UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Data-Driven State-of-Health Estimation and Optimal Energy Dispatch of Battery Energy Storage Systems

Permalink

https://escholarship.org/uc/item/3fz1j1pj

ISBN

9798193298496

Author

Ling, Long

Publication Date

2026-10-15

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA SAN DIEGO

Data-Driven State-of-Health Estimation and Optimal Energy Dispatch of Battery Energy Storage Systems

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Engineering Sciences (Mechanical Engineering)

by

Long Ling

Committee in charge:

Professor Jan Kleissl, Chair
Professor Zheng Chen
Professor Raymond de Callafon
Professor Yuanyuan Shi

2026

Copyright

Long Ling, 2026

All rights reserved.

The Dissertation of Long Ling is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2026

TABLE OF CONTENTS

Dissertation Approval Page iii

Table of Contents iv

List of Figures vii

List of Tables ix

List of Abbreviations x

Acknowledgements xii

Vita xiii

Abstract of the Dissertation xiv

Chapter 1 Introduction 1
1.1 Motivation 1
1.2 Research Questions 2
1.3 Approach and Overview of the Three Studies 2
1.4 Contributions 5
1.5 Organization of the Dissertation 6

Chapter 2 Lithium-Ion Battery State-of-Health Estimation Based on Probability Density Functions under Randomized Cycling 7
2.1 Introduction 7
2.2 Dataset Description 13
2.3 Proposed SoH Estimation Framework 14
2.3.1 Feature Extraction 15
2.3.2 Estimation Algorithm 24
2.4 Results and Discussion 27
2.4.1 Comparative Analysis of the Proposed Framework 27
2.4.2 Verification of Generalization Capability 31
2.5 Downsampling of the Dataset 32
2.6 Conclusion 34

Chapter 3 Masked-Autoencoder Pretraining for State-of-Health Estimation of Second-Life Battery Packs 36
3.1 Introduction 36
3.2 Related Work 41
3.2.1 SoH Estimation on Randomized Battery Usage Data 41
3.2.2 Self-Supervised and Pretrained Models for Battery Health 41
3.2.3 Second-Life Battery Screening 43

3.3 Datasets . . . 44
3.3.1 NASA Randomized Battery Usage Dataset . . . 45
3.3.2 Second-Life Automotive Battery-Pack Datasets . . . 46
3.3.3 Common Preprocessing . . . 50
3.3.4 Cross-Dataset Comparability and Data Release . . . 52
3.4 Proposed Framework . . . 53
3.4.1 Cell-Level Stratified Data Preparation . . . 54
3.4.2 Masked-Autoencoder Pretraining . . . 55
3.4.3 Supervised Fine-Tuning . . . 57
3.4.4 Supervised Baselines . . . 60
3.4.5 Cross-Corpus Transfer Experimental Design . . . 62
3.4.6 Hyperparameter Optimization . . . 63
3.4.7 Evaluation Metrics . . . 65
3.5 Results . . . 66
3.5.1 Pretraining: Masked Reconstruction . . . 66
3.5.2 Public-Benchmark Performance . . . 68
3.5.3 Transfer to Second-Life Systems . . . 72
3.5.4 Summary of Evidence . . . 78
3.6 Discussion . . . 79
3.7 Conclusion . . . 83

Chapter 4 An Optimal Dispatch Model for Economic Benefit Valuation of Grid Energy Storage . . . 86
4.1 Introduction . . . 86
4.2 Problem Formulation . . . 88
4.2.1 Objective Function . . . 88
4.2.2 Constraints . . . 89
4.2.3 Relaxation Techniques . . . 90
4.3 Case Study . . . 92
4.3.1 Price Time Series Data . . . 92
4.3.2 Battery Modeling . . . 93
4.3.3 Optimization Setup . . . 95
4.4 Results and Discussion . . . 96
4.5 Conclusion . . . 98

Chapter 5 Conclusions . . . 100
5.1 Summary of Contributions . . . 100
5.2 Synthesis: From Estimated Health to Dispatched Value . . . 102
5.3 Future Directions . . . 103

Appendix A Supplementary Material for Chapter 3 . . . 105
A.1 Pretraining Corpus Composition . . . 105
A.2 Per-Seed Results . . . 106

Bibliography 107

LIST OF FIGURES

Figure 2.1. Cycling procedures for the NASA RW dataset. 14

Figure 2.2. Representative cell current, voltage, and temperature data from the NASA RW dataset. 15

Figure 2.3. Flowchart of the proposed SoH estimation methodology. 16

Figure 2.4. Estimated and reconstructed PDFs of voltage and temperature data for battery RW9. 17

Figure 2.5. Feature extraction from sample PDFs. 20

Figure 2.6. SoH degradation of the RW9–RW12 batteries. 22

Figure 2.7. Variation of features extracted from PDFs for battery RW9. 24

Figure 2.8. Pearson's correlation between PDF features and SoH for RW9–RW11. . . . 25

Figure 2.9. SoH estimation results for RW9–RW11 with different feature sets. 31

Figure 2.10. Impact of downsampling on feature correlation and SoH estimation accuracy. 33

Figure 3.1. Second-life battery test rig. 47

Figure 3.2. Reference and aging test profiles for the lab packs. 50

Figure 3.3. SoH-label distributions of the four datasets. 52

Figure 3.4. Two-stage masked-autoencoder SoH estimation framework. 53

Figure 3.5. Masked reconstruction on held-out NASA segments at three degradation stages. 68

Figure 3.6. Predicted versus true SoH on held-out test units. 71

Figure 3.7. Transfer-learning ablation on the three second-life systems. 73

Figure 4.1. Rint battery model. 90

Figure 4.2. Hourly real-time electricity prices at the La Jolla node in 2022 and their four clusters. 93

Figure 4.3. BESSs deployed at UCSD. 94

Figure 4.4. *SoC*-OCV curves of the three batteries. 95

Figure 4.5. Daily revenue and cycles for four price clusters and three BESSs. 97

Figure 4.6. Dispatch trajectories for the most representative price cluster. 98

LIST OF TABLES

Table 2.1. Distribution parameters requiring identification. 22

Table 2.2. Comparative SoH estimation errors across algorithms and feature sets. 29

Table 2.3. SoH estimation results across four NASA RW test groups. 32

Table 2.4. Growth rates of SoH estimation errors under downsampling. 34

Table 3.1. Summary of the four datasets. 45

Table 3.2. Per-sample clip bounds applied at data-loading time. 51

Table 3.3. Two-stage fine-tuning schedule. 60

Table 3.4. Pretraining conditions evaluated per fine-tune target. 63

Table 3.5. Hyperparameter search space and selected values. 64

Table 3.6. Masked-reconstruction error on held-out NASA segments. 69

Table 3.7. SoH estimation on the NASA Randomized Battery Usage benchmark. 69

Table 3.8. Leave-one-cell-out generalization on NASA Randomized Battery Usage. . . 70

Table 3.9. Transfer-learning ablation on the three second-life systems. 72

Table 3.10. Seen-target control on the three second-life systems. 74

Table 4.1. Battery parameters and operating limits. 94

Table 4.2. Identified battery model parameters. 95

Table 4.3. Aggregated daily optimization results. 97

Table A.1. Composition of the combined four-source pretraining pool. 105

Table A.2. Per-seed test RMSE for all transfer configurations. 106

LIST OF ABBREVIATIONS

AC	alternating current
ANFIS	adaptive neural fuzzy inference system
BESS	battery energy storage system
BiLSTM	bidirectional long short-term memory
BMS	battery management system
CAISO	California Independent System Operator
CC	constant current
CCCV	constant current constant voltage
CI	confidence interval
CNN	convolutional neural network
CV	constant voltage
DC	direct current
DVC	differential voltage curve
ECM	equivalent circuit model
EM	expectation–maximization
EMA	exponential moving average
ESS	energy storage system
ETS	energy time shifting
EV	electric vehicle
FFNN	feed-forward neural network
GPR	Gaussian process regression
GRU	gated recurrent unit
HPO	hyperparameter optimization
ICC	incremental capacity curve
LCO	lithium-cobalt-oxide
LIB	lithium-ion battery
LMO	lithium-manganese-oxide
LSTM	long short-term memory
MAE	mean absolute error
masked AE	masked autoencoder

ME	maximum absolute error
MILP	mixed-integer linear programming
MLP	multilayer perceptron
NASA	National Aeronautics and Space Administration
NCA	nickel-cobalt-aluminum-oxide
OCV	open-circuit voltage
PCoE	Prognostics Center of Excellence
PDF	probability density function
RC	resistance–capacitance
RF	random forest
RMSE	root mean square error
RW	random walk
SoC	state of charge
SoH	state of health
SSS	stationary storage system
SVM	support vector machine
SVR	support vector regression
UCSD	University of California San Diego
V/I	voltage and current (input channels)
V/I/T	voltage, current, and temperature (input channels)

ACKNOWLEDGEMENTS

I would like to thank my advisor, Professor Jan Kleissl, for his great support and guidance throughout my doctoral studies. His door was always open, and his patience and encouragement gave me the confidence to keep going whenever the path ahead felt uncertain.

I am deeply grateful to Dr. Shijie Tong, who supervised my research from its earliest days. He taught me how to turn ideas into experiments and experiments into papers, and his perspective kept this work grounded in real batteries and real systems.

I also thank Professors Zheng Chen, Raymond de Callafon, Yuanyuan Shi, and Patricia Hidalgo-Gonzalez for serving on my committee and for their time and thoughtful feedback.

Finally, I thank my family. Their unconditional love and support carried me through many years of study and across an ocean. This dissertation belongs to them as much as it does to me.

Chapter 2 is based on the following publication: Long Ling, Shaojie Yang, and Shijie Tong, "Lithium-ion battery state of health estimation based on statistical features derived from voltage and temperature probability density functions under realistic randomized cycling conditions," *Journal of Energy Storage*, vol. 137, p. 118496, 2025. The dissertation author was the primary investigator and author of this paper.

Chapter 3 is based on the following manuscript, currently under review: Long Ling, Shaojie Yang, Shijie Tong, and Jan Kleissl, "Masked-Autoencoder Pretraining for State-of-Health Estimation of Second-Life Battery Packs," submitted to *IEEE Transactions on Transportation Electrification*. The dissertation author was the primary researcher and author of this material.

Chapter 4 is based on the following publication: Long Ling, Cristian Cortes-Aguirre, Shaojie Yang, and Shijie Tong, "An Optimal Dispatch Model for Economic Benefit Valuation of Grid Energy Storage Solutions," in *Proceedings of the IEEE Green Technologies Conference (GreenTech)*, Wichita, KS, USA, 2025, pp. 113–117. The dissertation author was the primary investigator and author of this paper.

VITA

2018	Bachelor of Science in Automotive Engineering, Chongqing University
2021	Master of Science in Automotive Engineering, Zhejiang University
2026	Doctor of Philosophy in Engineering Sciences (Mechanical Engineering), University of California San Diego

PUBLICATIONS

Long Ling, Shaojie Yang, and Shijie Tong, "Lithium-ion battery state of health estimation based on statistical features derived from voltage and temperature probability density functions under realistic randomized cycling conditions," *Journal of Energy Storage*, vol. 137, p. 118496, 2025.

Long Ling, Cristian Cortes-Aguirre, Shaojie Yang, and Shijie Tong, "An Optimal Dispatch Model for Economic Benefit Valuation of Grid Energy Storage Solutions," in *Proceedings of the IEEE Green Technologies Conference (GreenTech)*, Wichita, KS, USA, 2025, pp. 113–117.

Long Ling, Shaojie Yang, Shijie Tong, and Jan Kleissl, "Masked-Autoencoder Pretraining for State-of-Health Estimation of Second-Life Battery Packs," under review at *IEEE Transactions on Transportation Electrification*.

Shaojie Yang, Long Ling, Xiang Li, Jintong Han, and Shijie Tong, "Industrial Battery State-of-Health Estimation with Incomplete Limited Data Toward Second-Life Applications," *Journal of Dynamics, Monitoring and Diagnostics*, vol. 3, pp. 246–257, 2024.

Shaojie Yang, Haoyuan Meng, Long Ling, Xiaosi Gao, Xiang Li, Xianke Lin, and Shijie Tong, "Cycle Life Prediction for Lithium-Ion Batteries Using Data-Augmented XGBoost with Arbitrary 11-Cycle Data," *IEEE Transactions on Transportation Electrification*, 2026, doi: 10.1109/TTE.2026.3695061.

FIELDS OF STUDY

Major Field: Engineering Sciences (Mechanical Engineering)

Studies in Battery State-of-Health Estimation and Energy Dispatch Optimization
Professor Jan Kleissl

ABSTRACT OF THE DISSERTATION

Data-Driven State-of-Health Estimation and Optimal Energy Dispatch of Battery Energy Storage Systems

by

Long Ling

Doctor of Philosophy in Engineering Sciences (Mechanical Engineering)

University of California San Diego, 2026

Professor Jan Kleissl, Chair

Lithium-ion batteries anchor transportation electrification and grid-scale energy storage. Usable capacity fades with age. Safe operation, second-life screening, and investment decisions therefore largely depend on the capacity-based state of health (SoH). A reference capacity test measures it directly but is costly and interrupts service. This dissertation connects health estimation to economic value in three parts.

The first part estimates SoH for cells cycled under randomized loads mirroring real-world usage. Each cycling stretch is summarized by cell voltage and temperature probability density functions. The expectation–maximization algorithm identifies the assumed bimodal voltage and

normal temperature distribution parameters. Identified parameters and statistical descriptors most correlated with SoH feed a support vector regression. On the NASA random walk benchmark it reaches 2.63 % mean absolute error and 3.31 % root mean square error. Its maximum error, 6.77 %, is the lowest of forty algorithm-input combinations. Moreover, the accuracy extends to three more predictable test groups.

The second part carries SoH estimation to three retired electric-vehicle packs. These packs supply scarce labels and sense only voltage and current. A Transformer encoder is pretrained as a masked autoencoder on raw cycling data and fine-tuned on labeled stretches. A transfer study shows that corpus composition, not size, decides whether pretraining helps. A pool mixing a wide-SoH public NASA dataset and the second-life packs lowers error on the two lithium-manganese-oxide packs, which make up most of this pool, by up to 35.5 % versus no pretraining. On the chemistry-distinct nickel-cobalt-aluminum-oxide pack, however, the same pool raises error by 41.6 %. The NASA dataset alone is instead the most accurate and most stable source for that pack.

The third part turns from battery health to economic worth. A dispatch model maximizes energy-trading revenue subject to a battery model's constraints. Piecewise linear approximation and McCormick envelopes relax them into a mixed-integer linear program. Vanadium redox flow, sodium-ion, and lithium-ion systems, dispatched for energy time shifting against 2022 California real-time prices, earn aggregated daily revenues of $111.08, $67.63, and $31.17. Higher C-rate and stronger power capability let a system cycle more often and earn more per cycle.

Chapter 1

Introduction

1.1 Motivation

Lithium-ion batteries are a key enabling technology for transportation electrification and for grid-scale energy storage [1, 2]. Their usable capacity fades as they age. Safe operation, warranty decisions, second-life screening, and investment planning therefore all depend on the state of health (SoH). SoH describes battery degradation, which appears not only in capacity fade but also in internal resistance growth and open-circuit voltage shifts [3]. This dissertation focuses on the most common, capacity-based definition and leaves the other indicators outside its scope. Throughout this dissertation, SoH therefore denotes the ratio of a battery's actual capacity to its nominal capacity. A reference capacity test measures this SoH directly. The test, however, takes the battery out of service for hours and requires a controlled, near-full charge and discharge. Moreover, tracking SoH repeats this cost at every measurement and for every battery. The practical alternative is to estimate SoH from the operating data a battery already produces. Three properties of field data make this estimation hard. Usage is intermittent and diverse, and rarely offers a repeating charge or discharge pattern. Sensing is often limited to voltage and current. Labeled capacity measurements are scarce, because every label costs a reference test [4].

