XXXX

基于物理基线残差增强与一致性筛选的锂离子电池荷电状态估计

黎渤宇[1]，李祎婧[2]，付俊成[1]，宋政湘[3]，孟锦豪[3]，郭安详[4]，王若谷[4]
（[1]西安交通大学国家储能技术产教融合创新平台（中心），陕西 西安 710049；[2]西安交通大学未来技术学院，陕西 西安 710049；[3]西安交通大学电气学院，陕西 西安 710049；[4]国网陕西省电力有限公司电力科学研究院，陕西 西安 710100）

摘 要：针对单电芯有限真实训练样本条件下荷电状态（state of charge，SOC）估计模型跨电芯、跨工况误差增大的问题，构建一种物理基线残差增强与标签-响应一致性筛选方法。该方法以开路电压-SOC关系和欧姆内阻构成的轻量等效电路模型（equivalent circuit model，ECM）建立端电压基线，采用变分自编码器（variational autoencoder，VAE）学习基线未表征的电压残差并生成候选样本；随后，利用由真实训练集训练并冻结的SOC代理模型，根据代理输出与合成标签的偏差及电流方向筛除明显错配样本。普通VAE分支与两组基线残差筛选分支通过非负加权和一维仿射校准进行融合，相关参数仅由跨单体验证集确定。基于公开A123磷酸铁锂电池数据集的跨工况实验表明，3种随机种子下，融合模型的归一化均方根误差（root mean square error，RMSE）均值为0.05471，较普通VAE平均降低10.25%，较经相同验证集校准的VAE平均降低9.72%，且各随机种子下均取得更低RMSE。结果表明，在同型号电芯和有限源域样本条件下，物理基线残差生成、数据驱动一致性筛选与验证集定权能够形成互补，改善跨电芯、跨工况SOC估计性能。

关键词：锂离子电池；荷电状态估计；数据增强；物理基线；一致性筛选
doi：10.19799/j.cnki.2095-4239.2026.0515
中图分类号：TM 912　　文献标志码：A　　文章编号：2095-4239（XXXX）XX-1-14

Lithium-Ion Battery State-of-Charge Estimation via Physics-Baseline Residual Augmentation and Consistency Screening

Li Boyu[1], Li Yijing[2], Fu Juncheng[1], Song Zhengxiang[3], Meng Jinhao[3], Guo Anxiang[4], Wang Ruogu[4]
([1]National Innovation Platform (Center) for Industry-Education Integration of Energy Storage Technology, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, China; [2]School of Future Technology, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, China; [3]School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, China; [4]Electric Power Research Institute, State Grid Shaanxi Electric Power Co., Ltd., Xi'an 710100, Shaanxi, China)

Abstract: A physics-baseline residual augmentation and label-response consistency screening method is proposed to reduce cross-cell and cross-operating-regime errors in state-of-charge (SOC) estimation when only limited real samples from a single source cell are

收稿日期：2026-06-14；修改稿日期：2026-08-16。
基金项目：国家电网公司总部科技项目（4000-202499063A-1-1-ZN）。
第一作者：黎渤宇（2003—），男，硕士研究生，研究方向为数据-机理混合的大数据电池诊断预警，E-mail：lbyv587@stu.xjtu.edu.cn；通信作者：孟锦豪，副教授，研究方向为储能系统设计与能量管理、电池建模与SOX估计、数据-机理混合的大数据电池诊断预警、电池系统测试方法与性能评价，E-mail：jinhao@xjtu.edu.cn。
引用本文：黎渤宇, 李祎婧, 付俊成, 等. 基于物理基线残差增强与一致性筛选的锂离子电池荷电状态估计[J]. 储能科学与技术, XXXX, XX(XX): 1-14.
Citation：Li Boyu, Li Yijing, Fu Juncheng, et al. Lithium-Ion Battery State-of-Charge Estimation via Physics-Baseline Residual Augmentation and Consistency Screening[J]. Energy Storage Science and Technology, XXXX, XX(XX): 1-14.

available. A lightweight equivalent circuit model (ECM) based on the open-circuit voltage-SOC relationship and ohmic resistance provides the terminal-voltage baseline, while a variational autoencoder (VAE) learns the unmodeled voltage residuals to generate candidate samples. An SOC proxy model trained exclusively on the real training set is frozen and used, together with a current-direction check, to remove candidates whose proxy outputs are inconsistent with their synthetic labels. The ordinary VAE branch and two screened baseline-residual branches are then combined through non-negative weighting and one-dimensional affine calibration, with all fusion parameters determined solely from the cross-cell validation set. Experiments on a public A123 lithium iron phosphate battery dataset show that the fused model achieves a mean normalized root mean square error (RMSE) of 0.05471 across three random seeds, reducing RMSE by an average of 10.25% relative to the ordinary VAE and 9.72% relative to the VAE calibrated on the same validation set; lower RMSE is obtained for every seed. These results demonstrate that physics-baseline residual generation, data-driven consistency screening, and validation-set weighting provide complementary benefits for cross-cell and cross-operating-regime SOC estimation within the same cell type and limited-source setting.

Keywords: lithium-ion battery; state-of-charge estimation; data augmentation; physics baseline; consistency screening

锂离子电池凭借高能量密度、长循环寿命和持续下降的成本[1]，已成为电动汽车动力系统与固定式储能电站的重要储能单元。在新能源消纳与电力市场的发展驱动下，储能配置与运行优化中也需要电池状态的精确感知[2]。电池管理系统（battery management system，BMS）承担状态监测、健康评估、均衡控制、热管理与过充过放保护等任务[3-5]，而能量管理策略的有效执行通常以荷电状态（state of charge, SOC）的准确估计为前提。SOC估计偏差一旦累积，不仅会造成车辆的续航失准，还会影响储能系统调频/调峰能力。因此，SOC是储能系统安全与经济运行的核心状态量[6-7]，其估计精度直接决定了能量管理策略的有效性。

SOC表示电池剩余可用电量与额定容量之比，受电池内部状态难以直接测量的限制，SOC须借助端电压、电流和温度等外部观测量间接推断。现有SOC估计算法大体可分为模型驱动、数据驱动和混合方法三类。其中，模型驱动方法以等效电路模型（equivalent circuit model，ECM）和卡尔曼滤波为代表，参数物理意义明确。尹康涌等[8]通过动态噪声自适应无迹卡尔曼滤波减小噪声协方差预设误差对SOC估计的影响；吴华伟等[9]采用结合改进遗忘因子递推最小二乘法（improved forgetting factor recursive least squares，IFFRLS）与交互多模型无迹卡尔曼滤波（interacting multiple model unscented kalman filter，IMMUKF）联合算法实现商用车磷酸铁锂电池在线参数辨识与多模型切换，提升了工况切换下的估计适应性；Li等[10]基于ECM完成锂离子电池SOC估计，证明了低阶电路模型在状态估计中的可用性；黎冲等[11]比较两类ECM在SOC-SOH联合估计中的表现，指出模型结构和参数辨识会影响联合估计精度。总体来看，基于模型的方法计算量低、可解释性强，但其跨工况鲁棒性仍依赖参数辨识精度；温度变化和老化会改变ECM参数，低阶模型也往往难以完整表达长时间常数极化和扩散过程。数据驱动方法以神经网络为代表，可从实测样本中拟合端电压、电流、温度与SOC之间的非线性关系，其可靠性主要受模型结构与训练样本覆盖度影响[12]。混合方法进一步将物理先验引入数据驱动模型，以兼顾可解释性与拟合能力。

从阻抗角度看，ECM参数与电池内部欧姆内阻、界面反应、电荷转移和扩散过程相关。电化学阻抗谱（electrochemical impedance spectroscopy，EIS）可在频域表征这些过程的动态响应，但不同过程的弧段常相互重叠[13-14]。弛豫时间分布

