Online estimation of state of health for lithium-ion batteries in
electric vehicles based on voltage integral and temperature

Kailun Chen¹,²,³, Qizhong Li¹,²,³*, Yujiu Wang¹,²,³ and Miaohua Huang¹,²,³

¹ Hubei Key Laboratory of Advanced Technology for Automotive Components,
  Wuhan University of Technology, Wuhan 430070, China
² Hubei Collaborative Innovation Center for Automotive Components Technology,
  Wuhan University of Technology, Wuhan 430070, China
³ Hubei Research Center for New Energy & Intelligent Connected Vehicle,
  Wuhan University of Technology, Wuhan 430070, China
*Corresponding author’s e-mail: qizhongli@whut.edu.cn

Abstract. In order to achieve accurate online State of health estimation of power battery, this
paper proposes a method based on voltage integration and temperature. Based on the NASA
ternary lithium battery accelerated life test, three groups of battery experimental data are divided
according to the difference in temperature and discharge rate. According to the single battery
data and battery experimental data required to record in the Chinese national standard documents,
the feasibility of voltage integration as a characterization of battery health is verified. Then we
discussed the charging and discharging behavior of electric vehicles and combine the Particle
Swarm Optimization algorithm to select the voltage integral of the fixed voltage interval in the
charging stage as the equivalent health factor of the battery health state, and establish a battery
health state estimation model based on the Gaussian process regression with the temperature and
voltage integral as input. The proposed method is verified by experimental data. The proposed
method is suitable for online estimation of battery health status under different temperatures and
different discharge rates. The root mean square error and average absolute error of battery health
status estimation are both less than 1.4%.

1. Introduction
With the intensification of environmental pollution and energy crisis caused by traditional fuel vehicles,
new energy vehicles have received more and more attention. In 2019, pure electric vehicles in China
accounted for 80% of new energy vehicles [1]. Lithium batteries are favored by the electric vehicle (EV)
markets because of their low self-discharge rate, high energy density, long service life and portability.
However, in the process of repeated cycles of lithium batteries, the continuous consumption of the
positive and negative active materials will cause the battery to age. The two main effects of battery aging
are capacity reduction and resistance increase [2]. The former manifests itself as the continuous decay
of the range of EV, while the latter may cause safety problems such as auto-ignition. Therefore, real-
time monitoring of the state of health (SOH) of lithium batteries is of great significance for the safe
operation of EV.

Xu et al. [3] constructed the measured health indicators by checking the discharge voltage sequence
and time sequence and then estimated them with an improved particle filter algorithm. However, the
lithium battery of EV is very complex due to the complex vehicle environment during the discharge
process. There are few cases of constant current discharge, so it is not particularly accurate to extract the health factor based on the discharge data to estimate the state of health of the battery. Park et al. [4] Select the amount of charge corresponding to the fixed state of charge interval during the constant current charging process as health factor for health estimation. Zhou et al. [5] the voltage integral corresponding to the fixed voltage interval during constant current charging is selected as the health factor, but the above method does not consider the optimization of the voltage. Zheng et al. [6] used particle swarm optimization (PSO) algorithm and optimized the voltage interval of the constant current charging stage, but directly selected the charging capacity in the optimal voltage interval as the health factor. However, according to the Chinese national standard document the specified EV communication protocol and data format do not require the recording of the current of the single battery. Therefore, it is difficult to calculate the charging capacity of EV online through current data, and the above studies have not considered the influence of temperature on SOH degradation. Taking into account the influence of the EV data format and temperature specified by the Chinese national standard, this article provides a method for online estimation of the battery state of the power battery pack based on the voltage integration and temperature during the charging stage of the vehicle. Tian et al [7]. extracts the SOH estimation method of health indicators from surface temperature, uses the temperature difference (DT) curve in a specific voltage interval is used for SOH estimation. Although it reflects the influence of temperature on SOH well, the error of SOH estimation based on a single thermal model is too large.

