Energy is the foundation of all science and engineering technology, without which the human world will be difficult to operate [1]. Since the end of the 18th century and the beginning of the 19th century, the rise of industrialization has led to the demand for a large amount of energy to drive machinery and equipment, production lines and vehicles. Therefore, fossil energy began to be used in large quantities. The increasing use of other fossil fuels, such as coal used in the early days and oil and natural gas, has become a key factor supporting the modern society and economic system. However, the exhaustion of these fossil energy sources and a series of environmental problems, coupled with the acceleration of traditional energy demand, forced planners and decision makers to look for alternative energy sources [2,3]. Many renewable energy technologies, including solar energy, wind energy, tidal energy, biomass energy and hydropower, have been widely developed to reduce their dependence on fossil fuels.

However, most renewable energy sources, such as solar energy and wind energy, are intermittent in nature and rely on natural phenomena to generate electricity, so they must be stored and used on demand [4-5]. As an energy storage technology, rechargeable batteries have been widely used in aerospace, portable electronic equipment, electric vehicles and so on [6,7]. Among the existing rechargeable battery technologies, lithium battery is considered as the best energy storage mode because of its higher energy density, smaller volume, longer life and larger capacity. At the same time, with the demand of urban sustainable development and the application of a new generation of information technology, smart cities will become people's future model [8,9], and electric vehicles can well solve the problems of energy-saving development and environmental pollution in smart cities, and the development of new energy electric vehicles has become a global consensus [10,12]. As the best choice for new energy electric vehicles, rechargeable lithium-ion batteries have been widely used. However, we know that no matter what kind of application, lithium batteries will deteriorate with time, which is manifested in the loss of battery capacity and the increase of impedance [13]. Therefore, while promoting the development of electric vehicles, rechargeable lithium-ion batteries inevitably produce a series of problems, such as reduced battery life and insufficient power, and with the passage of time, the aging of lithium-ion batteries may cause safety accidents as Figure 1.

The degradation rate of the battery is affected by dynamic operating conditions, including different charging and discharging rates, different voltage operating limits and temperature fluctuations [14]. If we can predict the battery life before aging, it will bring new opportunities for battery production, use and optimization [15]. For example, manufacturers can speed up the cell development cycle, quickly verify new manufacturing processes, and classify/grade new cells according to life expectancy. Similarly, end users can estimate their battery life [16-18]. In addition, battery forecasting is crucial for expanding the recycling sector, enabling facilities to decide whether batteries should be recycled as scrap metal or used for less demanding "second life" applications. In short, accurate prediction of the current and future state of the battery will bring great opportunities for the manufacture, use and optimization of the battery [20, 21].

At present, the models used in battery life estimation can be mainly divided into the following three types: equivalent circuit model (ECM)[22-23], electrochemical model [24-26] or data-driven model [27-32]. The electrochemical model is similar to the chemical process in the battery during its operation, which requires detailed battery specification information and complex electrochemical knowledge. The equivalent circuit model uses circuit elements with empirical nonlinear parameters. However, the simple equivalent circuit can't completely simulate the chemical reaction inside the battery, and the complicated model has too much calculation, and the accuracy and robustness of the above two models are limited for the assumption of battery behavior. Therefore, these two models are not a good and feasible solution. On the contrary, data-driven method has a series of advantages, such as no need to understand the complex chemical reactions inside the battery, analysis of various battery degradation principles, no complicated process of establishing circuits, etc. So far, many studies have used machine learning tools to analyze battery life prediction and estimation.

With the development of research in recent years, it is found that noise in battery data set is inevitable, which mainly comes from environmental interference during charging and discharging, such as temperature change and humidity fluctuation. In addition, most public data sets are measured under experimental conditions, but the actual battery charge and discharge is incomplete. Therefore, it is closer to real life to study the battery data set with noise, and its robustness can be guaranteed when the model is extended to practical application.