Estimating SoH from operating data reveals the capacity a battery still holds. Some grid energy storage applications can turn this capacity into profit. In energy time shifting, for example, a system stores energy when electricity prices are low and delivers it when they are high. Beyond

lithium-ion, emerging battery technologies such as sodium-ion and vanadium redox flow are being developed for grid-scale storage [2]. How these technologies compete in such applications requires realistic quantification with an optimal dispatch model. This dissertation therefore starts from capacity estimation and extends to energy dispatch optimization and economic valuation.

1.2 Research Questions

The dissertation develops the chain from health estimation to the economic value of grid energy storage in three parts. Each part answers one question, and each question drops an assumption the previous one relied on.

1. Under randomized cycling, where no repeating charge or discharge template exists, which manually engineered features of the operating data still track SoH, and how accurately can they estimate it? (Chapter 2)

2. Second-life battery packs can sense only voltage and current, with no temperature channel. Moreover, the packs available for method development supply only scarce labeled reference tests, confined to the narrow SoH band each pack occupies. Under these constraints, can the abundant raw cycling voltage and current streams these packs produce be used to achieve accurate SoH estimation, and what impacts the accuracy and stability of the estimates across battery systems? (Chapter 3)

3. Once a storage system's capability is known, what is that capability worth when dispatched against electricity market prices, and which battery attributes drive the value? (Chapter 4)

1.3 Approach and Overview of the Three Studies

Chapter 2 addresses SoH estimation when the load is randomized. No repeating charge or discharge template therefore exists for features to align against. Classical protocol-locked features, such as incremental-capacity peaks extracted from complete constant-current constant-voltage charges, are unavailable under such usage. The chapter instead summarizes each stretch

of cycling by the probability density functions (PDFs) of its cell voltage and temperature measurements. Under randomized loads the voltage PDFs are near-bimodal and the temperature PDFs are near-normal. The expectation–maximization algorithm identifies the underlying distribution parameters. The identified parameters and statistical descriptors of the reconstructed PDFs most correlated with SoH form the feature set. A support vector regression (SVR) over these features reaches a mean absolute error (MAE) of 2.63 % and a root mean square error (RMSE) of 3.31 % on the NASA random walk benchmark. Its maximum error of 6.77 % is the lowest of all forty algorithm-input combinations. The framework maintains its accuracy across four test groups with different load profiles and ambient temperatures. Moreover, the features have low sensitivity to the sampling interval. Specifically, estimation error grows by less than 5 % when the interval is extended from 1 s to 5 s. Even at 180 s, absolute MAE and RMSE remain below 5 %. This tolerance permits lighter data collection in an embedded battery management system (BMS).

The statistical features of Chapter 2 are effective. The framework, however, rests on two premises. Specifically, it assumes instrumented single cells. Its feature selection and regression training also consume labeled reference tests across a wide SoH range. Retired electric-vehicle packs meet neither premise. Production BMS channels expose only group-level voltage and one shared string current. Signals at this level can smooth out cell-level characteristics. Moreover, each pack arrives with a real, heterogeneous first-life aging history for which no records exist. Even when a pack is tested on arrival, the new labels are scarce and confined to the narrow SoH band it occupies. Raw cycling data, however, is abundant and costs nothing extra, because packs produce it whenever they operate. Chapter 3 therefore replaces feature engineering with representation learning that pretrains on the abundant raw data and spends the scarce pack labels only on fine-tuning. Specifically, a Transformer encoder is pretrained as a masked autoencoder (masked AE) by reconstructing masked samples of raw voltage and current windows. Moreover, the NASA random walk windows carry cell temperature as a third channel, giving voltage, current, and temperature (V/I/T) inputs. Fine-tuning adds a cross-window attention pooler and a

regression head. The pooler aggregates the windows of each cycling segment into one embedding. The head maps that embedding to one SoH estimate. On the NASA random walk benchmark, the pretrained model reaches a test RMSE of 2.82 %SoH with no manually engineered features. This accuracy is in line with the strongest published cell-level result. The chapter's core result is a transfer ablation that crosses five pretraining sources with three genuinely retired packs at three seeds each. One pack is nickel-cobalt-aluminum-oxide (NCA), and the other two are lithium-manganese-oxide (LMO). Every source is compared with random initialization, that is, the same model trained on the target pack's labels with no pretraining. Pool composition, not pool size, decides whether pretraining helps or hurts. One source, the combined four-source pool, draws most of its windows from the two LMO packs. On both LMO packs it is the best source ($-35.5\,\%$ and $-28.1\,\%$ versus random initialization, 3/3 seeds each). On the chemistry-distinct NCA pack, however, the same checkpoint is the worst configuration ($+41.6\,\%$ versus random initialization). Wide-SoH NASA (V/I/T) pretraining is instead the most accurate and the most stable configuration on that narrow-band target. The target pack's own cycling data is neither sufficient nor necessary for transfer. Specifically, pools that include it can still fail, and the target-free NASA source transfers well without it. Moreover, removing the temperature channel from NASA pretraining raises mean transfer error by up to $+28.9\,\%$, although no target pack senses temperature during fine-tuning. These findings give a practitioner a rule that can be checked before any training run. Concretely, a pretraining corpus that spans a wide SoH range can be the more dependable choice for a narrow-band target.

An accurate SoH estimate quantifies how much capacity a battery still holds. It does not, however, quantify what that capacity is worth. For grid applications that earn revenue from electricity price differences, storage value depends on how capacity is dispatched against market prices over time. Chapter 4 closes the chain with an optimal dispatch model for economic benefit valuation of battery energy storage systems in price-based grid applications. The model maximizes energy-trading revenue. Its constraints represent battery behavior with the simplest equivalent circuit model, the Rint model. Two of these constraints are nonlinear. Specifically,

the open-circuit voltage is a nonlinear function of the state of charge, and the charging and discharging powers are each the product of terminal voltage and current. Piecewise linear approximation relaxes the voltage curve, and McCormick envelopes relax the power products. The relaxation yields a mixed-integer linear programming (MILP) problem that standard solvers handle efficiently. A case study evaluates three battery technologies in an energy time shifting application against 2022 real-time market prices from the California Independent System Operator. The technologies are lithium-ion and two of its emerging competitors, sodium-ion and vanadium redox flow. Each is characterized by in-house tests and then scaled to the same maximum terminal voltage and energy capacity for a fair comparison. The vanadium redox flow and sodium-ion systems earn higher aggregated daily revenue than the lithium-ion system ($111.08 and $67.63 versus $31.17). The reason is that higher C-rates (charge or discharge rates relative to ampere-hour capacity) and stronger power capability let a system cycle more often and capture more of each price swing. Usable capacity is the quantity that Chapters 2 and 3 estimate. The dispatch model therefore turns that quantity into revenue terms an investor or operator can act on.

1.4 Contributions

In summary, this dissertation makes four contributions.

1. A PDF-based feature extraction and SVR framework that estimates SoH under randomized cycling, with accuracy that is robust to load profile, ambient temperature, and sampling interval (Chapter 2).

2. A masked-AE pretraining and fine-tuning framework for second-life battery packs, together with a five-source transfer ablation showing that pool composition, not pool size, decides whether pretraining helps or hurts SoH estimation on a target pack, and how repeatable that outcome is across training runs (Chapter 3).

3. A MILP dispatch model for economic benefit valuation of battery energy storage systems in price-based grid applications, demonstrated on three battery technologies in an energy time shifting application against real-time market prices (Chapter 4).

4. A conceptual synthesis of the three studies into one workflow for a storage operator, in which the data regime selects the SoH estimator and the resulting estimate enters the dispatch model (Chapter 5).

1.5 Organization of the Dissertation

Chapters 2 through 4 present the three studies in the order of the chain, from health estimation under randomized cycling, to health estimation under deployment constraints, to the economic value of dispatched storage. Each of these chapters opens by placing its study within that chain and closes with the connection to the chapter that follows. Chapter 5 summarizes the contributions, synthesizes the three studies into one workflow for a storage operator, and outlines future directions.

Chapter 2

Lithium-Ion Battery State-of-Health Estimation Based on Probability Density Functions under Randomized Cycling

2.1 Introduction

This chapter is the first link in the dissertation's chain from estimated state of health (SoH) to dispatched economic value. It asks how SoH can be estimated from operating data alone when the load is randomized, and it answers in the most fully instrumented of the three settings studied in this dissertation. Single cells expose both voltage and temperature, and labeled reference tests are plentiful across a wide SoH range. Chapter 3 removes both supports. Chapter 4 then asks what the estimated capacity is worth.

Lithium-ion batteries (LIBs) have become a key enabling solution for electrifying transportation and supporting the renewable energy transition due to their advantages, including high energy density, longevity, and efficiency [1]. The global installed capacity of LIBs is projected to reach 1500 GWh annually by 2030, driven by applications in various industries, such as electric vehicles (EVs) and stationary storage systems (SSSs) [2]. As a LIB's capacity decreases during cycling, regular health evaluation is essential to ensure the battery's safe and efficient operation [3]. SoH and remaining useful life are two primary indices used for battery health assessment [5]. SoH reflects degradation in several battery properties, such as capacity, internal resistance,

and open-circuit voltage [3]. This work focuses on the capacity-based definition and leaves the other indicators outside its scope. SoH hereafter refers to the ratio of actual capacity to nominal capacity. Under this definition, SoH indicates the battery's maximum available capacity. Therefore, SoH is critical for safety and energy management, such as state of charge (SoC) estimation [6] and power scheduling in SSSs [7]. A reference capacity test measures SoH directly but interrupts operation. SoH must therefore be estimated from operating data, which remains challenging due to the complex interactions between battery degradation, varying operational loads, and environmental conditions [8].

Various methods have been proposed for SoH estimation, broadly classified into model-based and data-driven approaches. Model-based methods, including equivalent circuit and electrochemical models, primarily focus on modeling battery behavior [9]. These approaches often leverage observers and adaptive filters, such as sliding mode observers [10], Kalman filters [11], and particle filters [12], to optimize model parameters. However, accurately estimating capacity over the battery's lifespan is challenging due to the complex and coupled degradation mechanisms.

Data-driven methods, particularly those based on machine learning, circumvent the need for explicit knowledge of aging mechanisms. Instead, they rely on extracting health indicators strongly correlated with battery capacity from collected aging data. Machine learning algorithms, such as neural networks [13, 14], support vector regression (SVR) [15], and Gaussian process regression (GPR) [16], subsequently establish relationships between these indicators and SoH.

Data generated during repetitive charging and discharging cycles have been widely utilized for feature extraction in SoH estimation. Common approaches involve analyzing the peaks and valleys of incremental capacity curves (ICCs) and differential voltage curves (DVCs) throughout the complete constant current constant voltage (CCCV) charging process [17, 18]. Additionally, useful features can be extracted from purely constant current (CC) or constant voltage (CV) charging processes, such as the battery surface temperature change over a specific voltage range during CC charging [19] and the CV–CC time ratio [20]. Statistical metrics derived

from DVCs across cycles during CC discharging have also proven effective as SoH indicators [21]. Moreover, shifts in the probability density functions (PDFs) of battery voltage data have been shown to reflect battery degradation, as they share an equivalent mathematical foundation with ICCs and DVCs [4]. Specifically, the voltage PDF measures the share of sampling time the battery spends at each voltage level. Under constant current, that time is proportional to the charge transferred at that level. The voltage PDF therefore conveys the same charge-per-voltage information that ICCs plot and DVCs invert. For example, peak positions and regional charging durations (RCDs) between two specified voltages, extracted from voltage PDFs during CCCV charging, have been used as features [22]. Lin et al. modeled voltage data using a two-component Gaussian mixture and computed the Kullback–Leibler distance between identified PDFs across cycles as a feature [23]. Similarly, another study utilized two peak values and their corresponding voltage positions on the voltage PDFs during CC charging as degradation-sensitive features [24].

However, due to intermittent user behaviors, CCCV charging, CC charging, and CC discharging processes are typically incomplete in real-world applications. Batteries often start or stop charging or discharging at arbitrary intermediate SoC. For instance, operating conditions for batteries in SSSs may be continuously updated during day-ahead power scheduling, resulting in no discernible repetitive patterns [25]. Although CV charging is commonly implemented in EVs, corresponding data may still be incomplete, as the battery management system (BMS) may prevent certain cells in battery packs from reaching the cut-off current. Consequently, features extracted under complete CC and CV conditions may become unsuitable for SoH estimation in such scenarios. Therefore, it is essential to identify more practical features derived from partial CC/CV data or more general battery operating conditions (such as random walk (RW) charging and discharging) that effectively correlate with SoH.

Partial CC charging conditions have garnered substantial research attention, particularly in applications like EVs, where charging processes are typically more predictable compared to discharging. Wu et al. analyzed over 1200 EV charging cycles, concluding that the predominant charging voltage range lies between 3.8 V and 4.1 V. Leveraging this charging behavior, features

were extracted from capacity–voltage curves and ICCs within this voltage window [26]. Wei et al. proposed using RCDs and fuzzy entropy features from charging voltages to handle scenarios where the initial CC charging voltage exceeds 4 V [27]. Deng et al. segmented capacity curves during the CC charging process and estimated SoH based on the averages and standard deviations of differential capacities across these segments [28]. Li et al. randomly excluded 15 % of testing samples and estimated SoH using only the upper portion of the CC charging voltage–time curves [29]. Schaltz et al. introduced the partial charging method (PCM), which assesses SoH from partial charging capacities by establishing linear relationships between partial and full charging capacities while incorporating Coulombic efficiency [30]. Recognizing the influence of measurement conditions such as C-rate, temperature, and starting voltage on battery charges, Braco et al. expanded the PCM by introducing three correction factors to relate charges measured under varied conditions (different C-rates, starting voltages, and temperatures) to standardized conditions (C/3, 45 °C, 3.7 V) [31]. He et al. employed a Lorentzian function-based voltage–capacity model to fit ICCs from full-range CC charging data, enabling feature extraction from partial CC data using this parametrized model [32]. Kohtz et al. trained a GPR model using output voltage curves from a physics-based finite element model, linking partial charging segments' voltage curves to the solid electrolyte interface (SEI) layer thickness [33]. Subsequently, a co-kriging-based multi-fidelity model was utilized to correlate the SEI thickness with battery SoH.

The exploration of features also extends to CV charging and CC discharging processes. Deng et al. transformed voltage curves obtained under dynamic current discharging conditions into smoother open circuit voltage curves, from which the capacity and differential capacity curves were segmented [34]. Standard deviations of these segments were then used for SoH forecasting. Zhao et al. demonstrated statistical feature extraction from capacity, capacity differences, and temperature data collected during CC discharging within specific voltage ranges [35]. Yang et al. adopted an alternative approach by deriving the time constant from CV charging currents and utilized partial CV charging data to achieve SoH estimation with approximately

5 % error [36]. Park et al. leveraged context encoders to reconstruct missing information in partial-cycling datasets, enabling SoH estimation through a deep learning-based graphical method [37].

The primary limitation of SoH estimation methods based on partial CC or CV data is their dependence on specific voltage ranges, which may not be consistently available in practical scenarios. Additionally, methods that rely solely on charging data or solely on discharging data leave the remainder of the operating data unused. The studies reviewed above show, however, that charging and discharging data both carry degradation information. In summary, despite extensive research efforts, current features and operational conditions investigated in existing studies lack sufficient generality for accurate SoH estimation across diverse battery operating scenarios, such as those in EVs and SSSs. In contrast, this work addresses the issue of feature generality by developing novel features applicable under RW charging and discharging conditions.

