

# 基于 ARIMA 模型的工业锂电池剩余使用寿命预测<sup>①</sup>

陶耀东<sup>1</sup>, 李 宁<sup>1,2</sup>

<sup>1</sup>(中国科学院 沈阳计算技术研究所, 沈阳 110168)

<sup>2</sup>(中国科学院大学, 北京 100049)

**摘 要:** 研究了随机建模技术在锂电池剩余使用寿命预测中的应用. 基于此, 使用 Box-Jenkins ARIMA 模型模拟锂电池退化过程. 在 NASA PCoE 获取锂电池测量数据集, 采用 ADF 单根检验与差分法对锂电池容量原始数据平稳化处理. 结合自相关函数与偏自相关函数进行参数估计, 构建多个 ARIMA 模型, 并通过评估各种估计参数验证各个模型的有效性, 根据 AIC、SC 准则与正态化 BIC 选择最佳预测模型. 在对所选模型进行严格评估之后, ARIMA(2,1,2) 被识别为最佳拟合模型. 使用 ARIMA 模型获得了比较精确的预测结果, 结果表明 ARIMA 模型预测锂电池剩余使用寿命短期内具有较高的精确度和较强的可行性.

**关键词:** ARIMA; 剩余使用寿命预测; ADF

引用格式: 陶耀东, 李宁. 基于 ARIMA 模型的工业锂电池剩余使用寿命预测. 计算机系统应用, 2017, 26(11): 282-287. <http://www.c-s-a.org.cn/1003-3254/6067.html>

## Industrial Lithium Battery Remaining Useful Life Prediction Based on the ARIMA Model

TAO Yao-Dong<sup>1</sup>, LI Ning<sup>1,2</sup>

<sup>1</sup>(Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, China)

<sup>2</sup>(University of Chinese Academy of Sciences, Beijing 100049, China)

**Abstract:** The present study focuses on the application of stochastic modeling technique in analyzing the remaining useful life prediction of lithium battery. For this, the Box-Jenkins ARIMA model has been used for simulating the lithium battery degradation process. The lithium battery dataset has been collected from NASA PCoE. ADF unit root test and difference method are used to smooth the original data of lithium battery capacity. The parameters are estimated by analyzing autocorrelation function and partial autocorrelation function. Several ARIMA models have been generated and their validation has been verified by assessing various estimation parameters. According to AIC, SC criteria and normalized BIC, the optimal prediction model is selected. After rigorous evaluation of the selected models, the ARIMA (2,1,2) is identified as the best fit model. Satisfactory results have been obtained with the selected ARIMA models, indicating that the ARIMA model is highly accurate and feasible in the short term.

**Key words:** ARIMA; remaining useful life prediction; ADF

锂离子电池因其质量轻、低自放电率、循环使用寿命长等优点在手持终端等民用领域已经得到广泛的应用, 并开发出大量配套的电池管理系统, 具备完善的使用条件. 同样的, 在航空、航天、海上作业、汽车等前端工业领域也已得到推广, 但因工业领域复杂的工

作环境与工作负荷要求对锂离子电池提出更高的要求, 在工业领域不时有安全事故报出. 例如电池电源汽车着火爆炸, 美国太空实验 AFRL 实验失败事件, 2013 年多架波音 787 梦幻客机无限期停飞事件. 锂离子电池的可靠性和安全性成为工业应用中的关键问题, 建立

<sup>①</sup> 基金项目: “高档数控机床与基础制造装备”科技重大专项 (F14-056-7-00)

收稿时间: 2017-03-02; 修改时间: 2017-03-20; 采用时间: 2017-03-23

准确的锂离子电池容量退化模型,准确预测锂离子电池的剩余寿命,是保证锂离子电池运行可靠性和安全性的基础,也是目前锂离子电池预测与健康管理体系(PHM)技术的重要组成部分<sup>[1,2]</sup>.

