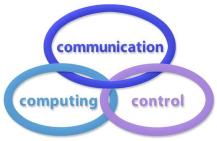
INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 20, Issue: 3, Month: June, Year: 2025
Article Number: 7063, https://doi.org/10.15837/ijccc.2025.3.7063



CCC Publications

Prediction of power battery health state based on data-driven Gaussian process regression algorithm

Wei Li, Chen Cheng, Qian Wang, Chang-song Ma

Wei Li*
School of Electronic Information Engineering
Geely University of China
Chengdu, 641423
*Corresponding author: li.wei.77@163.com

Chen Cheng
Yibin city Syzhou district Anju property services Co.
LTD,Yibin,644600

Qian Wang
School of Electronic Information Engineering
Geely University of China
Chengdu, 641423

Chang-song Ma
School of Electronic Information Engineering
Geely University of China
Chengdu, 641423

Abstract

As one of the key functions of the battery management system, the prediction of the state of health (SOH) of the power battery of new energy vehicles is crucial for the management and maintenance of the power battery system and its safe use. In order to improve the prediction accuracy and generalization ability of power battery health state, a SOH prediction method combining data-driven and gaussian process regression (GPR) based on data correlation was proposed. The normalized values of discharge capacity, voltage range, voltage variance, internal resistance range, internal resistance variance and final discharge temperature of lithium battery were selected to analyze the relevant characteristics of SOH. The aging model of power battery was constructed by GPR algorithm, and 6 health indexes were input into the model to predict the test data set of power battery. mean absolute error (MAE), root mean square error (RSME) and fitting coefficient R^2 were used as model evaluation indexes. Input 6 normalized values of the test set under different working conditions to verify the generalization ability of the model. The experimental results show that the error of the model is less than 1%. Both MAE and RSME values were within 0.04, and R^2 values were greater than 0.95. In generalization verification, the average error is less than 1%. MAE was 0.0238; The RMSE value was 0.0239. The R^2 value is 0.9241. The model has good generalization and application ability.

Keywords: State of health; Battery management system; Gaussian process regression; Average absolute error; Root mean square error; Generalization ability;

1 Introduction

The health of the power battery is critical to the performance and safety of electric vehicles. With the continuous use of the battery, its internal structure begins to gradually age, the active substance is reduced, the internal resistance is increased, the maximum available capacity and power performance are reduced, and the health of the battery is getting worse and worse. With the popularity of new energy electric vehicles in recent years, the efficient use of automotive power batteries has become increasingly prominent. As a key core component of new energy vehicles, the performance test and evaluation of power battery is an important link to ensure the power performance and safety performance of electric vehicles. Power battery state of health (SOH) indicators mainly include battery capacity, battery internal resistance, self-discharge rate, energy density and cycle life. Paying attention to the above health status indicators can make a better judgment on the health status of the battery, which is of great significance in extending the service life of the power battery. At the same time, it can effectively improve the safety performance of the power battery. The power battery health status prediction is an important research direction in the field of electric vehicles, whose purpose is to accurately evaluate the performance of the battery and predict the future performance attenuation trend of the battery, so as to provide decision support for battery maintenance and replacement. Common forecasting methods include model forecasting, data-driven forecasting and so on. Research based on these methods has made some progress, but still faces many challenges. Such as the uncertainty of the battery aging process, the complexity of the battery discharge process, the difference between different battery individuals, and the interference of different working environments. It can be predicted that future research will pay more attention to the research of battery aging mechanism, the generalization ability of prediction methods and the practical application ability.

2 Literature review

At present, many scholars have carried out research on battery SOH prediction methods. The research of lithium-ion battery management system has shifted from a single technological breakthrough to multi-disciplinary collaborative innovation, and its core tasks revolve around battery state estimation, optimization algorithm design and engineering application. Battery state estimation is the core research direction of battery management systems (BMS), covering the prediction of state of charge (SOC), State of health (SOH) and remaining useful life (RUL). Traditional methods such as nonlinear filtering (the Unscented Kalman filter is used in reference [1]) show high precision in the experimental model, but rely on precise electrochemical parameters. In recent years, data-driven methods have become the mainstream. In reference [2], Elman neural network is improved by particle swarm optimization (PSO) to achieve joint SOC and SOH estimation. In reference [3], an ensemble learning framework based on sparse Gaussian process regression (GPR) was proposed, and the SOH prediction error for fast-charging batteries was less than 3%. Reference [4] uses deep learning to extract aging features directly from historical data and achieves SOH error of less than 2% without additional experiments. Hybrid models (e.g., Reference [4] combining electrochemical impedance spectroscopy with cuckoo search optimization) further improve the robustness of life cycle health assessments. The current challenge lies in real-time optimization and small sample generalization capability under dynamic conditions (reference [5]). The meta-heuristic algorithm shows significant advantages in parameter tuning and resource allocation. Reference [1] combines PSO with fuzzy logic to optimize the maximum power point tracking efficiency of photovoltaic systems. The ant colony algorithm is improved to solve the dynamic optimization problem of new energy resource allocation in the cloud environment. In reference [6], a PSO-Least-square support vector machine (LS-SVM) hybrid model was proposed to dynamically correct the SOC estimation error to within 1.5%. The application potential of meta-heuristic methods such as genetic algorithm and cuckoo search in PID controller tuning is systematically reviewed in reference [7]. Reference [8] Optimization of cavity cold plate design through simulation to improve the thermal management efficiency of lithium-ion batteries. The trend indicates that multi-objective optimization (e.g., multi-agent reinforcement learning in reference [9]) and online real-time computing (e.g., Harris Hawks optimization algorithm in reference