Recent advancements in battery SoH estimation increasingly utilize automatic feature extraction approaches, such as deep autoencoders [38], convolutional autoencoders [39], and multi-head convolution attention blocks [40]. These methods effectively reduce reliance on manually designed features by directly learning latent representations from raw battery data, achieving high accuracy and robustness across varied conditions. However, despite these advantages, automatic feature extraction techniques typically involve complex model structures, resulting in higher computational complexity and an increased risk of overfitting, particularly with limited-size datasets.

In practical applications, individual input data sequences for estimation models can be voluminous, leading to substantial computational overhead and storage demands. Consequently, data downsampling has become an essential technique in battery health research, effectively reducing computational complexity and data volume without significantly compromising performance. For example, a proton exchange membrane fuel cell degradation study successfully applied downsampling by increasing sampling intervals, maintaining both data distribution integrity and accurate degradation trend analysis [41]. Similarly, downsampling in battery

SoH prediction enabled efficient extraction of hierarchical temporal patterns across multiple scales by reducing temporal resolution [42]. In rapid battery capacity estimation, interpolation-based downsampling effectively compressed input data, thereby lowering model complexity and computational load [43]. Incremental capacity analysis also benefited from a distance-based downsampling approach, significantly decreasing computation time [44]. Thus, downsampling techniques play a crucial role in enhancing the efficiency and practical applicability of battery health estimation methods.

The main contributions of this chapter are as follows:

1. An investigation conducted under RW charging and discharging conditions revealed that cell voltage and temperature data approximately follow bimodal and normal distributions, respectively. Distribution parameters were effectively identified using the expectation–maximization (EM) algorithm.

2. A systematic extraction of health-correlated features from PDFs, reconstructed using the identified distribution parameters, was performed. Subsequent SoH estimation using an SVR model based on these features achieved high accuracy, with a mean absolute error (MAE) of 2.63 % and a root mean square error (RMSE) of 3.31 %, and maintained robust performance across varying load profiles and temperature conditions.

3. Through comprehensive analysis, this study highlights the proposed features' notable insensitivity to changes in sampling intervals. This characteristic demonstrates that the features retain their effectiveness despite variations in data collection frequency, underscoring their potential for practical applications with reduced computing power.

The remainder of this chapter is organized as follows: Section 2.2 introduces the dataset utilized in this research. Section 2.3 describes the proposed framework, including feature extraction, feature selection, and algorithm selection. Section 2.4 presents comparative analyses evaluating the proposed framework against state-of-the-art studies, and examines the generaliza-

tion capability of our approach. Section 2.5 investigates the impact of downsampling on feature correlation and SoH estimation accuracy. Finally, Section 2.6 provides the conclusions.

2.2 Dataset Description

The data used in this research were collected by the NASA Ames Prognostics Center of Excellence (PCoE) [45]. The full dataset comprises seven groups, each consisting of four lithium cobalt oxide cells (2.11 Ah). Each group was cycled under different temperature and current conditions. Among these groups, only one (RW9–RW12) employed a randomized protocol, where both the decision to charge or discharge and the magnitude of current applied were determined randomly. This randomization closely mimics the unpredictable usage patterns observed in certain practical battery applications.

Fig. 2.1 illustrates the cycling protocol for cells RW9–RW12. These cells were operated at room temperature. Currents were randomly selected from the set $\{-4.5$ A, -3.75 A, -3 A, -2.25 A, -1.5 A, -0.75 A, 0.75 A, 1.5 A, 2.25 A, 3 A, 3.75 A, 4.5 A$\}$. Negative values represent charging currents, while positive values denote discharging currents. Each selected current was applied until either the battery voltage moved outside the range of 3.2–4.2 V or a 5-minute interval elapsed. Each such application of current is defined as one "step," and this cycling protocol is termed RW. After every 1500 RW steps, two reference tests were conducted. Each reference test consisted of CC charging at 2 A, followed by CV charging at 4.2 V, and finally a CC discharge at 2 A. These reference tests provided benchmark measurements for battery SoH.

To further verify the generalization capability of our proposed method, we included three additional battery groups (groups 1, 3, and 4) from NASA PCoE. These groups expand our analysis to other realistic conditions. As illustrated in Fig. 2.1, these additional datasets involve elevated ambient temperature conditions (40 °C) and varied charging and discharging behaviors. Specifically, these conditions include partial CC charging, full CCCV charging, and discharging

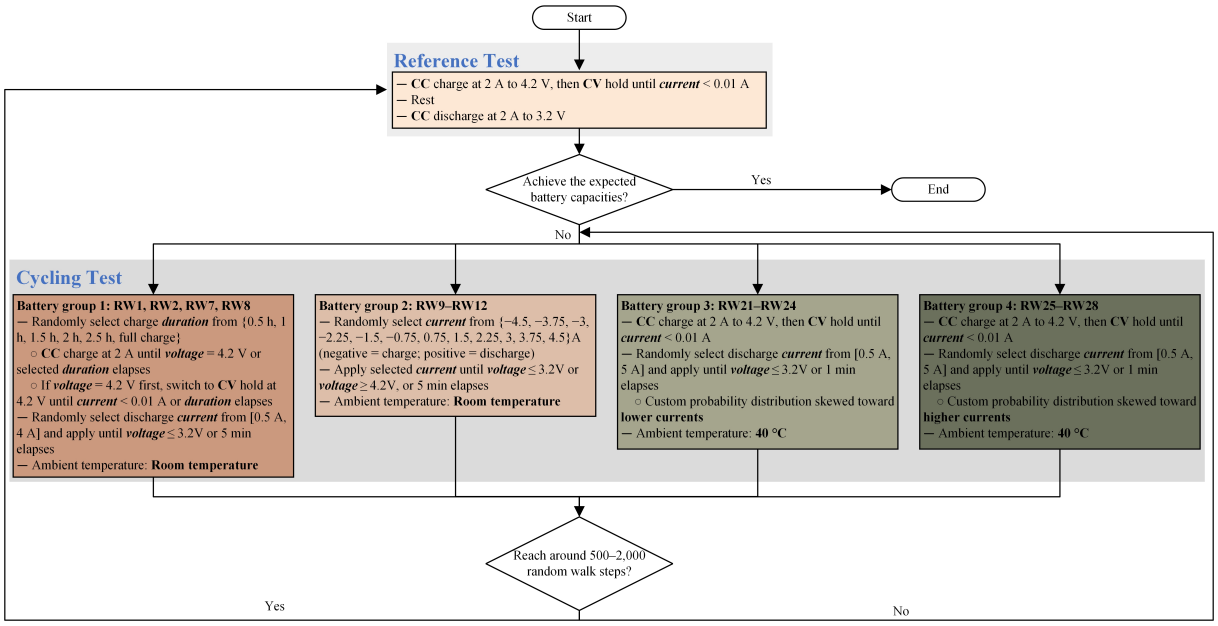


Figure 2.1. Flowchart detailing the cycling procedures for the NASA RW dataset.

under skewed current distributions (either lower or higher currents).

Fig. 2.2 presents exemplary cycling data for batteries RW1, RW9, RW21, and RW25. Specifically, this figure illustrates segments of cell current, voltage, and temperature data corresponding to RW steps 50–180, with measurements recorded at 1-second intervals.

With the dataset and its characteristics established, the next step was to develop the SoH estimation framework that leverages these real-world data patterns.

2.3 Proposed SoH Estimation Framework

The overall methodology employed in this study is illustrated in Fig. 2.3. The primary innovation of our approach lies in extracting battery health-related features directly from cycling test data, rather than from reference test data, as commonly done in existing studies. Relying on reference tests may limit effectiveness when such tests are unavailable. We hypothesized that the voltage and temperature data collected during cycling tests conformed to bimodal and normal distributions, respectively. Initially, we identified the essential parameters for these assumed distributions. Subsequently, we reconstructed the PDFs from these identified parameters and

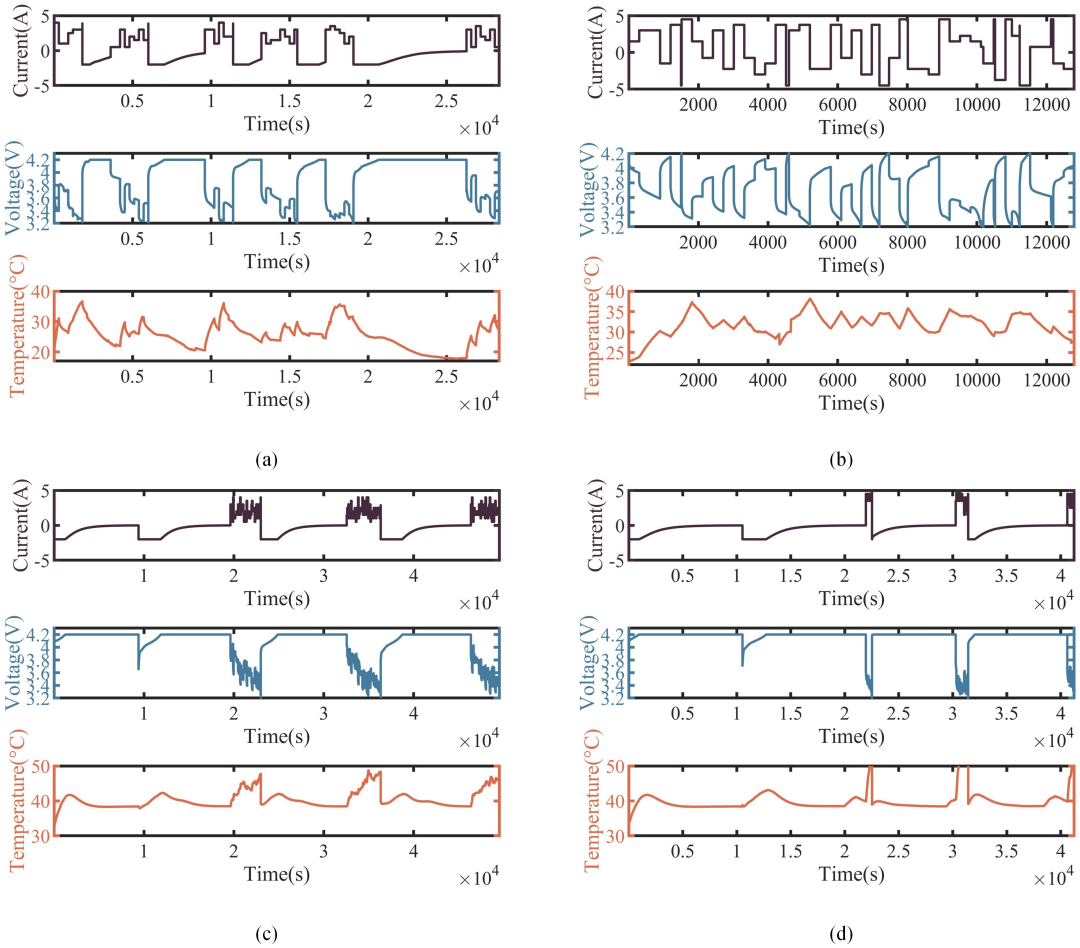


Figure 2.2. Representative data for four batteries from the NASA RW dataset showing cell current, voltage, and temperature profiles: (a) RW1; (b) RW9; (c) RW21; (d) RW25.

conduct feature engineering to extract features that are closely correlated with battery health. These features were then compiled into feature vectors. Finally, by inputting these feature vectors into an SVR model, we accurately estimated the batteries' SoH.

2.3.1 Feature Extraction

(1) Identification of PDFs

Feature extraction begins from the voltage and temperature data collected over every 1500 RW steps. Kernel density estimation constructs a PDF estimate for each such stretch with a normal kernel function and assumes no parametric form. The estimates therefore reveal

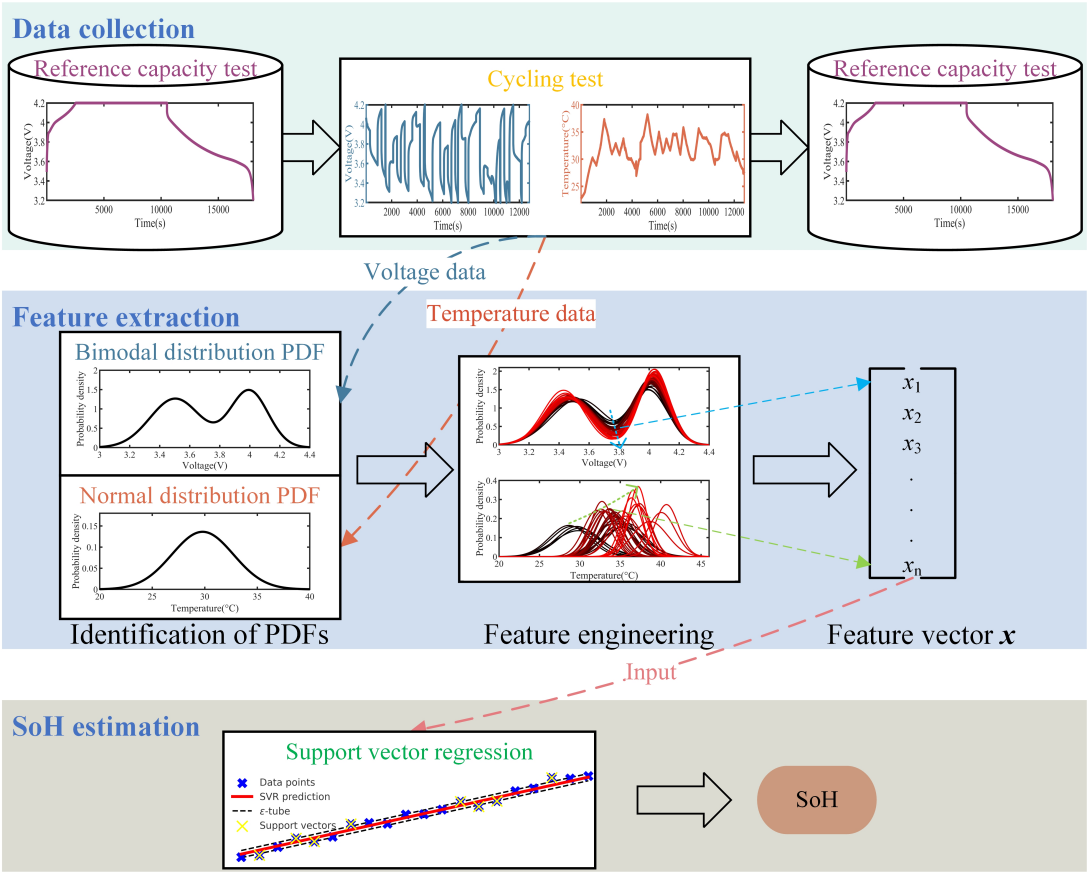


Figure 2.3. Flowchart of the proposed methodology for battery SoH estimation, illustrating data collection, PDF identification, feature extraction, and the final estimation using SVR.

the rough distribution shapes of the RW data. We initially applied kernel density estimation using the MATLAB *ksdensity* function, which Feng et al. validated for LIB SoH estimation under CCCV conditions [46]. As depicted in Fig. 2.4(a) and (b), each curve represents the estimated PDF of voltage data from 1500 RW steps. Cycling tests containing fewer than half of the standard number of RW steps (1500), ranging from only 127 to 653 RW steps, were classified as outliers. However, features identified in previous studies, such as peak positions, showed weak correlations with battery aging. Additionally, many small fluctuations within the PDFs complicate the identification of traditional features such as peaks and valleys. Therefore, we adopted a fitting method to generate smoother, more intuitive curves while preserving essential statistical information.