（distribution of relaxation times，DRT）将频域阻抗转换到弛豫时间尺度上，使不同过程以峰位、峰高和峰面积等特征呈现[15]。上述阻抗分析表明，电池端电压并非仅由开路电压和瞬时欧姆压降决定，还受到电荷转移、扩散、温度扰动以及老化状态变化等多种慢动态过程的影响。因此，低阶ECM未能显式描述的部分不能简单视为随机噪声，可理解为端电压响应中由多时间尺度极化、扩散和温度扰动等因素共同造成的残差项。这也为后续将端电压分解为ECM基线和残差提供了物理解释。

为弥补模型驱动方法在参数表达上的局限，数据驱动方法尝试从BMS实测样本中直接学习电压、电流与SOC之间的非线性关系。张远进等[16]提出自适应无迹卡尔曼滤波（adaptive unscented Kalman filter，AUKF）与反向传播（back propagation，BP）神经网络的联合估计方案，验证了滤波算法与神经网络互补的可行性。此类方法的可靠应用通常依赖覆盖充分、分布多样的训练样本，尤其在跨工况和跨电芯验证中需要足够样本支撑[17]。然而，在新电芯或新工况开发初期，可用真实样本数量有限，原有工况覆盖不足，电芯个体差异明显。数据稀缺以及跨电芯、跨工况和跨温度分布偏移共同限制了数据驱动SOC估计的泛化能力，使小样本条件下的SOC估计仍面临挑战。

数据增强是缓解小样本问题的常用方法。现有时序增强工作通常聚焦于提升实测样本的有效性。BMS采集端电压、电流和温度时会受到传感器噪声、采样不一致与电芯个体差异影响，因此，对已有样本施加局部变换是该类方法的主要思路。Wen等[18]系统归纳了加噪、缩放、裁剪、滑动窗口和时间扭曲等策略。Iwana等[19]通过大规模实验发现，时序增强收益高度依赖具体任务和数据分布。曹新宇等[20]采用改进天牛须搜索算法（improved beetle antennae search, IBAS）用于带外生输入的非线性自回归（nonlinear autoregressive with exogenous inputs, NARX）神经网络。训练样本扩充，提高了SOC估算的局部精度。Pohlmann等[21]研究了数据增强与特征选择对基于人工神经网络（artificial neural network，ANN）的SOC估计的影响。此类局部扰动增强可改善估计器对传感器噪声和小幅波动的鲁棒性，但其作用仍局限于已有样本邻域，难以构造跨工况条件下的电流-电压响应，也难以弥补源域样本对目标工况覆盖不足的问题。

与时序增强不同，生成式增强试图从真实训练样本中学习电池时序的底层分布，再从该分布中采样得到新的端电压、电流和温度窗口。Qiu等[22]提出条件生成对抗网络合成SOC估计样本，在小数据条件下改善了估计器精度。Hu等[23]将生成对抗网络（generative adversarial network，GAN）增强应用于跨电芯SOC估计，验证了生成样本在新电芯应用初期的适用性。Kingma等[24]提出的变分自编码器（variational auto-encoder，VAE）和Goodfellow等[25]提出的GAN为电池领域的生成式建模提供了基础架构。不过，电池窗口中的端电压响应严格受电流激励、极化过程与SOC共同约束。若直接生成完整端电压序列，可能出现电流方向与电压响应不匹配、SOC标签与电压波形不一致，或合成片段偏离EIS/DRT所揭示的极化特性。此类低质量样本进入混合训练集后，会削弱真实样本主导的电压-电流-SOC映射关系。

严格的物理约束方法通常将控制方程或守恒关系直接嵌入模型。Li等[26]将电极级反应方程嵌入神经网络，实现对电极内部状态的间接估计；Nascimento等[27]将基于Nernst与Butler-Volmer方程的降阶放电模型嵌入神经网络，并以数据驱动层表征非理想电压和老化相关参数。与上述方法不同，本文采用OCV-SOC关系和欧姆内阻构成的轻量ECM组织端电压基线，并由冻结SOC代理模型执行标签-响应一致性筛选；二者分别承担主体响应建模和候选样本质量控制，构成面向有限源域SOC增强的协同流程。

综上，有限源域样本下的SOC数据增强面临两个相互关联的问题：一是直接生成完整端电压时，生成模型需要同时学习可由低阶模型解释的主体响应和复杂残差；二是生成波形与合成SOC标签之间可能出现明显错配。

针对上述问题，本文的主要贡献概括如下：

（1）提出面向SOC时序增强的轻量物理基线-残差生成方法。利用OCV-SOC关系和欧姆内阻构造端电压主体响应，将生成对象由完整端电压转化为低阶ECM未表征的电压残差，使生成模型集中学习极化、温度扰动及老化差异共同形成的动态项；进一步通过零阶ECM与一阶RC ECM的受控基线替换，量化基线复杂度、参数可辨识性与跨域

稳定性之间的取舍。

（2）提出数据驱动的标签-响应一致性筛选方法。使用真实训练样本训练并冻结SOC代理模型，将代理输出与合成标签的偏差、电流方向与充放电响应关系转化为可执行的样本接受准则，在不改变下游SOC估计器结构的条件下，抑制标签与电压-电流响应明显错配的候选窗口进入训练集。

（3）针对单电芯有限样本下不同增强分支在跨电芯迁移中易受阈值、合成比例和初始化影响的问题，构建跨单体验证定权、跨工况测试严格隔离的融合评估流程。普通VAE分支与两组残差筛选分支的权重及仿射校准参数仅由未参与训练的跨单体验证集确定，测试集只用于最终评价；结合消融实验、ECM阶次对照、SOC区间分析、参数敏感性和随机种子重复实验验证分支互补性。在公开A123数据集的3种随机种子实验中，融合模型平均RMSE为0.05471，较普通VAE和经相同验证集校准的VAE分别降低10.25%和9.72%。

本文余下部分组织如下：第1节描述数据集划分与问题定义；第2节介绍物理基线残差增强、标签-响应一致性筛选及验证校准融合；第3节从整体误差、模块作用、ECM阶次、SOC区间、参数敏感性和随机种子重复性等方面进行分析；第4节总结主要结论及适用边界。

1 数据集划分与问题定义

1.1 公开数据集与样本构造

本研究采用Severson等[28]发布的公开锂离子电池数据集，数据对象为A123APR18650M1A磷酸铁锂电池。原始循环记录按电流符号和幅值划分为充电、放电和静置片段；本文仅保留具有明显电流激励的充放电片段，电流有效阈值为0.02A，最短片段长度为80个采样点。SOC标签由单循环内安时积分得到，并按该循环累计电量的最小值和最大值归一化到[0,1]。时序数据采用长度128、步长16的滑动窗口切分，输入通道为端电压、电流和温度，标签为窗口末端SOC。为模拟新电芯或新工况开发初期真实样本有限的情况，本文采用单电芯小样本训练、同工况跨单体验证和跨工况测试的划分方式。各工况簇及其用途见表1。

表1 数据集划分与用途

Table 1 Dataset partition and usage

工况簇	温度	充电倍率	放电倍率	电芯数量	本文用途
Regime 0	30℃	3.6C-8.0C	4.0C	18	源域；b1c2为真实训练电芯，其余电芯为跨单体验证集
Regime 1	35℃	4.8C-5.4C	4.0C	9	跨工况测试集
Regime 2	35℃	6.0C	4.0C	10	跨工况测试集
Regime 3	35℃	7.0C-8.0C	4.0C	9	跨工况测试集

训练样本按充放电片段和SOC分段进行分层抽样。具体而言，将SOC区间[0,1]按0.1间隔划分为10个子区间，并在不同充放电片段及SOC分箱内抽取训练窗口。当前实验中真实训练样本数为12,000，跨单体验证样本数为170,335，跨工况测试样本数为259,558。所有输入通道的标准化参数仅由真实训练集统计得到，并用于验证集、测试集和合成样本；验证集仅用于早停、分支权重搜索和仿射校准，测试集不参与模型选择。全部电芯均为同一公开数据集中的A123 APR18650M1A电芯，因此本文将实验范围界定为同型号、同一数据来源内部的跨电芯和跨工况泛化。