[10]) are the focus of the future, but the contradiction between algorithm complexity and hardware compatibility needs to be solved. Deep learning and reinforcement learning excel in the modeling and control of complex systems. Reference [11] Uses convolutional neural network (CNN) to realize automatic voltage stabilization in substation. Reference [12] proposes a multi-objective multi-agent reinforcement learning framework to reduce energy consumption of energy distribution tasks by 15% through knowledge transfer. References [13] and [14] use LSTM and end-cloud collaborative framework to predict battery SOH with high accuracy, respectively. Reference [15] uses deep learning to diagnose distribution network faults, with an accuracy rate of 98%. The application of reinforcement learning in control strategy optimization (such as comparison of model reference control methods in Reference [16]) provides a new idea for BMS dynamic response. However, insufficient model interpretability (such as the genetic testing study in reference [17]) and strong data dependence remain major bottlenecks. The interdisciplinary hybrid model significantly improves the prediction accuracy and engineering applicability. By combining electrochemical impedance spectroscopy (EIS), cuckoo search and Elman neural network, the whole life cycle health model was constructed, and the error was reduced by 30% compared with the single method. In reference [18], the battery decay mechanism was embedded in multi-core support vector regression (SVR) to extract health factors indirectly, and the RUL prediction error was less than 5%. In reference [19], an adaptive multi-Gaussian process regression (GPR) model was proposed, and the Harris Hawks algorithm was used to optimize the hyperparameters to achieve a strong correlation between voltage characteristics and health status. In reference [20], artificial neural network (ANN) is used to predict metal machining roughness, which provides a reference for battery manufacturing process optimization. In the future, it is necessary to further integrate electrochemical, thermodynamic and data-driven models (such as thermal-electrical coupling design in reference [21]) to promote the landing of the "mechanical-data" dual drive framework. The battery management technology is expanding to many fields: Reference [22] optimizes the positioning accuracy of medical electronic wristband by differential control, which provides reference for the design of miniaturized BMS; The cavity cold plate simulation model in reference [23] verifies the importance of thermal management on battery life. Reference [24] applies deep learning to power grid fault diagnosis, highlighting the necessity of battery-grid collaborative optimization. These studies demonstrate the evolution from single-component optimization (e.g., motor control in reference [25]) to system-level integration (e.g., end-cloud collaboration framework in reference [26]), but address challenges such as standardization of cross-system communication (e.g., reference [27]) and reliability verification. Several reviews systematically summarized technical progress and bottlenecks: References [28] and [29] emphasized that the combination of aging mechanism and data-driven is the key to improving the accuracy of RUL prediction; Reference [30] analyzes the engineering obstacles of battery life prediction technology from the perspective of laboratory to industrial implementation. Reference [31] compares the advantages and disadvantages of hybrid collaborative estimation methods and points out the application potential of edge computing and lightweight models. Together, these reviews call for standardized data sets (e.g., EV battery health analysis in reference [32]) and cross-domain collaboration frameworks (e.g., Reference [33]) to accelerate technology transition from theory to industry. Lithium-ion battery management research is shifting from a single technological breakthrough to multidisciplinary collaborative innovation, covering state estimation, optimization algorithms, deep learning, and engineering applications. Despite significant advances in accuracy and adaptability, real-time, interpretability, and system integration remain core challenges. In the future, it is necessary to focus on lightweight hybrid models, edge intelligent deployment and standardized data platforms to promote the wide application of BMS technology in large-scale energy storage and smart grids.

As you can see, although these methods are classified differently, these classifications have some things in common. Most methods are model-based prediction methods for machine learning methods. Machine learning methods mainly include neural network learning methods and statistical learning methods. Statistical learning is more suitable for the application of small samples, and has the advantages of fast computation and easy training. Currently, commonly used statistical learning algorithms include SVR, RVM, GPR, LSTM-RNN, BPNN, etc(Reference [34]). GPR is a method based on statistical principles and Bayesian theory, which has outstanding advantages in dealing with

small sample regression problems, and can output confidence intervals to represent the uncertainty of results. GPR model can give predicted results through posterior probability, which is simple and convenient to operate, suitable for high-dimensional small sample problems, and is a very widely used machine learning method. Because the GPR model uses a single kernel function, the single kernel function has the shortcomings of weak fitting and generalization ability, easy to fall into the local optimal, which makes the model difficult to accurately capture the capacity regeneration phenomenon.

In order to effectively solve the problem of weak generalization ability of GPR prediction model, this paper proposes a SOH prediction method that combines data-driven and Gaussian process regression based on data correlation. This method adopts data-driven, constructs multidimensional kernel function, and can highly fit SOH curve, effectively improve SOH prediction accuracy, and enhance the generalization ability of the model.

3 Research methodology

In this paper, the battery health status prediction model framework is shown in Figure 1 below. The model consists of three parts: training data set, test data set and generalization data set. The training data set consists of data filtering and GPR algorithm construction. In the GPR model, kernel function parameters σ^2 and l are used to control the fitting ability of the model. Selecting the appropriate σ^2 and l can effectively improve the prediction performance of the GPR model and avoid falling into overfitting or underfitting.

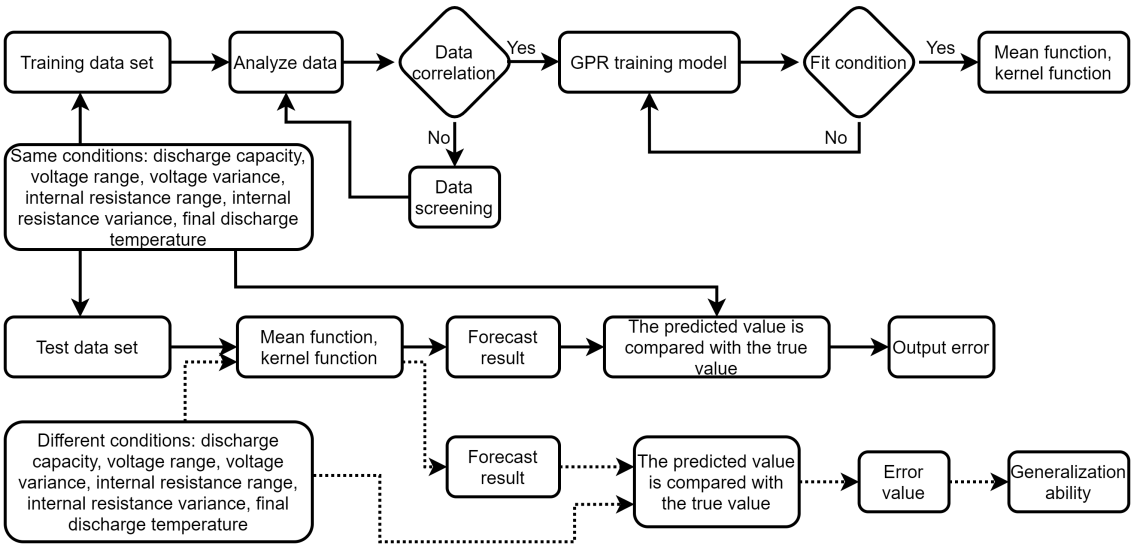


Figure 1: Model framework

Pearson correlation coefficient was used to measure the correlation degree between the health parameters of the power battery and the health state . The Pearson correlation coefficient is represented by the letter ρ and is defined as the quotient of the covariance and standard deviation between two variables, as follows:

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \tag{1}$$

The above equation defines the overall correlation coefficient. σ_x, σ_y is the standard deviation of the two variables. To estimate the covariance and standard deviation of the power battery health data sample, the Pearson correlation coefficient can be obtained, as shown below:

$$\rho = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}} \tag{2}$$

where X_i is the ith input battery health parameter, Y_i is the health value predicted by the i-th sampling time, n is the total number of battery tests, and $\bar{X}, \bar{Y}$ are the average values of the battery health status data samples. The stronger the correlation between X_i and Y_i, the closer the Pearson coefficient is to 1.

Gaussian process regression is used to fit the corresponding Gaussian process through finite high-dimensional data to predict the function value under any random variable. The Gaussian process expression is:

$$f(X) \sim \mathcal{N}[v(X), k(X, X\hat{})] \tag{3}$$

The Gaussian process consists of the mean function $v(X)$ and the kernel function $k(X, X\hat{})$, as follows:

$$v(X) = \mathbb{E}[f(X)] \tag{4}$$

$$k(X, X\hat{}) = \mathbb{E}\left\{[f(X) - v(X)][f(X\hat{}) - v(X\hat{})]\right\} \tag{5}$$

where $f(X)$ is the output target value function; X is multidimensional with n input vectors.

Gaussian regression is a non-parametric regression method based on Bayesian theory, which continuously updates the posterior probability distribution through the measured data until the posterior distribution essentially matches the true distribution. Setting the mean function value to 0, the kernel function is represented by the squared exponential function:

$$k(X, X\hat{}) = \sigma^2 e^{-\frac{(X - X\hat{})^2}{2l^2}} \tag{6}$$

where σ^2 is the output variance and l is the length scale. The distance of the sampling points determines their correlation. The closer the distance, the stronger the correlation; the farther the distance, the weaker the correlation.

The multidimensional kernel function is as follows:

$$k(X, X\hat{}) = \frac{1}{(2\pi)^{m/2}} \sigma^2 e^{-\frac{\left(X - X\hat{}\right)^2}{2l^2}} \tag{7}$$

Where m is the dimension of the kernel function.

Noise is added to the kernel function, and the noise satisfies the Gaussian distribution. At this time, the input and output can be expressed as follows:

$$Y = f(X) + \xi \tag{8}$$

The Gaussian process at this time is:

$$f(X) \sim N[0, k(X, X\hat{}) + \sigma_n^2 L_n] \tag{9}$$

Where, $\xi \sim N(0, \sigma^2)$ is the white noise with variance σ_n^2, and Ln is the n-dimensional identity matrix. ‘ The identity matrix is the identity matrix of the kernel function, as follows:

$$K = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & ... & 0 \\ 0 & 0 & 0 & \sigma_n^2 \end{bmatrix} \tag{10}$$

In the identity matrix of the kernel function, there is A hyperparameter $\zeta = [\sigma, l, \sigma_n]$, which can be solved by maximizing likelihood estimation, expressed as:

$$\begin{aligned} L = \log(Y \mid X, \xi) = -\frac{1}{2} \log\{\det[k(X, X^{\wedge}) + \sigma_n^2 l_n]\} \\ - \frac{1}{2} Y [k(X, X^{\wedge}) + \sigma_n^2 l_n]^{-1} Y - \frac{n}{2} \log 2\pi \end{aligned} \tag{11}$$

The maximum value of the objective function is obtained by derivation of the log-likelihood function, calculated as follows:

$$\frac{\delta}{\delta\varsigma_i}\log(Y \mid X,\varsigma) = \frac{1}{2}tr\begin{bmatrix}\alpha\alpha^T - [k(X,X^\wedge)+\sigma_n^2 l_n]^{-1} \\ \\ \frac{\delta[k(X,X^\wedge)+\sigma_n^2 l_n]}{\delta\varsigma_i}\end{bmatrix} \tag{12}$$

Where, $\alpha = [k(X,X^\wedge)+\sigma_n^2 l_n]^{-1}Y$, tr() represents the trace of the matrix, and the joint prior distribution of the measured value Y and the predicted value Yˆ in the test data set Xˆ is:

$$\begin{bmatrix}Y \\ Y^\wedge\end{bmatrix} \sim N(0,[\begin{matrix}k(X,X)+\sigma_n^2 & k(X,X^\wedge) \\ k(X,X^\wedge)^T & k(X^\wedge,X^\wedge)\end{matrix}]) \tag{13}$$

The Bayes formula is as follows:

$$p(X \mid Y) = \frac{p(Y \mid X)p(x)}{p(y)} = \frac{p(Y \mid X)p(x)}{\int p(Y \mid X)p(X)dX} \tag{14}$$

Where, $p(X|Y)$ is the posterior probability; $p(X)$ is the prior probability; $p(Y|X)$ is the likelihood probability.