2. Health factor extraction

2.1. China National Standard Document

According to the barrel effect of the battery pack, we can estimate the life of the battery pack by estimating the shortest life of the single battery. According to the Chinese national standard document GB/T 32960.3-2016, the real-time report information of EV can be effectively used to estimate the SOH of the power battery is shown in Table 1. It can be seen that there is no need to record the current of a single battery, but the extreme values of the voltage and temperature are recorded. Therefore, it is not feasible to calculate SOH online by selecting current as the equivalent health factor. Consider selecting voltage integral to characterize battery SOH.

| Data content                                      | Length/byte | Type of data |
|--------------------------------------------------|-------------|--------------|
| Maximum & Minimum voltage battery subsystem number | 1           | BYTE         |
| Maximum & Minimum voltage battery cell code       | 1           | BYTE         |
| Maximum & Minimum voltage value                   | 2           | WORD         |
| Maximum & Minimum temperature battery subsystem number | 1           | BYTE         |
| Maximum & Minimum temperature battery cell code    | 1           | BYTE         |
| Maximum & Minimum temperature value               | 1           | BYTE         |

Table 1. Single battery related data

2.2. NASA ternary lithium battery accelerated life test

NASA testers set up many groups of ternary lithium battery accelerated life tests according to different temperatures and discharge currents. Here, the following nine battery data under different experimental conditions are selected and divided into three groups as data sets for analysis, the battery parameters and test conditions are shown in Table 2.

| Group | Battery number | Temperature (°C) | charging current(A) | Discharge current(A) | cut-off voltage(V) | Capacity (Ah) | battery model |
|-------|----------------|-----------------|---------------------|----------------------|-------------------|---------------|--------------|
| A     | A1, A2, A3     | 24              | 1.5                 | 2                    | 2,7,2.5,2.2      |               | NCA          |
| B     | B1, B2, B3     | 43              | 1.5                 | 4                    | 2,0,2,2,2,5      | 2             | NCA          |
| C     | C1, C2, C3     | 4               | 1.5                 | 1,2,4                | 2,2,2,5,2,7      |               | NCA          |

Table 2. Experimental conditions of the three groups of batteries
A complete cycle consists of a charging process and a discharging process. Taking group A as an example. First, keep the battery at about 24°C, charge with a constant current of 1.5A until the voltage reaches 4.2V, and then charge at a constant voltage until the current reduce to 20mA, and then start to discharge with a current of 2A until the voltage drops to 2.7V, 2.5V, and 2.2V. Figure 1 shows the current and voltage change diagram of A1 battery during a charge and discharge cycle.

Figure 1. A complete charging and discharging process

2.3. Health factor extraction
Since there is rarely constant current discharge during the actual operation of EV, the health factor is chosen to be extracted from the charging data. Based on the previous discussion on China's national standard documents, the voltage integral (VI) is initially selected as the equivalent health factor. Calculated as follows:

\[ VI = \int_{T_1}^{T_2} V \, dx \]  

In formula (1), \( T_1 \) and \( T_2 \) are the moments when the voltage reaches \( V_1 \) and \( V_2 \). The relationship between VI and the number of charge and discharge cycles and SOH of 9 batteries is shown in Figure 2. It can be seen that VI can effectively characterize SOH, and the two are approximately linearly related.

Figure 2. (a) the relationship between VI and cycle, (b) the relationship between VI and SOH

For the two linearly related variables of VI and SOH, the Pearson correlation coefficient (PCC) and Spearman correlation coefficient (SCC) can be used to measure the degree of linear correlation between them. It is generally considered that the correlation coefficient is greater than 0.8 as a very strongly correlated. Calculated as follows:

\[ r_p = \frac{E(XY) - E(X)E(Y)}{\left[ E(X^2) - (E(X))^2 \right]^{1/2} \left[ E(Y^2) - (E(Y))^2 \right]^{1/2}} \]  

(2)
The calculation results are shown in Table 3. It can be seen that the PCC and SCC between VI and SOH of the nine batteries are all greater than 0.8. Considered that there is a strong correlation between VI and SOH, so the effectiveness of selecting VI as an equivalent health factor is guaranteed.