## 1 锂电池剩余使用寿命预测模型研究

锂离子电池的寿命指标涵盖使用寿命、循环寿命和储存寿命.其中,循环寿命的应用最为广泛,其研究价值也最为重要,本文的主要研究内容就是锂电池剩余使用寿命,电池剩余使用寿命(RUL)指在一定的充放电条件下,电池性能或健康状态退化到不能满足设备继续工作或者是规定的最低工作状况(阈值)之前达到的充放电循环次数<sup>[3]</sup>.锂电池剩余使用寿命预测模型研究包括两个步骤,一是退化状态训练,即利用电池模型方法、数据驱动方法或二者融合的方法对锂离子电池退化状态识别;二是利用第一步识别的模型进行预测,设定退化阈值(根据美国电气和电子工程师协会(Institute of Electrical and Electronics Engineers, IEEE)规定电池容量下降到原始容量的80%即应更换新电池),当预测时间与退化阈值相等时认为电池的使用寿命终止,预测剩余使用寿命即为预测时间与当前时间的差值.

国内外针对电池健康管理研究主要研究两个问题,其一是荷电状态(State of Charge, SOC)要求预测单次充电放电的终止点(End of Discharge, EOD),解决的是当前周期内充放电剩余时间问题;其二是健康状态(State of Health, SOH),SOH针对电池在长期使用过程中逐渐累积的衰退过程,预测电池寿命消耗到不合要求的点,也就是电池寿命终止点(End of Life, EOL),预测电池剩余使用寿命,解决的是电池整个的生命过程何时终止的问题.

锂电池RUL分为三类<sup>[4]</sup>,包括基于锂电池物理化学模型的预测,方法包括电流积分法、放电试验法、开路电压法、负载电压法、电化学阻抗谱、内阻法;数据驱动RUL预测方法,包括模糊逻辑、回归分析、神经网络、支持向量机,相关向量机(RVM);融合型RUL预测方法,包括粒子滤波算法(PF),卡尔曼滤波算法及其相应改进算法.马里兰大学先进寿命周期工程中心(CALCE)对锂电电子电池寿命预测实验及锂电池退化过程研究处于世界领先地位;美国航空航天局重点研究航空航天锂电池退化机理,IMS李杰团队最近几年致力于车用充电锂电池的PHM研究.国内哈尔滨

工业大学刘大同、周建宝等人对锂电池退化过程有较深入的研究.

本文将ARIMA模型应用于锂离子电池剩余使用寿命预测,使用自相关分析图与单位根检测,对NASA PCoE提供的电池数据进行平稳性检验,运用差分法对原始数据进行平稳性处理;依据最小信息化准则多次试验测试选定ARIMA最优参数;选择平均绝对误差MAE、预测误差均方根RMSE量化其预测结果.

## 2 ARIMA模型

锂电池退化具有动态、非线性性质,其退化原理复杂,是正极材料、负极材料、隔膜、有机电解液老化分解,相互作用等众多因素所致.在宏观上锂电池容量数据表现出非线性特征,而ARIMA是一种针对非线性平稳时间序列的良好预测方法,将锂电池容量变化数据整理汇总得到可应用于ARIMA模型的时间序列.