图1(a)展示了该数据集全生命周期的容量衰减曲线，可见容量随循环次数增加呈单调下降趋势，不同工况簇（Regime 0~3）对应循环区间内电池处于不同的老化阶段。图1(b)进一步对比了循环前期与后期的电压-SOC关系曲线，随着老化加剧，充放电曲线在SOC中段的电压平台区出现明显偏移与形变，表明跨循环阶段SOC-电压映射存在显著分布差异。图1(c)给出了一个典型的输入窗口示例，窗口长度为128个采样点，包含端电压、电流和温度三条时序通道，窗口末端SOC作为回归标签。

上述划分用于模拟单电芯源域训练以及同型号电芯跨电芯、跨工况评价：开发阶段仅使用一个电芯在有限工况下的真实窗口，部署评价则涉及同一数据集中的其他电芯和未参与训练的工况。在此设置下，生成式增强用于补充源域窗口的SOC与电压响应多样性。若直接生成完整端电压序列，可能出现电流方向、端电压响应与SOC标签错配。本

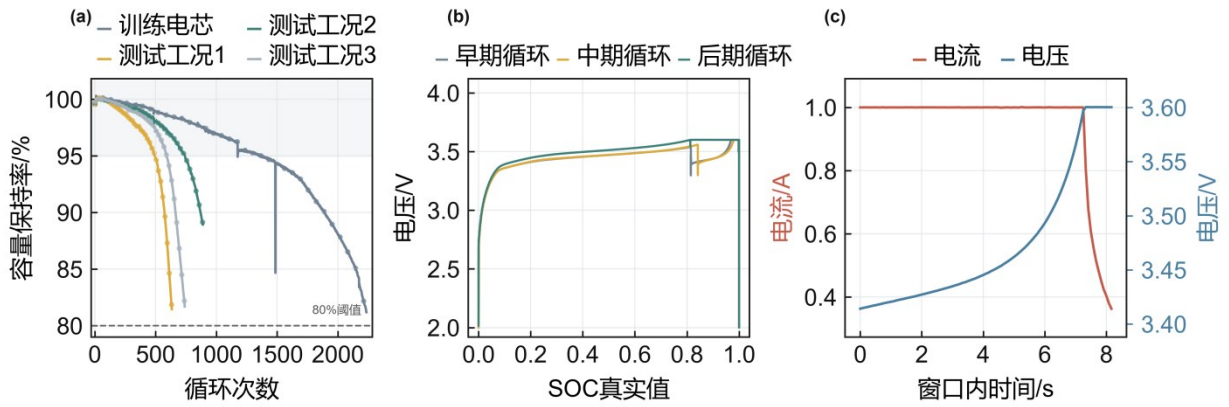

图1　生命周期容量衰减曲线、电压-SOC 曲线与典型输入窗口

Fig. 1　Lifecycle curve, voltage-SOC profiles, and a representative input window

文因此采用ECM基线-残差生成-数据驱动一致性筛选的流程。

设第i个样本的输入窗口为

$$\mathbf{X}_i = \left[\mathbf{V}_i, \mathbf{I}_i, \mathbf{T}_i\right] \in \mathbb{R}^{L\times 3} \tag{1}$$

其中N=128，V_i、I_i和T_i分别表示窗口内的端电压、电流和温度序列；对应SOC标签为$y_i\in[0,1]$。真实训练集记为

$$\mathcal{D}_{\text{real}} = \left\{\left(\mathbf{X}_i, y_i, s_i\right)\right\}_{i=1}^{N_r} \tag{2}$$

其中$s_i\in${充，放}为样本所属片段类型。跨单体验证集和跨工况测试集分别记为$\mathcal{D}_{val}$和$\mathcal{D}_{test}$。

本文的目标是在固定真实训练集规模下生成并筛选候选样本，构建真实-合成混合训练集，使下游SOC估计模型在未见电芯和未见工况上获得较低误差。经标签-响应一致性筛选后得到的合成样本集统一采用公式（3）中的符号表示，混合训练集定义为

$$\mathcal{D}_{\text{mix}} = \mathcal{D}_{\text{real}} \cup \mathcal{D}_{\text{syn}}^{\text{sample}} \tag{3}$$

其中，公式（3）所示合成样本集的样本数量由合成比例r控制。下游SOC估计器采用均方误差训练，并可降低合成样本的损失权重，以减弱低置信度候选对真实样本主导关系的干扰。

2　物理基线残差增强与一致性筛选方法

本文方法包括轻量ECM基线构造、残差VAE候选样本生成、数据驱动标签-响应一致性筛选以及混合训练与验证校准融合4个环节，如图2所示。ECM给出与SOC和电流相关的端电压主体响应，VAE学习基线之外的残差分布；候选窗口进入下游训练前，由冻结SOC代理模型和电流方向规则筛除明显错配样本。

2.1　ECM基线与电压残差表示

从阻抗建模角度看，ECM是对电池动态过程的低阶时域近似。本文主流程采用由OCV-SOC关系和欧姆内阻R_0构成的零阶基线，用于稳定表征端电压主体响应。零阶结构参数少、计算开销低，中频极化及更慢的扩散过程则保留在残差中。为比较不同基线复杂度下的残差增强效果，本文设置零阶ECM与一阶RC ECM对照：固定VAE生成的残差项、合成标签、候选数量、筛选模型和下游训练设置，仅改变候选端电压重组所采用的ECM基线，以考察新增极化支路对电压重构、跨域基线误差和下游SOC估计稳定性的影响。

对于给定的SOC和电流序列，零阶ECM端电压基线表示为

$$V_{\text{ECM}}(t) = \text{OCV}\left(y(t)\right) - I(t)R_0\left(y(t)\right) \tag{4}$$

一阶RC对照在零阶结构上增加极化支路。其离散极化状态和端电压基线分别为

$$U_p(t) = a_{\text{RC}}U_p(t-1) + \left(1-a_{\text{RC}}\right)R_1 I(t),\quad a_{\text{RC}} = \exp\left(-\frac{\Delta t}{\tau_1}\right) \tag{5}$$

$$V_{\text{1RC}}(t) = \text{OCV}\left(y(t)\right) - I(t)R_0 - U_p(t) \tag{6}$$

其中，Δt为采样间隔，τ_1为极化时间常数。R_0和R_1采用约束非负最小二乘辨识，取值均限制在[0， 0.05] Ω；a_{RC}从{0.50， 0.70， 0.85， 0.93，

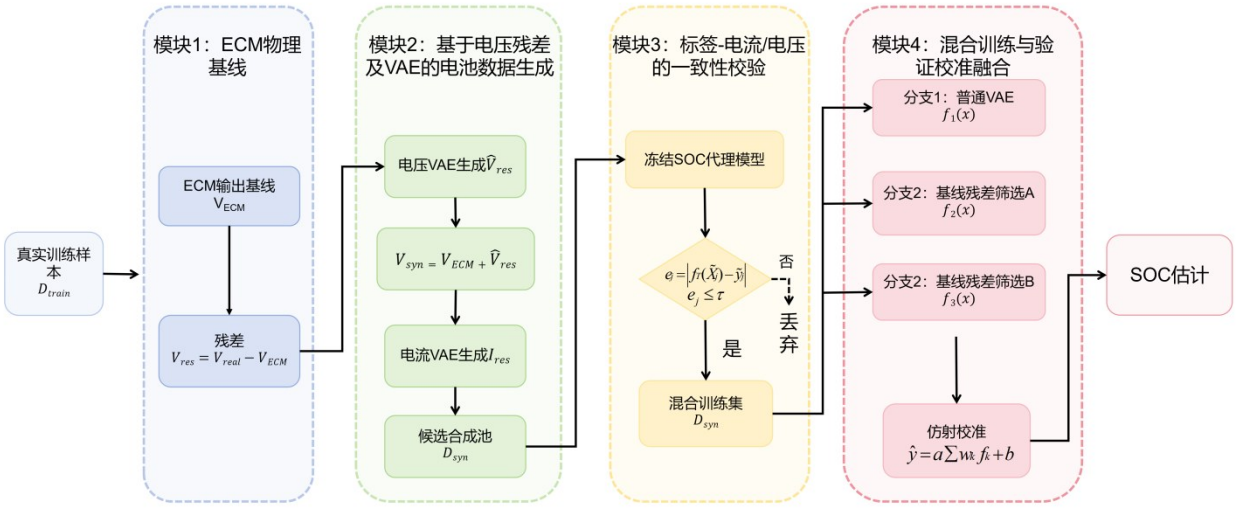

图2 物理基线残差增强与一致性筛选方法流程

Fig. 2 Workflow of physics-baseline residual augmentation and consistency screening