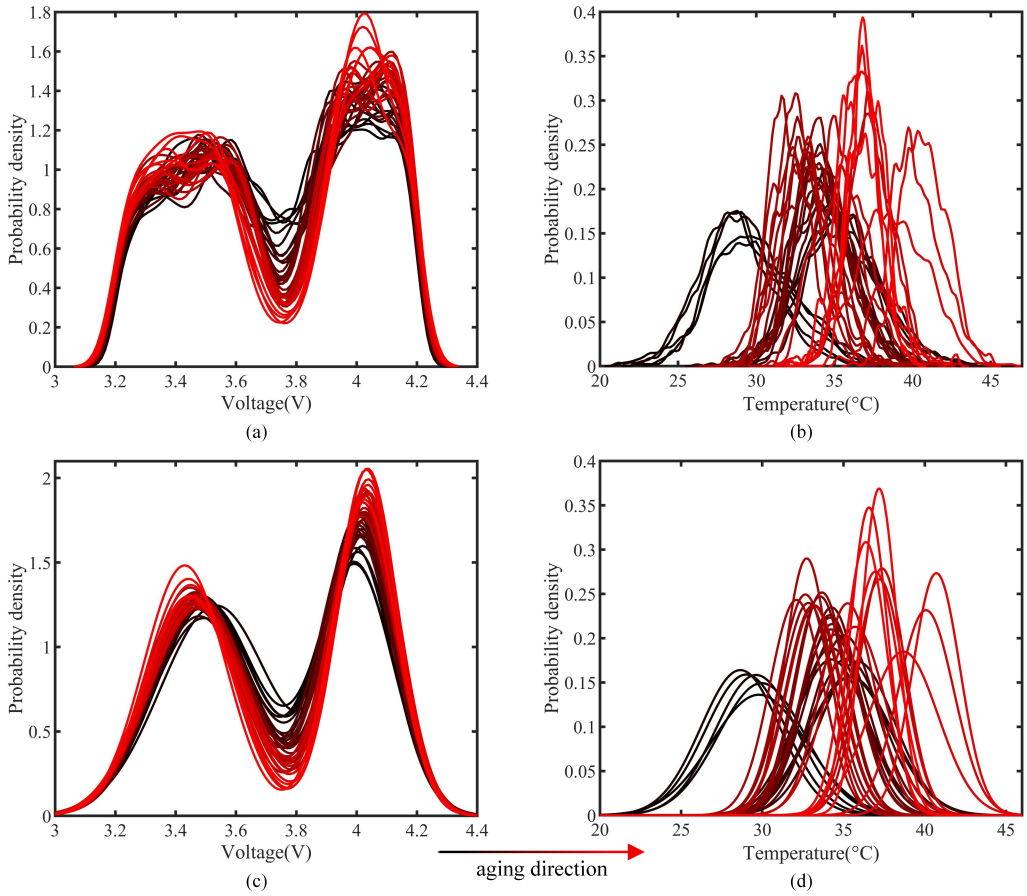


Figure 2.4. PDFs of battery RW9 from the NASA RW dataset: (a, b) Kernel density estimates of voltage data and temperature data; (c, d) PDFs of voltage data and temperature data reconstructed from the parameters identified by the EM algorithm and by distribution fitting. The color progression from black to red indicates battery aging.

Inspired by observations of multimodal distributions, we assumed that the RW voltage data conformed to bimodal distributions. Similarly, the RW temperature data were presumed to follow normal distributions. The EM algorithm, widely employed in image clustering [47, 48], was then applied to identify the parameters of the most likely bimodal distribution underlying the voltage data.

Let $x = (x_1, x_2, \ldots, x_n)$ be a sample of n independent observations drawn from a mixture

of two univariate normal distributions. Its PDF, $\mathrm{p}(x)$, is defined in Eq. (2.1).

$$\mathrm{p}(x) = \sum_{m=1}^{2} \pi_m \, \mathrm{p}(x \mid \theta_m), \quad x \mid \theta_m \sim \mathcal{N}\left(\mu_m, \sigma_m^2\right) \tag{2.1}$$

where $\pi_m \in (0,1)$ is the mixing proportion of the m-th normally distributed component, satisfying $\pi_1 + \pi_2 = 1$. The mixing proportion is the probability that an observation originates from component m. $\theta_m = (\mu_m, \sigma_m)$ specifies the mean and standard deviation of the corresponding component.

Let $z = (z_1, z_2, \ldots, z_n)$ be the latent variables that determine the component from which each observation originates. The pseudo-code of the EM algorithm is outlined in Algorithm 1, where $f\left(x_i; \mu_m^{(k)}, \sigma_m^{(k)}\right)$ represents the m-th normal density evaluated at x_i. The initial parameter guess is $\theta^{(0)} = (0.54, 3.50, 0.01, 0.46, 3.99, 0.01)$. Iterations are indexed by k, and convergence is determined using a threshold ε. In the expectation step, we compute the membership probability $r_{m,i}^{(k)}$, which quantifies how strongly component m accounts for observation x_i. These values inform the auxiliary function $Q\left(\theta \mid \theta^{(k)}\right)$, representing the expected log-likelihood under the current membership probabilities. In the maximization step, we update each mixing proportion π_m as the average of membership probabilities, and recompute μ_m and σ_m as the weighted mean and weighted root-mean-square deviation, respectively. Finally, we evaluate ΔQ, the change in Q from iteration $k-1$ to k, and iterate the procedure until $\Delta Q \leq \varepsilon$.

In addition, the parameters of the most likely normal distribution underlying the temperature data had to be identified. The identified parameters, the mean and standard deviation, serve as candidate features and as the basis for PDF reconstruction. Distribution fitting identifies the parameter values of an assumed distribution form under which the observed data are most likely. The MATLAB *fitdist* function implements this fitting for a specified distribution form. We therefore fitted normal distributions to the temperature data with this function. The fitted PDFs of voltage and temperature data are presented in Fig. 2.4(c) and (d). Compared to the curves in Fig. 2.4(a) and (b), the fitted PDFs are smoother while retaining crucial statistical information.

Algorithm 1. EM algorithm

Require: $x = (x_1, x_2, \ldots, x_n)$, $\theta^{(0)} = \left(\pi_1^{(0)}, \mu_1^{(0)}, \sigma_1^{(0)}, \pi_2^{(0)}, \mu_2^{(0)}, \sigma_2^{(0)}\right)$, $\varepsilon = 10^{-3}$, $\Delta Q = 10^3$, $k = 0$

Ensure: $\theta^{(k)} = \left(\pi_1^{(k)}, \mu_1^{(k)}, \sigma_1^{(k)}, \pi_2^{(k)}, \mu_2^{(k)}, \sigma_2^{(k)}\right)$

while $\Delta Q > \varepsilon$ **do**

  Expectation step:

  Membership probabilities:

$$r_{m,i}^{(k)} := \mathrm{P}\left(z_i = m \mid x_i; \theta^{(k)}\right) = \frac{\pi_m^{(k)} f\left(x_i; \mu_m^{(k)}, \sigma_m^{(k)}\right)}{\sum_{m=1}^{2} \pi_m^{(k)} f\left(x_i; \mu_m^{(k)}, \sigma_m^{(k)}\right)}$$

  Log likelihood function:

$$Q\left(\theta \mid \theta^{(k)}\right) = \sum_{i=1}^{n} \sum_{m=1}^{2} r_{m,i}^{(k)} \left[\log \pi_m - \log \sigma_m - \frac{(x_i - \mu_m)^2}{2\sigma_m^2} - \frac{1}{2}\log(2\pi)\right]$$

  Maximization step:

  Update mixing proportions: $\pi_m^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n} r_{m,i}^{(k)}$

  Update means: $\mu_m^{(k+1)} = \frac{\sum_{i=1}^{n} r_{m,i}^{(k)} x_i}{\sum_{i=1}^{n} r_{m,i}^{(k)}}$

  Update standard deviations: $\sigma_m^{(k+1)} = \sqrt{\frac{\sum_{i=1}^{n} r_{m,i}^{(k)} \left(x_i - \mu_m^{(k+1)}\right)^2}{\sum_{i=1}^{n} r_{m,i}^{(k)}}}$

  Termination step:

  if $k \geq 1$ **then**

    $\Delta Q = \left|Q\left(\theta \mid \theta^{(k)}\right) - Q\left(\theta \mid \theta^{(k-1)}\right)\right|$

  end if

  $k = k + 1$

end while

For example, the PDFs in Fig. 2.4 display similar contours, but the modes in Fig. 2.4(c) and (d) are easier to recognize.

Smooth PDFs of voltage and temperature data were reconstructed based on the identified distribution parameters. Feature extraction was then carried out on these reconstructed PDFs. For voltage data exhibiting near-bimodal distributions, summary statistics such as anti-mode, bimodality amplitude, and bimodality coefficient demonstrated strong correlations with battery aging. Similarly, the mode of the nearly normally distributed temperature data exhibited a close

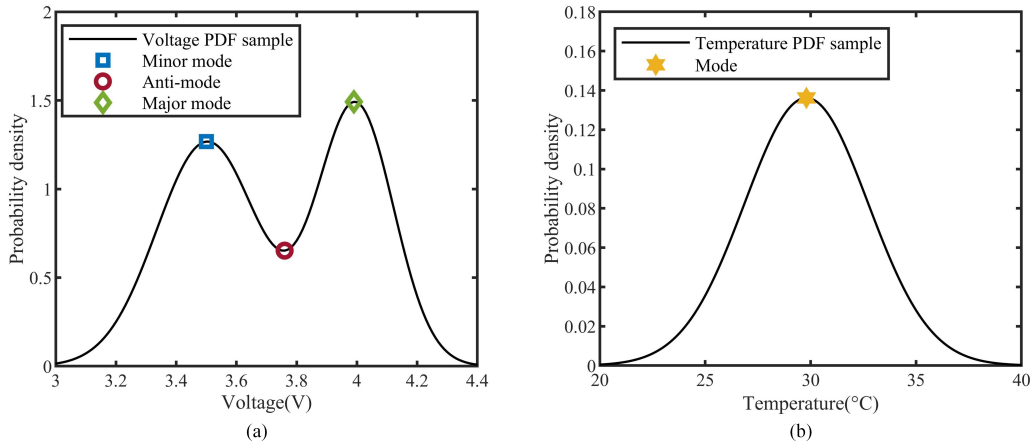


Figure 2.5. Illustration of feature extraction from sample PDFs: (a) The major mode (the larger of the two modes), the minor mode (the smaller mode), and the anti-mode (the least frequent value between them) identified from voltage data, with amplitudes M_a, M_i, and A_n, respectively; (b) The mode identified from temperature data.

relationship with battery degradation. Moreover, certain key identified parameters that are crucial for defining the distributions showed a significant association with battery capacity. A detailed exploration of these features is provided in Section 2.4.

(2) Definition of Features

a) Summary statistics

In Fig. 2.5(a), the fundamental summary statistics (modes) are marked on a sample voltage PDF. More complex features are subsequently constructed based on these modes:

i) Voltage PDF modes

When the two modes of a voltage PDF differ in amplitude, the larger mode is termed the major mode, while the smaller is termed the minor mode. The least frequent value between these two modes is designated as the anti-mode. In this context, M_a, M_i, and A_n represent the amplitudes of the major mode, minor mode, and anti-mode, respectively.

ii) Bimodality amplitude

Bimodality amplitude is defined in Eq. (2.2).

$$B_{\mathrm{i}} = \left(M_{\mathrm{i}} - A_{\mathrm{n}}\right) / M_{\mathrm{i}} \tag{2.2}$$

Because the anti-mode is the least frequent value between the two modes, its amplitude cannot exceed the minor mode amplitude ($A_{\mathrm{n}} \leq M_{\mathrm{i}}$). B_{i} therefore lies between 0 and 1 and cannot become negative.

iii) Bimodality ratio

Similar to bimodality amplitude, bimodality ratio is defined in Eq. (2.3).

$$B_{\mathrm{a}} = A_{\mathrm{n}} / M_{\mathrm{a}} \tag{2.3}$$

iv) Bimodality coefficient

Bimodality coefficient is defined in Eq. (2.4), where γ and κ denote the skewness and kurtosis of the sample, respectively.

$$B_{\mathrm{c}} = \left(\gamma^2 + 1\right) / \kappa \tag{2.4}$$

v) Temperature PDF mode

The mode (which is also the mean) of a temperature PDF, denoted as M_{t}, is illustrated in Fig. 2.5(b).

b) Identified distribution parameters

As described in Section 2.3.1, each voltage PDF comprises two normal distributions weighted by their mixing proportions, and each temperature PDF is approximately normally distributed. The distribution parameters requiring identification are summarized in Table 2.1.

Having defined a comprehensive set of features from the RW data, we next assessed which of these features were most strongly related to battery SoH through a correlation analysis.

Table 2.1. Distribution parameters requiring identification for voltage and temperature PDFs.

Voltage data	Mixing proportion π_m
	Mean μ_m
	Standard deviation σ_m
Temperature data	Mean μ_t
	Standard deviation σ_t

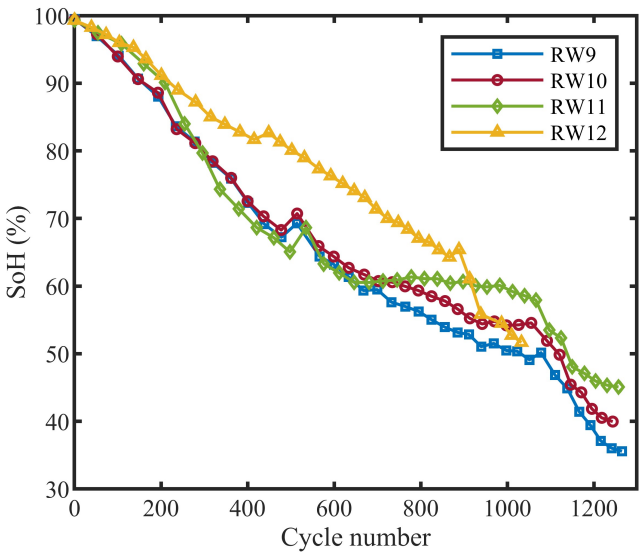


Figure 2.6. Comparison of SoH degradation among the RW9 to RW12 batteries tested under NASA RW charging and discharging conditions.

(3) Feature Correlation Analysis

Coulomb counting, combined with CC discharging data from reference tests, was employed to calculate battery capacities. To mitigate accidental errors, NASA PCoE performed two reference tests before and after every 1500 RW steps, and we subsequently averaged these results. SoH values were then determined by dividing these averaged capacities by the nominal capacity of each battery (2.11 Ah). Two adjacent SoH values were averaged to represent the SoH for each 1500 RW steps. Fig. 2.6 shows the SoH variation as the four batteries aged.

Since the raw dataset does not provide explicit cycle count information, we estimated the number of cycles using the following procedure. First, we integrated the current over time to

calculate the ampere-hour value for each time interval. Next, we summed these ampere-hour values between each pair of reference tests and divided the total by the battery's capacity prior to the RW steps. This result was subsequently divided by two, assuming one complete cycle comprises a full charge and discharge, to obtain the estimated number of cycles for the RW steps between each pair of reference tests.

Notably, RW12 was identified as an outlier due to its distinct aging pattern compared to the other three batteries. This divergence might be attributed to production inconsistencies or experimental errors. Consequently, subsequent analyses focused exclusively on RW9, RW10, and RW11.

The SoH of RW9 to RW11 recovers slightly around cycle 500 before the fade resumes. Temporary recovery of this kind is a known regeneration phenomenon in battery aging data [49]. The campaign logs of all three cells record the same roughly 125-hour rest, the longest interruption of the campaign, immediately before the reference tests near cycle 530. Rest periods temporarily restore capacity, and the recovered capacity grows with the rest duration [50]. The reported reversible mechanisms include slow lithium exchange with the passive anode overhang [51] and the partial reversal of lithium plating [50].

Feature values from the three battery samples were scaled to a range of 0 to 1 to ensure consistent evaluation across features. Fig. 2.7 visualizes the feature variations for battery RW9. Pearson's correlation coefficient r_{xy}, defined in Eq. (2.5), was employed to quantify the linear relationship between the extracted features and battery SoH.

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}} \tag{2.5}$$

where x_i represents the i-th data point of the sample; n is the sample size, and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ denotes the mean of the x sample; similarly, y_i and $\bar{y}$ represent the data points and mean of the y sample, respectively.