所谓ARIMA(BoxeJenkins Autoregressive Integrated Moving Average (ARIMA) model)模型又称自回归移动平均模型,由博克思(Box)和詹金斯(Jenkins)于70年代初提出的一著名时间序列预测方法,又称为box-jenkins模型.是时间序列分析中常用模型,具有较高的预测精度;ARIMA模型仅仅考虑单个变量,不以锂电池背后复杂的物理化学理论为依据,以锂电池容量变化描述锂电池退化规律,进而使用该规律外推以实现预测<sup>[5]</sup>.建模的基本思想若锂电池原始数据是平稳的则直接使用ARMA模型,否则使用差分法对非平稳时间序列进行平稳化处理,通过观察相关函数截尾和拖尾特征自回归阶数( $p$ )、差分次数( $d$ )和移动平均阶数( $q$ )3个参数,进而对时间序列进行预测分析<sup>[6]</sup>.用如下公式定义自回归滑动平均模型 $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$ ,该公式存在如下两种特殊情形,其一 $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t$ ,称为自回归形式(AR),该式表示当前 $t$ 时刻的锂电池容量值由过去的 $p$ 个历史值 $Y_{t-1}$ 、 $Y_{t-2}$ 、 $\dots$ 、 $Y_{t-p}$ 的加权和组成, $\phi_1$ 、 $\phi_2$ 、 $\dots$ 、 $\phi_p$ 表示其权重, $\varepsilon_t$ 表误差项;另一种是移动平均模型(MA), $Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$ ,该式表示 $t$ 时刻的电池容量由若干个白噪声的加权平均和组成,其中 $\varepsilon_t$ 、 $\varepsilon_{t-1}$ 、 $\dots$ 、 $\varepsilon_{t-q}$ 表高斯白噪声, $\theta_1$ 、 $\dots$ 、 $\theta_q$ 表其权重<sup>[7]</sup>.将AR与MA组合经过积分或者差分生成ARIMA( $p$ ,  $d$ ,  $q$ )模型.该模型同时也可表示为

$$(1 - B)^d Y_t = \frac{\theta(B)}{\phi(B)} \varepsilon_t, B \text{ 表延迟算子, 其中 } \theta(B) = 1 - \sum_{k=1}^p \theta_k B^k,$$

$$\phi(B) = 1 - \sum_{k=1}^p \phi_k B^k.$$

锂离子电池 ARIMA 模型预测的关键步骤如图 1.

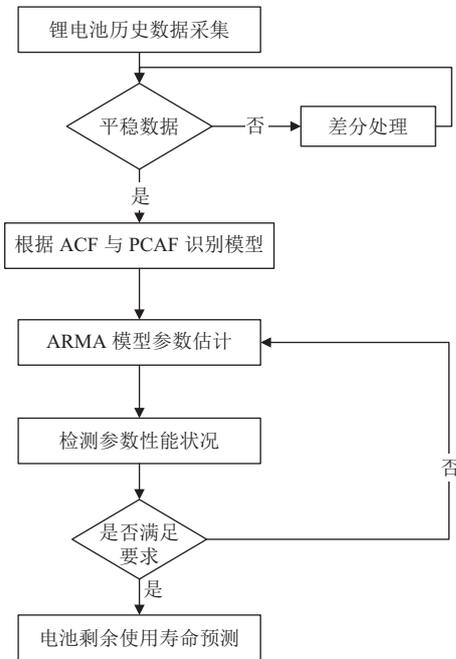


图 1 ARIMA 模型建模流程

因 ARIMA 模型只适用于平稳时间序列, 在应用该模型时必须做平稳性检测; 根据锂离子电池容量序列的散点图、自相关函数分析图以及 ADF 单位根检验, 对序列的平稳性进行识别.

(1) 锂电池容量数据平稳化处理: 包括差分, 取对数等方法, 直到数据满足平稳性条件.

(2) 锂电池 ARIMA 模型参数估计, ARIMA 模型包含 AR(p), MA(q), ARMA(p, q) 三种形式, ARIMA(p, d, q) 参数估计就是指确定这三个参数. 参数 d 识别, 较简单, 对锂电池容量数据进行单位根检验, 若存在单位根, 则对其差分后的序列进行判断, 若 d 阶差分后序列满足平稳条件, 则可确定 d. 若一直达不到平稳条件或者达到平稳但对锂电池容量的研究已经无意义, 则不能适用 ARIMA 模型. 参数 q 识别需要使用自相关函数 ACF 和偏自相关函数 PACF 进行判断. 即根据自相关函数与偏自相关函数的截尾拖尾特性选择合适的 p, q 值, 根据差分次数确定 d 值.

