提纲1提纲2主要区别是提纲1使用了一张电池背景图分析，提纲2 则是将电池的应用和电池寿命预测分为两张图

**提纲1：**

1. 引入带噪声的电池寿命预测，图1分析（新能源崛起，锂电池应用广泛，电池寿命预测的意义）
2. LS、TLS在降噪方面的广泛运用，提出存在不同数据集的噪声分布不同这个问题
3. 总结本文贡献，算法大致结构，图2分析
4. 所使用的特征分析（三个特征数据图，分别分析它们的趋势等）
5. 噪声水平增大，图5分析（详细讲述算法改进，算法优势）
6. 训练集比例增大，图6分析（算法在训练集较少情况下仍然有效）
7. 总结全文

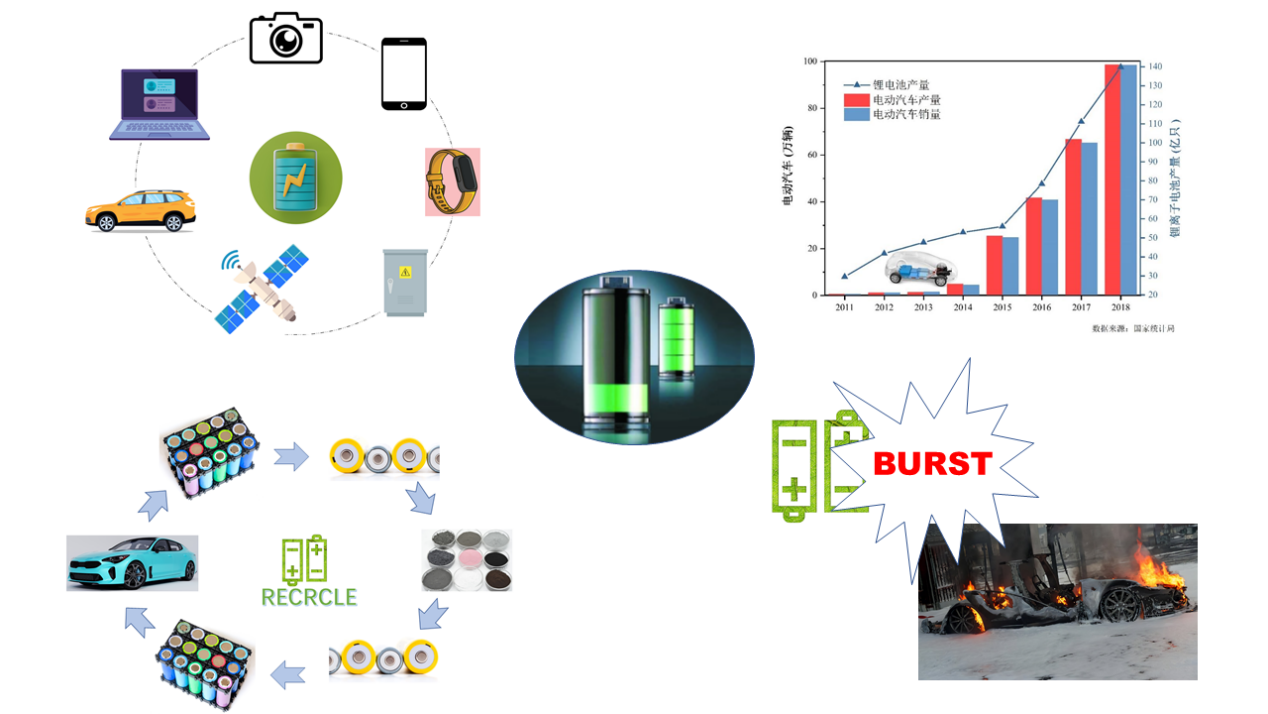


图1

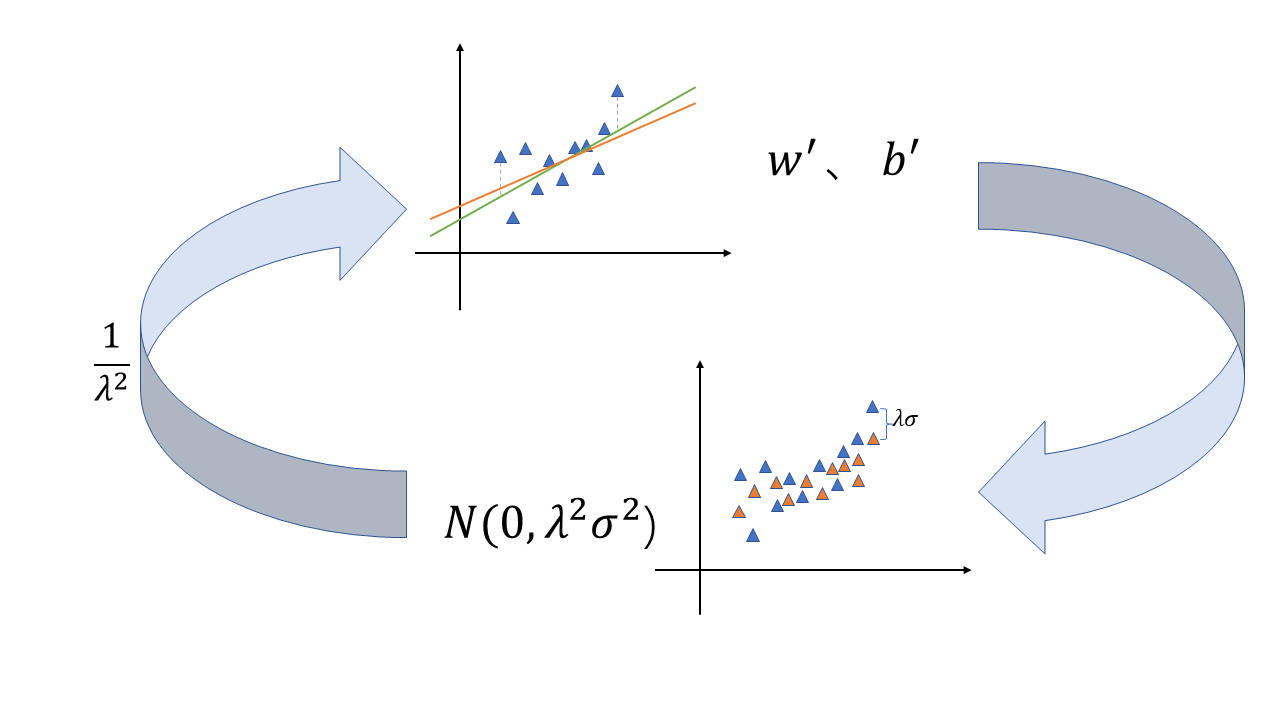


图2

