The industrialization revolution led to a large demand for energy, and the depletion of fossil energy and environmental pollution promoted the rise of new energy [1-3]. However, most renewable energy sources, such as solar and wind, are intermittent in nature and rely on natural phenomena to generate electricity, at which point energy storage becomes a new demand [4-10]. Due to their high density, high capacity, and long lifespan, lithium-ion batteries have become the primary energy storage device for portable electronic devices, electric vehicles (EVs), and many other applications[11,12]. However, in the use of lithium batteries, it cannot be ignored that the decline in battery performance will cause problems such as reduced battery life, insufficient power, and battery explosion, as shown in Figure 1. If the battery life can be predicted before the battery aging, the above problems can be avoided, and the battery development cycle can be accelerated, new processes can be validated, and the battery can be rerecycled [13-15], which will bring major opportunities for the manufacturing, use and optimization of the battery [16-18].

At present, the models used for battery life estimation can be mainly divided into equivalent circuit model (ecm)[19-20], electrochemical model [21-23] and data-driven model [24-29]. The accuracy and robustness of the electrochemical model and the equivalent circuit model are limited, so these two models are not a good viable solution. In contrast, data-driven models do not need to understand the complex chemical reactions inside the battery, there is no complex process to build the circuit and other advantages are widely used by researchers. At the same time, with the development of research in recent years, it is found that the noise in the battery data set is inevitable, which is mainly due to the environmental interference during the charging and discharging process, such as temperature change and humidity fluctuation. In addition, experimental conditions do not fully simulate reality, so more and more research has begun to focus on the prediction of battery life with noise.

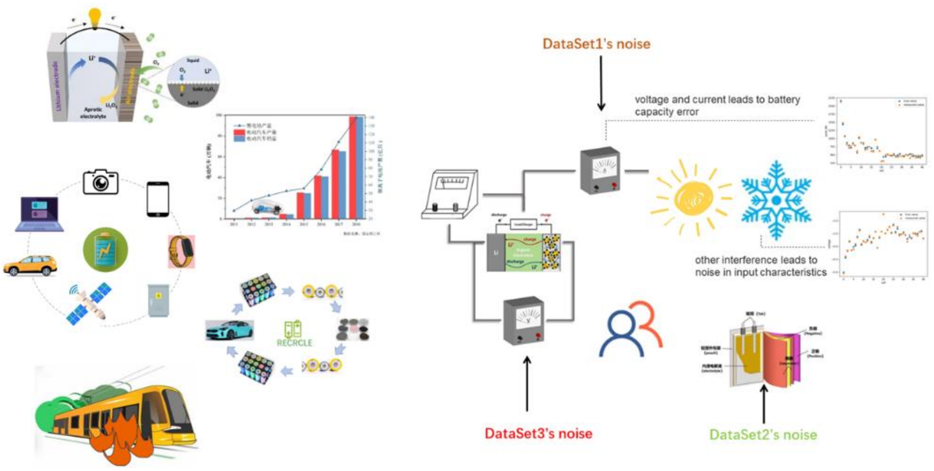

Figure1: Lithium battery applications and hidden dangers

Linear parameter estimation problems arise in a wide range of scientific disciplines such as signal processing [30-31]. As shown in [32] and [33], total least squares is the best choice for parameter estimation when all variables of interest have parametric linear relationships and all measurements are noise polluted.

However, in the actual situation, the battery information data sets 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 obey the same distribution, and the direct use of TLS/OLS cannot establish a good battery life prediction model. Therefore, this paper made improvements in the establishment of a linear model to calculate battery life. After weighted battery samples with different noise distributions, TLS/OLS was used for prediction. The standard deviation of noise distribution could be accurately calculated through cyclic iteration, and a prediction model adapted to different noise distributions was established to predict battery life. The prediction results show that our method is better than the traditional TLS/OLS method.

In this paper, three public datasets based on commercial lithium-ion batteries, called "Dataset 1", "Dataset 2" and "Dataset 3", are composed of 41, 43 and 40 samples, respectively[17]. Although these datasets extract a total of 20 features, we further manually select subsets of 3 important features based on domain expertise. The specific meaning is shown in table1. We divide the three data sets according to 9:1 and combine them as training sets and test sets. Each experiment randomly scramps the sample order, and each run independently randomly generates training and test data sets. A median of 1,000 RMSE values is reported for each method so that the error measure is not strongly biased by random fluctuations.

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

As shown in Figure 2, we set four different noise ratios. Figure a, b, c, and d show the experimental results with the increase of noise level: (i) With the increase of noise, the effects of TLS and OLS become significantly worse, while the improved algorithm is not significantly affected by noise level and has strong stability. (ii) The improved algorithms TLS_EM and OLS_EM combined with EM have better effects than the traditional algorithms TLS and OLS, indicating that the improved algorithm is more suitable for the battery data set with noise. (iii) The effect of TLS_EM is better than that of OLS_EM, and the effect of TLS is also better than that of OLS, In the case that all the measured values receive noise pollution, TLS has greater advantages than LS.

