The rise of industrialization led to a large demand for energy, and the depletion of fossil energy and environmental pollution promoted the rise of new energy [2,3]. 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]. As an energy storage technology, rechargeable lithium-ion batteries have been widely used due to their advantages such as high density, large capacity and long life [11,12]. However, widespread use also brings problems such as battery life, lack of 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 [16-18], which will bring major opportunities for the manufacturing, use and optimization of the battery [19-21].

At present, the models used for battery life estimation can be mainly divided into equivalent circuit model (ecm)[22-23], electrochemical model [24-26] and data-driven model [27-32]. 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.

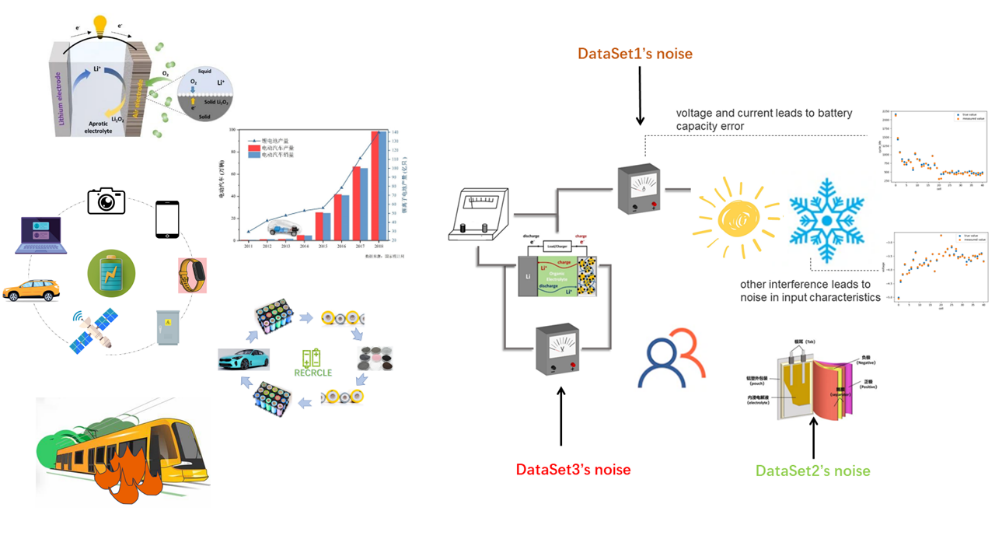


Figure1: Lithium battery applications and hidden dangers

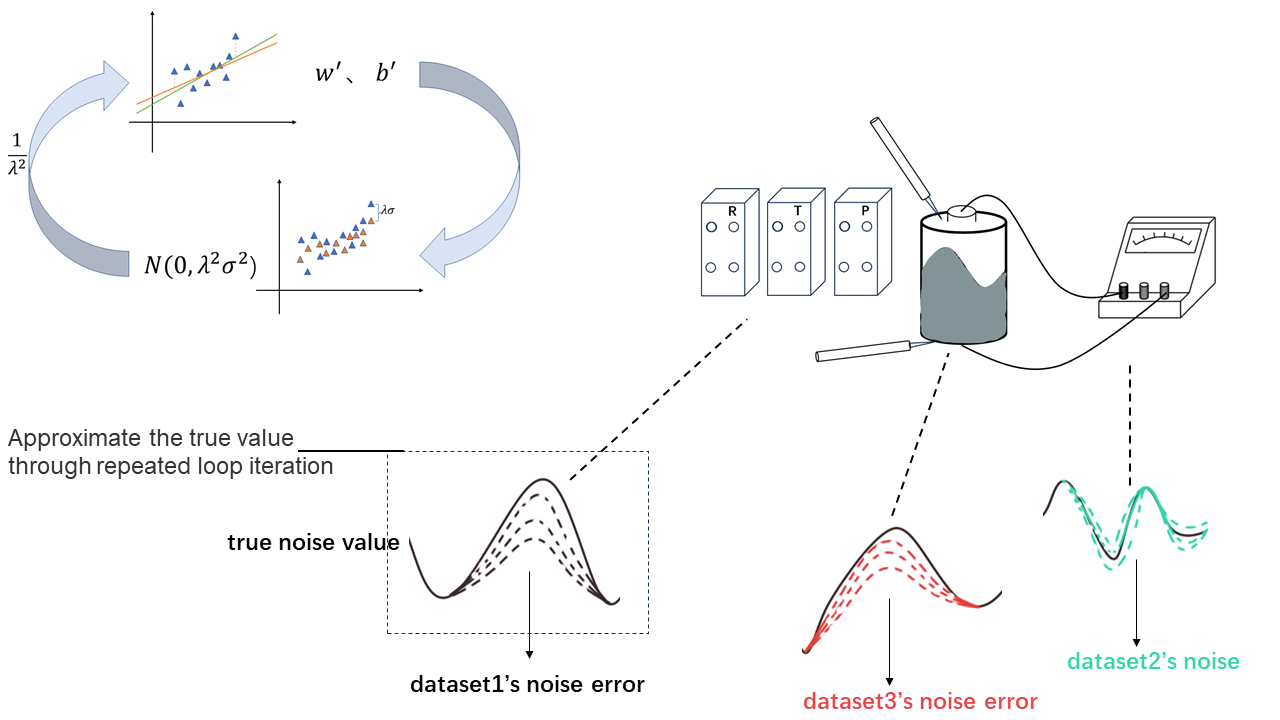


Figure 2: Improved TLS/OLS algorithm

Linear parameter estimation problems arise in a wide range of scientific disciplines such as signal processing [33-34]. As shown in [35] and [36], 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.

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.

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.

1 , ()

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.

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. 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 3, we set four different noise ratios. Figure a, b, c, and d show the experimental results with the increase of noise level: 1) 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. 2) The improved algorithms TLS\_EM and OLS\_EM combined with EM idea 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. 3) 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.

Figures e ,f ,g and h are the experimental results of increasing the proportion of training sets: 1) With the increase of the proportion of training sets, the four methods have better effects, and with more training data, the prediction ability of the model is improved. 2) 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 idea. 3) 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 4 shows the iterative noise convergence process of the algorithm.