$Y_t$  与  $Y_{t-k}$  的 ACF 定义为:  $\rho_k = \frac{\text{cov}(Y_t, Y_{t-k})}{\sqrt{\text{var}(Y_t)\text{var}(Y_{t-k})}$ ,  $Y_t, \dots, Y_{t-k}$  的偏自相关函数定义为  $\varphi_{kk} = \frac{E(Y_t Y_{t-k})}{\sqrt{E(Y_t^2)E(Y_{t-k}^2)}}$ ; 总结模

型识别方法如表 1 所示.

表 1 ARIMA 模型识别图形判断方法总结

	AR(p)	MA(q)	ARMA(p, q)
ACF	拖尾/震荡	q步截尾	拖尾/震荡
PACF	p步截尾	拖尾/震荡	拖尾/震荡

(3) 选择最优模型参数, 预测锂电池剩余使用寿命. 最优模型参数选择依据其一是选择最小信息准则, AIC、SC 与正态化 BIC 越小, 模型的拟合优度更好, 模型对原始数据的拟合具有更好的“精确性”和“简单性”, 若 AIC、SC 与正态化 BIC 差别不明显, 可进一步比较  $R^2$ ,  $R^2$  越大, 模型的拟合效果越好; 其二是“简约原则”, 模型设立单一, 滞后期较小的模型.

(4) 采用最小二乘法进行参数估计, 检验 p, d, q 是否具有统计意义.

(5) 进行假设检验, 诊断残差序列是否为白噪声, 若残差序列的自相关系数和偏自相关系数在统计上均不显著, 则认为模型是可接受的<sup>[8]</sup>.

(6) 利用已通过检验的模型进行预测分析.

### 3 锂离子电池剩余使用寿命预测

#### 3.1 实验数据获取

以 NASA 官网的 PCOE datasets 提供的锂离子电池老化测量数据为检测数据, 使用锂离子电池一个循环后电容量为参考标准, 经初步整理后, 采样间隔表示一次循环, 取 B0005 中前 100 组数据为训练数据, 后 40 组数据作为预测数据. 经初步处理后数据具有如图 2 所示形式: 很明显锂电池的容量会随着慢慢减少, 并不是平稳的时间序列, 也即是锂电池容量原始数据并不能直接使用 ARMA 模型, 需要进行平稳化处理.

#### 3.2 数据平稳化处理

应用 ARIMA 模型预测锂电池剩余使用寿命的前提是采集的时间序列必须是平稳序列, 文献[9]使用游程检验, 游程检验实际上是检测数据是否是随机序列, 并不是严格的平稳检验; 部分文献使用图形判断, 如锂电池原始数据散点图, 锂电池容量的的自相关和偏自相关系数图判断, 这类检测方式往往是经验主义判断, 带有主观臆断色彩, 并不严密, 且散点图从视觉判断, 视觉判断不免会受到制图比例的影响, 而偏自相关系数的计算方式在不同软件上存在差异, 例如将少量的锂电池容量变化数据在 matlab 与 spss 在处理时, 计算

的 PACF 就有很大差异. 所以需要严密的数学理论进行验证, 最为科学的平稳性检验方法是单位根检验. 单位根检验包括 Dickey-Fuller(DF) 检验, Augmented Dickey-Fuller(ADF) 检验, Phillips-Perron 检验等, 这里选择 ADF 检验方法检验锂电池容量数据的平稳性. 选取前 100 个数据作为训练数据, 经过反复检验, 一阶差分后, 为平稳时间序列. 一阶差分的 ADF 单位根检验结果图 3. 原假设为存在单位根, 检验结果显示只有 0.0001 支持该假设, 即拒绝存在单根假设, 得出结论该组锂电池容量退化数据一阶差分后为平稳时间序列, 能够应用 ARIMA 模型, 并可以确定参数  $d = 1$ .