Figure e ,f ,g and h are the experimental results of increasing the proportion of training sets: (i) With the increase of the proportion of training sets, the four methods have better effects because the increase in training data is the more important information for the model. (ii) Regardless of the proportion of the training set, the improved algorithm is superior to the traditional algorithm, which shows the effectiveness of the algorithm with EM. (iii) In most cases (the training set accounts for more than

25%), TLS_EM is better than OLS_EM, which shows that TLS_EM is more applicable than OLS_EM .

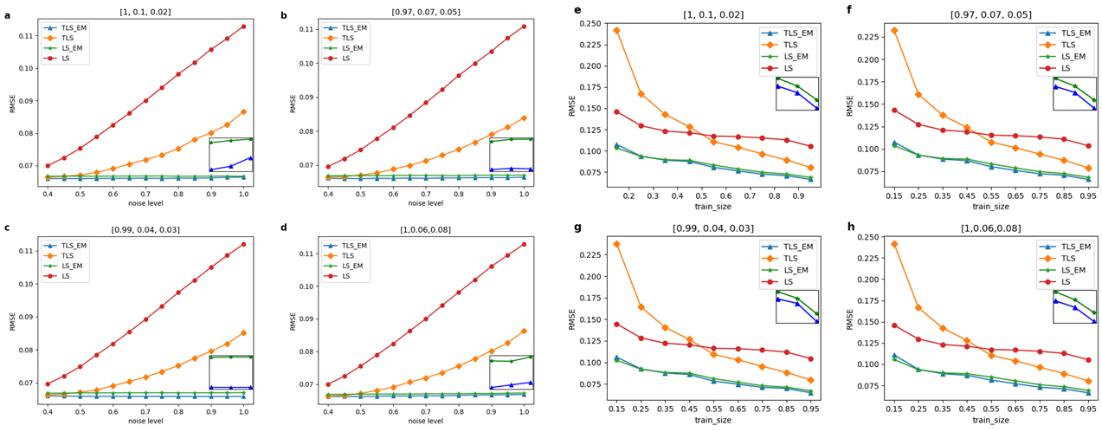

Figure 2: experimental result

The main flow of this algorithm is as follows. We consider that when dataset samples obey noise of different distributions, $y_i = X_i w + \epsilon_i$, we can know that our goal is to minimize the error when the samples obey different noises：

$$\min \sum_i (y_i - X_i w)^2,$$

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

$$\min \sum_i \alpha_i (y_i - X_i w)^2,$$

Assuming that the noise obeys a Gaussian distribution with zero mean and different variances: $\epsilon_i \sim N\left(0, \lambda_i^2 \sigma^2\right)$, then y_i obeys a Gaussian distribution $y_i \sim N(X_i w, \lambda^2 \sigma^2)$. The likelihood function is:

$$l(w) = \prod_i p(y_i|w) = \prod_i \frac{1}{\sqrt{2\pi}\lambda_i \sigma} \exp\left[-\frac{(y_i - X_i w)^2}{2\lambda_i^2 \sigma^2}\right],$$

Maximizing the likelihood function is equivalent to minimizing the objective function $\sum_{\mathrm{i}} \frac{1}{\lambda_i^2} (y_i - X_i w)^2$, and we can get $\alpha_i = \frac{1}{\lambda_i^2}$ by comparison.

EM algorithm[34] 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.

$$\min \sum_i \frac{1}{\lambda_i^2}(y_i - X_i w)^2 \ ,$$

3. The samples are weighted according to the formula, and a new round of model coefficients are obtained by using TLS[35]/OLS[36].

4. Repeat steps 2 and 3 until convergence.

Figure 3 gives the general flow of the improved algorithm, and the convergence process of the standard deviation of noise distribution is shown in Figure 4.