The problem of linear parameter estimation appears in a wide range of scientific disciplines such as signal processing [33-34]. It starts with a linear (in-parameter) model, which represents process variables that can be measured or inferred from other measurements or calculated by nonlinear transformation; All variables are affected by measurement noise; Contains parameters that represent the basic relationship of process variables. As shown in [35] and [36], the total least square method is the best choice for parameter estimation when all the variables of interest have parameter linear relations and all the measured values are polluted by noise.

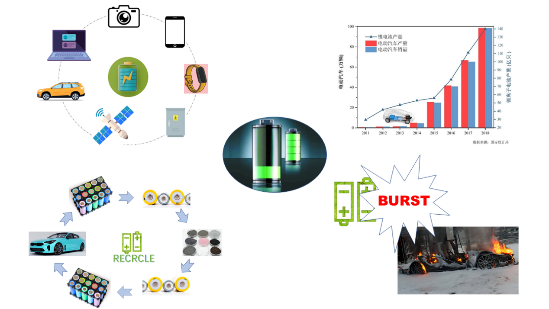


Figure 1: Application of Lithium Battery and Significance of Life Prediction

However, in the actual situation, the data sets of battery information provided by battery manufacturers come from different sources, so the errors caused by temperature, human interference and sensors are very different. At this time, it is impossible to simply assume that the noise of the data sets obeys the same distribution. At this time, directly using TLS/OLS can't establish a battery life prediction model. Therefore, this paper improves the linear model to calculate the battery life. After weighting the battery samples with different noise distributions, TLS/OLS is used to predict the battery life. Through cyclic iteration, the standard deviation of noise distribution can be accurately calculated, and a prediction model suitable for different noise distributions can be established to predict the battery life. The prediction results show that our method is better than the traditional TLS/OLS method.

Considering a set of features **x** = [*x*1 *x*2 ⋅⋅⋅ *xN*]*T*, where *N* denotes the total number of features and the superscript *T* represents the transpose of a vector or matrix, our objective is to learn a mapping from **x** to the battery lifetime *y* [23],[29]:

1 , (1)

where *gm*(**x**) is the *m*-th basis function, *wm* denotes the *m*-th model coefficient, *εi* stands for the modeling error following a zero-mean Gaussian distribution *N*(0, *σ*y2), and *M* represents the total number of basis functions.

Given a set of data samples {(**x***k*, *y*k); *k* =1, ⋅⋅⋅, *K*}, the model coefficients {*wm*; *m* =1, ⋅⋅⋅, *M*} may be determined by minimizing the total squared error [14]:

2 , (2)

where ||•||2 denotes the L2 norm of a vector and

3 (3)

4 (4)

5 . (5)

The aforementioned approach is referred to as OLS regression in the literature and it aims to find the maximum-likelihood solution **w** for the unknown model coefficients [14].

In practice, each basis function *gm*(**x**) may be noisy, as the elements in feature vector **x** is obtained through the physical measurements of batteries (e.g., voltage, current, temperature, etc.), which are usually associated with measurement errors. In this case, Eq. (1) should be re-written as:

6 , (6)

where *εg*,*m* represents the measurement error associated with the *m*-th basis function *gm*(**x**). In this paper, we assume that *εg*,*m* follows a zero-mean Gaussian distribution *N*(0, *σg*,*m*2).

In order to compute the solution of the unknown model coefficientsin (6), we formulate an optimization problem of **w** based on maximum-likelihood estimation. To achieve this goal, we make two further assumptions. First, both *εy* and {*εg*,*m*; *m* =1, ⋅⋅⋅, *M*} can be normalized to standard Gaussian distribution *N*(0, 1) by appropriately scaling *y* and {*gm*(**x**); *m* =1, ⋅⋅⋅, *M*} respectively. Second, *εy* and {*εg*,*m*; *m* =1, ⋅⋅⋅, *M*} are statistically independent.