0.97，0.99}中通过训练组内80%/20%划分选择，极化状态按$U_p(0)=R_1I(0)$初始化。上述参数仅使用12,000个源域真实训练窗口辨识，跨单体验证集和跨工况测试集均不参与。

实际端电压与ECM基线之间的差值定义为

$$V_{\mathrm{res}}(t)=V(t)-V_{\mathrm{ECM}}(t) \tag{7}$$

零阶ECM保留OCV和瞬时欧姆压降，残差包含倍率相关极化滞后、温度扰动及更长时间尺度的扩散等未建模响应。参考DRT的多时间尺度解释，这些残差不宜简单视为随机噪声。一阶RC能够进一步吸收部分极化动态，但也增加R_1和时间常数的辨识要求，其实际收益需由严格隔离的验证与测试结果判断。

2.2 基于电压残差与VAE的候选样本生成

从频-时域关系看，DRT可将频域阻抗表征为不同弛豫时间尺度的响应。零阶基线未显式建模的中低频极化与扩散过程会进入残差，因此VAE学习的是确定性未建模动态与随机扰动的综合统计表征。一阶RC对照进一步将一个可辨识极化状态移入基线，用于检验残差范围缩小能否转化为下游SOC估计收益。两种阶次均使用相同真实训练源域、残差候选、标签和筛选模型，以避免把生成随机性误认为基线阶次效应。

VAE不直接学习完整端电压，而是学习式（7）定义的残差窗口。编码器将残差窗口映射为潜变量分布，解码器从潜变量恢复残差序列。其训练目标为

$$\mathcal{L}_{\mathrm{VAE}}=\mathcal{L}_{\mathrm{rec}}+\beta_{\mathrm{KL}}\mathcal{L}_{\mathrm{KL}} \tag{8}$$

其中，重构项约束生成残差与真实残差之间的差异，KL散度约束潜变量分布接近标准正态分布，β_{KL}为权重系数。生成阶段先在潜变量空间采样得到残差窗口，再与ECM端电压基线重组，得到候选窗口及其SOC标签。

为补充源域训练电芯中覆盖不足的SOC区间，候选样本按0.1的SOC间隔分段生成。普通VAE增强、ECM残差增强和ECM残差一致性筛选均使用同一真实训练源域构造候选样本池，差异仅在于是否引入ECM基线以及是否执行数据驱动一致性筛选。

2.3 标签-响应一致性筛选

生成样本能否用于下游SOC估计，不仅取决于波形是否接近真实样本，还取决于SOC标签是否与电压、电流和温度响应保持一致。本文仅使用真实训练样本训练SOC代理模型$f_T(\cdot)$，并在训练结束后冻结参数。对于第j个候选窗口及其合成标签，定义代理输出与合成标签之间的偏差为

$$e_j=\left|f_T(\tilde{\mathbf{X}}_j)-\tilde{y}_j\right| \tag{9}$$

当$e_j\le\tau$时，候选窗口通过标签-响应一致性筛选；否则不参与下游训练。除SOC偏差约束外，本文还依据电流符号检查候选窗口的充放电片段类型，剔除片段标记与电流方向不一致的样本。该过程以真实训练样本学习到的SOC映射关系对候选

标签与响应进行代理判别，并结合电流方向规则提高进入下游训练集的样本一致性。该筛选属于基于经验映射的样本质量控制，ECM物理基线与代理筛选分别承担主体响应建模和候选样本判别作用。本文方法面向固定、有限且非零的真实训练集，一致性筛选以现有样本能够训练具有基本判别能力的SOC代理模型为应用前提。

2.4 混合训练与验证校准融合

通过一致性筛选的合成样本按设定比例加入真实训练集。当真实训练窗口数为N_r、合成比例为r时，每次实验使用的合成窗口数为

$$N_s = \lfloor rN_r \rfloor \tag{10}$$

当请求数量超过筛选后候选池规模时，将N_s限制为候选池大小，以避免过多重复采样。混合训练损失定义为

$$\mathcal{L}_{\mathrm{mix}} = \frac{1}{N_r + N_s} \sum_{i=1}^{N_r+N_s} w_i \left[f_\theta(\mathbf{X}_i) - y_i \right]^2 \tag{11}$$

其中，真实样本权重设为1，合成样本权重记为λ_{syn}并由分支配置给定。最终模型包含普通VAE分支、基线残差筛选分支A（CS-A）和分支B（CS-B）。CS-A采用r=0.75、τ=0.08和λ_{syn}=0.50；CS-B采用r=0.50、τ=0.12和λ_{syn}=0.25。三分支预测通过

$$\hat{y}_{\mathrm{fuse}} = \alpha \sum_{k=1}^{3} \omega_k f_k(\mathbf{X}) + \beta, \quad \omega_k \geqslant 0, \sum_{k=1}^{3} \omega_k = 1 \tag{12}$$

进行融合与一维仿射校准。在单电芯有限样本条件下，不同增强分支对阈值、合成比例和初始化的敏感性会引起跨电芯性能波动。为此，非负权重ω_k以0.01为步长在单纯形上搜索，每组候选权重对应的α和β仅由跨单体验证集拟合。验证集RMSE最低的组合固定后直接用于跨工况测试集，测试数据不参与模型选择。该设置利用跨单体验证误差表征分支互补性，在严格测试集隔离条件下抑制单一增强分支的不稳定性。

3 电池SOC估计验证与结果分析

本节将面向新工况、跨电芯部署时的小样本场景，在跨工况测试集上对所提方法与各类基线方案进行对比，从整体精度、模块贡献、ECM阶次、SOC区间表现、参数敏感性以及种子重复性六个维度评估合成数据增强的有效性。所有方法均使用相同源域真实训练样本与标准化统计量，验证集仅用于早停、分支权重搜索与一维仿射校准，测试集严格隔离。

3.1 实验设置与评价指标

本文以SOC估计器在跨工况测试集上的误差评价合成样本的有效性。主估计器采用一维卷积特征提取层与GRU组合，输入为长度128的端电压、电流和温度三通道的归一化窗口，输出为单个SOC估计值。卷积部分包含两个一维卷积层，通道数分别为32和64，卷积核宽度均为5；GRU隐藏维度为64。为考察方法对不同估计器结构的适用性，另采用两层LSTM，隐藏维度为128，dropout为0.1。所有模型均采用AdamW优化器训练，学习率为1×10-3，批量大小为256，最大训练轮数为100。模型以跨单体验证集MSE作为早停指标。

ECM阶次对照采用GRU估计器，固定合成比例r=0.50、代理筛选阈值τ=0.10和合成损失权重λ_{syn}=0.50，并使用随机种子7、21和42重复训练。两种阶次均由同一12,000个真实训练窗口辨识基线，使用相同22,000个候选窗口的残差项、SOC标签、电流与温度窗口；跨单体验证集和跨工况测试集均不参与ECM参数辨识。

对比实验包含6类训练设置，具体含义见表2。GRU估计器用于比较全部设置，LSTM估计器用于考察筛选后的合成样本对另一类时序模型的影响。其中，PB-RVAE-VF表示仅使用验证集进行非负加权融合，PB-RVAE-VCF表示在验证加权基础上进一步进行一维仿射校准；PB专指物理基线（physics baseline）。

模型评价使用平均绝对误差（mean absolute error，MAE）和均方根误差（root mean square error，RMSE）：

$$\mathrm{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{13}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \tag{14}$$