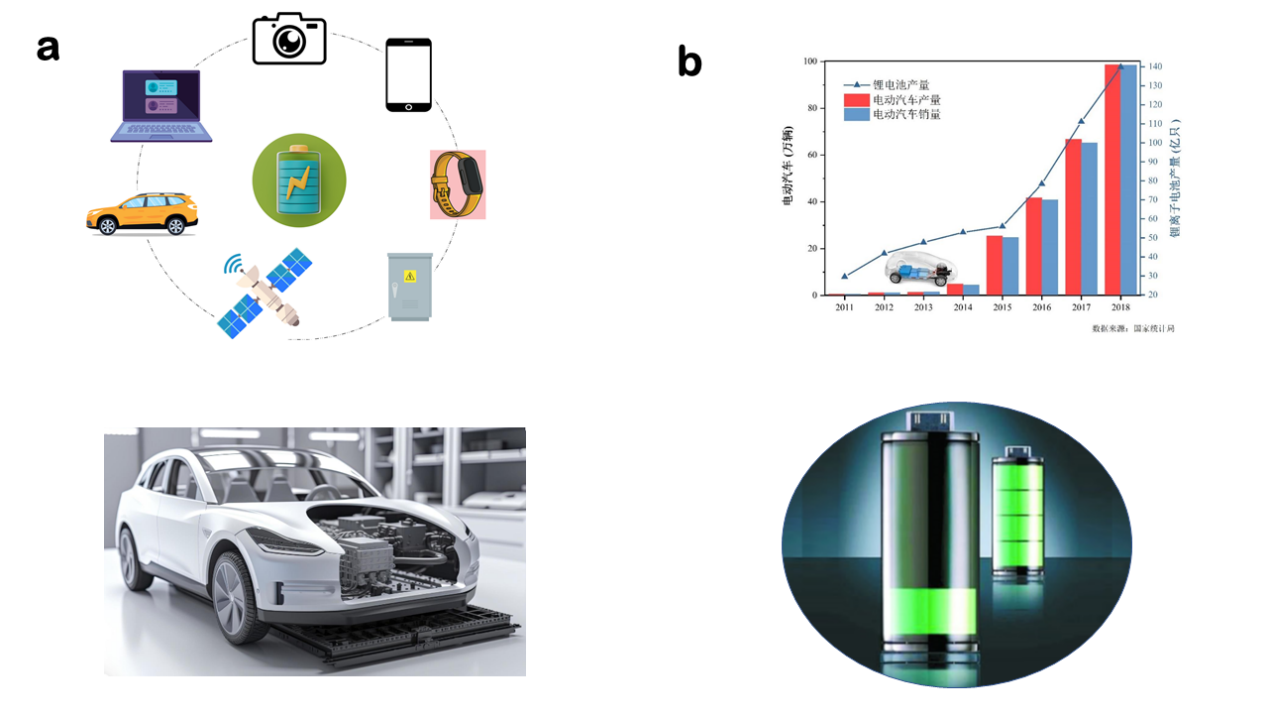


图3

图4

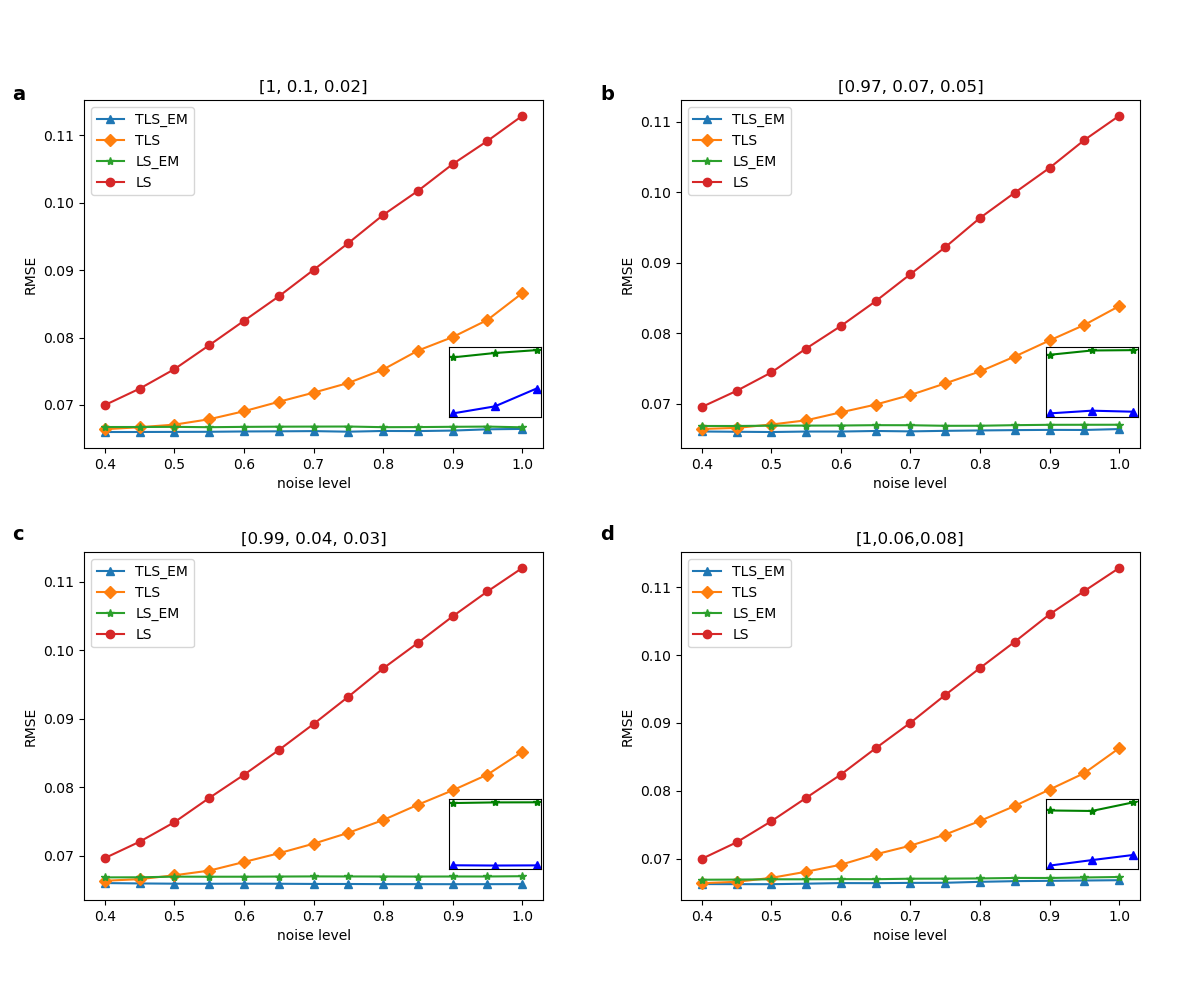


图5

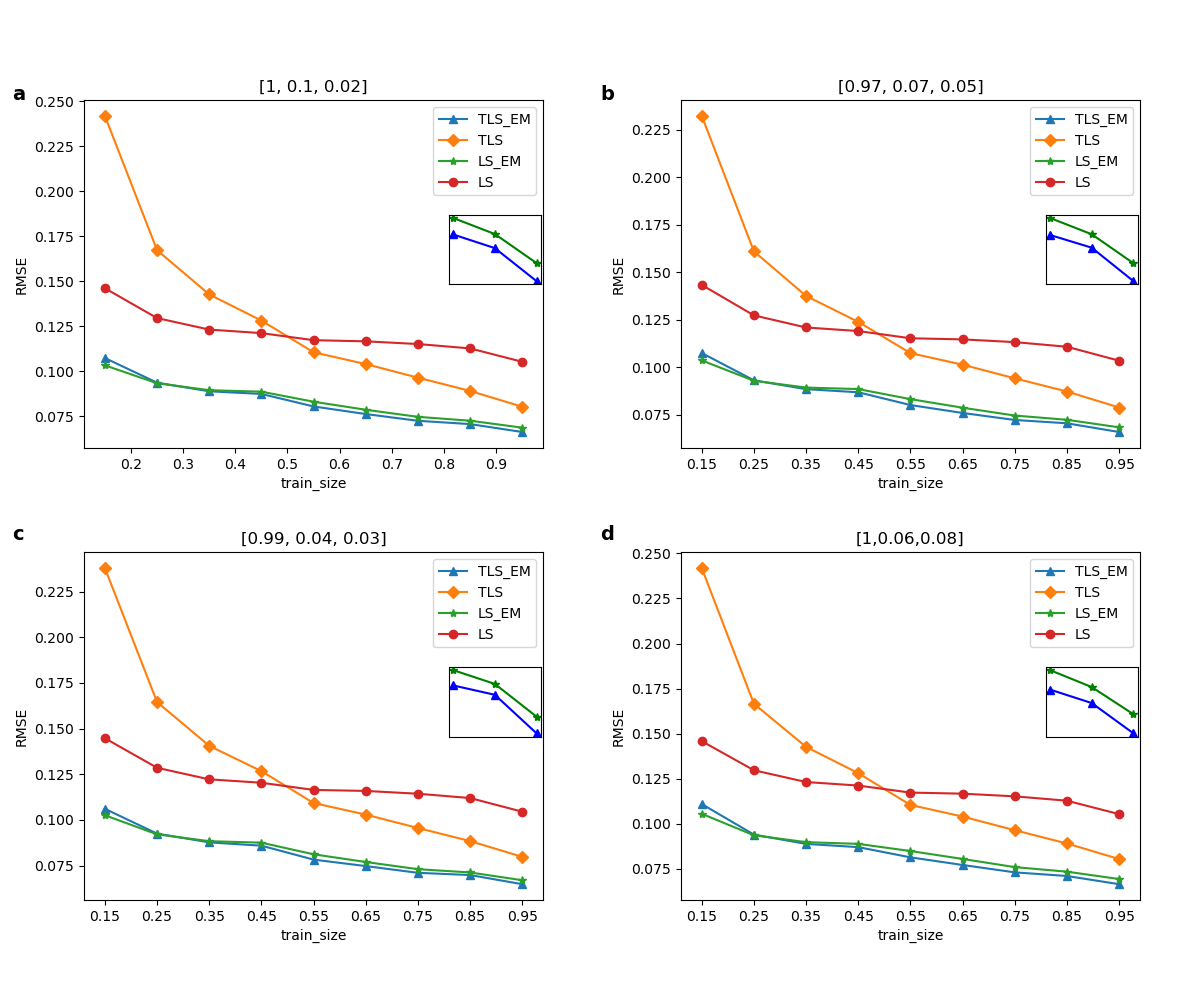


图6

化石能源的问题-》新能源崛起-》锂电池能源的广泛应用-》锂电池的问题-》寿命预测的必要性、

能源是所有科学和工程技术中的基础，没有能源人类世界将难以运转[1]。工业化的兴起导致了对大量能源的需求，但化石能源的枯竭以及随之产生的一系列环境问题，加上传统能源需求的加速迫使规划人员和决策者寻找替代能源[2,3]许多可再生能源技术已经得到了广泛的发展。然而，大多数可再生能源，如太阳能和风能，本质上是间歇性的，依靠自然现象来发电，必须储存和按需使用[4-5]。