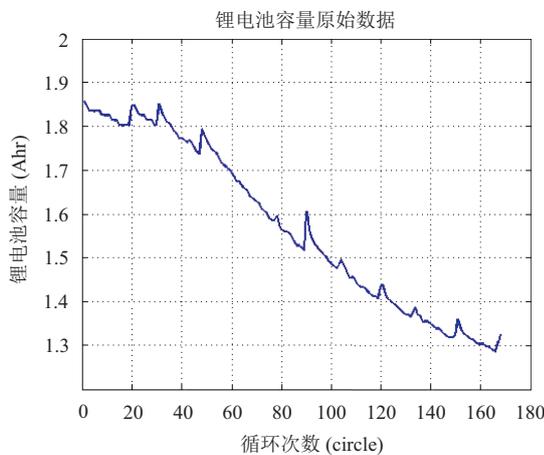


图 2 锂离子动力电池容量原始数据

Null Hypothesis: D (CAPACITY) has a unit root  
 Exogenous: Constant  
 Lag Length: 0 (Automatic-based on SIC, maxlag=12)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-11.46369	0.0001
Test critical values: 1% level	-3.498439	
5% level	-2.891234	
10% level	-2.582678	

\*MacKinnon (1996) one-sided p-values.

图 3 锂电池容量序列一阶差分单位根检验结果

### 3.3 模型参数估计

利用 SPSS 或 Eviews 统计分析软件生成锂电池容量一阶差分数据的 ACF 与 PACF 如图 4: ACF 与 PACF 后期虽都有震荡, 但 ACF 在滞后 3 期, PACF 在滞后 4 期接近 0, 可构建 ARIMA(3,1,1), ARIMA(3,1,2),

ARIMA(2,2,2) 模型, 依据 AIC 与 SC 最小化原则,  $R^2$  最大化的原则, 选取最合适的模型.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	-0.155	-0.155	2.4449	0.118	
2	-0.125	-0.153	4.0560	0.132	
3	-0.050	-0.100	4.3114	0.230	
4	-0.004	-0.053	4.3130	0.365	
5	-0.018	-0.052	4.3459	0.501	
6	0.004	-0.024	4.3473	0.630	
7	0.023	0.006	4.4038	0.732	
8	-0.015	-0.018	4.4279	0.817	
9	-0.059	-0.066	4.8117	0.850	
10	0.018	-0.010	4.8496	0.901	
11	0.045	0.027	5.0818	0.927	
12	0.085	0.098	5.9084	0.921	
13	0.005	0.054	5.9113	0.949	
14	-0.034	0.011	6.0444	0.965	
15	-0.030	-0.009	6.1527	0.977	
16	-0.068	-0.073	6.7130	0.978	

图 4 锂电池容量一阶差分序列 ACF 与 PACF 系数

表 2 比较各组参数值模型的 AIC, SC, BIC 等信息, ARIMA(3,1,2) 与 ARIMA(2,2,2) 具有较好的特性, 选择两组参数作为后期残差检验与预测模型参数.

表 2 模型 AIC、SC 与正态化 BIC 检验值

模型	R-square	AIC	SC	正态化BIC	残差序列相关性
ARIMA (3,1,2)	0.163	-5.61	-5.45	-8.145	不相关
ARIMA (3,1,1)	0.063	-5.52	-5.39	-8.202	不相关
ARIMA (2,1,2)	0.155	-5.63	-5.5	-8.202	不相关
ARIMA (2,1,1)	0.062	-5.55	-5.44	-8.259	不相关

ARIMA(2,1,2) 与 ARIMA(3,1,2) 的残差相关图如下: 从图中可以看出, 所有滞后都在 95% 置信水平内, 残差是随机的, 残差序列的自相关系数和偏自相关系数在统计上均不显著, 模型可较好地适用于锂电池容量退化数据.