其中，y_i为真实SOC标签，$\hat{y}_i$为估计值，N为评价窗口数。由于SOC已归一化至[0，1]，本文MAE和RMSE均为归一化SOC误差。除整体误差外，还按0.1的SOC间隔统计分箱RMSE。

3.2 跨工况的电池SOC估计结果

表3给出了不同训练设置在真实跨工况测试集

表2 对比训练设置说明

Table 2 Description of comparative training settings

训练设置	真实样本	合成样本来源	是否引入ECM残差	是否执行一致性筛选	用途
REAL	使用	无	否	否	真实小样本基线
NoiseAug	使用	真实窗口小幅噪声扰动	否	否	传统局部扰动增强基线
VAE	使用	普通 VAE 候选样本	否	否	生成式增强基线
PB-RVAE	使用	ECM 基线与残差 VAE 重组样本	是	否	检验残差生成本身的作用
PB-RVAE-CS	使用	通过标签-响应校验的 ECM 残差样本	是	是	检验数据驱动一致性筛选效果
PB-RVAE-VF	使用	VAE 与两组 PB-RVAE-CS 分支	是	是	验证集非负加权融合模型（未校准）
PB-RVAE-VCF	使用	VAE 与两组 PB-RVAE-CS 分支	是	是	最终验证校准融合模型

上的整体SOC估计误差。测试集仅包含Regime 1、Regime 2和Regime 3的真实窗口样本，未混入合成数据。所有方法均使用相同真实训练源域和标准化统计量。仅使用真实样本训练时，GRU的MAE为0.05568，RMSE为0.07593；NoiseAug可使RMSE降至0.07272，但降幅有限。

普通VAE增强后，GRU跨工况测试MAE为0.04263，RMSE为0.06382。未筛选的PB-RVAE的RMSE为0.07554，表明物理基线主要负责组织生成形式；加入一致性筛选后，调参配置的RMSE降至0.05974，说明标签-响应质量控制能够将残差生成转化为下游收益。

因此，本文将普通VAE分支与两组基线残差筛选分支进行验证校准融合。3种随机种子下，最终模型MAE均值为0.03857±0.00091，RMSE均值为0.05471±0.00221，较同种子普通VAE的RMSE平均降低10.25%，较校准VAE降低9.72%，且每个随机种子均取得更低RMSE。该结果说明，在当前数据划分下，分支互补与验证集定权能够降低单一增强分支的波动，非负加权和仿射校准在本文中主要承担稳定化作用。不同模型的整体误差对比如图3所示。

在LSTM估计器验证中，REAL的跨工况测试MAE为0.05121，RMSE为0.07407；使用PB-RVAE-CS后，MAE降至0.04905，RMSE降至0.06736，说明经一致性筛选的合成样本在另一类时序估计器上仍可降低RMSE。

表3 跨工况SOC估计误差

Table 3 SOC estimation errors under cross-operating-regime testing

下游模型	训练设置	关键参数	MAE	RMSE	RMSE 相对 REAL
GRU	REAL	无合成样本	0.05568	0.07593	-
GRU	NoiseAug	r=0.5	0.05176	0.07272	4.23%
GRU	VAE	r=1.0	0.04263	0.06382	15.95%
GRU	PB-RVAE	r=1.0	0.04694	0.07554	0.52%
GRU	PB-RVAE-CS	r=0.5, 默认阈值	0.04774	0.06723	11.45%
GRU	PB-RVAE-CS	r=0.5, τ=0.10，λ_{syn}=0.50	0.04103	0.05974	21.33%
LSTM	REAL	无合成样本	0.05121	0.07407	-
LSTM	PB-RVAE-CS	r=1.0, 默认阈值	0.04905	0.06736	9.05%
GRU	PB-RVAE-VF	VAE+CS-A+CS-B，验证加权（未校准）	0.04013±0.002794	0.05758±0.003706	较VAE均值降低5.64%
GRU	PB-RVAE-VCF	VAE+CS-A+CS-B，验证校准	0.03857±0.00091	0.05471±0.002207	较VAE均值降低10.25%

3.3 物理基线与一致性筛选的模块作用

表4对普通VAE、ECM残差生成和一致性筛选进行消融。普通VAE不使用物理基线和一致性筛选，跨工况RMSE为0.06382。PB-RVAE将端电压生成转化为物理基线上的残差补偿，但未筛选时RMSE升至0.07554，说明残差生成若不控制标签-响应关系，仍可能削弱真实样本主导的映射。

加入一致性筛选后，默认设置的RMSE降至0.06723。当阈值收紧至0.10且合成样本损失权重为0.5时，RMSE进一步降至0.05974。该配置在

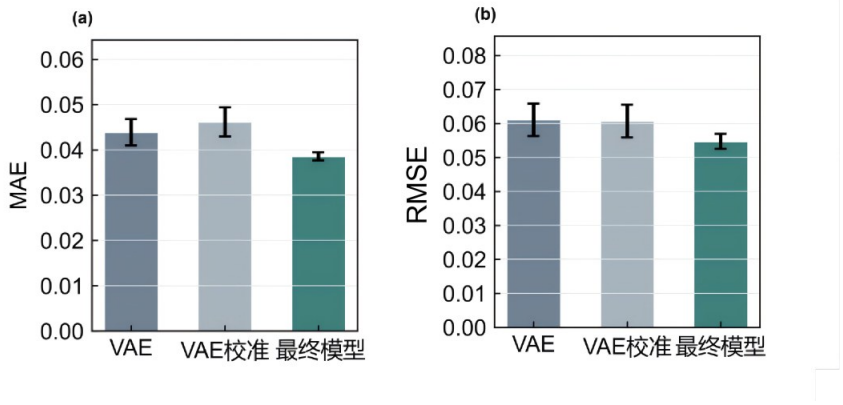

图3　跨工况测试误差对比

Fig. 3　Cross-operating-regime testing error comparison

22,000 个候选样本中保留 8,907 个，保留率为 40.49%，表明一致性筛选可在训练前剔除部分低质量候选样本。

鉴于单一残差筛选分支会随筛选阈值、合成比例和样本权重呈现性能差异，本文采用非负加权与仿射校准进行验证集定权：验证集表现较优的分支获得更高权重，组合固定后再用于测试集，测试集始终不参与选择。

表4　模块消融结果

Table 4　Module ablation results

方法	ECM 物理基线	一致性筛选	关键参数	MAE	RMSE
VAE	否	否	r=1.0	0.04263	0.06382
PB-RVAE	是	否	r=1.0	0.04694	0.07554
PB-RVAE-CS	是	是	r=0.5, 默认阈值	0.04774	0.06723
PB-RVAE-CS	是	是	r=0.5, τ=0.10，λ_{syn}=0.50	0.04103	0.05974

表4进一步展示了一致性筛选的作用。被丢弃样本主要表现为片段类型与电流方向不一致，或冻结 SOC 代理模型输出明显偏离合成标签。可能原因包括：VAE 在残差空间外推时生成了偏离源域极化趋势的电压片段；SOC 分箱采样将不同电流幅值下相似的电压形态错配到同一 SOC 区间；充放电片段标签与实际电流符号不一致。该结果表明，数据驱动筛选基于真实样本学习的经验映射执行样本质量控制，能够识别并移除标签-响应明显错配的候选窗口。

3.4　ECM 阶次对照与基线取舍

为隔离基线阶次的影响，本文固定 VAE 生成的残差项、合成标签、电流与温度窗口、22,000 个候选样本规模以及数据驱动筛选模型，仅将端电压从零阶 ECM 基线重组替换为一阶 RC ECM 基线。两种设置均采用 r=0.50、τ=0.10 和 λ_{syn}=0.50，并在相同的跨单体验证/跨工况测试隔离协议下使用 3 种随机种子训练 GRU。相关电压重构与 SOC 估计结果见表5。

表5　不同 ECM 基线复杂度下的电压重构与 SOC 估计结果

Table 5　Voltage reconstruction and SOC estimation with different ECM baseline orders