Table 3. Correlation between VI and SOH

| Number | r_p  | r_s  |
|--------|------|------|
| A1     | 0.996| 0.987|
| A2     | 0.986| 0.984|
| A3     | 0.993| 0.992|
| B1     | 0.965| 0.961|
| B2     | 0.942| 0.946|
| B3     | 0.889| 0.886|
| C1     | 0.753| 0.822|
| C2     | 0.848| 0.851|
| C3     | 0.955| 0.959|

3. Selection of voltage window

3.1. Particle swarm optimization

PSO is an evolutionary computation technology, originated from the study of bird predation behavior. The basic idea of that is to find the optimal solution through collaboration and information sharing between individuals in the group. Its operation is similar to genetic algorithm (GA), but there is no crossover and mutation operation [8]. Because it is easy to implement and requires few parameters to be adjusted, it is widely used in the field of function optimization. The specific process is as follows:

Step1: Define the objective function and initialize the particles.
Step2: Update the velocity and position of particles.
Step3: Evaluate the fitness function value of the particle.
Step4: If the end condition is met, output the result, otherwise return to the step 2.

The formula for updating particle velocity and position by PSO is as follows:

\[
\begin{align*}
V_i &= \omega \cdot v_i + c_1 \cdot r_1 \cdot (p_b - x_i) + c_2 \cdot r_2 \cdot (g_b - x_i) \\
x_i &= x_i + v_i
\end{align*}
\]

(4)

In this article, set the initial value \( \omega = 0.8 \), \( c_1 = c_2 = 0.5 \).

3.2. Selection of voltage window

According to statistical analysis, due to "range anxiety", drivers were most likely to charge their vehicles when SOC was between 50% and 60%, and only a few drivers will charge their vehicles when SOC was between 0% and 20% [9]. At the same time, many car owners have a large charging time span.

Therefore, when selecting a certain voltage interval to calculate the VI, the SOC corresponding to the lower limit of the voltage interval should not be too low. And the span of the voltage range should not be too large, otherwise the SOC of the battery cannot be effectively estimated under most charging conditions. Preliminarily select 3.85V as the lower limit of the voltage interval, 4.15V as the lower limit of the voltage interval, and 0.2V as the span of the voltage interval. The starting voltage \( V_A \) and ending voltage \( V_B \) of the charging interval should meet the following conditions:

\[
\begin{align*}
3.85 &< V_A < 3.95, \\
4.0 &< V_B < 4.15
\end{align*}
\]
use $V_A$ and $V_B$ as the input of the PSO algorithm, the optimal voltage segment.

Fit the VI and the $VI_{AB}$ of the fixed voltage interval by the least square method:

$$VI_{i,k} = VI_{i,k}^{AB} \cdot k_i + b_i$$  \hspace{1cm} (5)

Where $VI_{i,k}$ is the voltage integral of the i-th battery during the k-th cycle charging stage, and $VI_{i,k}^{AB}$ is the voltage integral of the fixed voltage interval of the charging stage. $k_i$ is the slope and $b_i$ is the intercept.

The average slope and average intercept of all batteries in the training set are calculated as follows:

$$\bar{k} = (k_1 + k_2 + \cdots + k_i + \cdots + k_n)/n,$$
$$\bar{b} = (b_1 + b_2 + \cdots + b_i + \cdots + b_n)/n.$$  \hspace{1cm} (6)

Use the average slope and average intercept to estimate the voltage integral of each battery as:

$$VI_{i,k} = VI_{i,k}^{AB} \cdot \bar{k} + \bar{b}$$  \hspace{1cm} (7)

Then for the optimal voltage segment of the i-th battery, the objective function can be constructed as:

$$G_i = \left[ \frac{1}{m} \sum_{k=1}^{m} \left( \frac{VI_{i,k}}{VI_{i,k}^{AB}} - 1 \right)^2 \right]^{1/2}$$  \hspace{1cm} (8)

Consider all the batteries in the training set and construct the final objective function as:

$$G = \frac{1}{n} \sum_{i=1}^{n} G_i$$  \hspace{1cm} (9)

4. SOH estimation model based on GPR

GPR is a machine learning regression method. The purpose of regression is to find a function to describe a given set of data points as closely as possible. For a set of data, there may be an infinite number of functions available for fitting. The Gaussian process (GP) assigns a probability value to each such function, and then performs Bayesian inference in the function space, then the mean value of this probability distribution is the most likely representation of this set of data. To establish the desired distribution, first determine the mean function $\mu$ and the covariance matrix $\Sigma$:

$$\begin{align*}
\mu &= m(x) = E[f(x)], \\
\Sigma &= k(x,x') = E[(f(x) - m(x))(f(x') - m(x'))].
\end{align*}$$  \hspace{1cm} (10)

Where $x, x' \in \mathbb{R}^d$ are arbitrary random variables, then GP can be defined as $f(x) \sim GP(m(x), k(x,x'))$.