可充电电池作为一种储能技术，已经被人们广泛应用于航空航天、便携式电子设备、电动汽车等领域【6,7】。同时智慧城市将成为人们未来的模式[8,9]，而电动汽车能够很好的解决智慧城市的节能型发展和环境污染问题，发展新能源电动汽车已成为全球共识[10,12]。可充电锂离子电池由于拥有更高的能量密度，更小的体积、更长的寿命、更大的容量等优点，作为新能源电动汽车的最佳选择，已经得到了广泛应用。

但锂电池会随着时间的推移而退化，具体表现为电池容量的丧失和阻抗的增加[13]。因此可充电锂离子电池在满足促进电动汽车发展的同时也比不可避免的产生了一系列的问题，比如电动汽车的续航减少、动力不足等，且随时间推移，锂离子电池的老化可能会造成安全事故，如图1所示。

电池的降解速率受动态运行条件的影响，包括不同的充放电速率、不同的电压运行极限和温度波动[14]。如果我们能在电池老化之前对其寿命进行预测将为电池生产、使用和优化带来新的机遇[15]。例如，制造商可以加快细胞开发周期，对新的制造工艺进行快速验证，并根据预期寿命对新细胞进行分类/分级。同样，终端用户可以估计他们的电池寿命[16-18]。此外，电池预测对扩大回收部门至关重要，使设施能够决定电池是作为废金属回收还是用于要求较低的“第二生命”应用。总之，对电池当前和未来状态的准确预测将为电池的制造、使用和优化带来巨大的机会【20、21】。

目前的电池寿命估计所采用的模型主要可以分为以下三种，等效电路模型(ecm)[22-23]、电化学模型[24-26]或数据驱动模型[27-32]。电化学模型和等效电路模型的准确性和鲁棒性有限。因此这两种模型并不是一个很好的可行解决方案。相反，数据驱动的方法有着一系列的优势，比如不需要了解电池内部的复杂化学反应，分析各种电池降解原理，没有复杂的建立电路的过程等迄今为止，许多研究都使用机器学习工具来分析电池寿命预测估计。

随着近几年的研究展开，发现了电池数据集中带有噪声是不可避免的，这主要源于充放电过程中的环境干扰，如温度变化、湿度波动的影响。此外，大多公用数据集都是在实验条件下完全充放电测量得到，但实际情况下的电池充放电是不完全的，因此，研究带有噪声的电池数据集才更贴近实际生活，将模型推广到实际应用时才能保证其健壮性。

线性参数估计问题出现在信号处理等广泛的科学学科中[33-34]。它从线性(参数内)模型开始，其中表示可以测量或可以从其他测量推断或可以通过非线性变换计算的过程变量;所有变量都受到测量噪声的影响;包含表征流程变量的基础关系的参数。如[35]和[36]所示，在所有感兴趣的变量都具有参数线性关系且所有测量值都受到噪声污染的情况下，总最小二乘法是参数估计的最佳选择。