基线	新增参数	训练电压 RMSE/V	验证电压 RMSE/V	测试电压 RMSE/V	测试SOC MAE	测试SOC RMSE
零阶 ECM	—	0.11484	0.14809	0.16029	0.04981±0.00431	0.07005±0.00442
一阶 RC ECM	R_1，a_{RC}	0.11352	0.15388	0.16650	0.04991±0.00835	0.06864±0.00829

一阶 RC 使源训练集电压 RMSE 由 0.11484 V 降至 0.11352 V，降幅为 1.15%，但跨单体验证和跨工况测试电压 RMSE 分别由 0.14809 V、0.16029 V 升至 0.15388 V、0.16650 V，表明额外

极化参数对源训练瞬态存在一定拟合能力，但未改善未见电芯和工况上的基线泛化。下游SOC估计中，一阶RC使验证集RMSE由0.08329±0.00431降至0.07845±0.00359，跨工况测试RMSE由0.07005±0.00442小幅降至0.06864±0.00829；然而测试MAE由0.04981±0.00431变为0.04991±0.00835，且seed 42的测试RMSE由0.07191升至0.07530。

一阶RC在部分种子下可降低较大误差样本对RMSE的影响，但平均MAE没有同步改善，测试RMSE标准差也由0.00442增至0.00829。在当前仅含单一训练电芯、恒流片段占比较高且动态激励有限的源域中，R_1和时间常数的辨识更容易吸收源域特有瞬态。综合参数量、逐时状态递推开销、跨域电压基线误差和下游种子稳定性，这一结果支持将零阶ECM作为当前数据条件下兼顾复杂度与跨域稳定性的主流程基线。

3.5 SOC区间误差分析

为观察不同SOC区间的影响，本文按0.1间隔统计GRU跨工况测试集RMSE，结果见表6。与REAL相比，调参后的PB-RVAE-CS在0.0-0.8的多数SOC区间降低RMSE，其中0.1-0.2、0.2-0.3、0.3-0.4区间降幅分别为39.7%、38.7%和41.6%。

与VAE相比，PB-RVAE-CS在0.1-0.4和0.6-0.8区间取得更低RMSE，而0.0-0.1、0.5-0.6和0.9-1.0区间仍由VAE取得更低RMSE，表明该方法具有明确的SOC区间依赖性。分段RMSE对比如图4所示。

SOC分段误差体现了基线-残差结构的区间依赖性。在0.1-0.4 SOC区间，OCV曲线相对平缓且欧姆内阻变化较小，零阶基线可近似端电压主体响应；在0.6-0.8 SOC区间，筛选后的残差样本补充了源域覆盖。0.5-0.6和0.9-1.0 SOC区间的退化则表明，未由零阶ECM描述的极化动态可能被VAE误学为可泛化残差。ECM阶次对照显示，一阶RC能够改善部分随机种子的SOC RMSE，但并未同步改善平均MAE，且种子间波动增大，表明更高阶基线的局部RMSE收益需要与平均MAE和种子稳定性综合评价。

表6 SOC分段RMSE对比

Table 6 Comparison of RMSE across SOC intervals

SOC区间	REAL	VAE	PB-RVAE-CS	相对REAL	相对VAE
0.0-0.1	0.0337	0.0183	0.0209	38.2%	−13.8%
0.1-0.2	0.0642	0.0433	0.0387	39.7%	10.6%
0.2-0.3	0.0683	0.0553	0.0419	38.7%	24.3%
0.3-0.4	0.0789	0.0617	0.0461	41.6%	25.4%
0.4-0.5	0.0846	0.0584	0.0610	28.0%	−4.4%
0.5-0.6	0.0878	0.0572	0.0813	7.4%	−42.1%
0.6-0.7	0.1104	0.1099	0.0987	10.7%	10.2%
0.7-0.8	0.1236	0.1395	0.1050	15.0%	24.7%
0.8-0.9	0.1010	0.1115	0.1015	−0.5%	8.9%
0.9-1.0	0.0327	0.0308	0.0358	−9.6%	−16.4%

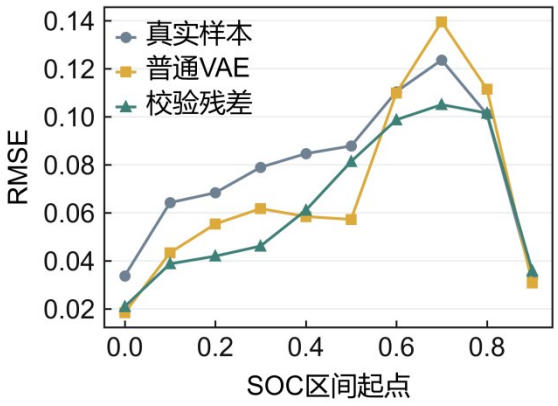

图4 SOC分段RMSE对比

Fig. 4 RMSE comparison across SOC intervals

3.6 参数敏感性与选参口径

为分析参数敏感性，本文在PB-RVAE-CS下扫描3个超参数：合成比例r、SOC筛选阈值τ和合成损失权重λ_{syn}。其中，r表示加入训练集的合成窗口数与真实训练窗口数之比；τ表示冻结SOC校验模型预测值与候选样本SOC标签之间允许的最大偏差；λ_{syn}表示下游SOC估计器训练中合成样本损失相对于真实样本损失的权重。结果表明，GRU跨工况测试RMSE在不同配置之间波动较大，最低值为0.05974，较高配置可达到0.08677，说明合成样本规模并非越大越好。

为避免测试集参与模型选择，表7第一行仅作为事后敏感性分析候选，用于说明单一物理分支在某些参数组合下可能达到的测试误差，该行不参与最终模型权重和校准参数确定。实际选参仅依据跨单体验证集RMSE，较优配置为r=0.75、τ=0.08、λ_{syn}=0.5，对应测试MAE为0.04442，RMSE为0.06268。最终，融合模型不采用事后测试表现最佳的单一分支，而用验证集非负权重降低偶然性。

参数变化对测试RMSE的影响如图5所示。

上述选参方式将参数选择与最终测试严格分离，验证集负责判断分支互补性与校准偏差，测试集仅用于报告固定模型在未见工况上的真实误差，由此，避免将单次测试集最优配置误认为方法的稳定优势。

表7　选参口径对比

Table 7　Comparison of parameter-selection criteria

选参口径	r	τ	λ_{syn}	验证RMSE	测试MAE	测试RMSE
测试集最优候选	0.50	0.10	0.50	0.07781	0.04103	0.05974
验证集选参候选	0.75	0.08	0.50	0.06586	0.04442	0.06268

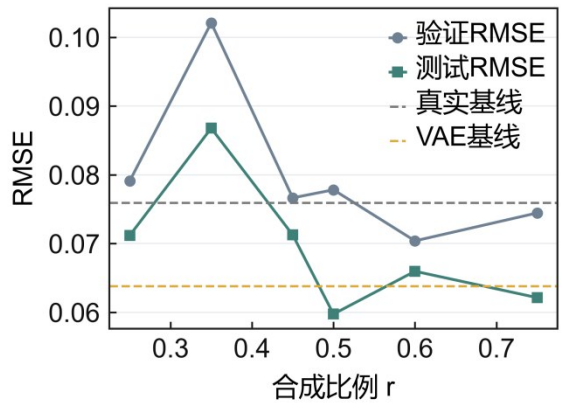

图5　参数变化对RMSE的影响

Fig. 5　Influence of parameters on RMSE

综上，基线残差生成与数据驱动一致性筛选在部分参数组合下降低了跨工况RMSE，但单一分支收益对合成比例、筛选阈值和初始化敏感。常规验证校准融合根据跨单体验证集表现分配普通VAE分支与残差筛选分支的贡献，用于提高当前重复实验的稳定性。

3.7　随机种子重复性验证

为判断最终模型在当前设置下的重复性，本文采用3种随机种子重复实验，结果见表8。随机种子控制模型初始化、候选样本采样和训练批次顺序；每个种子分别训练普通VAE分支、基线残差筛选分支A和B，并仅利用跨单体验证集拟合非负融合权重和一维仿射校准参数，测试阶段不再调整参数。