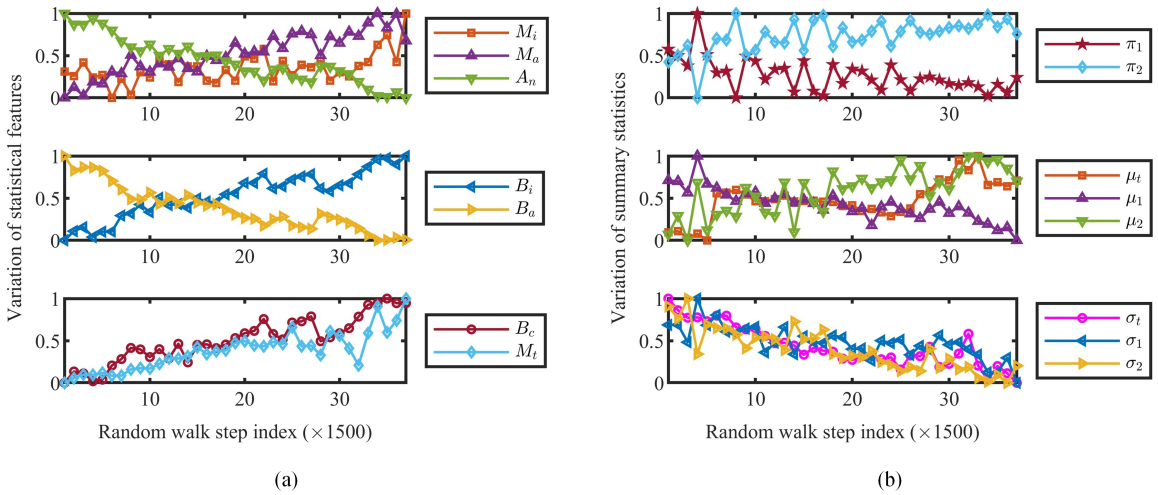


Figure 2.7. Variation of features extracted from PDFs for battery RW9 from the NASA RW dataset: (a) Summary statistics; (b) Identified distribution parameters.

The absolute values of these correlation coefficients are shown in Fig. 2.8. Features with r_{xy} below 0.3 are considered loosely correlated or irrelevant. Values of r_{xy} between 0.3 and 0.5 indicate a low correlation, values between 0.5 and 0.8 indicate a moderate correlation, and values above 0.8 indicate highly correlated features with battery SoH. Notably, the temperature data for RW11 range from 18 °C to 36 °C, differing from the 20 to 46 °C range observed for RW9 and RW10. Variations in experimental setup or measurement errors may explain the high correlation of μ_t observed for RW9 and RW10 but its weak correlation for RW11.

To effectively leverage these identified health indicators, an appropriate estimation algorithm was selected, as discussed in Section 2.3.2.

2.3.2 Estimation Algorithm

SVR is a supervised learning method derived from support vector machine (SVM). Unlike SVM, which primarily addresses classification problems involving discrete categories, SVR is designed to predict continuous real-valued outputs. SVR achieves competitive accuracy when compared with algorithms such as least squares and relevance vector machines [34], works effectively with small datasets, and avoids intensive mathematical operations [52, 53]. Due to

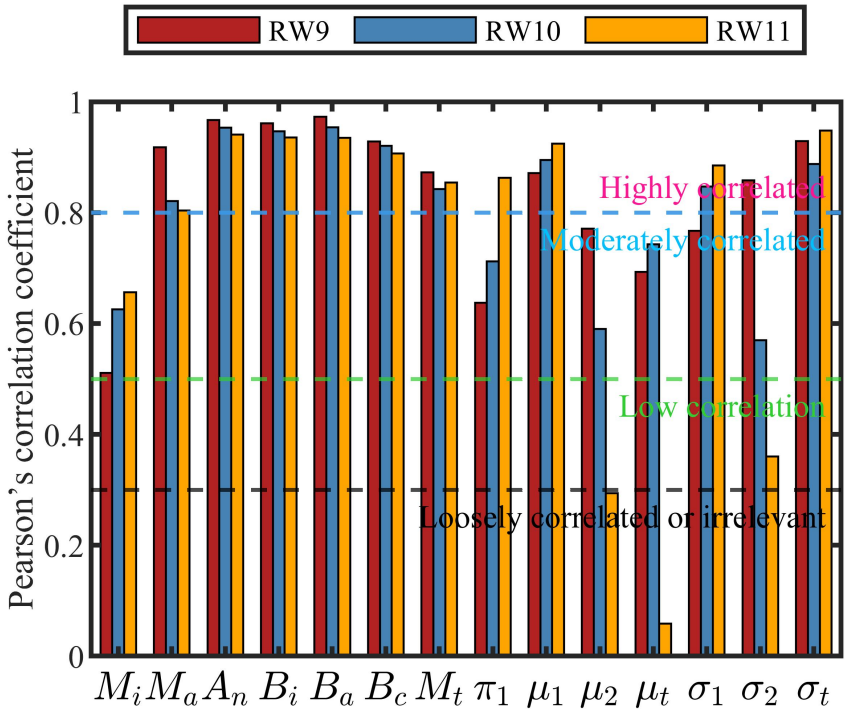


Figure 2.8. Pearson's correlation analysis between features extracted from PDFs and the SoH for batteries RW9–RW11 from the NASA RW dataset.

these advantages, SVR was adopted in this study to model the relationship between the input health indicators and the batteries' SoH.

Given a set of training data $\{x_i, y_i\}$, where $x_i \in \mathbb{R}^p$, $y_i \in \mathbb{R}$, and $i = 1, 2, \ldots, N$, SVR aims to find predicted values $\widehat{y}_i$ that differ from the actual values y_i by no greater than a specified margin ε for each training point x_i, as shown in Eq. (2.6).

$$\widehat{y}_i = \beta^{\mathrm{T}} \varphi(x_i) + b \tag{2.6}$$

where β represents the coefficient vector, b is a bias parameter, and $\varphi(\bullet)$ denotes a transformation function that maps a low-dimensional nonlinear problem into a high-dimensional linear space. The primal form of this optimization problem is defined by Eq. (2.7). However, directly solving for $\widehat{y}_i$ under these constraints may be challenging. To address this, slack variables (ξ_i and ξ_i^*)

are introduced to relax the regression constraints, resulting in the reformulated primal problem shown in Eq. (2.8).

$$\min_{\beta} \frac{1}{2}\beta^{\mathrm{T}}\beta, \text{ s.t.} \, |y_i - \widehat{y}_i| \leq \varepsilon \tag{2.7}$$

$$\begin{aligned} &\min_{\beta} \frac{1}{2}\beta^{\mathrm{T}}\beta + \lambda \sum_{i=1}^{N} (\xi_i + \xi_i^*), \\ &\text{s.t.} \begin{cases} -\varepsilon - \xi_i^* \leq y_i - \widehat{y}_i \leq \varepsilon + \xi_i \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \tag{2.8}$$

where the term $\frac{1}{2}\beta^{\mathrm{T}}\beta$ regularizes the model, as in regularized least squares, and thereby reduces the risk of overfitting. The penalty factor λ balances the training error, captured by the slack variables, against this regularization. Solving this optimization problem is simplified by converting it into its dual problem. A detailed derivation of this dual problem can be found in Ref. [54].

The features outlined in Section 2.3.1 were grouped into four sets based on their Pearson's correlation with SoH: (1) features with $r_{xy} \geq 0.8$, (2) features with $r_{xy} \geq 0.5$, (3) features with $r_{xy} \geq 0.3$, and (4) all features. Each set's criteria had to be satisfied by all three battery samples. These feature sets were input into the SVR model to estimate SoH. During training, 70 % of the 109 observations were randomly selected as training samples, and a 10-fold cross-validation procedure was used to fine-tune the regression model's hyperparameters, including ε and the kernel function. The remaining 30 % of observations formed the test dataset, and the errors from ten separate runs were averaged to obtain the final result.

In the following section, we evaluate the performance of the proposed framework by comparing it against a state-of-the-art method and seven alternative algorithms. We also assess its generalization capabilities across diverse operational conditions.

2.4 Results and Discussion

2.4.1 Comparative Analysis of the Proposed Framework

For clarity in the subsequent discussion, key terms are defined as follows: algorithms refer to computational techniques (e.g., SVR, GPR); combinations of algorithms with specific input features constitute methods (e.g., SVR with selected features); and the overall structured approach proposed in this study, encompassing feature extraction, algorithm selection, and validation, is termed the framework.

To demonstrate the effectiveness and superiority of our proposed framework, we conducted comparative studies from two perspectives. First, we reviewed published research using the same dataset (RW9–RW12). We found that most existing studies only used reference test data (CC and CV) for feature extraction, model training, and evaluation. Since such methods might become ineffective in real-world field applications, where reference test data are typically unavailable, they are not reproduced here. Liu et al. proposed a practical alternative by using statistical features from cycling data collected between reference tests [55]. Thus, we included the adaptive neural fuzzy inference system (ANFIS) method proposed by Liu et al. in Table 2.2. Another relevant method by El-Dalahmeh et al. employed adaptive empirical wavelet transform to extract voltage features from RW tests [56]. However, their method was evaluated using only one battery cell (RW9). Our assessment covered 30 % of the RW9–RW11 dataset, leading us to exclude their method from our comparative analysis.

Second, we expanded our evaluation by incorporating seven additional state-of-the-art machine learning algorithms frequently used in the recent literature. These algorithms were identified from a review of 384 recent publications up to October 2024. They include feed-forward neural network (FFNN), convolutional neural network (CNN), gated recurrent unit (GRU), and long short-term memory (LSTM) models, XGBoost, random forest (RF), and GPR [57, 58]. All eight algorithms (including ours) were implemented using the four feature sets detailed in Section 2.3.2 to demonstrate our feature selection strategies. Additionally,

we examined the direct use of PDFs identified in our study as inputs to further validate the effectiveness of the extracted features.

The comparative results, including detailed SoH estimation errors, are comprehensively summarized in Table 2.2. The three error metrics play distinct roles in this comparison. The maximum absolute error (ME) captures the single largest deviation, so it measures worst-case behavior. MAE characterizes a method's ability to consistently deliver accurate estimations across the entire dataset. RMSE also aggregates over the entire dataset. However, it penalizes large errors most heavily, which aligns with established practices in battery health assessments [59, 60]. In practical applications, such as EVs and SSSs, substantial errors in battery SoH estimation can result in significant safety and reliability concerns. A suitable method must therefore combine a low worst-case error with low average errors.

Reading the three metrics together identifies SVR with the $r_{xy} \geq 0.3$ feature set, highlighted in bold in Table 2.2, as the strongest method. It attains the lowest ME of all forty algorithm-input combinations, 6.77 %. The only combination with a lower RMSE is FFNN with $r_{xy} \geq 0.5$ (3.14 % versus 3.31 %), which also has a lower MAE (2.34 % versus 2.63 %). Its ME, however, is 9.72 %, which is 44 % larger than that of SVR. Every other combination leaves its maximum error above 7.4 %, regardless of the input feature set. SVR is therefore the only method that pairs near-best average errors with the best worst-case error, the combination that safe battery management requires. Two properties of SVR explain this robustness. First, the dataset contains only 109 observations. The ε-insensitive margin and the penalty factor λ in Eq. (2.8) bound the model complexity, which guards against overfitting to the limited training samples. The more flexible neural network-based algorithms show the expected opposite behavior. Their average errors can approach SVR's, but their worst-case errors remain larger. Second, the extracted features are low-dimensional statistical descriptors that already correlate strongly with SoH (Fig. 2.8). A kernel regression over these features only needs to capture the remaining nonlinearity. Specifically, the 10-fold cross-validation procedure tunes ε and the kernel function directly for this feature set.

Table 2.2. Comparative analysis of SoH estimation errors obtained using the proposed method, one state-of-the-art method (ANFIS), and seven advanced algorithms under various feature sets. Bold marks the proposed method (SVR with $r_{xy} \geq 0.3$). For ANFIS, ME was not reported in [55].

Algorithm/Method	**Algorithm input**	**ME (%)**	**MAE (%)**	**RMSE (%)**
GPR	$r_{xy} \geq 0.8$	9.45	2.68	3.47
	$r_{xy} \geq 0.5$	9.98	2.77	3.59
	$r_{xy} \geq 0.3$	9.65	2.59	3.38
	All features	7.90	2.67	3.39
	Identified PDFs	20.67	6.86	8.85
FFNN	$r_{xy} \geq 0.8$	11.29	3.26	4.29
	$r_{xy} \geq 0.5$	9.72	2.34	3.14
	$r_{xy} \geq 0.3$	10.29	2.66	3.71
	All features	12.43	2.60	3.60
	Identified PDFs	14.18	3.06	4.04
CNN	$r_{xy} \geq 0.8$	10.79	3.24	4.05
	$r_{xy} \geq 0.5$	11.26	3.65	4.59
	$r_{xy} \geq 0.3$	9.35	3.41	4.34
	All features	8.11	3.24	3.93
	Identified PDFs	12.47	4.35	5.52
LSTM	$r_{xy} \geq 0.8$	8.07	3.17	4.00
	$r_{xy} \geq 0.5$	8.24	3.18	4.07
	$r_{xy} \geq 0.3$	8.74	3.04	3.93
	All features	8.90	3.08	3.98
	Identified PDFs	11.62	3.16	4.25
GRU	$r_{xy} \geq 0.8$	7.69	3.08	3.86
	$r_{xy} \geq 0.5$	7.48	3.11	3.88
	$r_{xy} \geq 0.3$	8.95	3.19	4.08
	All features	9.52	3.14	4.02
	Identified PDFs	12.67	2.99	4.18
RF	$r_{xy} \geq 0.8$	10.43	3.51	4.31
	$r_{xy} \geq 0.5$	9.91	3.56	4.38
	$r_{xy} \geq 0.3$	9.96	3.57	4.41
	All features	10.05	3.50	4.36
	Identified PDFs	9.52	3.36	4.09
XGBoost	$r_{xy} \geq 0.8$	12.92	3.55	4.40
	$r_{xy} \geq 0.5$	10.34	3.44	4.41
	$r_{xy} \geq 0.3$	9.17	3.44	4.26
	All features	10.60	3.15	4.20
	Identified PDFs	12.20	3.68	4.65
SVR	$r_{xy} \geq 0.8$	7.61	2.86	3.66
	$r_{xy} \geq 0.5$	8.93	3.01	3.85
	$r_{xy} \geq 0.3$	**6.77**	**2.63**	**3.31**
	All features	9.75	3.24	4.22
	Identified PDFs	7.88	2.98	3.72
ANFIS [55]			2.57	3.73

Table 2.2 also validates the extracted features against the raw identified PDFs used directly as inputs. For six of the eight algorithms, at least one extracted-feature set outperforms the identified PDFs on all three metrics simultaneously. The exceptions are RF, which performs best with the raw PDFs, and GRU, whose MAE alone is slightly lower with the raw PDFs. These results confirm the effectiveness of our feature extraction strategy.

The ANFIS study did not report ME. The comparison with it therefore rests on MAE and RMSE. The proposed method attains a lower RMSE (3.31 % versus 3.73 %), and its MAE is 0.06 percentage points higher (2.63 % versus 2.57 %). RMSE weighs the large errors that matter most for safe battery management, and on that decisive metric our method outperforms ANFIS.