With these two assumptions, it is straightforward to show that the likelihood of observing a sample (**x***k*, *yk*) is equal to:

7, (7)

where *εy*,*k* and *εg*,*m*,*k* denote the *k*-th samples for *εy* and *εg*,*m* respectively. Furthermore, by assuming that all samples in the dataset {(**x***k*, *yk*); *k* =1, ⋅⋅⋅, *K*} are statistically independent, the likelihood for observing these *K* samples is equal to:

8 . (8)

Hence, the maximum-likelihood solution **w** can be found by solving the following optimization problem:

9 , (9)

where ||•||F denotes the Frobenius norm of a matrix and

10 (10)

11 . (11)

Note that minimizing the cost function in (9) is equivalent to maximizing the likelihood in (8). Such an approach is referred to as the TLS regression in the literature [12].

According to the sample relationship of data set, we can know that our goal is to minimize the error when the samples obey different noises：

6 , (6)

To solve this problem, we give each sample a different weight:

6 , (6)

Assuming that the noise obeys a Gaussian distribution with zero mean and different variances: , then obeys a Gaussian distribution . The likelihood function is:

6 , (6)

Maximizing the likelihood function is equivalent to minimizing the objective function , and we can get by comparison.

EM algorithm can solve the problem with hidden variables well. In this paper, we assume that the standard deviation of noise obeys the distribution as hidden variables, and improve the traditional OLS/TLS algorithm as Figure 2. The specific steps are as follows:

1. Initialize model coefficients **w,b**

2. Predict the battery life according to the model coefficient, and calculate the error between the predicted value and the real value to update the noise standard deviation of the three data sets.

6 , (6)

3. The samples are weighted according to the formula, and a new round of model coefficients are obtained by using TLS/OLS.

4. Repeat steps 2 and 3 until convergence.

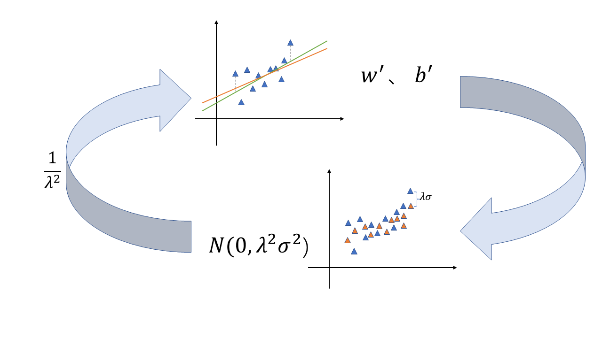


Figure 2: Algorithm improvement process

The dataset, referred to as “Dataset”, was generated by Severson et al. [1], which consists of 124 commercial LiFePO4/graphite batteries cycled to EOL under fast-charging conditions. During the cycling test of these batteries, several important metrics, such as voltage, current, discharge capacity, temperature, impedance, charge time, etc., are measured in real time. Based on the availability of measurement data and domain expertise, three features in total are extracted for regression modeling, which are indexed by *x*1, *x*2 and *x*3. Note that all these three features are available for Dataset . The feature names, physical meanings and their availabilities are summarized in Table 1. To reduce the nonlinearity of our modeling task, we take the logarithm for both the battery lifetime and the first feature *x*1, following the common practice in the literature [1]. With these nonlinear transformations, we adopted a linear model template

28 (28)

for Dataset 1. To improve numerical stability, we normalize the predicted outcome log(*y*) and all features {log(*x*1), *x*2, *x*3, *x*4, *x*5} so that they have zero mean and unit variance over the training dataset.