结果表明，PB-RVAE-VCF在seed 7、seed 21和seed 42下的测试RMSE分别为0.05248、0.05474和0.05690，均低于同种子的普通VAE。3种随机种子下，最终模型MAE均值为0.03857±0.00091，RMSE均值为0.05471±0.00221；与同种子普通VAE相比，RMSE平均降低10.25%。这表明，在单电芯有限训练样本条件下，基于跨单体验证集的定权与校准能够稳定利用普通VAE与基线残差筛选分支的互补信息，且未使用跨工况测试样本参与参数选择。不同随机种子下的RMSE和分支权重分别如图6和图7所示。

表8　随机种子重复结果

Table 8　Results of repeated experiments with random seeds

随机种子	VAE测试RMSE	VAE校准RMSE	融合权重(VAE;CS-A;CS-B)	校准参数(α,β)	MAE	RMSE	RMSE相对VAE
Seed 7	0.05721	0.05624	0.42;0; 0.58	0.93, 0.035	0.03769	0.05248	8.27%
Seed 21	0.05959	0.06014	0.19;0.37;0.44	0.945,0.03	0.0385	0.05474	8.14%
Seed 42	0.06644	0.06579	0.14;0.86;0	0.938,0.032	0.03951	0.05690	14.36%
均值±标准差	0.06108±0.004790	0.06072±0.004798	-	-	0.03857±0.00091	0.05471±0.002207	10.25%

随机种子重复实验中，三组融合权重存在明显差异。seed 7的权重集中于VAE与CS-B分支，seed 21的三分支权重相对均衡，seed 42则主要依赖CS-A分支。三组权重差异表明，各增强分支的相对优势会随初始化与候选采样发生变化，因此需要由跨单体验证集完成定权。跨单体验证集对三分支进行统一评价后，权重会倾向当前种子下验证误差更低的分支，并由仿射校准修正残余偏差。该过程体现的是验证集隔离条件下的常规模型选择作用。

综合上述实验，物理基线负责组织残差生成空间，一致性筛选负责剔除明显错配窗口，跨单体验证定权进一步利用普通VAE与两组残差筛选分支的互补性；三者协同使3种随机种子下融合模型相对普通VAE的RMSE平均降低10.25%。ECM阶次对照表明，在当前单电芯有限源域条件下，零阶基线

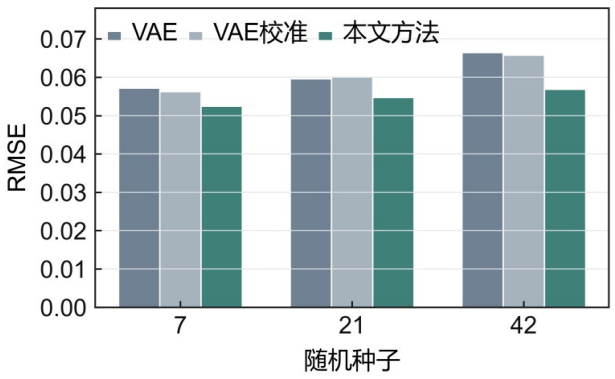

图6 不同随机种子下RMSE对比

Fig. 6 RMSE comparison under different random seeds

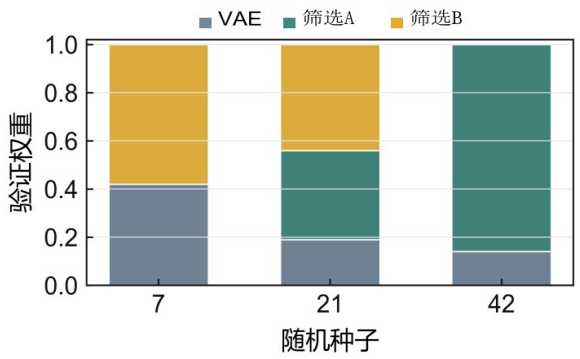

图7 不同随机种子下分支权重

Fig. 7 Branch weights under different random seeds

在参数开销、跨域电压误差与种子稳定性之间取得了更合适的平衡；SOC区间与敏感性分析则明确了筛选分支的主要收益区间和验证集选参依据。据此，本文将实验结论界定为同型号A123电芯、固定非零真实训练集条件下的跨电芯和跨工况SOC估计。

4 结论与展望

本文面向单电芯有限真实训练样本下的跨电芯、跨工况SOC估计，构建了物理基线残差增强与一致性筛选流程PB-RVAE-VCF。该流程以零阶ECM形成端电压主体响应，由VAE学习基线之外的残差并生成候选窗口，再利用真实训练集训练并冻结的SOC代理模型执行标签-响应一致性筛选；普通VAE分支与两组残差筛选分支的非负权重和仿射校准参数只由跨单体验证集确定。

公开A123磷酸铁锂电池数据集的跨工况测试表明，PB-RVAE-VCF在3种随机种子下的平均MAE为0.03857、平均RMSE为0.05471，较普通VAE和经相同验证集校准的VAE分别降低10.25%和9.72%，且各随机种子下均取得更低RMSE。在固定残差项的ECM基线替换对照中，一阶RC的平均测试RMSE由0.07005±0.00442小幅降至0.06864±0.00829，但平均MAE没有改善，测试RMSE波动同步增大，跨工况电压基线RMSE也高于零阶结果。因此，本文在当前有限源域条件下保留零阶ECM作为兼顾复杂度与跨域稳定性的轻量基线。

本研究的验证范围为同型号A123电芯、固定非零真实训练集条件下的跨电芯和跨工况SOC估计。一致性筛选以有限真实样本能够训练具有基本判别能力的SOC代理模型为应用前提。后续将结合包含脉冲与动态工况激励的独立数据，分别重建不同ECM阶次下的残差生成器，并进一步评估温度、倍率、老化程度和电芯型号变化下的泛化性能。

参 考 文 献

[1] WARD L, BABINEC S, DUFEK E J, et al. Principles of the Battery Data Genome[J]. Joule, 2022, 6(10): 2253-2271. DOI: 10.1016/j.joule.2022.08.008.

[2] 崔嘉雁, 杨苹, 李丰能, 等. 电力市场环境下基于IGDT的风电场储能优化配置[J]. 电力工程技术, 2025, 44(3): 18-29. DOI: 10.12158/j.2096-3203.2025.03.002.
CUI Jiayan, YANG Ping, LI Fengneng, et al. Optimal configuration of wind farm energy storage based on IGDT in the power market environment[J]. Electric Power Engineering Technology, 2025, 44(3): 18-29. DOI: 10.12158/j. 2096-3203.2025.03.002.

[3] 董晓红, 董进波, 王明深, 等. 基于新型健康特征的锂电池健康状态快速估计方法[J]. 电力工程技术, 2025, 44(1): 136-142, 206. DOI: 10.12158/j.2096-3203.2025.01.014.
DONG Xiaohong, DONG Jinbo, WANG Mingshen, et al. Rapid estimation method of lithium battery state of health based on novel health feature[J]. Electric Power Engineering Technology, 2025, 44(1): 136-142, 206. DOI: 10.12158/j. 2096-3203.2025.01.014.

[4] 郑启达, 赵谡, 汪彪, 等. 基于在线顺序极限学习机模型的锂离子电池健康状况预测[J]. 电力工程技术, 2026, 45(2): 51-59. DOI: 10.12158/j.2096-3203.2026.02.006.
ZHENG Qida, ZHAO Su, WANG Biao, et al. Lithium-ion battery health prediction based on online sequential extreme learning machine model[J]. Electric Power Engineering Technology, 2026, 45(2): 51-59. DOI: 10.12158/j.2096-3203.2026.02.006.