Therefore, from the joint prior distribution of Y, the posterior distribution $p(Y^\wedge|X,Y,Y^\wedge)$ can be calculated as follows:

$$p(Y^\wedge|X,Y,Y^\wedge) = \mathcal{N}[\bar{Y}^\wedge, \mathrm{cov}(\bar{Y}^\wedge)] \tag{15}$$

In the formula, covariance $\mathrm{cov}(\bar{Y}^\wedge)$ and forecast mean $\bar{Y}^\wedge$ are calculated as follows:

$$\mathrm{cov}(\bar{Y}^\wedge) = k(X^\wedge,X^\wedge) - k(X,X^\wedge)^T[k(X,X)+\sigma_n^2 I_n]^{-1}k(X,X^\wedge) \tag{16}$$

$$\bar{Y}^\wedge = k(X,X^\wedge)^T[k(X,X)+\sigma_n^2 I_n]^{-1}Y \tag{17}$$

At this point, the construction of the mean function and kernel function is completed, and the establishment of the Gaussian process regression model is completed.

There are many ways to define SOH, such as the definition of battery internal resistance, battery capacity, battery power, and remaining cycle number. This paper adopts the capacity definition method and the internal resistance definition method, and the definition expressions are shown as follows:

$$SOH = \frac{B_{\text{actual}} - B_{\text{finish}}}{B_{\text{initiate}} - B_{\text{finish}}} \times 100\% \tag{18}$$

$$SOH = \frac{R_{\text{finish}} - R_{\text{actual}}}{R_{\text{finish}} - R_{\text{initiate}}} \times 100\% \tag{19}$$

Where, B_{actual} is the current actual battery capacity; B_{finish} is the capacity value at the end of the battery life (less than 80% capacity); B_{initiate} is the new battery capacity value; R_{actual} is the current actual internal resistance of the battery; R_{finish} is the internal resistance at the end of battery life (less than 80% capacity); R_{initiate} is the internal resistance of the new battery.

The type of power battery in the experiment is a lithium battery used as a test sample. The lithium battery is divided into four groups, namely Li01, Li02, Li03, and Li04. Battery working conditions are divided into four types. One type is constant current charging, constant voltage charging, and constant discharge at the ambient temperature of 25°C. The second is constant current charging, constant voltage charging, and constant discharge at the ambient temperature of 40°C. The third is constant current charging, constant voltage charging, and constant discharge at the ambient temperature of 10°C. The fourth is constant current charging, constant voltage charging, and constant discharge at the ambient temperature of -5°C.

The cross-current charging current is 1A; the constant voltage charging voltage is 4.2V; the cross-flow discharge current is 1.5A; and the minimum discharge current is 18mA. The minimum discharge voltage is 2.5V.

After testing, the SOH change curve of the four groups of lithium batteries is shown in Figure 2.

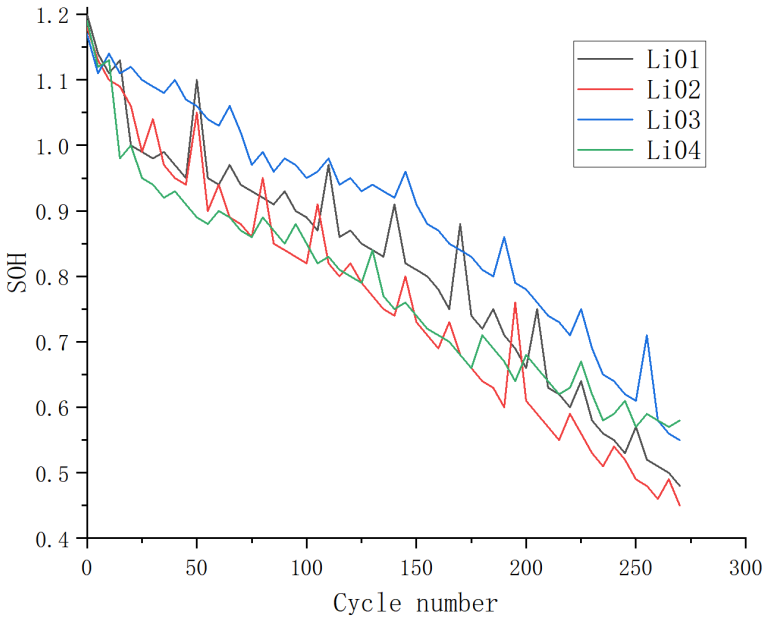


Figure 2: SOH curves under different working conditions

As can be seen from the figure, when the capacity of lithium battery has a certain regeneration ability, the SOH trend is nonlinear decline with the increase of the number of cycles. This trend makes it difficult to predict the SOH of power batteries. Therefore, this paper directly uses battery discharge voltage, discharge temperature, discharge capacity, discharge internal resistance, discharge voltage range and time as the model prediction inputs. These parameters have certain certainty and robustness, which provide a good guarantee for the accuracy of the prediction model.

Taking experimental battery Li01 as an example, as the number of cycles increases, the time to reach the minimum discharge voltage becomes smaller, and the change curve is shown in Figure 3 below. It shows that the discharge voltage is different under different working conditions.

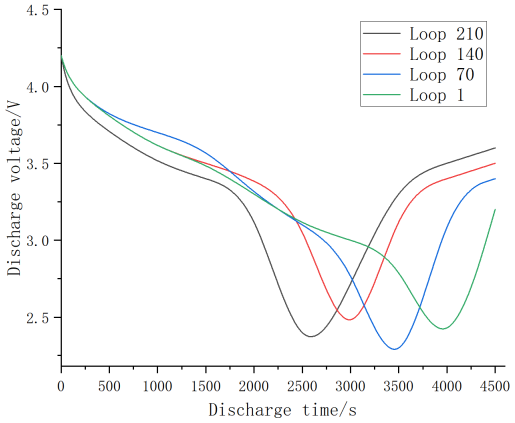


Figure 3: Discharge voltage curves of different cycles

With the increase of the number of cycles, the time to reach the maximum discharge temperature of battery Li01 gradually decreases. The change curve is shown in Figure 4 below. It shows that the discharge temperature is different under different working conditions.