但实际情况中电池厂商提供的电池信息数据集来源不同，故由于温度、人为干扰以及传感器等造成的误差大不相同，此时便不能简单的假设数据集的噪声服从同一个分布。此时直接使用TLS/OLS并不能很好的建立电池寿命预测模型，故本文在建立线性模型计算电池的寿命时进行改进，对带有不同噪声分布的电池样本进行加权之后，使用TLS/OLS进行预测，经循环迭代能够准确的计算出噪声分布的标准差的同时建立适应不同噪声分布的预测模型对电池寿命进行预测，预测结果显示我们的方法比传统的TLS/OLS方法更好。

Energy is the foundation of all science and engineering technology, without which the human world will be difficult to operate [1]. The rise of industrialization has led to a large demand for energy, but the exhaustion of fossil energy and a series of environmental problems, coupled with the acceleration of traditional energy demand, have forced planners and decision makers to find alternative energy sources [2,3]. Many renewable energy technologies have been widely developed. 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 other fields [6,7]. At the same time, smart cities will become the future model of people [8,9], and electric vehicles can solve the energy-saving development and environmental pollution problems of smart cities well, and the development of new energy electric vehicles has become a global consensus [10,12]. Rechargeable lithium-ion batteries have been widely used as the best choice for new energy electric vehicles because of their higher energy density, smaller volume, longer life and larger capacity.

However, 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. 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 a word, 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 accuracy and robustness of electrochemical model and equivalent circuit model are limited. Therefore, these two models are not a good and feasible solution. On the contrary, the 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.

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.

When dataset samples obey noise of different distributions,, 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. 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.

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 |

数据集由三个不同来源的三个小数据集组成，我们对每个小数据集按照9:1比例划分训练集和测试集，再将训练集和测试集分别合并组成最终的训练集和测试集。实验重复1000次，每次运行均独立随机生成训练和测试数据集。为每种方法报告1000个RMSE值的中位数，以便误差度量不会因随机波动而产生强烈偏差。

图3显示了实验中不同噪声水平下四种方法TLS、OLS、改进的TLS(TLS\_EM)和改进的OLS(OLS\_EM)的RMSE，为了模拟现实情况，我们为三个小数据集添加的噪声水平不一致，但无论是a,b,c,d哪种噪声比例模式，都可以看出随着噪声水平的增加，OLS\_EM和TLS\_EM 优势明显，且在放大的图片里可以观察到TLS\_EM比OLS\_EM效果更好，结果说明我们提出方法的有效性。

图4显示了实验中不同训练集大小四种方法TLS、OLS、改进的TLS(TLS\_EM)和改进的OLS(OLS\_EM)的RMSE，为了模拟现实情况，我们为三个小数据集添加的噪声水平不一致，但无论是a,b,c,d哪种噪声比例模式，都可以看出，第一，随着训练集比例的增加，四种方法的RMSE都在降低，因为更多的训练样本是可用的，并且因此，更多的信息被结合用于模型训练。第二，OLS\_EM和TLS\_EM 的效果始终优于OLS和TLS，且在放大的图片里可以观察到TLS\_EM比OLS\_EM效果更好，结果说明我们提出方法的有效性。第三，仅仅在训练集比例非常小（15%）时，TLS\_EM效果差于OLS\_EM，说明在绝大部分情况下TLS\_EM比OLS\_EM更准确。

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.

本文考虑了当数据集噪声服从不同分布时的电池寿命预测问题，使用结合了EM思想的改进的OLS和TLS算法对其进行预测，预测结果显示，改进的方法效果显著。

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.

What is the expectation maximization algorithm?

Jing Song, G., Wen Wang, Q. On the weighted least-squares, the ordinary least-squares and the best linear unbiased estimators under a restricted growth curve model. *Stat Papers* **55**, 375–392 (2014). <https://doi.org/10.1007/s00362-012-0483-9>

B. De Moor and J. Vandewalle, "A unifying theorem for linear and total linear least squares," in IEEE Transactions on Automatic Control, vol. 35, no. 5, pp. 563-566, May 1990, doi: 10.1109/9.53523.