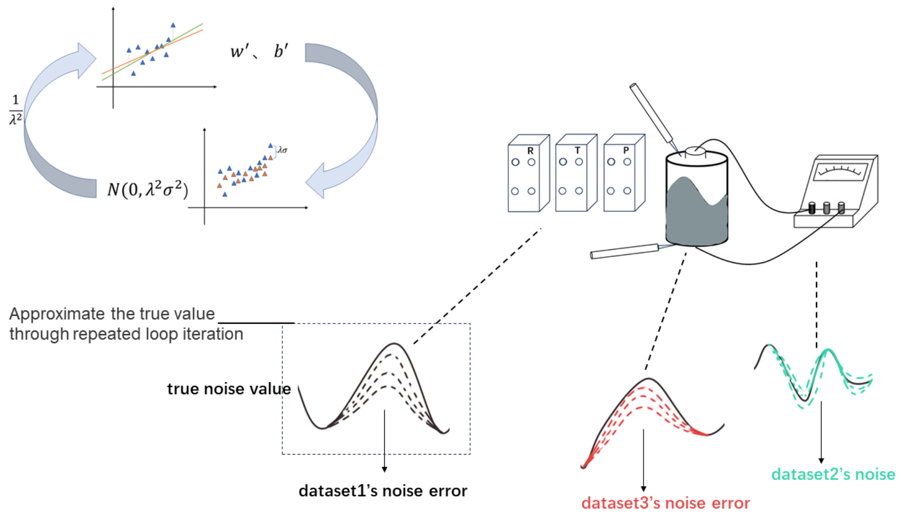

Figure 3: Improved TLS/OLS algorithm

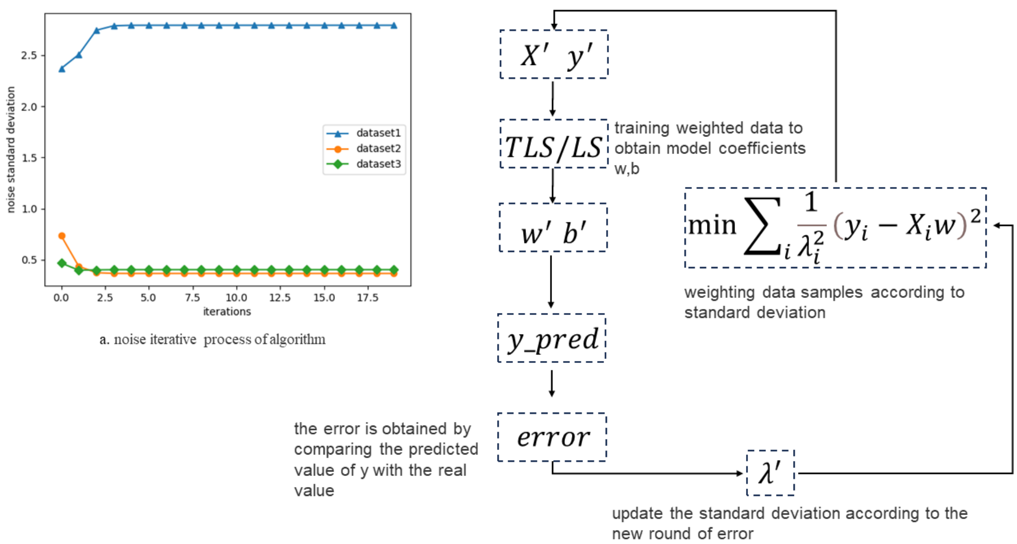

Figure 4: noise convergence process

In this paper, the battery samples with different noise distributions are weighted and predicted by TLS/OLS. Through cyclic iteration, the standard deviation of noise distribution can be accurately calculated, and then the battery life can be accurately predicted. Experimental data prove the effectiveness of TLS_EM/OLS_EM algorithm.

[1] M. Ashraf Chaudhry, R. Raza, S.A. Hayat, Renewable energy technologies in Pakistan: Prospects and challenges, Renewable and Sustainable Energy Reviews, Volume 13, Issues 6–7,2009, Pages 1657-1662,

[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] H. Liu, X. X. Liu, W. Li, X. Guo, Y. Wang, G. X. Wang, D. Y. Zhao, Adv. Energy Mater. 2017, 7, 1700283. https://doi.org/10.1002/aenm.201700283

[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] Huiyuan Xiong, Huan Liu, Ronghui Zhang, Limin Yu, Zhijian Zong, Minghui Zhang, Zhu Li, An energy matching method for battery electric vehicle and hydrogen fuel cell vehicle based on source energy consumption rate, International Journal of Hydrogen Energy, Volume 44, Issue 56, 2019, Pages 29733-29742,

[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]. 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).

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

[15]. 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)

[16]. Zhang, Y., Tang, Q., Zhang, Y. *et al.* Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning. *Nat Commun* **11**, 1706 (2020).

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

[18]. 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).

[19] 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.

[20] 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.

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

[22] 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.

[23] 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.

[24] Roman, D., Saxena, S., Robu, V. *et al.* Machine learning pipeline for battery state-of-health estimation. *Nat Mach Intell* **3**, 447–456 (2021). https://doi.org/10.1038/s42256-021-00312-3

[25] 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.

[26] 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.

[27] 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.

[28] 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

[29] 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.

[30] B. Huang, "Detection of abrupt changes of total least squares models and application in fault detection," in IEEE Transactions on Control Systems Technology, vol. 9, no. 2, pp. 357-367, March 2001, doi: 10.1109/87.911387.

[31] 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

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

[33] B. Huang, "System identification based on last principal components analysis," in Proc. IFAC'99 World Congr., Beijing, China, July 1999, pp. 211–216.

[34] Do, C., Batzoglou, S. What is the expectation maximization algorithm?. Nat Biotechnol 26, 897–899 (2008).

[35] 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.

[36] 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