Next, we provide a detailed analysis of SVR estimation results with various feature sets. Table 2.2 shows that SVR consistently maintains MAE and RMSE values below 5 %. Incorporating moderately and weakly correlated features with highly correlated features improves SoH estimation accuracy. However, accuracy decreases when only moderately correlated features or all features are included. This reduction occurs because additional features, although informative, can introduce noise and confusion due to their varied correlation levels. Therefore, adding more features does not necessarily enhance accuracy. Additionally, Fig. 2.9 reveals that most estimation errors remain within a 5 % range, though a few outliers are present. These outliers primarily occur within the 50–70 % SoH range, accounting for approximately 57 % of the total observations, with one additional outlier observed at 37 % SoH. The elevated errors in the 50–70 % SoH range likely result from a more complex aging pattern emerging after a capacity recovery event around 70 % SoH (as shown in Fig. 2.6), complicating subsequent estimations. The isolated outlier at 37 % SoH is mainly due to insufficient training data in this lower SoH region. Only three out of the 109 observations in the dataset fall below 40 % SoH, two of which were utilized as test samples. In practical scenarios, batteries with an SoH below 40 % are typically retired, indicating that this single low-SoH outlier has minimal impact on the practical applicability of the proposed method.

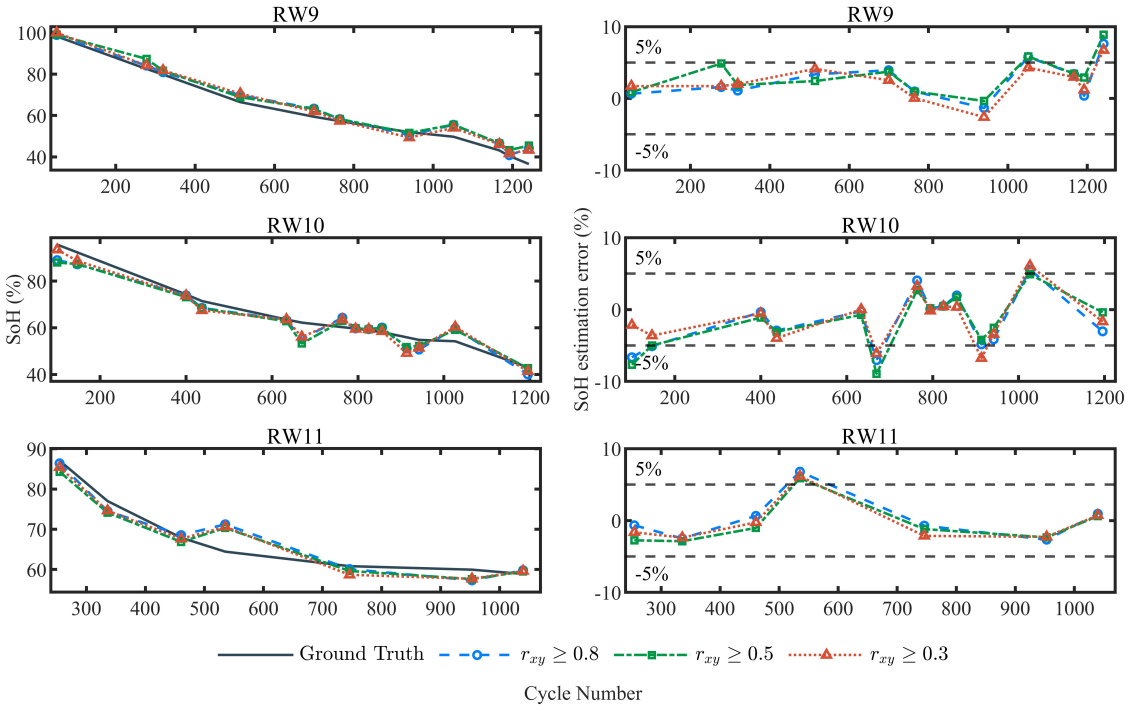


Figure 2.9. SoH estimation results for batteries RW9–RW11 from the NASA RW dataset using features with different Pearson's correlation magnitudes.

2.4.2 Verification of Generalization Capability

The comparative analysis in Section 2.4.1 demonstrated that SVR, utilizing features with Pearson's correlation greater than 0.3, achieved optimal performance under uniformly distributed RW charging and discharging conditions at room temperature. Given that ambient temperature and load profiles significantly influence battery aging, we further verified the generalization capability of our proposed method (SVR using features with $r_{xy} \geq 0.3$) using datasets collected under a different ambient temperature and varied load profiles. Specifically, battery group 1 featured randomly selected CC charging durations and CV charging. Battery groups 3 and 4 both employed full CCCV charging and featured lower-skewed and higher-skewed RW discharging currents, respectively. RW tests for groups 3 and 4 lasted 1 min compared to 5 min for groups 1 and 2. Groups 3 and 4 operated at an elevated ambient temperature (40 °C). These variations present distinct challenges to our proposed method.

Table 2.3. SoH estimation results obtained by an SVR using features with Pearson's correlation greater than 0.3 across four NASA RW test groups (Fig. 2.1).

Dataset	ME (%)	MAE (%)	RMSE (%)
Battery group 1	6.84	1.72	2.22
Battery group 2	6.77	2.63	3.31
Battery group 3	2.84	1.82	1.96
Battery group 4	5.36	1.95	2.65

The estimation results for different battery groups are listed in Table 2.3. Almost all new results, except the ME for group 1, show improved performance compared to those of group 2. This demonstrates that our method is robust to variations in ambient temperature and load profiles, which explains why our primary analysis focused on group 2. The randomized protocol used by group 2, in which both the decision to charge or discharge and the current magnitude were randomly determined, inherently complicates SoH estimation. Since the MAE and RMSE for group 1 are smaller than those for group 2, the larger ME for group 1 may be attributed to data outliers.

While the proposed framework demonstrates robust accuracy under varied conditions, practical deployment also requires data efficiency. Therefore, Section 2.5 investigates how reducing data sampling frequency impacts feature correlations and SoH estimation accuracy.

2.5 Downsampling of the Dataset

As detailed in Section 2.3.1, each PDF is constructed based on data from 1500 RW steps. Given the sampling interval (Δt) of one second, each set of input data used for estimation contains more than 3×10^5 entries. To mitigate storage and computational demands, we investigated the impact of data downsampling on SoH estimation accuracy. We increased Δt incrementally to 5 s, 10 s, 20 s, 40 s, 60 s, 120 s, and 180 s, and recalculated the PDFs under each condition. The variation in absolute correlation values between selected features and SoH with increasing Δt

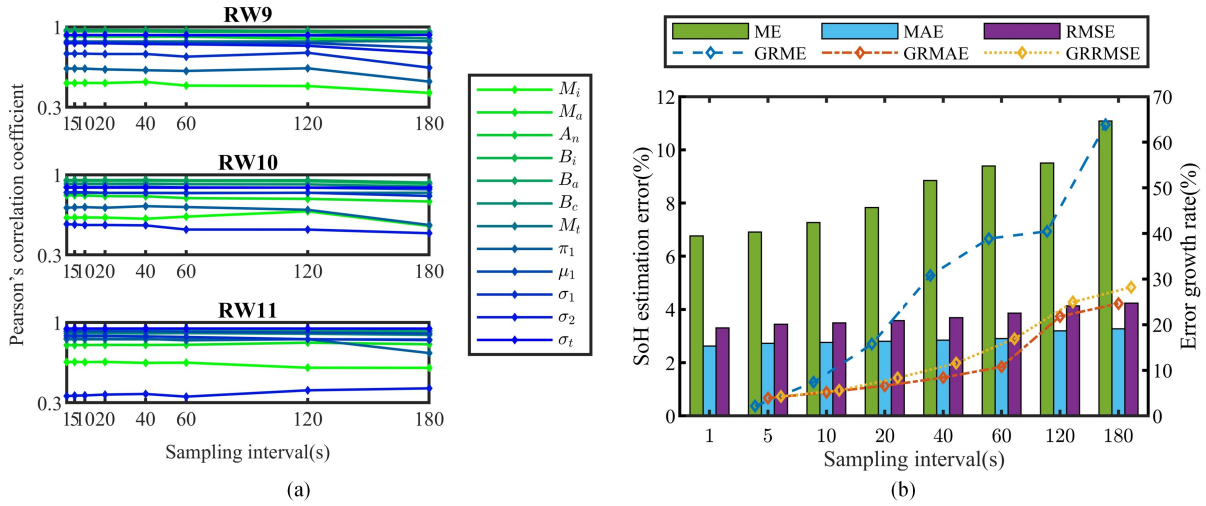


Figure 2.10. Analysis of the downsampling impact on feature correlation and SoH estimation accuracy: (a) Variation of feature correlation under different sampling intervals; (b) SoH estimation errors under different sampling intervals. GRME, GRMAE, and GRRMSE indicate the growth rates of ME, MAE, and RMSE, respectively.

is illustrated in Fig. 2.10(a). Features with correlations (r_{xy}) below 0.3 were excluded from the analysis due to their limited contribution to estimation accuracy.

As shown in Fig. 2.10(a), approximately half of the features maintain high correlations even when Δt increases to 180 s. There are no significant correlation changes for almost all the features when Δt is less than 10 s. Correlation begins to noticeably decline for some features as Δt surpasses 10 s, with significant reductions becoming evident beyond 120 s.

Fig. 2.10(b) compares SoH estimation errors at various sampling intervals. The bar graphs represent the SoH estimation error indicators, and the dashed lines illustrate the growth rates of these error indicators relative to the baseline ($\Delta t = 1$ s). Overall, all error indicators increase proportionally with larger sampling intervals. According to Table 2.4, when Δt increases to 5 s, the growth rates of all error indicators remain below 5 %. Even at $\Delta t = 10$ s, the growth rates remain below 10 %. These results highlight the low sensitivity of the proposed features to increased sampling intervals, suggesting that this PDF-based SoH estimation approach can effectively streamline data collection, storage, and computational efforts in practical applications.

To further evaluate the practical suitability of our method for real-time BMS, we analyzed

Table 2.4. Growth rates of SoH estimation errors relative to the baseline sampling interval ($\Delta t = 1$ s).

Δt (s)	**5**	**10**	**20**	**40**	**60**	**120**	**180**
ME (%)	2.17	7.37	15.74	30.79	38.87	40.47	63.87
MAE (%)	3.91	5.18	6.62	8.40	10.77	21.81	24.64
RMSE (%)	4.26	5.60	8.31	11.56	16.91	24.95	28.22

the storage requirements for PDF identification and feature extraction processes. Using RW9–RW11 data, the average storage requirement at a sampling interval of 1 s was approximately 3.18 MB. By increasing the sampling interval to 5 s, this requirement could be significantly reduced to approximately 0.64 MB, with acceptable degradation in SoH prediction accuracy (error growth rates remained below 5 %). This amount of data (0.64–3.18 MB) can be comfortably managed by typical BMS hardware. Additionally, in Section 2.4.2, we validated the effectiveness of our method under partial CC charging conditions, which commonly occur in practical battery operations, and under a full CCCV charging condition, which represents a comprehensive test scenario. These validations, conducted using battery groups 1, 3, and 4 from the NASA RW dataset, further support the method's applicability in real-time BMS environments.

2.6 Conclusion

This study introduces a novel PDF-based SoH estimation approach that extracts distribution parameters and statistical descriptors from voltage and temperature PDFs derived from RW cycling conditions, realistically representing real-world battery operations. Voltage PDFs were innovatively identified using the EM algorithm under the assumption of bimodal distributions, while temperature PDFs were obtained via distribution fitting assuming normal distributions. Leveraging these extracted features, an SVR model achieved high SoH prediction accuracy (MAE 2.63 %, RMSE 3.31 %) and demonstrated robust performance under varying load profiles and temperature conditions. Furthermore, the downsampling analysis confirmed the

method's low sensitivity to sampling frequency, showing less than 5 % increase in MAE and RMSE when the sampling interval was extended from 1 s to 5 s. This characteristic highlights the potential of the proposed method to reduce computational costs in practical applications. Although the proposed method was validated on a limited NASA RW dataset, these promising results underscore the practical applicability of the proposed PDF-based method for battery health management. Future validation with expanded datasets and more frequent reference tests could further strengthen feature-health correlations and enhance estimation reliability.

Within the dissertation, this chapter establishes the first regime of SoH estimation. The features that make it accurate are engineered. The designer selects which channels enter the PDFs and which statistics of them track degradation, choices that presume the instrumented single-cell setting of this chapter. Chapter 3 turns to retired EV battery packs, where those premises fail, and learns the representation from the data instead.

Acknowledgments

Chapter 2 is based on the following publication: Long Ling, Shaojie Yang, and Shijie Tong, "Lithium-ion battery state of health estimation based on statistical features derived from voltage and temperature probability density functions under realistic randomized cycling conditions," *Journal of Energy Storage*, vol. 137, p. 118496, 2025. The dissertation author was the primary investigator and author of this paper.

Chapter 3

Masked-Autoencoder Pretraining for State-of-Health Estimation of Second-Life Battery Packs

3.1 Introduction

The previous chapter estimated the capacity-based SoH from engineered statistical features, in a setting where single cells expose voltage and temperature and reference tests supply plentiful labels across a wide SoH range. This chapter moves to the setting those features cannot reach. Retired electric-vehicle packs expose only group-level voltage and one shared string current, and offer scarce labeled tests within the narrow SoH band each pack occupies on arrival. The representation is therefore learned from unlabeled cycling data rather than manually engineered.

Retired electric-vehicle traction batteries typically leave first-life service with 70–80 % of their initial capacity [61]. At fleet scale, second-use capacity alone is projected to be sufficient to meet the short-term grid-storage demand of 3.4–19.2 TWh by 2050 [62]. Whether each retired pack is redeployed in a stationary storage application [7] or routed to recycling hinges on its state of health (SoH), assessed through indicators such as remaining capacity, internal resistance, and open-circuit voltage [63]. This work focuses on the capacity indicator and leaves the others outside its scope. SoH hereafter denotes the ratio of remaining capacity to nominal capacity.

That screening decision is only the first time SoH must be known. A redeployed pack must then be tracked as it continues to age in service for safe operation and energy management. A reference capacity test measures SoH directly, but it suits neither use. For screening it cannot scale as each test runs for hours, one pack at a time. For in-service tracking it can only be done infrequently. Measuring capacity requires a controlled, near-full charge/discharge cycle. The partial, irregular cycling of normal second-life duty rarely supplies such a cycle. Each measurement therefore requires taking the pack out of service. Predicting SoH from operating data is the alternative, but conventional supervised learning still needs SoH labels. In practice, during algorithm development those labels come from reference tests on the second-life packs. The trained model must then serve packs beyond this development set. Labels collected this way, however, are narrow-band and pack-specific. A supervised regressor therefore never sees enough of the degradation range, or enough distinct packs, to generalize across SoH levels and chemistries.

Data-driven SoH estimation has progressed from statistical and feature-based shallow estimators [64, 65] to deep neural estimators [66], and most recently to self-supervised and foundation-model approaches that pretrain on multi-source cycling data before transferring to the target [67–69]. The accuracy advantage of self-supervised pretraining is largest when labeled data are scarce [70–72]. Second-life screening is appealing. Unlabeled voltage/current streams are abundant, since every cycler and battery management system (BMS) produces them. SoH labels, however, require reference tests. Self-supervised learning converts this abundant unlabeled stream into a training signal. Among its objectives, this paper adopts the masked autoencoder (masked AE) [73], which masks part of a short window of the measured signals and trains an encoder–decoder pair to reconstruct it. Two properties motivate this choice. First, reconstructing the masked samples forces the encoder to model the voltage–current dynamics from which SoH must be inferred. Second, masked pretraining suits the channel asymmetry of second-life screening. Laboratory cells are usually fully instrumented with voltage, current, and temperature (V/I/T). Many deployed packs, however, expose only voltage and current (V/I). An

encoder can pretrain on the full V/I/T channels and transfer to V/I-only targets, a property this paper quantifies directly (Section 3.5.3).