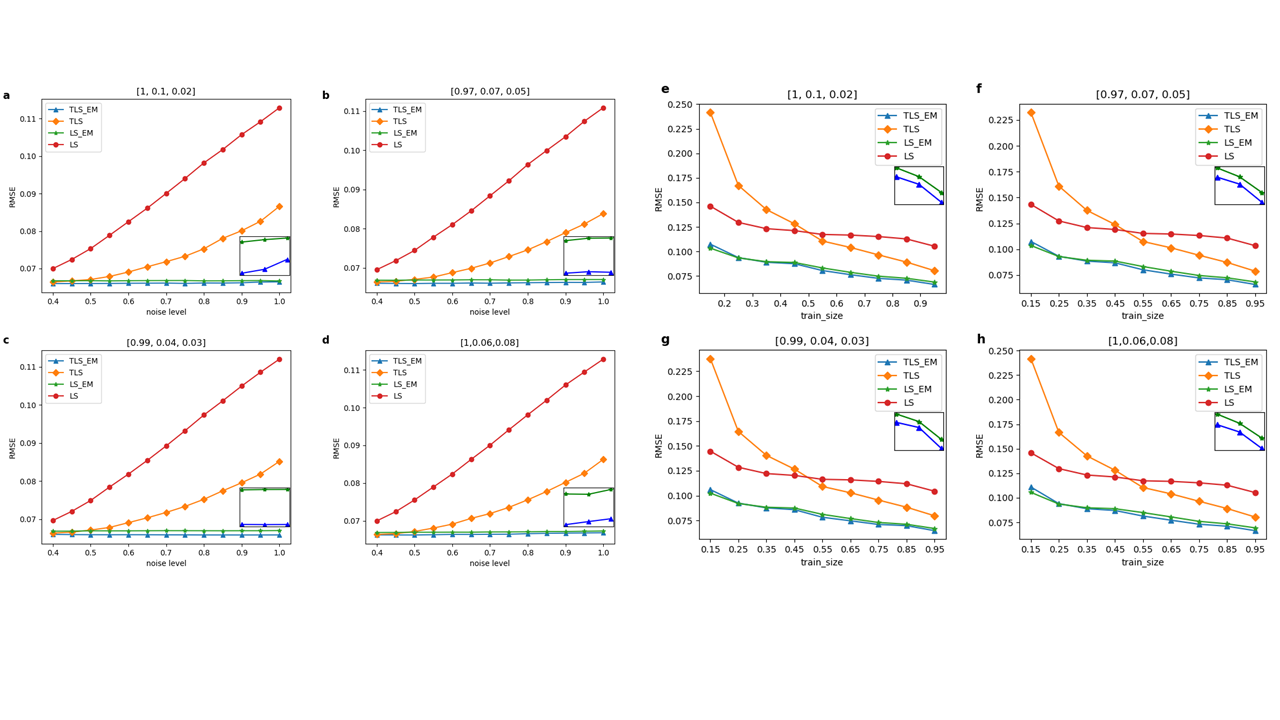


Figure 3: experimental result

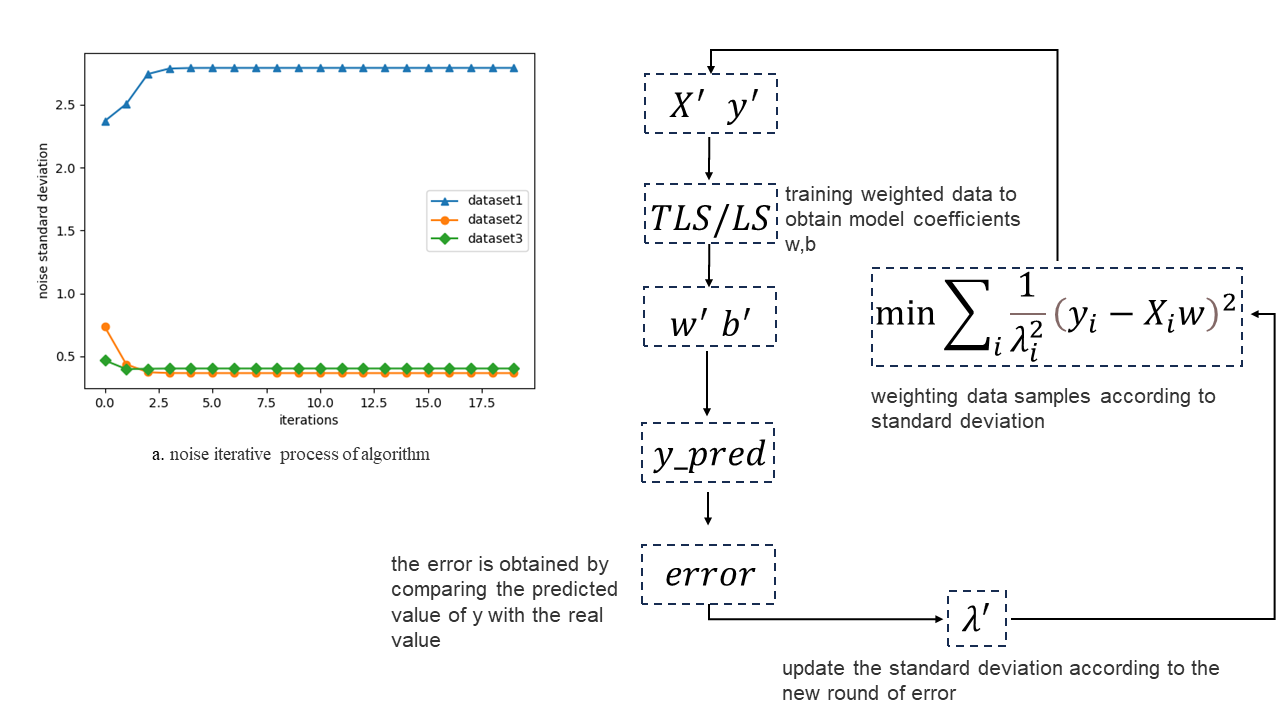


Figure 4: noise convergence process

In this paper, the linear model was established to calculate the battery life. After the battery samples with different noise distributions were weighted, TLS/OLS was used for prediction. After cyclic iteration, the standard deviation of noise distribution could be accurately calculated while the prediction model adapted to different noise distributions could be established to predict the battery life. The results show that our method has better results.