Energy is the foundation of all science and engineering technology, without which the human world will be difficult to operate [1]. The rise of industrialization has led to a large demand for energy, but the exhaustion of fossil energy and a series of environmental problems, coupled with the acceleration of traditional energy demand, have forced planners and decision makers to find alternative energy sources [2,3]. Many renewable energy technologies have been widely developed. 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 other fields [6,7]. At the same time, smart cities will become the future model of people [8,9], and electric vehicles can solve the energy-saving development and environmental pollution problems of smart cities well, and the development of new energy electric vehicles has become a global consensus [10,12]. Rechargeable lithium-ion batteries have been widely used as the best choice for new energy electric vehicles because of their higher energy density, smaller volume, longer life and larger capacity.

However, 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. 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 a word, 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]. This is shown in Figure 1.

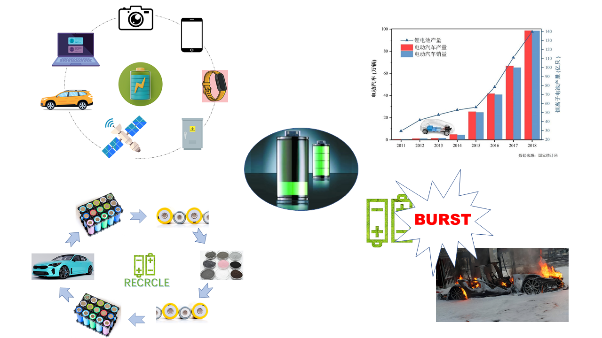


Figure1: Lithium battery applications and hidden dangers

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 accuracy and robustness of electrochemical model and equivalent circuit model are limited. Therefore, these two models are not a good and feasible solution. On the contrary, the 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.

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.

When dataset samples obey noise of different distributions,, we can know that our goal is to minimize the error when the samples obey different noises：

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

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

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

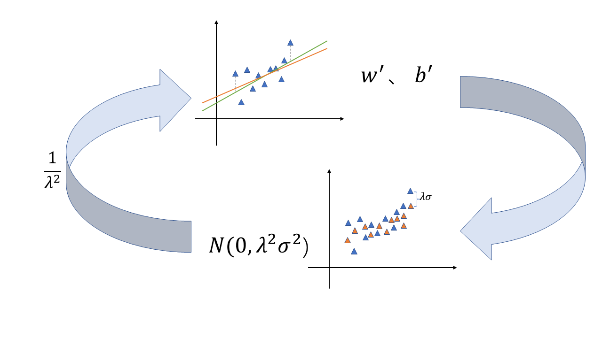


Figure 2: Improved TLS/OLS algorithm

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, The algorithm flow is shown in 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.

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.

The dataset, referred to as “Dataset”, was generated by Severson et al. [15], 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 [15]. With these nonlinear transformations, we adopted a linear model template

for Dataset .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 |

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.

