minimize the variance between the actual value and predicted value. Probability distribution and distribution density have been used for a wide range of data to test the regression model of optimized-to-be parameters. Different approaches including Multivariate Regression and Support Vector Machine (SVM) have been applied and will be compared when other variables such as temperature are involved. More importantly, we need to know to which extent ER and temperature impose on the performance of the battery, although pre-heating system has been widely applied into the on-board battery.

Fig. 8. Errors and Accuracy

9. Conclusion

We observe the behavior of the coulombic efficiency and thus create a model of polynomial regression. Inspired by the previous works and citing the open-access data, we train and test the model for the prediction of the cells' life span which finally reaches a relatively high percentage accuracy for a majority of cells.

Non-linear regression heavily depends on the data, some of which are disproportionately model-friendly while others not, thereby causing overfitting or underfitting [5]. The change of parameters is thus deployed for parameter optimization. Our model, though not following the trajectory of the fall, is still able to forecast the life expectancy with RMS 27.576 for 88 cells. The secondary test has been done with a continued predictability of the regression model and the errors of the prediction can be decreased to within 5% for a wider range, 118 tested cells.

Complex event processing (CEP), including discharge capacity with regard to temperature, material, ER, etc. is required to align the generally decreasing trend of a cell's degradation to the data visualization and associated usage [6]. The model of non-linear regression passes the primary and secondary tests, both with high performance. Despite the models of SVM and multiple regression that may not reach our expectation, the fitness level of the life cycle of a particular cell can arrive at r-square 0.9030 and adjusted r-squared 0.9024 with multivariate statistics. To wrap up, this paper interprets, integrates and illustrates powerful algorithms, statistical modes, physical formula and coding in python, and deploys ML models to obtain the above results, which will be upgraded for future impacts and implications.

ACKNOWLEDGEMENTS

The authors would like to thank Susan, an MBA female who had a dream of building an ecological system or a green community with self-contained energy chain in Africa and who gave her inspiration of waste recycling and renewable energy into this team.

REFERENCES

[1] Cui S et al. Large-Scale Modification of Commercial Copper Foil with Lithiophilic Metal Layer for Li Metal Battery. Nano Micro Small. 2019. [2023-09-09]. DOI: 10.1002/smll.201905620.

[2] K. A. Severson et al. Data-driven prediction of battery cycle life before capacity degradation. Nat. Energy, 2019, 4, 383.

[3] Yang F et al. A coulombic efficiency-based model for prognostics and health estimation of lithium-ion batteries. Energy, 2019, 117:1173-1182.

[4] Liu, WM. Early Prediction of Battery Cycle Life Using Machine Learning. Published online, University of Victoria 2021.

[5] Krishnan K. Building Big Data Applications. Academic Press, 2020: 113-125.2021.

[6] Ahmed M. Data Analytics: Concepts, Techniques, and Applications. CRC Press, 2019: 33-54.