In the GP, it is generally assumed that the mean function is equal to zero. The key is how to set the covariance function $\Sigma$.

Considering noise, the regression model based on Gaussian process is:

$$y = f(x) + \epsilon$$  \hspace{1cm} (11)

Where $x$ is the input data vector, $\epsilon \sim N(0, \sigma^2)$ is the noise variance, and $y$ is the measured value under the influence of noise.

The joint prior distribution of the measured value $y$ and the predicted value $f^*$ is:

$$\begin{bmatrix} y \\ f^* \end{bmatrix} \sim N \begin{bmatrix} K(X,X) + \sigma_n^2 I_n \\ K(X,x^*) \\ K(x^*,x^*) \end{bmatrix}$$  \hspace{1cm} (12)

Where $K(X,X) = K_n = (k_{ij})$ is a symmetric positive definite covariance matrix of order $n \times n$, and the Matrix element $k_{ij} = k(x_i, x_j)$ is used to measure the correlation between $x_i$ and $x_j$. $K(X,x^*) = K(x^*,X)^T$ is the $n \times 1$ order covariance matrix between the point $x^*$ to be estimated and the input $X$ of the training set. $I_n$ is an $n$-dimensional identity matrix. $k(x^*,x^*)$ is the covariance of the point $x^*$ to be estimated.
Based on Bayesian inference, the posterior distribution of the predicted value $f^*$ can be calculated as:

$$f^* \sim (\mu', k')$$  \hspace{1cm} (13)

The mean and error of the predicted value $f^*$ are as follows:

$$\mu' = K(x, X) \cdot [K(X, X) + \sigma_n^2 I_n]^{-1}y$$  \hspace{1cm} (14)

$$k' = K(x, x) - K(x, X) \times [K(X, X) + \sigma_n^2 I_n]^{-1}K(X, x)$$  \hspace{1cm} (15)

According to the above model, the data set form is $\{x_i, y_i\}_{i=1}^n$, where $x_i = [V_i, T_i]$ is the input, $V_i$ and $T_i$ are respectively the VI and temperature in the voltage window of the i-th charge and discharge cycle. The output is $y_i$, which is the SOH of the i-th cycle.

5. Experiment

5.1. Evaluation indicators

This paper compares the Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Coefficient of Determination (R2) between the estimated SOH and the real SOH in the training set to evaluate the accuracy of the SOH estimation results:

$$\text{RMSE} = \left( \frac{1}{n} \sum_{i=1}^{n} \left( \frac{\text{SOH}_{i}}{\text{SOH}_{i}'} - 1 \right)^2 \right)^{1/2}$$  \hspace{1cm} (16)

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\text{SOH}_{i}}{\text{SOH}_{i}'} - 1 \right|$$  \hspace{1cm} (17)

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (\text{SOH}_{i} - \text{SOH}_{i}')^2}{\sum_{i=1}^{n} (\text{SOH}_{i} - \text{SOH})^2}$$  \hspace{1cm} (18)

Where SOHi is the true value of SOH in the i-th period, SOHi' is the estimated SOH value in the i-th period, and SOH is the average value of the true value of SOH in n periods.

5.2. Experimental Results and Analysis

Select the first two batteries of all groups in Table 1 as the training set, and the third battery as the test set. The training set data is used to solve the SOH estimation model represented by Eq. (11). Extract the voltage integral and ambient temperature from the test set data as input, and substitute it into the trained SOH estimation model to estimate the SOH of the battery corresponding to the charge and discharge cycle. Finally, evaluate the pros and cons of the model by comparing with the real value.