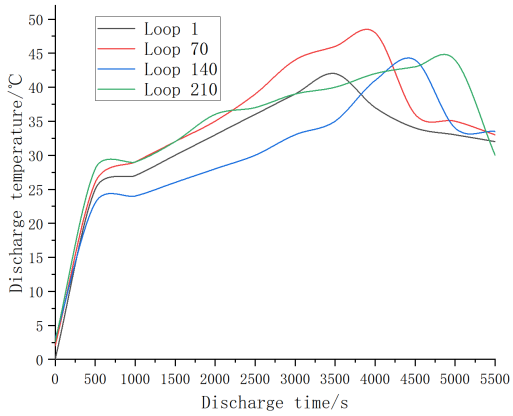


Figure 4: Discharge temperature curves of different cycles

With the increase of the number of cycles, the time required for the battery Li01 discharge capacity to reach 80% is gradually reduced. The change curve is shown in Figure 5 below. It shows that the discharge capacity is different under different working conditions.

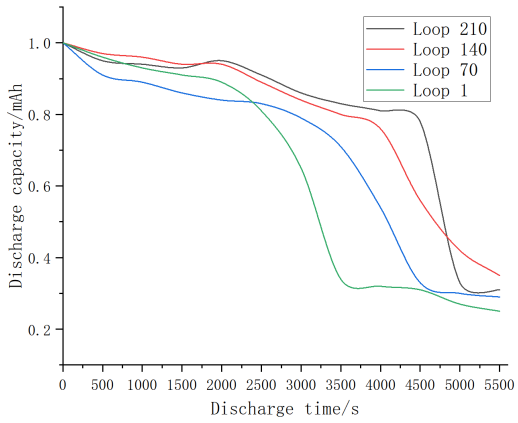


Figure 5: Discharge capacity curves of different cycles

With the increase of cycle times, the internal discharge resistance of battery Li01 increases gradually. The SOC increases significantly in the range of $0 \sim 35\%$. The change curve is shown in Figure 6 below. It shows that the internal resistance of discharge is different under different working conditions.

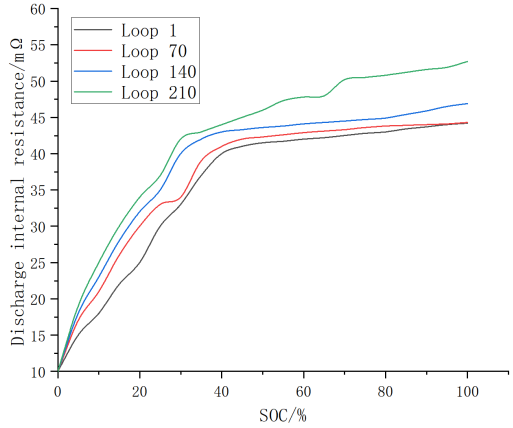


Figure 6: Internal resistance curve of discharge

The voltage range is at a trough when SOC is 10%, and the battery Li01 voltage range gradually decreases with the increase of cycles. This difference is mainly caused by the inconsistent attenuation

of cell capacity, resulting in the formation of DC internal resistance gradient. The voltage range curve is shown in Figure 7.

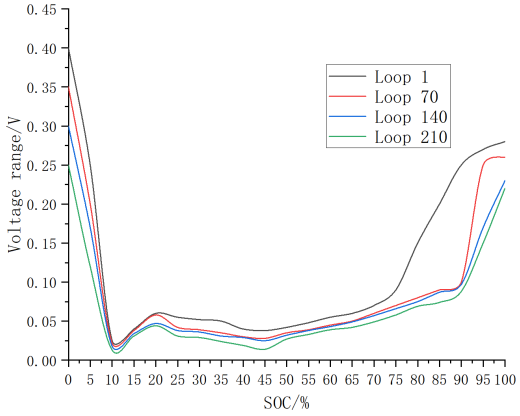


Figure 7: Voltage range curve

Based on the above analysis, the battery health indicators were selected as the health indicators of the model, including discharge capacity, voltage range, voltage variance, internal resistance range, internal resistance variance and final discharge temperature.

6 health indicators are brought into formula (2) to obtain Pearson correlation coefficient, as shown in Table 1 below.

Table 1: Pearson correlation coefficients of different health indicators

Health index	Li01	Li02	Li03	Li04
Discharge capacity	0.9187	0.9072	0.8741	0.8927
Voltage range	0.9233	0.9109	0.8933	0.9021
Voltage variance	0.9481	0.8996	0.8896	0.8579
internal resistance is very poor	0.9312	0.9201	0.9028	0.8406
Internal resistance variance	0.9328	0.9111	0.8674	0.8937
Final discharge temperature	0.9417	0.8936	0.8925	0.9015

As can be seen from Table 1, Pearson correlation coefficients are all greater than 0.85. Therefore, the above 6 parameters are selected as the input values of the Gauss process review model.

4 Results and discussion

The normalized data sets of discharge capacity, voltage range, voltage variance, internal resistance range, internal resistance variance and final discharge temperature were obtained and divided into two groups. The first group was trained by Gaussian process regression model and the output model was obtained. Use the second set as the test set. The test results of the test battery Li01 are shown in Figure 8. The predicted value fluctuates around the true value, and the volatility is more obvious on individual data. The overall error is less than 1%.

The errors of test batteries Li01, Li02, Li03 and Li04 are shown in Table 2 below.

Table 2: Model error

Battery number	Li01	Li02	Li03	Li04
error	0.963%	0.985%	0.979%	0.998%

The error is less than 1%. It shows that the Gaussian process regression model has better predictive ability in the test data set.

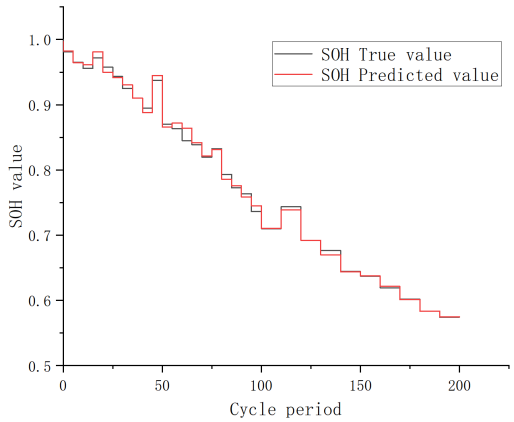


Figure 8: Voltage range curve

SVR(support vector regression) model, LSTM-RNN (short term memory recurrent neural network model), BPNN (back propagation neural network) model and Gaussian process regression model (GPR) in this paper were used to predict the health state of batteries Li01, Li02, Li03 and Li04 respectively. Taking battery Li01 as an example, the measurement results are shown in Figure 9.