Table 1: Features for battery lifetime modeling

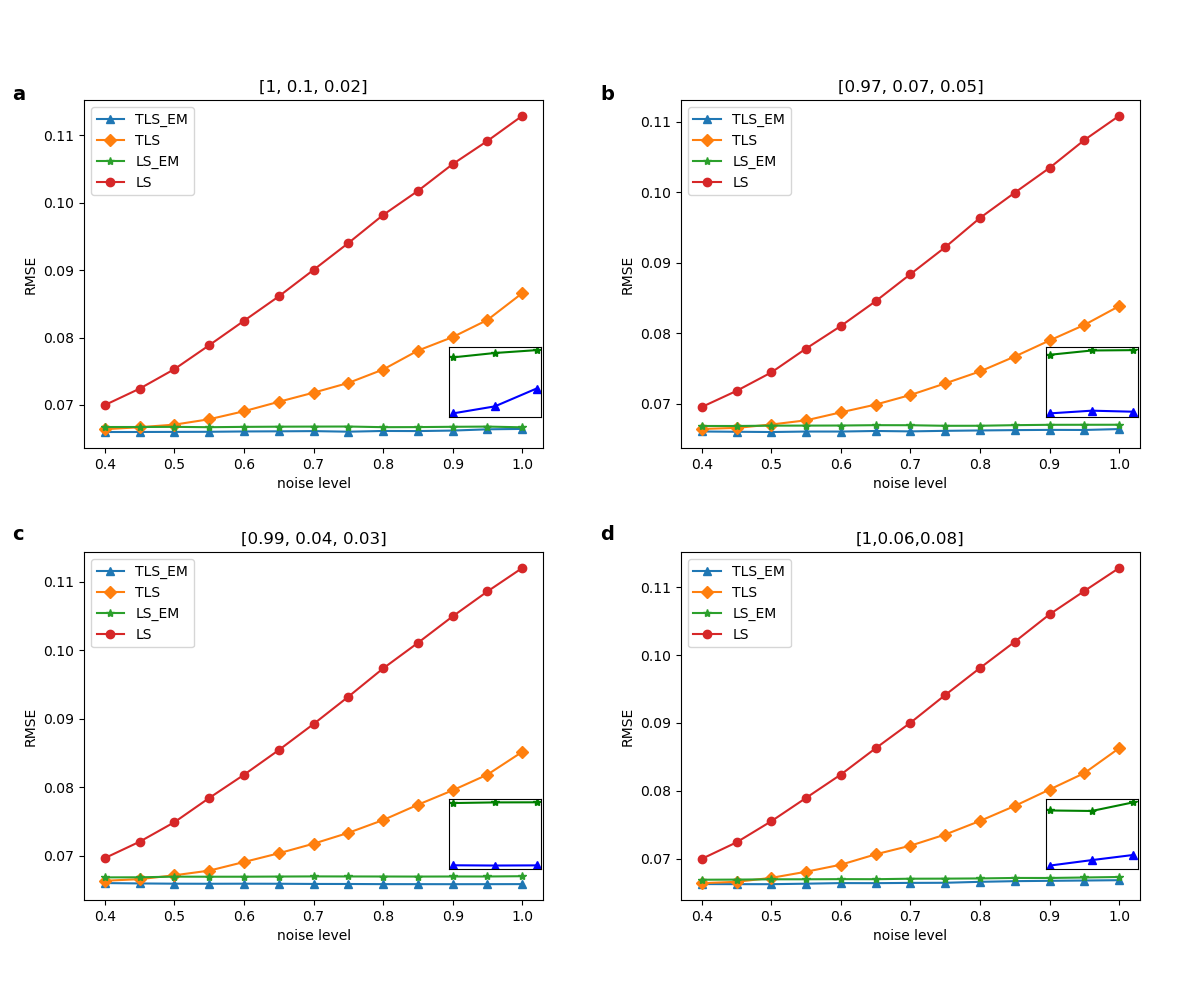
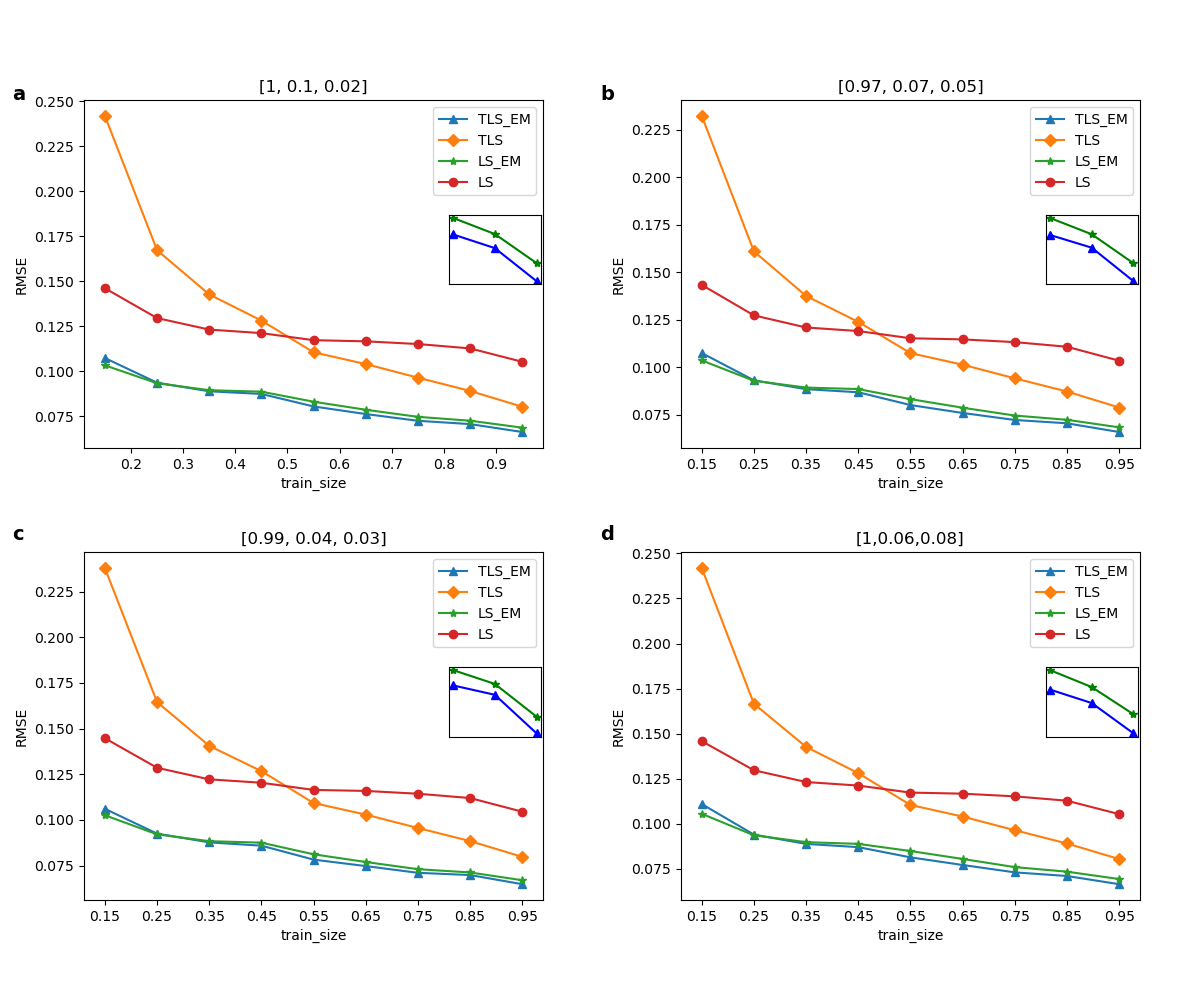
| Feature Name | Description |
| --- | --- |
| *x*1 | Variance of the difference in the discharge capacity curves as a function of voltage between the 10-th and 100-th cycles |
| *x*2 | Slope of the capacity fade curve fitted by a linear function |
| *x*3 | Discharge capacity of the 2-nd cycle |
| *x*4 | Integral of temperature over time from the 2-nd cycle to the 100-th cycle |
| *x*5 | Difference of internal resistances between the 2-nd cycle and the 100-th cycle |