Despite this appeal, published studies of this approach stop short of the second-life setting in two respects. First, self-supervised SoH estimation has been evaluated almost entirely on datasets with predictable cycling, collected from laboratory single cells or fleet charging records. No masked-AE-pretrained method has been benchmarked on the NASA Randomized Battery Usage dataset (NASA RW) [74], whose random-walk (RW) load profiles lack the cycle-scale regularity that data-driven methods usually rely on. Concretely, there is no repeating charge/discharge template along which features can be aligned from one cycle to the next. We adopt this dataset because that irregularity exemplifies real-world challenges, not as a proxy for any particular drive cycle. Evidence from our earlier work supports this choice [65]. A model able to estimate SoH under such inconsistent usage transferred readily to the more regular charging protocols that dominate controlled testing (full or partial constant current constant voltage (CCCV) charging, with discharge still randomized). None of the recent battery foundation models include this dataset in their corpora [67–69].

Second, and more consequentially, validation on connected second-life systems is absent. Retired packs arrive as series–parallel assemblies whose built-in sensing is sparse by construction. A series string shares one current sensor, since the same current flows through every cell. Cells wired in parallel share one pair of terminals, so a parallel group exposes a single common terminal voltage that no sensor placement can resolve into per-cell values. Voltage sensing on such a pack is therefore group-level by construction. Temperature sensing is sparse for a different reason. Many packs instrument only a few locations rather than every cell, because a per-cell thermal harness would add sensor and wiring cost. These sensing constraints compound the irregular-cycling constraint described above. We therefore deliberately quantify how well SoH can be estimated from V/I data alone, collected from irregular cycling where a reference capacity test is not applicable. Whether self-supervised features survive the transfer from instrumented single cells to these deployment-realistic measurement conditions is untested.

Beyond these two gaps, the field operates on the assumption that more pretraining data helps. That assumption has not been checked against chemistry-mismatched targets. The two closest prior efforts transfer SoH estimators across batteries. One adapts an estimator to retired second-life cells through deep generative transfer [75]. The other transfers features from public single cells to in-service vehicle packs [76]. Neither examines how the composition of the source data shapes transfer. Moreover, neither separates the inductive (unseen-target) from the transductive (seen-target) transfer regime [77].

This paper closes both gaps, and tests the more-data assumption, with a single masked-AE framework evaluated across one public benchmark and three in-house second-life systems. The systems are two retired lithium-manganese-oxide (LMO) packs and a nickel-cobalt-aluminum-oxide (NCA) pack. The same Transformer encoder is pretrained by masked reconstruction on raw V/I or V/I/T windows. It is then fine-tuned under a cross-window attention pooler and a regression head. The pooler attends over all windows of a cycling segment, the stretch of cycling between two consecutive capacity measurements, and the head maps the pooled segment to a single SoH estimate. The two-stage design follows from the economics of the data. Unlabeled cycling streams and wide-range SoH labels are abundant on instrumented single cells, where deployment does not happen. On connected packs, where deployment does happen, both are scarce. Pretraining draws on the abundant regime so that fine-tuning can spend the costly pack-level labels efficiently. The five pretraining sources are a combined four-source pool, leave-one-out lab pools, NASA with and without temperature, and a random-initialization control that skips pretraining altogether. These are crossed with three second-life targets at three seeds each. The resulting transfer ablation is the core result of the paper.

The contributions are fourfold:

- **Public benchmark.** On NASA RW under a stratified-by-minimum-SoH cell-level split, the masked-AE-pretrained model reaches 2.82 ± 0.28 %SoH test root-mean-square error (RMSE, $n = 3$ seeds). LightGBM, bidirectional long short-term memory (BiLSTM), and

PatchTST baselines trail at 1.5–3.2× this error. A 14-fold leave-one-cell-out validation yields a median of 3.91 %SoH (Section 3.5). The identical architecture trained from random initialization matches the pretrained model's accuracy. The margin over the three external baselines therefore traces to the encoder architecture that both variants share, not to the pretrained weights. That architecture is itself a product of the pretraining pipeline, since it is the configuration selected by the pretraining hyperparameter search.

- **Composition over scale.** The combined four-source pool attains the lowest mean RMSE of any pretraining source on both LMO packs. On the NCA pack, however, it attains the highest mean RMSE. Pool composition, not size, thus governs the sign of transfer. Every configuration that fails there draws on a corpus saturated by one narrow SoH band. The failure is therefore diagnosable in advance from corpus composition. A seen-target control (NASA plus the target's own pack) indicates that having the target's own data in the corpus is not sufficient for the LMO gain. The target-free NASA and leave-one-out sources indicate it is not necessary for transfer (Section 3.5.3).

- **Wide-range pretraining as a stabilizer.** On the narrow-band, seed-unstable NCA pack, single-cell NASA (V/I/T) pretraining is at once the most accurate and the most stable configuration across seeds. The result identifies a stability role for wide-SoH-range pretraining that single-seed evaluation cannot detect (Section 3.5.3).

- **Deployment-realistic evidence and release.** All transfer results hold under production-BMS sensing constraints (V/I only, shared pack current). The three second-life datasets, code, and trained checkpoints are released.

The remainder of the paper is organized as follows. Section 3.2 positions the work against prior SoH estimation on randomized-usage data, battery foundation models, and second-life screening. Section 3.3 describes the four datasets and their preprocessing. Section 3.4 details the masked-AE framework. Section 3.5 reports the benchmark and transfer results, Section 3.6

interprets them, and Section 3.7 concludes.

3.2 Related Work

3.2.1 SoH Estimation on Randomized Battery Usage Data

The NASA RW dataset is widely used as a public proxy for field-like loading. Reported accuracies on it are comparable only under a common evaluation protocol, especially the train/test cell split. Early work established cell-level evaluation. Richardson *et al.* [64] reached a normalized capacity RMSE of ≈4.3 % with Gaussian-process transition models over 26 cells. Aliberti *et al.* [66] trained a one-dimensional convolutional neural network on per-step aggregate features, reporting 3.5 %SoH RMSE with one cell fully held out. The strongest published cell-level result is the multifeature linear regression of Ospina Agudelo *et al.* [78], ≈2.45 %SoH averaged exhaustively over all 256 possible 12/4 train/test cell splits of four groups. That pipeline relies on hand-engineered incremental-capacity and dynamic-resistance features rather than a learned representation. Many other studies evaluate on within-cell splits, in which segments of one cell appear in both training and test. This leaks the cell's degradation context into training. Their numbers are therefore not comparable with cell-level results. Moreover, several studies predict capacity-curve trajectories or remaining useful life rather than per-segment SoH. Our previous work [65] used support vector regression on voltage/temperature probability-density features over three RW cells. It sits in the feature-engineering lineage that the present raw-window approach replaces.

3.2.2 Self-Supervised and Pretrained Models for Battery Health

Self-supervised pretraining has entered battery health estimation along three lines. The first uses masked or denoising reconstruction on cycling signals from the deployment domain itself. Shen *et al.* [79] pretrain a masked AE on unlabeled field charging data and fine-tune with 20 % of the usual labels. Wang *et al.* [80] combine patch-based masking with cross-

variate attention on one year of data from 41 vehicles. Yuan *et al.* [81] integrate reconstruction, denoising, and prediction objectives for multi-step SoH forecasting. These works demonstrate label efficiency. Shen *et al.* and Yuan *et al.* pretrain and fine-tune within a single domain. Wang *et al.* evaluate chemistry transfer, but between vehicle fleets rather than across measurement regimes. None of the three reaches connected second-life packs.

A second line targets dynamic-profile data directly. Chen [82] learns a Transformer-based health index from unlabeled V/I sequences under real-driving profiles, using a loss that encodes the irreversibility of capacity degradation. Of the three lines, this one is closest in spirit to RW usage. It remains, however, single-cell.

The third line is battery foundation models that pretrain once on many datasets. FMAE [67] learns channel-flexible representations across eleven datasets. LiPM [83] adds physics-aware masking over eight. PPT [84] scales a 1.9M-parameter convolutional transformer to hundreds of commercial cells. Two concurrent models, PBT [68] and the unified model of Chan *et al.* [69], push the corpus to 13–20 datasets. These works share a scale-first premise. Specifically, they aggregate as much cycling data as possible, pretrain once, and fine-tune everywhere. None includes RW data in its corpus. LiPM's "NASA" source is the CCCV batch, not NASA RW. Moreover, the dataset enumerations of FMAE, PBT, and Chan *et al.* list no NASA source at all. None evaluates on connected second-life systems either. FMAE's system-level tasks use first-life electric-vehicle fleets and one first-life grid-storage system. Chan *et al.* [69] include one laboratory parallel-module set. The other downstream targets are instrumented single cells under scripted protocols. The present study is complementary in scope and corrective in one conclusion. With matched architecture and hyperparameters across pretraining sources, we find that corpus composition dominates corpus size. Specifically, the largest pool is the worst choice for a chemistry-mismatched target (Section 3.5.3).

3.2.3 Second-Life Battery Screening

Second-life deployment shifts the measurement regime from instrumented cells to series–parallel packs observed through production BMS channels (group-level voltages, one shared current, and frequently no usable temperature). Prior work from our group addressed the label-scarcity half of this problem with domain-adversarial transfer from a public first-life cell dataset to retired second-life batteries [85]. That work operates on derived dQ/dV or feature curves rather than raw BMS streams. Moreover, it does not test whether self-supervised pretraining survives the topology and channel constraints of connected packs.

More broadly, battery SoH transfer work spans cross-cell domain adaptation [86] and test-time-adaptive foundation models [87], as surveyed in [88]. Within this breadth, two recent efforts come closest to our setting. Tao *et al.* [75] estimate the remaining capacity of heterogeneous retired cells by deep generative domain-adaptive transfer using a small fraction of target data. Rahman and Deb [76] transfer incremental-capacity and differential-voltage features from public single cells to in-service vehicle packs. Neither, however, pretrains an encoder by self-supervision on a *public single-cell* corpus and transfers it to *connected* second-life packs under V/I-only, shared-current sensing. This is our setting. Both adapt to a target whose data is partly in hand, a seen-target regime. To our knowledge, this is the first public release of a second-life battery-pack corpus for evaluating self-supervised transfer under these sensing constraints (Section 3.3). The claim concerns the release of the data, not the first study of such packs. What has not been available before is a public corpus that records retired packs through production BMS channels, with voltage and current only and one shared string current, together with the reference-capacity labels a transfer study requires. Moreover, the transfer ablation of Section 3.4.5 is built directly on this corpus.

3.3 Datasets

The data used in this study comprise one public single-cell benchmark and three in-house second-life battery-pack datasets. Table 3.1 summarizes the four datasets along the dimensions that drive the experimental design of this study: cathode chemistry, topology, channel availability, sampling rate, cycle regime, and operating SoH range. The two dataset families are complementary by construction, reflecting the asymmetry that shapes the entire study. Specifically, NASA RW supplies relatively abundant SoH labels that span nearly the full degradation range. Each second-life pack supplies far fewer labeled cycling segments (stretches of cycling, each bounded by two consecutive capacity measurements), confined to the narrow SoH band it occupies on arrival. No pack carries records of its first-life cycling. Within this design, NASA RW serves three distinct functions. It is the public benchmark anchor, the only dataset on which the framework can be compared with published methods. This connects the study to our previous work on statistical-feature-based SoH estimation under RW conditions [65]. It is the wide-band, fully instrumented pretraining source, supplying what the lab packs structurally cannot: a 35.6–100.0 %SoH label range, a temperature channel, and RW usage diversity. And it is the out-of-domain control in the transfer ablation of Section 3.5. The three in-house packs likewise serve two functions. They are the deployment-realistic fine-tune targets. Specifically, they are genuine retired electric-vehicle packs that carry real, heterogeneous first-life aging histories. They are sensed as production systems are, through group-level voltages and a single shared pack current. They also serve as pretraining sources in the combined and leave-one-out pools. This dual role enables the corpus-composition comparison of Section 3.5.3.

Together the four datasets span an overall SoH range of 35.6–100.0 %, three cathode chemistries (lithium-cobalt-oxide (LCO), NCA, and LMO), three topology classes (single cell, series-connected brick pack, series-connected pouch pack), and two cycle regimes (RW and predefined). Additionally, the asymmetric channel availability (temperature is retained only for the NASA dataset, and only the NASA dataset provides per-cell current) is a structural feature of

Table 3.1. Summary of the four datasets. The cell-group unit is the physical unit behind a single observed voltage channel.

	NASA RW	NCA	LMO-A	LMO-B
Cathode chemistry	$LiCoO_2$ (LCO), 18650	$LiNi_{0.8}Co_{0.15}Al_{0.05}O_2$ (NCA), 18650	$LiMn_2O_4$ (LMO), pouch	$LiMn_2O_4$ (LMO), pouch
Cell-group unit	1 cell	1 brick (64 cells ‖)	1 pouch	1 pouch
Voltage channels per pack	1	12 ($= 6 \times 2$ in series)	16 ($= 8 \times 2$ in series)	16 ($= 8 \times 2$ in series)
Current channel	per-cell	shared pack current	shared pack current	shared pack current
Temperature channel	per-cell	not retained	not retained	not retained
Sampling rate	1 Hz	1 Hz	1 Hz	1 Hz
Cycle regime	random walk + reference	predefined + reference	predefined + reference	predefined + reference
SoH range (%)	35.6–100.0	86.4–96.3	63.7–74.2	60.2–74.0
Cycling segments	271 (over 14 cells)	49 (over 12 bricks)	94 (over 16 pouches)	105 (over 16 pouches)

the corpus that the proposed masked-AE framework is specifically designed to accommodate.

3.3.1 NASA Randomized Battery Usage Dataset

The NASA RW dataset was collected by the NASA Ames Prognostics Center of Excellence [74]. The full dataset comprises seven release groups, 28 cells in total. Each group contains four cells of the same commercial LCO cathode / graphite-anode chemistry [89], cycled under a distinct combination of ambient temperature, current bound, and randomization scheme. This study draws on four of the seven groups, which supply 16 cells. Two cells are excluded as outliers, leaving the 14-cell NASA RW corpus. Cross-cell differences therefore arise from test conditions rather than cell type. In every group the applied current varies in randomized steps within a bounded range. Which operations are randomized differs by group. This randomization mimics the unpredictable usage patterns of many practical battery applications. RW cycling was periodically interrupted for reference capacity tests. These tests provide the capacity measurements from which per-segment SoH labels are derived. The full per-group schedules are given in the dataset documentation [74]. Voltage, current, and surface temperature were recorded at 1 Hz throughout the RW cycling.

We define SoH as measured reference capacity divided by a single fixed reference capacity (2.107 Ah), following the standard rated-capacity convention [90] rather than per-cell initial capacity. This places all 14 cells and the lab packs on one absolute SoH scale. The denominator is taken from the cells that were characterized at standard temperature and begin

essentially at their rated capacity. The eight cells of the two warmer groups (median surface temperature $\approx$45 °C versus $\approx$31 °C elsewhere) measure marginally ($\leq 2\,\%$) above nominal at their first reference test. This reflects batch variation [91] and the reduced cell polarization reported at elevated temperature [92], not measurement error. We clip the labels to the physically meaningful range $[0, 100]\,\%$, which removes this beginning-of-life artifact and matches the bounded output of the regression head used at fine-tuning time (Section 3.4). The 14 cells then span SoH from 100.0 % (beginning of life) down to 35.6 % (end of the test program). The same corpus and labels serve both masked-AE pretraining (Section 3.4) and the in-distribution fine-tune evaluation (Section 3.5).

3.3.2 Second-Life Automotive Battery-Pack Datasets

Three in-house datasets were acquired in 2020–2023 on a laboratory cycling rig at the University of California, San Diego. The packs were retired from automotive service after their first life. We anonymize the source-vehicle provenance. The datasets are therefore described throughout this paper by their cathode chemistry and topology only, as NCA, LMO-A, and LMO-B.