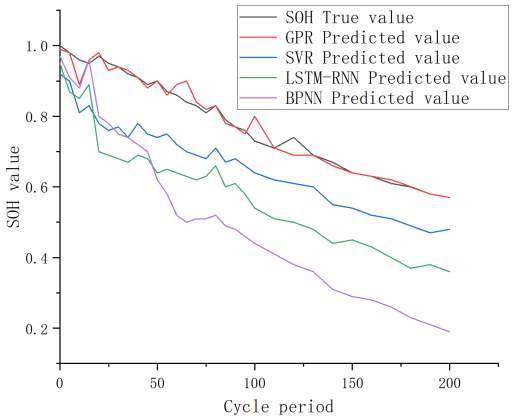


Figure 9: Model comparison test curve

The average absolute error between the predicted value and the true value (MAE), the sample standard deviation (RMSE) of the difference between the predicted value and the true value, and the fitting coefficient (R^2) were used as three quantities for model evaluation. In the evaluation, the smaller the error of MAE and RMSE, and the larger the R2 value, the better the model. The definition is as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - Y_i^{\wedge}| \tag{20}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(|Y_i - Y_i^{\wedge}|)^2} \tag{21}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - Y_i^{\wedge})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2} \tag{22}$$

Where, n is the number of experiments; Y_i is the real measured value; $Y_i^{\wedge}$ is the predicted value; $\bar{Y}_i$ is the average of the actual measured values.

For the health status of batteries Li01, Li02, Li03, and Li04, MAE results are calculated as shown in Figure 10.

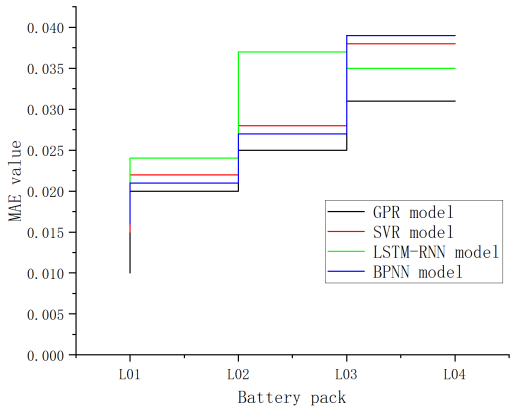


Figure 10: Comparison of MAE values of four forecasting models

For the health status of batteries Li01, Li02, Li03 and Li04, the RMSE calculation result is shown in Figure 11.

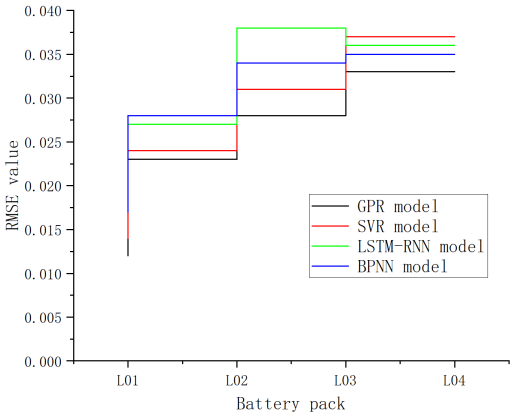


Figure 11: Comparison of RMSE values of four forecasting models

For the health status of batteries Li01, Li02, Li03, and Li04, the results of calculating R^2 are shown in Figure 12.

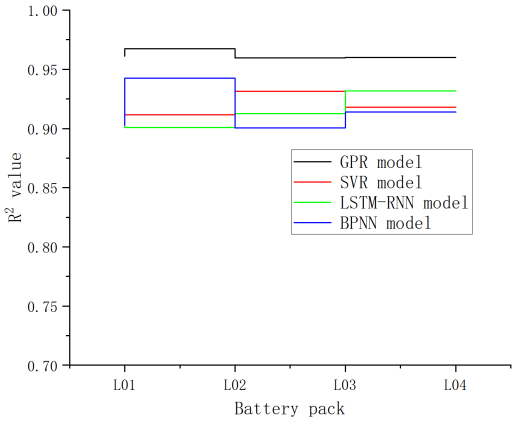


Figure 12: Comparison of R^2 values of four forecasting models

As can be seen from the above Figure 10, 11, and 12, MAE values and RMSE values of the four models are all within 0.04, and R^2 values are all greater than 0.9. However, the distribution of the GPR model is more uniform, the fluctuation is small, and the R^2 value is better than the other three groups of algorithms. The prediction accuracy of the GPR model is better than the SVR model, LSTM-RNN model, and BPNN model.

Table 3: Model error

temperature	-5°C	10°C	25°C	40°C
error	0.989%	0.985%	0.978%	0.996%

Added lithium battery Li05. There are four battery working conditions: constant current charging, constant voltage charging, and constant discharge at -5°C, 10°C, 25°C, and 40°C. The cross-current charging current is 1.5A; the constant voltage charging voltage is 4.2V; the cross-flow discharge current is 2A; and the minimum discharge current is 18mA. The minimum discharge voltage is 2.5V. The predictive ability of the model is shown in Figure 13.

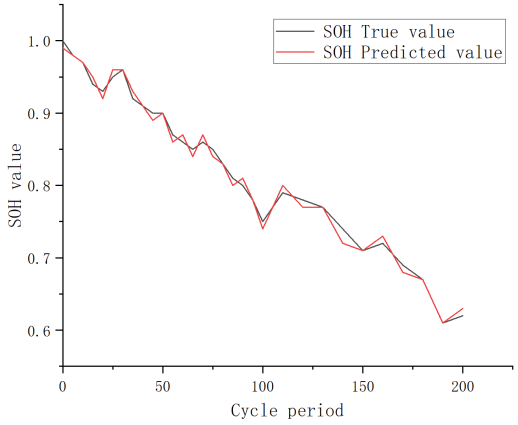


Figure 13: Model generalization ability

Errors at -5°C, 10°C, 25°C and 40°C are shown in Table 3 below.