The data set consists of three small data sets from three different sources. We divide each small data set into training set and test set according to the ratio of 9:1, and then merge the training set and test set to form the final training set and test set. The experiment was repeated for 1000 times, and the training and test data sets were generated independently and randomly for each run. Report the median of 1000 RMSE values for each method, so that the error measurement will not be strongly biased due to random fluctuations.

Figure 3 shows the RMSE of four methods TLS, OLS, improved TLS(TLS\_EM) and improved OLS(OLS\_EM) under different noise levels in the experiment. In order to simulate the real situation, the noise levels we added to the three small data sets are different, but no matter which noise ratio mode is A, B, C and D, we can see that with the increase of noise levels, OLS\_EM and TLS\_EM have obvious advantages.

Figure 4 shows the RMSE of four methods TLS, OLS, improved TLS(TLS\_EM) and improved OLS(OLS\_EM) with different training sets in the experiment. In order to simulate the real situation, the noise levels we added to the three small data sets are different, but no matter which noise proportion mode is A, B, C and D, it can be seen that, firstly, with the increase of the proportion of training sets, the four methods Secondly, the effects of OLS\_EM and TLS\_EM are always better than OLS and TLS, and it can be observed that TLS\_EM is better than OLS\_EM in the enlarged picture. The results show the effectiveness of our proposed method. Thirdly, TLS\_EM is worse than OLS\_EM only when the proportion of training set is very small (15%), which shows that TLS\_EM is more accurate than OLS\_EM in most cases.

In this paper, the problem of battery life prediction when the noise of data sets obeys different distributions is considered, and the improved OLS and TLS algorithms combined with EM idea are used to predict it. The prediction results show that the improved method is effective.



[1]Renewable energy technologies in Pakistan: Prospects and challenges

[2] Mirza UK, Ahmad N, Majeed T, Harijan K. Wind energy development in Pakistan. Renewable and Sustainable Energy Reviews 2007;11(9): 2179–90.

[3] Tiwari, GN, Ghosal, MK. Renewable Energy Resources: Basic Principles and Applications. Alpha Science Int’l Ltd., 2005. ISBN 1-84265-125-0

[4] B. Obama, Science 2017, DOI: 10.1126/science.aam6284.

[5]Porous Carbon Composites for Next Generation Rechargeable Lithium Batteries

[6] P. Poizot, S. Laruelle, S. Grugeon, L. Dupont, J. M. Tarascon, Nature 2000, 407, 496.

[7] B. Dunn, H. Kamath, J. M. Tarascon, Science 2011, 334, 928.

Recent progress of magnetic field application in lithium-based batteries

[8] Raza MQ, Khosravi A. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. Renew Sustain Energy Rev 2015;50:1352e72.

[9] Xu J, Zhang R. CoMP meets smart grid: a new communication and energy cooperation paradigm. IEEE Trans Veh Technol 2013;64(6):2476e88.

[10]An energy matching method for battery electric vehicle and hydrogen fuel cell vehicle based on source energy consumption rate

[11] Wang G, Xu Z, Wen F, et al Traffic-constrained multiobjective planning of electric-vehicle charging stations.IEEE Trans Power Deliv 2013;28(4):2363e72.

[12] Hu J, Zheng L, Jia M, et al Optimization and model validation of operation control strategies for a novel dual-motor coupling-propulsion pure electric vehicle. Energies 2018;11.

[13]Predicting the State of Charge and Health of Batteries using Data-Driven Machine Learning

[14]MACHINE LEARNING PIPELINE FOR BATTERY STATE OF HEALTH ESTIMATION

15Data-driven prediction of battery cycle life before capacity degradation

16. Peterson, S. B., Apt, J. & Whitacre, J. F. Lithium-ion battery cell degradation resulting from realistic vehicle and vehicle-to-grid utilization. J. Power Sources 195, 2385–2392 (2010).

17. Ramadesigan, V. et al Modeling and simulation of lithium-ion batteries from a systems engineering perspective. J. Electrochem. Soc. 159, R31–R45 (2012).