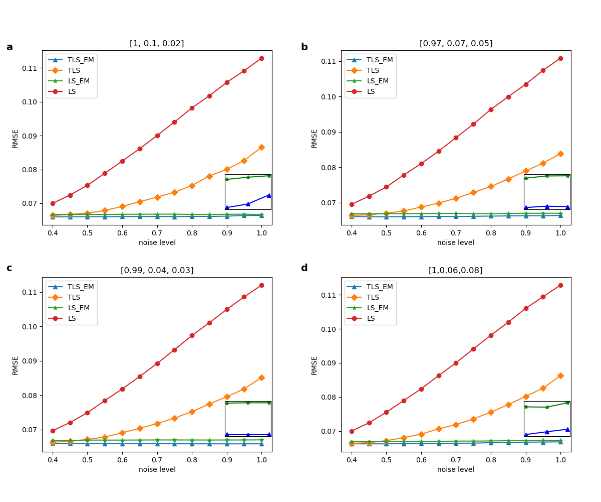


Figure 3: RMSE with increased noise level

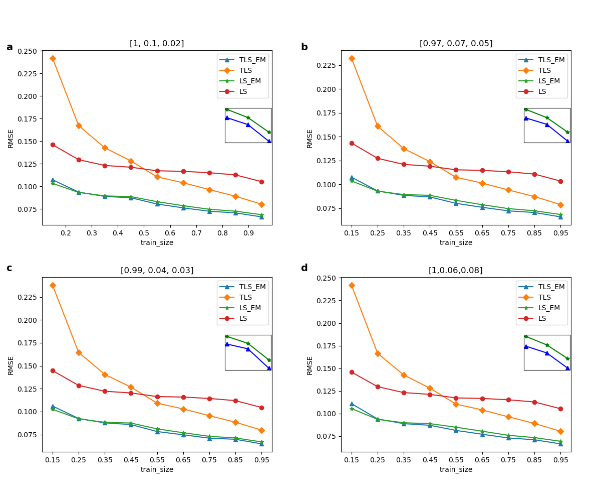


Figure 4: RMSE with increased training set proportion

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] M.A. Chaudhry, et al. ISO4, 13(6), 1657-1662 (2009).

[2] Mirza UK, Ahmad N, et al. ISO4,11(9),2179–90(2007).

[3] G.N. Tiwari, M.K. Ghosal, Renewable Energy Resources: Basic Principles and Applications, 2005.

[4] Barack Obama.*Science***355**,126-129(2017).

[5] LIU H, LIU X, et al. Advanced Energy Materials, 2017, 7(24): 1700283.

[6] P. Poizot, S. Laruelle, et al. Nature 2000, 407, 496.

[7] B. Dunn, H. Kamath, et al. Science 2011, 334, 928.

[8] Raza MQ, Khosravi A. Renew Sustain Energy Rev 2015;50:1352e72.

[9] Xu J, Zhang R. IEEE Trans Veh Technol 2013;64(6):2476e88.

[10] H. Xiong, H. Liu, et al. International Journal of Hydrogen Energy, vol. 44, no. 56, pp. 29733-29742, 2019.

[11] Wang G, Xu Z, et al. IEEE Trans Power Deliv 2013;28(4):2363e72.

[12] Hu J, Zheng L, et al . Energies 2018;11.

[13] Ng, MF., Zhao, J., Yan, Q. et al.  Nat Mach Intell 2, 161–170 (2020).

[14] Roman, D., Saxena, S., Robu, V. et al.  Nat Mach Intell 3, 447–456 (2021). [15] Severson, K.A., Attia, P.M., Jin, N. et al.  Nat Energy 4, 383–391 (2019).

16. Peterson, S. B., et al. J. Power Sources 195, 2385–2392 (2010).

17. Ramadesigan, V. et al .J. Electrochem. Soc. 159, R31–R45 (2012).

18. Waag, W., Fleischer, C. & Sauer, D. U. J. Power Sources 258, 321–339 (2014)

19 Zhang, Y., Tang, Q., Zhang, Y. et al.  Nat Commun 11, 1706 (2020).

20. Severson, K. A. et al .Nat. Energy 4, 383–391 (2019).

21. Nuhic, A., Terzimehic, T., et al. J. Power Sources 239, 680–688 (2013).

[22] Tianheng Feng, Lin Yang, et al. Journal of Power Sources, 281:192–203, 2015.

[23] D Andre, M Meiler,et al. Journal of Power Sources, 196(12):5349–5356, 2011.

[24] Daigle,et al. “Electrochemistry-based Battery Modeling for Prognostics.” (2013).

[25] Brian Bole Moffett Field United States, 2014.

[26] Githin K Prasad Journal of power sources, 232:79–85, 2013.

[27] Kristen A Severson Nature Energy, 4(5):383, 2019.

[28] Bhaskar Saha IEEE Transactions on instrumentation and measurement, 58(2):291–296, 2008.

[29] Kai Goebel, IEEE instrumentation & measurement magazine, 11(4):33–40, 2008.

[30] Xiaosong Hu, IEEE Transactions on Industrial Electronics, 63(4):2645– 2656, 2015.

[31] Verena Klass, Journal of Power Sources, 270:262–272, 2014

[32] Peter M Attia, et al. 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, Philadelphia, PA: SIAM, 1997

[35] S. Van Huffel, Frontiers in Applied Mathematics. PA: SIAM, 1991.

[36] S. Van Huffel, Ed., Philadelphia, PA: SIAM, 1997.