For the battery under different working conditions, the model also has good prediction ability, and the model has good generalization application ability. The evaluation error was less than 1%. MAE was 0.0238; The RMSE value was 0.0239. The R^2 value is 0.9241, which is slightly smaller than the fitting coefficient value of the test set, due to the reversibility of the battery power.

5 Conclusion

This paper mainly studied a SOH prediction method based on the combination of data-driven and Gaussian process regression based on data correlation. The capacity, voltage, internal resistance and temperature data of the battery were selected as the sample input, and the model was trained to obtain the prediction model. The accuracy of the prediction model is improved by controlling the fitting ability of the model by means of the mean function and the multi-kernel function parameters. The experimental results show that the proposed method is better than SVR model, LSTM-RNN model and BPNN model, and the error can be controlled within 1%, and it has good generalization application ability.

References

[1] Abdolrasol, M. G., Ayob, A., Mutlag, A. H., & Ustun, T. S. (2023). Optimal fuzzy logic controller based PSO for photovoltaic system. Energy Reports, 9, 427-434. https://doi.org/10.1016/j.egyr.2022.11.039.

[2] Du, H. & Chen, J. (2023). An Improved Ant Colony Algorithm for New energy Industry Resource Allocation in Cloud Environment. Tehnički vjesnik, 30 (1), 153-157. https://doi.org/10.17559/TV-20220712164019.

[3] Moon, J., Son, M., Oh, B., Jin, J., Shin, Y.: Automatic Voltage Stabilization System for Substation using Deep Learning. Computer Science and Information Systems, Vol. 21, No. 2, 437–452. (2024), https://doi.org/10.2298/CSIS220509050M.

[4] Gupta, S., Naik, A. K., Dey, R., Singh, A. K., Popa, M., & Popa, A. (2025). State-of-Charge Estimation of an Experimentally Identified Lithium-ion Cell Model using Advanced Nonlinear Filters. International Journal of Computers Communications & Control, 20(2). https://doi.org/10.15837/ijccc.2025.2.6896.

[5] Samanta, A., & Williamson, S. (2023). Machine learning-based remaining useful life prediction techniques for lithium-ion battery management systems: A comprehensive review. IEEJ Journal of Industry Applications, 12(4), 563-574. https://doi.org/10.1541/ieejjia.22004793.

[6] Li Wei, Meng Xeu-qin, Ma Chang-song. Optimal Differential Control for Accurate Positioning of Medical Electronic Wristband. Tehnicki Vjesnik-Technical Gazette, 2024, 31(3): 792-799. https://doi.org/10.17559/TV-20230809000866.

[7] Almeida, V. N. d., Alegre, L. N., Bazzan, A. L. C.: Knowledge Transfer in Multi-Objective Multi-Agent Reinforcement Learning via Generalized Policy Improvement. Computer Science and Information Systems, Vol. 21, No. 1, 335–362. (2024). https://doi.org/10.2298/CSIS221210071A.

[8] Grimaldi, A., Minuto, F. D., Perol, A., Casagrande, S., & Lanzini, A. (2023). Ageing and energy performance analysis of a utility-scale lithium-ion battery for power grid applications through a data-driven empirical modelling approach. Journal of Energy Storage, 65, 107232. https://doi.org/10.1016/j.est.2023.107232.

[9] Miranda, M. H., Silva, F. L., Lourenço, M. A., Eckert, J. J., & Silva, L. C. (2023). Particle swarm optimization of Elman neural network applied to battery state of charge and state of health estimation. Energy, 285, 129503. https://doi.org/10.1016/j.energy.2023.129503.

[10] Xiong, R., Wang, S., Takyi-Aninakwa, P., Jin, S., Fernandez, C., Huang, Q., ... & Zhan, W. (2024). Critical Review on Improved Electrochemical Impedance Spectroscopy-Cuckoo Search-Elman Neural Network Modeling Methods for Whole-Life-Cycle Health State Estimation of Lithium-Ion Battery Energy Storage Systems. Protection and Control of Modern Power Systems, 9(2), 75-100. https://doi.org/10.23919/PCMP.2023.000234.

[11] Chen J. Q., Guan J. F., Liu J. D., Wu J. Q., Chen W. R. (2024). Simulation Study on Dynamic Envelope and Interference Check of Stitch Wire. Int. Journal of Simulation Modelling, Vol. 23, No. 2, p. 311-322.10.2507/IJSIMM23-2-687.

[12] Qu, X., Shi, D., Zhao, J., Tran, M. K., Wang, Z., Fowler, M., ... & Burke, A. F. (2024). Insights and reviews on battery lifetime prediction from research to practice. Journal of Energy Chemistry. https://doi.org/10.1016/j.jechem.2024.03.013. https://doi.org/10.1007/s10008-023-05650-3.

[13] Cao, J., Wang, S., & Fernandez, C. (2024). Multi-kernel support vector regression optimization model and indirect health factor extraction strategy for the accurate lithium-ion battery remaining useful life prediction. Journal of Solid State Electrochemistry, 28(1), 19-32. https://doi.org/10.1007/s10008-023-05650-3.

[14] Das, K., Kumar, R., & Krishna, A. (2024). Analyzing electric vehicle battery health performance using supervised machine learning. Renewable and Sustainable Energy Reviews, 189, 113967. https://doi.org/10.1016/j.rser.2023.113967.

[15] Alsuwian, T., Ansari, S., Zainuri, M. A. A. M., Ayob, A., Hussain, A., Lipu, M. H., ... & Hindi, A. T. (2024). A Review of Expert Hybrid and Co-Estimation Techniques for SOH and RUL Estimation in Battery Management System with Electric Vehicle Application. Expert Systems with Applications, 123123. https://doi.org/10.1016/j.eswa.2023.123123.

[16] Li, F., Min, Y., Zhang, Y., Zhang, Y., Zuo, H., & Bai, F. (2024). State-of-health estimation method for fast-charging lithium-ion batteries based on stacking ensemble sparse Gaussian process regression. Reliability Engineering & System Safety, 242, 109787. https://doi.org/10.1016/j.ress.2023.109787.