18. Waag, W., Fleischer, C. & Sauer, D. U. Critical review of the methods for monitoring of lithium-ion batteries in electric and hybrid vehicles. J. Power Sources 258, 321–339 (2014)

19Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning

20. Severson, K. A. et al Data-driven prediction of battery cycle life before capacity degradation. Nat. Energy 4, 383–391 (2019).

21. Nuhic, A., Terzimehic, T., Soczka-Guth, T., Buchholz, M. & Dietmayer K. Health diagnosis and remaining useful life prognostics of lithium-ion batteries using datadriven methods. J. Power Sources 239, 680–688 (2013).

[22] Tianheng Feng, Lin Yang, Xiaowei Zhao, Huidong Zhang, and Jiaxi Qiang. Online identification of lithium-ion battery parameters based on an improved equivalent-circuit model and its implementation on battery state-of-power prediction. Journal of Power Sources, 281:192–203, 2015.

[23] D Andre, M Meiler, K Steiner, H Walz, T Soczka-Guth, and DU Sauer. Characterization of high-power lithium-ion batteries by electrochemical impedance spectroscopy. ii: Modelling. Journal of Power Sources, 196(12):5349–5356, 2011.

[24] Matthew J Daigle and Chetan Shrikant Kulkarni. Electrochemistry-based battery modeling for prognostics. 2013.

[25] Brian Bole, Chetan S Kulkarni, and Matthew Daigle. Adaptation of an electrochemistry-based li-ion battery model to account for deterioration observed under randomized use. Technical report, SGT, Inc. Moffett Field United States, 2014.

[26] Githin K Prasad and Christopher D Rahn. Model based identification of aging parameters in lithium ion batteries.

Journal of power sources, 232:79–85, 2013.

[27] Kristen A Severson, Peter M Attia, Norman Jin, Nicholas Perkins, Benben Jiang, Zi Yang, Michael H Chen, Muratahan Aykol, Patrick K Herring, Dimitrios Fraggedakis, et al Data-driven prediction of battery cycle life before capacity degradation. Nature Energy, 4(5):383, 2019.

[28] Bhaskar Saha, Kai Goebel, Scott Poll, and Jon Christophersen. Prognostics methods for battery health monitoring using a bayesian framework. IEEE Transactions on instrumentation and measurement, 58(2):291–296, 2008.

[29] Kai Goebel, Bhaskar Saha, Abhinav Saxena, Jose R Celaya, and Jon P Christophersen. Prognostics in battery health management. IEEE instrumentation & measurement magazine, 11(4):33–40, 2008.

[30] Xiaosong Hu, Jiuchun Jiang, Dongpu Cao, and Bo Egardt. Battery health prognosis for electric vehicles using sample entropy and sparse bayesian predictive modeling. IEEE Transactions on Industrial Electronics, 63(4):2645– 2656, 2015.

[31] Verena Klass, Mårten Behm, and Göran Lindbergh. A support vector machine-based state-of-health estimation method for lithium-ion batteries under electric vehicle operation. Journal of Power Sources, 270:262–272, 2014

[32] Peter M Attia, Aditya Grover, Norman Jin, Kristen A Severson, Todor M Markov, Yang-Hung Liao, Michael H Chen, Bryan Cheong, Nicholas Perkins, Zi Yang, et al Closed-loop optimization of fast-charging protocols for batteries with machine learning. Nature, 578(7795):397–402, 2020.

[33]Detection of Abrupt Changes of Total Least Squares Models and Application in Fault Detection

[34] S. Van Huffel, “Tls applications in biomedical signal processing,” in Recent Advances in Total Least Squares Techniques and Error-in-Variables Modeling, S. Van Huffel, Ed. Philadelphia, PA: SIAM, 1997

[35] S. Van Huffel and J. Vandewalle, Frontiers in Applied Mathematics: The Total Least Squares Problem—Computational Aspects and Analysis. Philadelphia, PA: SIAM, 1991.

[36] S. Van Huffel, Ed., Recent Advances in Total Least Squares Techniques and Errors-In-Variables Modeling. Philadelphia, PA: SIAM, 1997.