### 3.4 预测

分析锂电池容量变化情况, 选择 ARIMA(2,1,2) 和 ARIMA(3,1,2) 对锂电池失效的循环次数进行预测, 利用前面 100 个数据作为训练数据, 剩下的数据作为预测对比数据, 结果如图 7、图 8. 由表 3 与图 7、8 可知, ARIMA 模型的绝对误差均值与预测误差均方根都较小, 预测失效值在第 123 循环, 实际失效在第 124 循环, 说明使用 ARIMA 模型预测锂电池剩余使用寿命短期内是具有较高的精确度和较强的可行性, 但是随着时间延长, 预测的置信上限和下限变大, 说明长期预测仍是不够准确的.

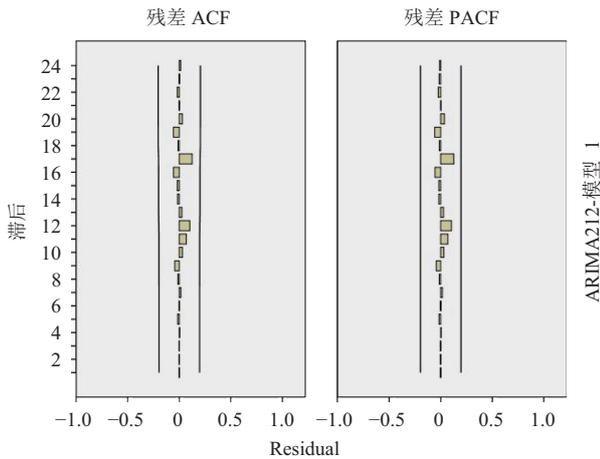


图5 ARIMA(2,1,2) 模型残差序列自相关分析图

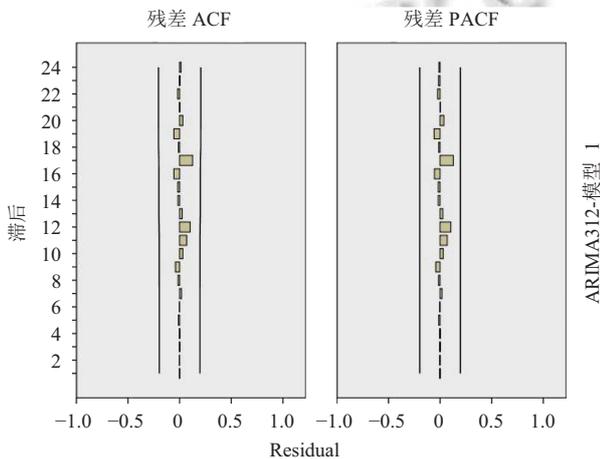


图6 ARIMA(3,1,2) 模型残差序列自相关分析图

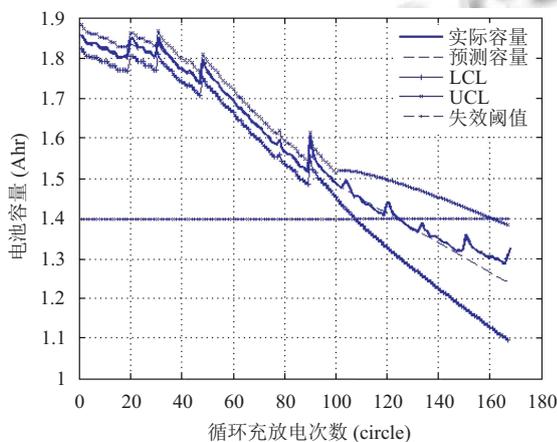


图7 ARIMA(2,1,2) 预测结果

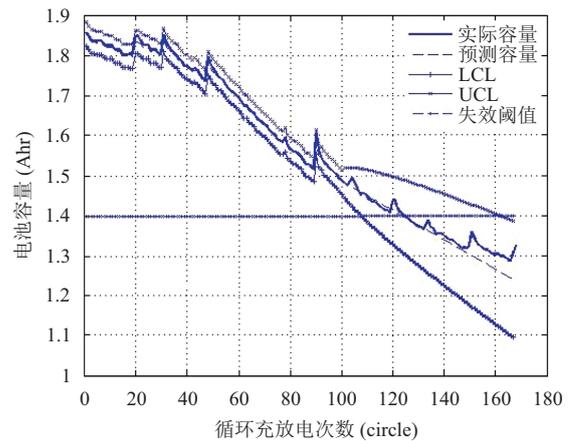


图8 ARIMA(3,1,2) 预测结果

表3 ARIMA 模型预测结果比较

模型	预测失效 循环次数	绝对误差 均值 MAE	预测误差 均方根 RMSE <sup>[10]</sup>
ARIMA (3,1,2)	123	0.007	0.015
ARIMA (2,1,2)	123	0.007	0.015

#### 4 总结

工业锂电池剩余使用寿命预测对于锂电池在工业领域的推广布局,提高能源利用率,环境保护方面具有重要意义.基于锂电池物理化学模型的扩展卡尔曼滤波算法需要考虑锂电池复杂的内部老化过程,而锂离子电池的老化过程是非线性的,退化因子非常多,用这类模型预测结果不理想,并且建模过程复杂.本文针对如何预测锂电池剩余使用寿命,不考虑锂电池老化的复杂过程,应用时间序列模型 ARIMA 对锂电池的老化过程进行建模预测,详细阐述了 ARIMA 模型在锂电池剩余使用寿命建模的数学原理以及工程实现.最后采用 NASA 提供数据仿真实践,仿真结果表明,在短期内 ARIMA 模型具有较高的预测精度和较强的适应能力.

#### 参考文献

- 刘大同,周建宝,郭力萌,彭宇. 锂离子电池健康评估和寿命预测综述. 仪器仪表学报, 2015, 36(1): 1-16.
- 朱亮标. 基于数据驱动的锂离子电池剩余寿命预测模型及软件实现[硕士学位论文]. 广州: 华南理工大学, 2014: 1-2.