[17] Villalobos, J., Martell, F., Sanchez, I. Y.: Comparison of model reference control schemes for motor speed control under variable load torque. Studies in Informatics and Control, Vol. 32, No. 2, 63–72. (2023), https://doi.org/10.24846/v32i2y202306.

[18] Buchanan, S., & Crawford, C. (2024). Probabilistic lithium-ion battery state-of-health prediction using convolutional neural networks and Gaussian process regression. Journal of Energy Storage, 76, 109799. https://doi.org/10.1016/j.est.2023.109799.

[19] Guo, Y., Yu, X., Wang, Y., & Huang, K. (2024). Health prognostics of lithium-ion batteries based on universal voltage range features mining and adaptive multi-Gaussian process regression with Harris Hawks optimization algorithm. Reliability Engineering & System Safety, 244, 109913. https://doi.org/10.1016/j.ress.2023.109913.

[20] Zhou, Y., Wang, S., Xie, Y., Zhu, T., & Fernandez, C. (2024). An improved particle swarm optimization-least squares support vector machine-unscented Kalman filtering algorithm on SOC estimation of lithium-ion battery. International journal of green energy, 21(2), 376-386. https://doi.org/10.1080/15435075.2023.2196328.

[21] Xiao, Y., Wen, J., Yao, L., Zheng, J., Fang, Z., & Shen, Y. (2023). A comprehensive review of the lithium-ion battery state of health prognosis methods combining aging mechanism analysis. Journal of Energy Storage, 65, 107347. https://doi.org/10.1016/j.est.2023.107347.

[22] Lu, J., Xiong, R., Tian, J., Wang, C., & Sun, F. (2023). Deep learning to estimate lithium-ion battery state of health without additional degradation experiments. Nature Communications, 14(1), 2760. https://doi.org/10.1038/s41467-023-38458-w.

[23] Kim, Y. T. & Han, S. Y. (2023) Cooling Channel Designs of a Prismatic Battery Pack for Electric Vehicle Using the Deep Q-Network Algorithm. Applied Thermal Engineering. 219(C), 119610. https://doi.org/10.1016/j.applthermaleng.2022.119610.

[24] Athilakshmi, R., Jacob, S. G., Rajavel, R.: Automatic Detection of Biomarker Genes through Deep Learning Techniques: A Research Perspective. Studies in Informatics and Control, Vol. 32, No. 2, 51–61. (2023),https://doi.org/10.24846/v32i2y202305.

[25] Avram, F. T., Țarcă, D. I., Noje, D., Vesselenyi, T., & Țarcă, R. (2025). Surface Roughness Determination With the Help of Artificial Neural Networks as Enabler of Metal Machining Process Controlling System. International Journal of Computers Communications & Control, 20(2). https://doi.org/10.15837/ijccc.2025.2.7028.

[26] Fayti, M., Mjahed, M., Ayad, H., Kari, A. E.: Recent Metaheuristic-Based optimization for system modeling and PID controllers tuning. Studies in Informatics and Control, Vol. 32, No. 1, 57–67. (2023), https://doi.org/10.24846/v32i1y202306.

[27] Wang T., Zhang X., Zeng Q. L., Jiang S. B., Zhang Y. N. (2022). Modelling and Simulation on Cavity Cold Plate for Li-Ion Battery Thermal Management. Int. Journal of Simulation Modelling, Vol. 21, No. 1, p. 65-76. https://doi.org/10.2507/IJSIMM21-1-588.

[28] Liu, J., Qu, Q., Yang, H., Zhang, J., & Liu, Z. (2024). Deep learning-based intelligent fault diagnosis for power distribution networks. International Journal of Computers Communications & Control, 19(4). https://doi.org/10.15837/ijccc.2024.4.6607.

[29] Qiang, X., Tang, Y., Wu, L., & Lyu, Z. (2024). Li-Ion Battery State of Health Estimation Using Hybrid Decision Tree Model Optimized by Bayesian Optimization. Energy Technology, 12(3), 2301065. https://doi.org/10.1002/ente.202301065.

[30] Lyu, Z., Tang, Y., Wu, Z., Wu, L., & Qiang, X. (2024). Online state of health estimation for Li-ion batteries in EVs through a data-fusion-model method. Journal of Energy Storage, 100, 113588. https://doi.org/10.1016/j.est.2024.113588.

[31] Çarkıt, T., & Alçı, M. (2024). Investigation of V oc and SoH on Li-ion batteries with an electrical equivalent circuit model using optimization algorithms. Electrical Engineering, 106(2), 1781-1792. https://doi.org/10.1007/s00202-021-01484-2.

[32] Qiang, X., Liu, W., Lyu, Z., Ruan, H., & Li, X. (2024). A data-fusion-model method for state of health estimation of Li-ion battery packs based on partial charging curve. Green Energy and Intelligent Transportation, 3(5), 100169. https://doi.org/10.1016/j.geits.2024.100169.

[33] Saqli, K., Bouchareb, H., M'sirdi, N. K., & Bentaie, M. O. (2023). Lithium-ion battery electro-thermal modelling and internal states co-estimation for electric vehicles. Journal of Energy Storage, 63, 107072. https://doi.org/10.1016/j.est.2023.107072.

[34] Tian, J., Liu, X., Li, S., Wei, Z., Zhang, X., Xiao, G., & Wang, P. (2023). Lithium-ion battery health estimation with real-world data for electric vehicles. Energy, 270, 126855. https://doi.org/10.1016/j.energy.2023.126855.

[35] Xiong, R., Sun, Y., Wang, C., Tian, J., Chen, X., Li, H., & Zhang, Q. (2023). A data-driven method for extracting aging features to accurately predict the battery health. Energy Storage Materials, 57, 460-470. https://doi.org/10.1016/j.ensm.2023.02.034.

Cite this paper as:

Li, W.; Cheng, C.; Wang, Q.; Ma, C.-S. (2025). Prediction of power battery health state based on data-driven Gaussian process regression algorithm, *International Journal of Computers Communications & Control*, 20(3), 7063, 2025.
https://doi.org/10.15837/ijccc.2025.3.7063