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，不同数据集因各种原因导致噪声分布不同）

电池预测现状，有噪声预测现状，LS、TLS在降噪方面的广泛运用

1. 总结本文贡献，算法大致结构，图2分析
2. 噪声水平增大和训练集比例增大，图3分析（详细讲述算法改进，算法优势，算法在训练集较少情况下仍然有效）
3. 详述算法内部通过迭代不断接近真实噪声，图4分析
4. 总结全文

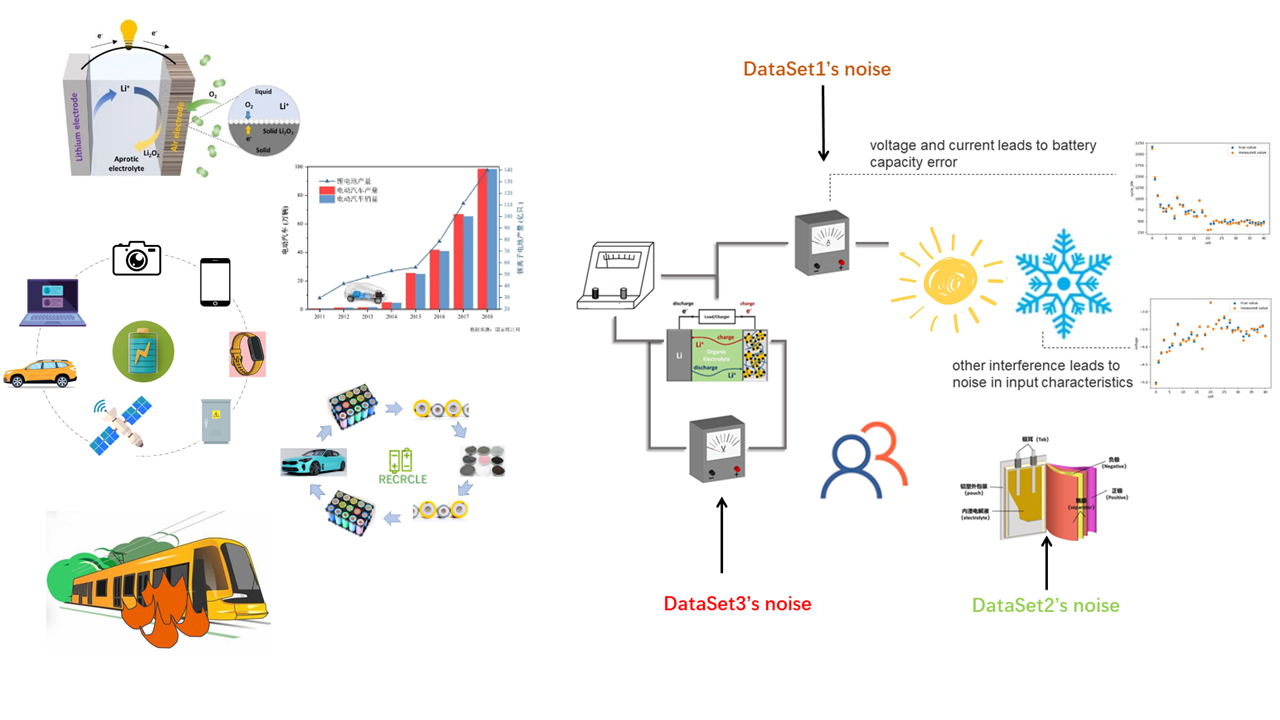


图1

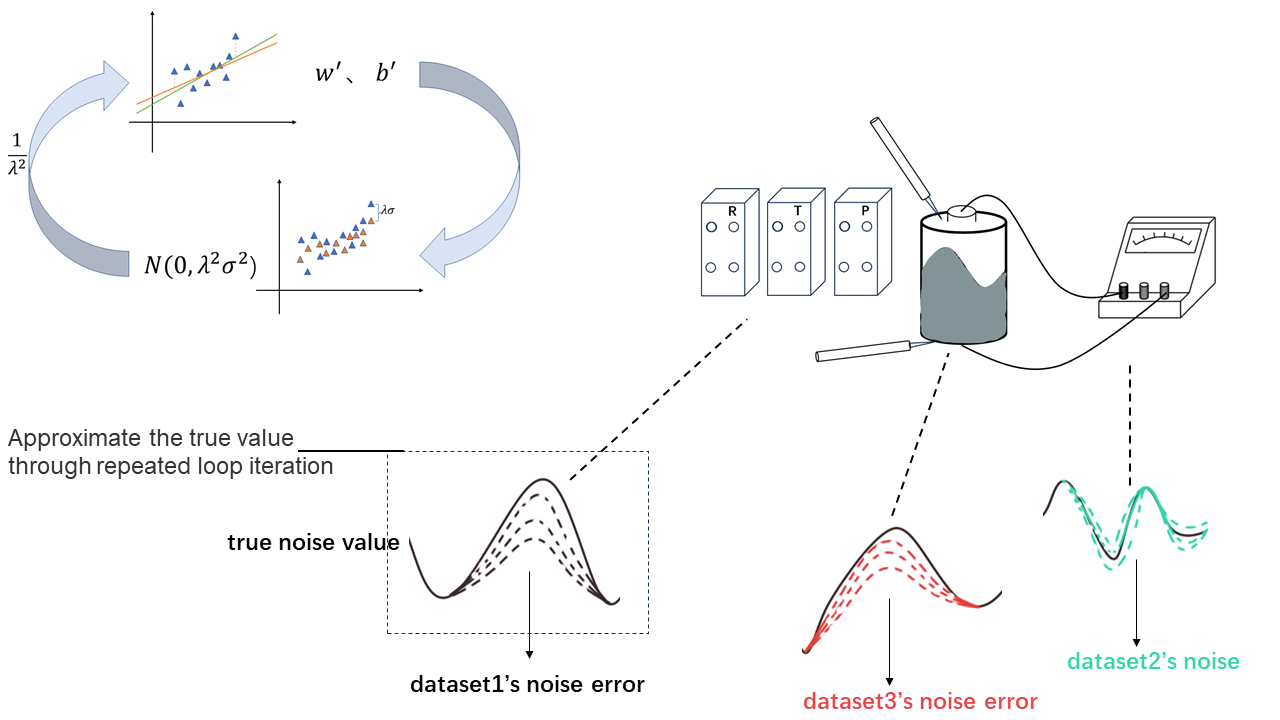


图2

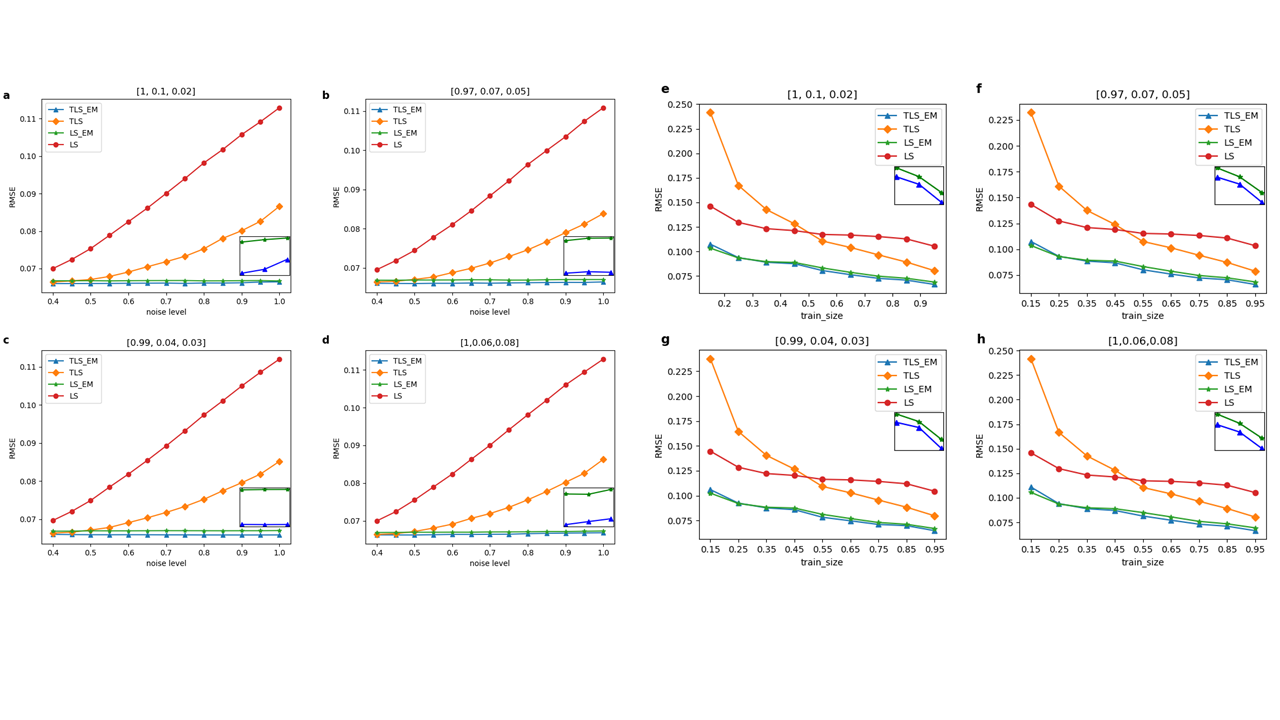


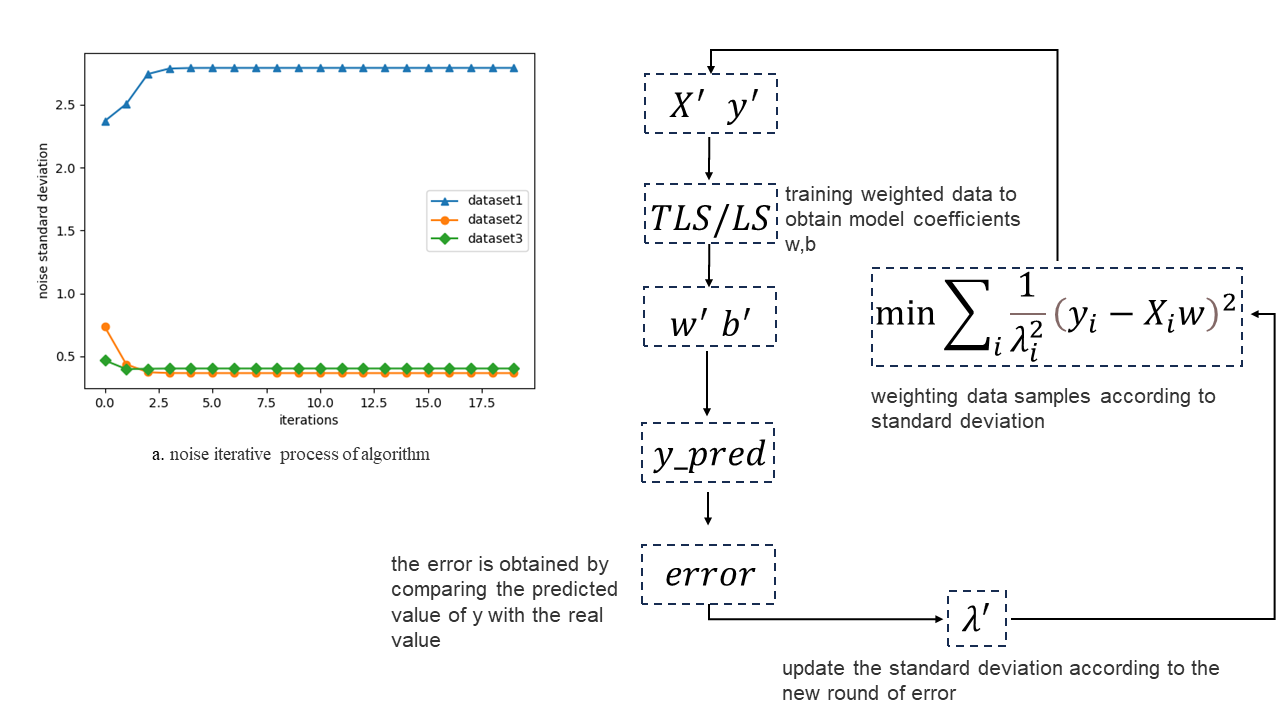
图3

图4

### 一、背景

化石能源的劣势导致新能源的兴起。可充电电池作为储能技术被广泛应用，锂电池则是智慧城市、新能源汽车发展必不可少的一项技术。但其广泛应用的同时带来了环境污染，电池老化等问题，因此对电池寿命预测是非常有必要的。

工业化的兴起导致了对能源的大量需求，化石能源的枯竭以及环境污染促进了新能源的崛起[2,3]。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，此时能源存储便成为新的需求 [4-10]。可充电锂离子电池作为一种储能技术，由于拥有高密度、大容量、长寿命等优点，已经得到了广泛应用[11,12]。但广泛使用的同时也带来了续航、动力不足、电池爆炸等问题，如图1所示。如果能在电池老化之前对电池寿命进行预测的话，在避免上述问题的同时还可以加快电池开发周期、验证新的工艺、对电池进行二次回收[16-18]，将为电池的制造、使用和优化带来重大机遇[19-21]。