[5] 邓小欣, 高思航, 陈佳佳, 等. 基于LSTM-XGBoost的锂电池健康状

态与剩余寿命联合评估方法[J]. 电力工程技术, 2026, 45(5): 148-156. DOI: 10.12158/*j*.2096-3203.2026.05.014.
DENG Xiaoxin, GAO Sihang, CHEN Jiajia, et al. A joint LSTM-XGBoost-based method for predicting the state of health and remaining useful life of lithium batteries[J]. Electric Power Engineering Technology, 2026, 45(5): 148-156. DOI: 10.12158/j.2096-3203.2026.05.014.

[6] 付诗意, 吕桃林, 闵凡奇, 等. 电动汽车用锂离子电池SOC估算方法综述[J]. 储能科学与技术, 2021, 10(3): 1127-1136. DOI: 10.19799/*j*.cnki.2095-4239.2021.0013.
FU Shiyi, LYU Taolin, MIN Fanqi, et al. Review of estimation methods on SOC of lithium-ion batteries in electric vehicles[J]. Energy Storage Science and Technology, 2021, 10(3): 1127-1136. DOI: 10.19799/j.cnki.2095-4239.2021.0013.

[7] LING C. A review of the recent progress in battery informatics[J]. npj Computational Materials, 2022, 8: 33. DOI: 10.1038/s41524-022-00713-x.

[8] 尹康涌, 孙磊, 李浩秒, 等. 基于动态噪声自适应无迹卡尔曼滤波的锂离子电池SOC估计[J]. 储能科学与技术, 2024, 13(11): 4065-4077. DOI: 10.19799/*j*.cnki.2095-4239.2024.0546.
YIN Kangyong, SUN Lei, LI Haomiao, et al. SOC estimation of lithium-ion batteries based on DN-AUKF[J]. Energy Storage Science and Technology, 2024, 13(11): 4065-4077. DOI: 10.19799/j.cnki.2095-4239.2024.0546.

[9] 吴华伟, 何成泽, 洪强, 等. 基于IFFRLS-IMMUKF的商用车磷酸铁锂电池SOC估算[J]. 储能科学与技术, 2025, 14(10): 3996-4008. DOI: 10.19799/*j*.cnki.2095-4239.2025.0194.
WU Huawei, HE Chengze, HONG Qiang, et al. IFFRLS-IMMUKF-based estimation of the state of charge of lithium iron phosphate batteries for commercial vehicles[J]. Energy Storage Science and Technology, 2025, 14(10): 3996-4008. DOI: 10.19799/j.cnki.2095-4239.2025.0194.

[10] 李清波, 张懋慧, 罗英, 等. 基于等效电路模型融合电化学原理的锂离子电池荷电状态估计[J]. 储能科学与技术, 2024, 13(9): 3072-3083. DOI: 10.19799/*j*.cnki.2095-4239.2024.0594.
LI Qingbo, ZHANG Maohui, LUO Ying, et al. Lithium-ion battery state of charge estimation based on equivalent circuit model[J]. Energy Storage Science and Technology, 2024, 13(9): 3072-3083. DOI: 10.19799/j.cnki.2095-4239.2024.0594.

[11] 黎冲, 韩毅, 黄凤保, 等. 基于两种等效电路模型的锂离子电池SOC和SOH联合估计算法性能对比[J]. 储能科学与技术, 2022, 11(5): 1574-1582.
LI Chong, HAN Yi, HUANG Fengbao, et al. Performance comparison of joint estimation algorithms for SOC and SOH of lithium-ion batteries based on two equivalent circuit models[J]. Energy Storage Science and Technology, 2022, 11(5): 1574-1582.

[12] NG M F, ZHAO J, YAN Q, et al. Predicting the state of charge and health of batteries using data-driven machine learning[J]. Nature Machine Intelligence, 2020, 2(3): 161-170. DOI: 10.1038/s42256-020-0156-7.

[13] CIUCCI F, CHEN C. Analysis of electrochemical impedance spectroscopy data using the distribution of relaxation times: A Bayesian and hierarchical Bayesian approach[J]. Electrochimica Acta, 2015, 167: 439-454. DOI: 10.1016/j.electacta.2015.03.123.

[14] PAUL T, CHI P W, WU P M, et al. Computation of distribution of relaxation times by Tikhonov regularization for Li ion batteries: usage of L-curve method[J]. Scientific Reports, 2021, 11: 12624. DOI: 10.1038/s41598-021-91871-3.

[15] MERTENS A, GRANWEHR J. Two-dimensional impedance data analysis by the distribution of relaxation times[J]. Journal of Energy Storage, 2017, 13: 401-408. DOI: 10.1016/j.est.2017.07.029.

[16] 张远进, 吴华伟, 叶从进. 基于AUKF-BP神经网络的锂电池SOC估算[J]. 储能科学与技术, 2021, 10(1): 237-241. DOI: 10.19799/*j*.cnki.2095-4239.2020.0285.
ZHANG Yuanjin, WU Huawei, YE Congjin. Estimation of the SOC of a battery based on the AUKF-BP algorithm[J]. Energy Storage Science and Technology, 2021, 10(1): 237-241. DOI: 10.19799/j.cnki.2095-4239.2020.0285.

[17] HARRIS S J, NOACK M M. Statistical and machine learning-based durability-testing strategies for energy storage[J]. Joule, 2023, 7(5): 920-934. DOI: 10.1016/j.joule.2023.03.008.

[18] WEN Q, SUN L, YANG F, et al. Time Series Data Augmentation for Deep Learning: A Survey[C]//Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence. 2021: 4653-4660. DOI: 10.24963/ijcai.2021/631.

[19] IWANA B K, UCHIDA S. An empirical survey of data augmentation for time series classification with neural networks [J]. PLOS ONE, 2021, 16(7): e0254841. DOI: 10.1371/journal.pone.0254841.

[20] 曹新宇, 彭飞, 李立伟, 等. 基于IBAS-NARX神经网络的锂电池荷电状态估计[J]. 储能科学与技术, 2021, 10(6): 2342-2351. DOI: 10.19799/*j*.cnki.2095-4239.2021.0291.
CAO Xinyu, PENG Fei, LI Liwei, et al. SOC estimation of lithium battery based on IBAS-NARX neural network model[J]. Energy Storage Science and Technology, 2021, 10(6): 2342-2351. DOI: 10.19799/j.cnki.2095-4239.2021.0291.

[21] POHLMANN S, MASHAYEKH A, KUDER M, et al. Data Augmentation and Feature Selection for the Prediction of the State of Charge of Lithium-Ion Batteries Using Artificial Neural Networks[J]. Energies, 2023, 16(18): 6750. DOI: 10.3390/en16186750.

[22] QIU X, WANG S, CHENG K. A conditional generative adversarial network-based synthetic data augmentation technique for battery state-of-charge estimation[J]. Applied Soft Computing, 2023, 142: 110281. DOI: 10.1016/j.asoc.2023.110281.

[23] HU C, CHENG F, ZHAO Y, et al. State of charge estimation for lithium-ion batteries based on data augmentation with generative adversarial network[J]. Journal of Energy Storage, 2024, 80: 110004. DOI: 10.1016/j.est.2023.110004.

[24] KINGMA D P, WELLING M. Auto-Encoding Variational Bayes[C]// Proceedings of the 2nd International Conference on Learning Representations. 2014.

[25] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative Adversarial Nets[C]//Advances in Neural Information

Processing Systems 27. 2014: 2672-2680.

[26] LI W, ZHANG J, RINGBECK F, et al. Physics-informed neural networks for electrode-level state estimation in lithium-ion batteries[J]. Journal of Power Sources, 2021, 506: 230034. DOI: 10.1016/j.jpowsour.2021.230034.

[27] NASCIMENTO R G, CORBETTA M, KULKARNI C S, et al. Hybrid physics-informed neural networks for lithium-ion battery modeling and prognosis[J]. Journal of Power Sources, 2021, 513: 230526. DOI: 10.1016/j.jpowsour.2021.230526.

[28] SEVERSON K A, ATTIA P M, JIN N, et al. Data-driven prediction of battery cycle life before capacity degradation[J]. Nature Energy, 2019, 4(5): 383-391. DOI: 10.1038/s41560-019-0356-8.