- Bagul YG, Zeid I, Kamarthi SV. Overview of remaining useful life methodologies. ASME 2008 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Brooklyn, New York, USA. 2008. 1391-1400.

- 4 陶耀东, 李宁. 工业锂电池退化过程研究与剩余使用寿命预测. 计算机系统应用, 2017, 26(2): 235–239. [doi: [10.15888/j.cnki.csa.005587](https://doi.org/10.15888/j.cnki.csa.005587)]
- 5 马慧慧. EViews 统计分析与应用. 3 版. 北京: 电子工业出版社, 2016: 142–158.
- 6 孙轶轩, 邵春福, 计寻, 等. 基于 ARIMA 与信息粒化 SVR 组合模型的交通事时序预测. 清华大学学报 (自然科学版), 2014, 54(3): 348–353, 359.
- 7 Taneja K, Ahmad S, Ahmad K, *et al.* Time series analysis of aerosol optical depth over New Delhi using Box-Jenkins ARIMA modeling approach. Atmospheric Pollution Research, 2016, 7(4): 585–596. [doi: [10.1016/j.apr.2016.02.004](https://doi.org/10.1016/j.apr.2016.02.004)]
- 8 郑荣, 颜七笙. 基于 ARIMA 与 SVM 的国际铀资源价格预测. 计算机工程与应用, 2016, 52(1): 146–150.
- 9 李秀琴, 梁满发. 基于 ARIMA 模型的股票行情预测. 长春教育学院学报, 2013, 29(14): 47, 49.
- 10 Hassan J. ARIMA and regression models for prediction of daily and monthly clearness index. Renewable Energy, 2014, (68): 421–427. [doi: [10.1016/j.renene.2014.02.016](https://doi.org/10.1016/j.renene.2014.02.016)]