Take the average of the optimal voltage interval obtained by the particle swarm algorithm training and keep two decimal places, and determine that 3.86V-4.02V is the optimal voltage interval. Then the divided data set is substituted into the built model for training, RBF + WhiteKernel is used as the kernel function to converge the noise in the data. Kernel function width $r=0.1$. The obtained SOH estimation results and evaluation indicators are shown in Table 4 and Figure 3. Table 4 records the SOH estimation errors of the training set and the test set, and Figure 3 shows the SOH fitting results of the test set (A3, B3, C3). It can be seen from Table 4 and Figure 3 that the R2 of the three batteries in the test set are all greater than 0.85, indicating that VI and temperature as model inputs can be used to estimate SOH. At the same time, the RMSE of the three batteries are all less than 2%, and the MAE is less than 1%.

| Number | Voltage range | RMSE | MAE | R2   |
|--------|---------------|------|-----|------|
| A1     | [3.86V, 4.02V]| 2.53% | 2.19% | 0.9638|
| A2     |               | 2.96% | 2.17% | 0.9593|
| A3     |               | 1.62% | 1.14% | 0.9679|
| B1     |               | 0.49% | 0.39% | 0.9797|
6. Conclusion
This paper proposes a data-driven online estimation method for the state of health of electric vehicle batteries. Firstly, based on the data format required to upload by national standard for electric vehicles, VI was selected as the equivalent health factor that could be used for online calculation. Taking ternary lithium battery as the research object, the feasibility of voltage integral to represent SOH was verified based on the accelerated life test of ternary lithium battery of NASA. Then it discusses the charging and discharging behavior of electric vehicle drivers. It is believed that a practical battery charging voltage range is generally from the intermediate SOC charging to the nearly full state. Therefore, an objective function is constructed for battery data, and the optimal voltage interval is determined by the PSO algorithm. Finally, based on GPR, a battery SOH estimation model with voltage integration and temperature as input is established, which can be used to estimate battery SOH at different temperatures.

According to the battery test data of different temperatures and discharge currents, the training set and the test set are divided. The results show that the SOH is accurately estimated with an overall RMSE of 1.35% and MAE of 0.94%. Compared with the SOH estimation methods in other papers, the method proposed in this paper is suitable for battery SOH estimation under different temperatures and different discharge rates. At the same time, it has high accuracy, which ensures the robustness and effectiveness of the method proposed in this paper.

Acknowledgments
This research was sponsored by the National Key Technology Support Program of China (Grant No. 2018YFE0105500).

References
[1] Shen, G.Y. (2020) Economic performance of the auto industry in 2019. Fine and Specialty Chemicals, 2020, 28(03):27.
[2] Barre, A., Deguilhem, B., Grolleau, S., Gerard, M., Suard, F., Riu, D. (2013) A review on lithium-ion battery ageing mechanisms and estimations for automotive applications. Journal of Power Sources, 2013, 241:680-689.
[3] Xu C., Li L.W., Yang Y.X. State of Health Estimation of Lithium-Ion Battery Based on Improved Particle Filter. Automobile Technology, 2020, 12:19-24.

[4] Park, M., Seo, M., Song, Y., Kim, S. (2019) Capacity Estimation of Lithium-ion battery using Constant Current Charging Voltage. In: the IEEE International Conference on Asia Pacific Conference on Circuits and Systems (APCCAS). Bangkok. pp. 202-204.

[5] Zhou, Y.P., Huang, M.H., Pecht, M. (2018) An Online State of Health Estimation Method for Lithium-ion Batteries Based on Integrated Voltage. In: the IEEE International Conference on International Conference on Prognostics and Health Management (ICPHM). Seattle. pp. 1-5.

[6] Zheng, Y.J., Wang, J.J., Qin, C., Lu, L.G., Han, X.B., Ouyang, M.G. (2019) A novel capacity estimation method based on charging curve sections for lithium-ion batteries in electric vehicles. ENERGY, 2019, 185:361-371.

[7] Tian, J.P., Xiong, R., Shen, W.X. (2020) State of health estimation based on differential temperature for lithium ion batteries. IEEE Transactions on Power Electronics, 2020, 35(10):10363-10373.

[8] Li, A.G., Qin, Z., Bao, F., He, S.P. (2002) Particle swarm optimization algorithms. Computer Engineering and Applications, 2002, 38(021):1-3.

[9] Liu, Y.J., Hong, Y., Hu, C. (2019) Research on Big Data Mining Technology of Electric Vehicle Charging Behaviour. ELEKTRONIKA IR ELEKTROTECHNIKA, 2019, 25(06):55-61.