Renewable energy technologies in Pakistan: Prospects and challenges

Wind energy development in Pakistan. Renewable and Sustainable Energy Reviews

Renewable Energy Resources: Basic Principles and Applications

]Porous Carbon Composites for Next Generation Rechargeable Lithium Batteries

Recent progress of magnetic field application in lithium-based batteries

A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings

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

Traffic-constrained multiobjective planning of electric-vehicle charging stations

Optimization and model validation of operation control strategies for a novel dual-motor coupling-propulsion pure electric vehicle

锂电池寿命预测方法主要分为三大类，其中，数据驱动的方法因其独特的优势（不需要了解电池内部化学反应，不需要基础的物理化学知识）作为当今热点研究方法。在测量时的不稳定性、环境变换、人为干扰等一系列因素导致电池数据集含有噪声。如果能在考虑噪声影响的同时准确的对电池寿命进行预测，就可以提高模型的稳定性，也能更好的拟合现实中电池衰退。

目前的电池寿命估计所采用的模型主要可以分为等效电路模型(ecm)[22-23]、电化学模型[24-26]和数据驱动模型[27-32]三种。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，数据驱动模型不需要了解电池内部的复杂化学反应，没有复杂的建立电路的过程等优势被研究者广泛应用。同时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。此外实验条件并不能完全模拟现实，因此，越来越多的研究开始关注带有噪声的电池寿命预测。

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

Mechine llearning piptline for battery state of health estimation

Data-driven prediction of battery cycle life before capacity degradation

Lithium-ion battery cell degradation resulting from realistic vehicle and vehicle-to-grid utilization

Modeling and simulation of lithium-ion batteries from a systems engineering perspective

Critical review of the methods for monitoring of lithium-ion batteries in electric and hybrid vehicles

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

Data-driven prediction of battery cycle life before capacity degradation

Health diagnosis and remaining useful life prognostics of lithium-ion batteries using datadriven methods

Online identification of lithium-ion battery parameters based on an improved equivalent-circuit model and its implementation on battery state-of-power prediction

Characterization of high-power lithium-ion batteries by electrochemical impedance spectroscopy

Matthew J Daigle and Chetan Shrikant Kulkarni. Electrochemistry-based battery modeling for prognostics

Adaptation of an electrochemistry-based li-ion battery model to account for deterioration observed under randomized use. Technical report

Model based identification of aging parameters in lithium ion batteries

Data-driven prediction of battery cycle life before capacity degradation

Prognostics methods for battery health monitoring using a bayesian framework.

Prognostics in battery health management.

Battery health prognosis for electric vehicles using sample entropy and sparse bayesian predictive modeling

A support vector machine-based state-of-health estimation method for lithium-ion batteries under electric vehicle operation

Closed-loop optimization of fast-charging protocols for batteries with machine learning. Nature

LS/TLS作为线性参数估计问题的经典解法，能够在测量数据含有噪声情况下保持较高的预测精准度，是预测带噪声数据集的不二选择。

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]。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。

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

in Recent Advances in Total Least Squares Techniques and Error-in-Variables Modeling

Frontiers in Applied Mathematics: The Total Least Squares Problem—Computational Aspects and Analysis. Philadelphia

Recent Advances in Total Least Squares Techniques and Errors-In-Variables Modeling

但在实际情况下，不同的电池数据集的噪声来源一般是不同的，造成了不同数据集的噪声分布不同，此时对LS/TLS进行改进以适应噪声分布不同的电池寿命预测问题。总结本文的改进。

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，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。

### 二、算法主要流程

介绍研究问题的背景，样本所携带的噪声标准差不同，对每个样本进行加权以适应不同的标准差，对比最大化似然函数得到权值。则算法中存在模型系数、噪声标准差两个变量，使用EM算法思想通过循环迭代拟合噪声标准差，写出具体循环过程，算法流程，算法伪代码。

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.

### 三、结果分析

本文使用三个基于商用锂离子电池的公共数据集，三个数据集分别称为“数据集1”、“数据集2”和“数据集3”，分别由41个、43个和40个样本组成，虽然这些数据集总共提取了20个特征，但我们进一步根据领域专业知识手动选择3个重要特征的子集，具体含义见table1，我们将三个数据集按照9:1划分之后合并作为训练集和测试集，每次实验随机打乱样本顺序，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

Data-driven prediction of battery cycle life before capacity degradation

如图3所示，我们设置了四种不同的噪声比例。图a,b,c,d是噪声水平增大的实验结果：（1）随着噪声增大，TLS和OLS效果明显变差，而改进的算法受噪声水平的影响不明显，具有较强的稳定性。（2）结合EM思想改进的算法（TLS\_EM、OLS\_EM）比传统算法（TLS、OLS）效果更佳,说明了改进的算法更能适应带有噪声的电池数据集。（3）TLS\_EM效果优于OLS\_EM，（TLS效果也优于OLS），在所有测量值都收到噪声污染的情况下，TLS比LS有更大的优势。

图e,f,g,h是训练集比例增大的实验结果：（1）随着训练集占比增大，四种方法效果更好，有了更多的训练数据，模型预测能力提升。（2）不论训练集比例大小，改进的算法优于传统算法，说明了融入EM思想的算法有效性 。（3） 在绝大部分情况下（训练集占比大于25%）TLS\_EM效果优于OLS\_EM，说明了TLS\_EM比OLS\_EM适用性更强。

图4展示了算法经过循环迭代噪声收敛过程。（算法通过TLS/OLS拟合样本数据得到模型系数和，根据模型系数和对样本数据进行预测得到新一轮的电池寿命预测值，将预测值和真实值对比得到新一轮的误差，由此求出标准差对样本数据加权后通过TLS/OLS求出下一轮的模型系数和）

### 四、总结

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，经循环迭代能够在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，进而对电池寿命进行预测。预测结果显示我们的方法有着更好的效果。