Fig. 3.1 illustrates the test rig architecture. Each test rail comprises a programmable power supply (tester), a relay board, the battery pack under test, and a dedicated BMS (Orion BMS 2, 24-cell variant). The relay board serves as a hardware safety interlock. Specifically, the BMS continuously monitors per-pouch (LMO) or per-brick (NCA) voltage against protection thresholds. Upon detecting an over-voltage, under-voltage, or over-temperature condition, it pulls a hardware enable line that opens the relay and physically isolates the pack from the tester, independent of the host-PC software state. A shared logging PC records pack-level voltage, current, and per-pouch/brick voltage from all three BMS units via a controller area network (CAN) bus. The logger oversamples at 3–11 Hz with 1 s timestamp resolution. Every channel is downsampled to a uniform 1 Hz, keeping the first sample of each timestamp second, before release.

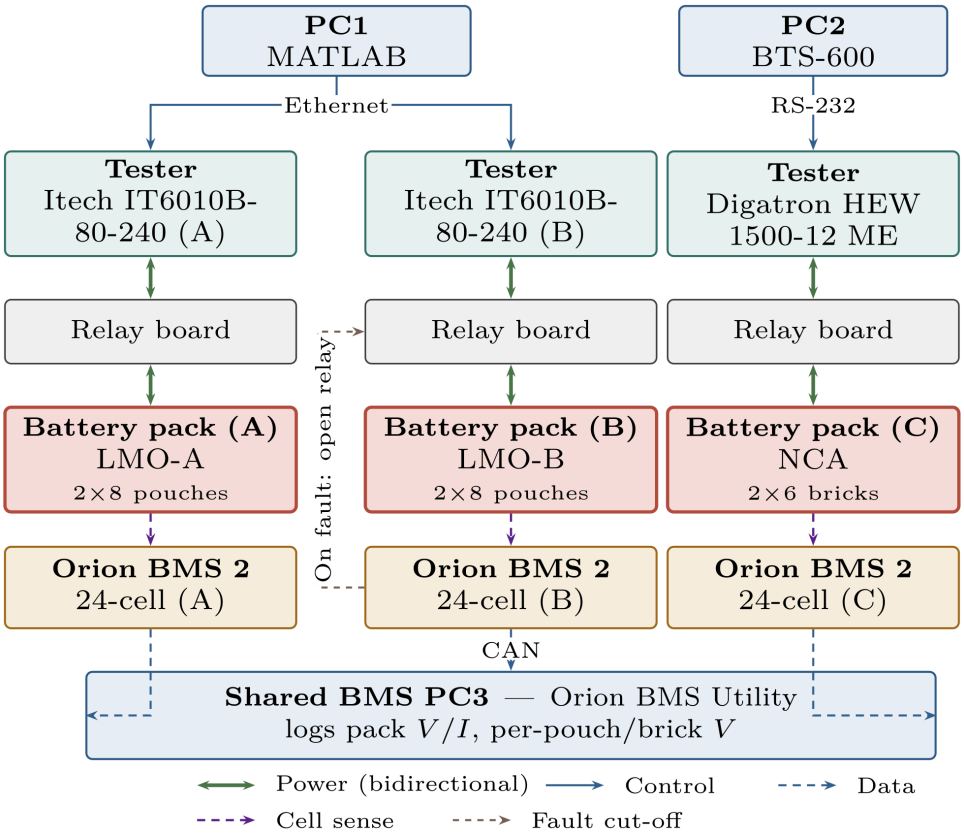


Figure 3.1. Architecture of the second-life battery test rig.

(1) NCA Dataset

The NCA dataset records a connected pack of 18650 NCA ($LiNi_{0.8}Co_{0.15}Al_{0.05}O_2$) cathode cells with graphite anodes. The pack topology aggregates 64 cells in parallel into a single brick (nominal brick capacity 198.4 Ah). Six bricks are connected in series to form a module. Two modules in series form the test pack. The measured signals comprise 12 brick-level voltage channels and a single pack-current channel shared by all bricks. Per-cell voltages within a brick are not observable. The brick is therefore the smallest unit visible to any model trained on this dataset. The dataset contains 49 cycling segments over the 12 bricks, spanning SoH 86.4–96.3 %. This is a narrow operating band at the high-SoH end of the corpus, consistent with the moderate-cycling-history provenance of the source pack.

(2) LMO-A and LMO-B Datasets

The LMO-A and LMO-B datasets record two units of the same retired second-life LMO automotive pack. Each pack consists of eight prismatic-pouch cells ($LiMn_2O_4$ cathode, graphite anode, nominal per-pouch capacity 56.3 Ah) connected in series to form a module. Two modules in series form the test pack. Per-pouch voltage is observable directly (16 voltage channels per pack). A single pack-current channel is shared across all pouches. Although the two units share a pack design, they entered testing at different states of degradation and aged at different rates under the same nominal cycling protocol. Each pack was cycled on its own test rail, with its own tester and BMS (Fig. 3.1). Divergent aging among nominally identical cells is well documented [91]. They are therefore treated throughout this paper as independent datasets. Transfer between them probes unit-to-unit variation within one chemistry and topology rather than cross-design generalization. LMO-A contains 94 cycling segments spanning SoH 63.7–74.2 %. LMO-B contains 105 cycling segments spanning SoH 60.2–74.0 %. Both packs operate in the deep-degradation region of their nominal life. Among the three second-life packs they occupy the lowest and the widest SoH bands. Both bands lie at or below the 70–80 % window in which traction packs typically retire from first-life service, whereas the NCA pack's band lies well above it. The widest coverage in the corpus as a whole, however, comes from NASA RW, whose labels span 35.6–100.0 % and contain the LMO band in full (Table 3.1).

(3) Cycling Protocol and Channel Availability

All three lab datasets followed a predefined cycling protocol. Each pack's campaign begins with a reference capacity test and then alternates aging tests with further reference tests. The aging tests emulate energy storage system (ESS) operation in stacked grid applications. Specifically, the charge/discharge schedule follows a 7-day pattern derived from an optimal ESS dispatching algorithm that co-optimizes five utility-grid services using California electricity-market pricing data [93]. The resulting current profiles therefore reflect deployment-realistic power demands rather than idealized constant-current (CC) cycling. Both test types nonetheless

open with the same initial CC charge/discharge cycles, from which a capacity is obtained at every test (Fig. 3.2). A reference test then runs only a few additional cycles. An aging test instead continues with the long ESS profile.

We deliberately retain only the voltage and current channels in the three lab datasets. This V/I-only sensing matches the channel availability of many production BMSs. The temperature channel is therefore excluded from the released lab artifacts and from all modeling experiments. The resulting asymmetry between NASA RW (V, I, and T at the cell level) and the three lab datasets (V and I only, with shared current) motivates several of the cross-corpus transfer experiments reported in Section 3.5.

(4) Reference Capacity Measurement

Each reference capacity test (Fig. 3.2) follows a structured protocol designed to yield a repeatable capacity measurement: (i) two conditioning CC charge–discharge cycles at the rated level (20 A for LMO, 40 A for NCA), which relax concentration gradients left by the prior aging test and establish a repeatable voltage-hysteresis state; (ii) a CCCV charge to a precise full-charge state independent of the aged internal resistance, followed by a rest; (iii) an impedance pulse ($\pm$4.9 A NCA, $\pm$5 A LMO) and rest, after which the settled terminal voltage V_1 approximates the open-circuit voltage (OCV) at the high state-of-charge (SoC) reference point; (iv) a full CC discharge with coulomb counting yielding the accumulated charge ΔAh, followed by a rest; and (v) an identical pulse-and-rest sequence giving the low-SoC reference V_2. The cell-group-unit capacity is then

$$C = \frac{\Delta\text{Ah}}{\text{SoC}(V_1) - \text{SoC}(V_2)}, \tag{3.1}$$

where $\text{SoC}(V)$ maps each near-OCV voltage to SoC via the chemistry-specific OCV–SoC curve. Anchoring the measurement at two OCV-approximated reference points removes the influence of ohmic and diffusion overpotentials on the capacity estimate. This consideration grows in importance as internal resistance rises with aging. The SoH label used throughout this study is

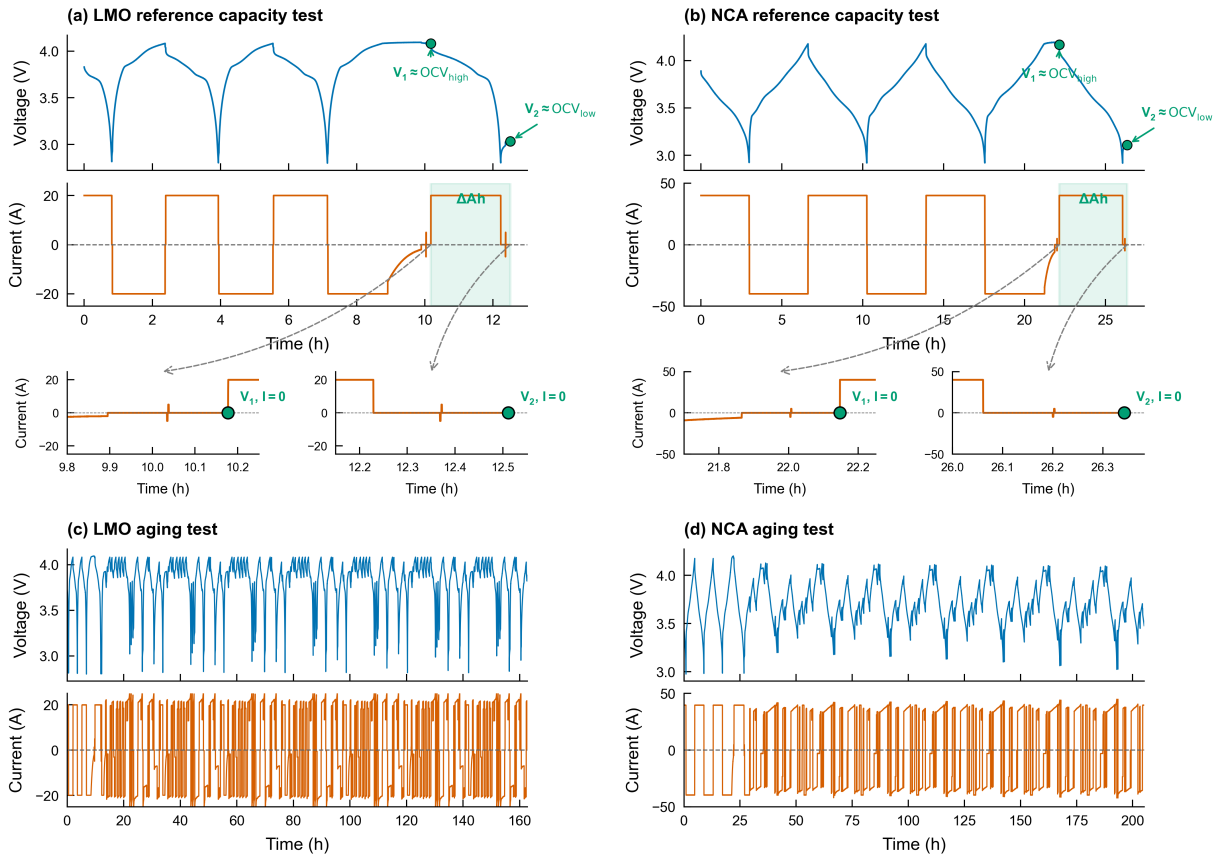


Figure 3.2. Reference and aging test profiles for the lab packs. Each column is one pack. **(a)–(b)** A reference capacity test. **(c)–(d)** An aging test, i.e. one cycling segment.

the ratio of this measured capacity to the nominal capacity.

3.3.3 Common Preprocessing

Four preprocessing steps were applied to all four datasets. First, per-sample voltage and current clipping (a sample is one 1 Hz timestep) was applied at the dataset-specific bounds listed in Table 3.2. Clipping rejects measurement artifacts and BMS transients while preserving the charge/discharge extrema required for SoH inference. For the three lab packs the bounds are the 0.1–99.9 percentiles of each channel's pooled data, rounded outward to the nearest 0.1 V and 0.5 A. The two LMO packs share the same voltage bounds and upper current bound. Their lower current bounds differ by a single 0.5 A rounding step because their different aging states slightly shift the realized cycling currents. For NASA RW the current bounds follow the same percentile

rule. Its voltage bounds retain the benchmark's known physical operating range.

Table 3.2. Per-sample clip bounds applied at data-loading. The NASA RW temperature channel is additionally clipped to $[17, 58]\,^{\circ}\mathrm{C}$.

Dataset	$V_{\min}$ (V)	$V_{\max}$ (V)	$I_{\min}$ (A)	$I_{\max}$ (A)
NASA RW	3.0	4.2	-4.5	5.0
NCA	2.9	4.2	-59.5	61.5
LMO-A	2.8	4.2	-26.0	38.5
LMO-B	2.8	4.2	-25.5	38.5

Second, a segment-level hard-reject filter was applied at data loading. Any cycling segment in which a voltage sample fell outside the physical range [1.0, 5.0] V or a current sample exceeded ± 100 A (LMO) or ± 150 A (NCA) was dropped in its entirety. These thresholds are unreachable in normal operation and can only be explained by instrumentation faults. Extreme but genuine samples are instead handled by the per-sample clipping of the first step. The hard-reject filter catches synchronous BMS-dropout events that per-sample clipping cannot recover. On the NCA pack the hard-reject removed one segment whose voltage log drops out synchronously across all 12 bricks. It removed nothing from the LMO packs or NASA RW. Third, each channel is standardized to zero mean and unit variance. The statistics for the standardization are computed on the training split of each experiment (Section 3.4.1) and reused for its validation and test splits. Therefore, no test information enters the scaling. Finally, each cell-group unit's reference-capacity measurements form a boundary sequence, with $k+1$ measurements bounding k cycling segments. Per-sample SoH is linearly interpolated between a segment's two bounding measurements. Each cycling segment is then assigned a single label, attached to the full window sequence of that segment. The label is the clipped segment-closing reference-capacity measurement. This per-segment label contract is the substrate on which the cross-window pooler of Section 3.4 operates.

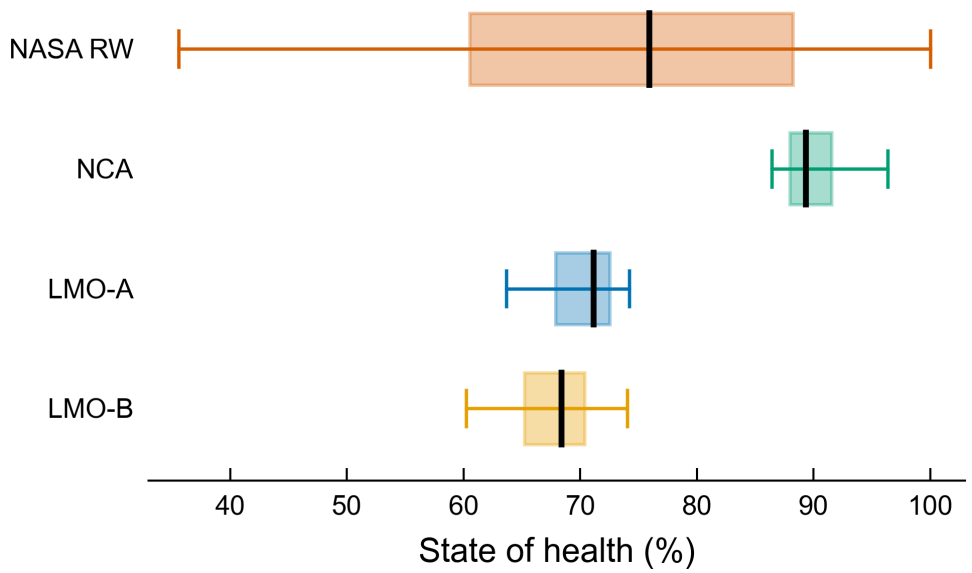


Figure 3.3. SoH-label distributions of the four datasets on a common axis (box: interquartile range; line: median; whiskers: full min–max range over all labeled SoH values). NASA RW spans nearly the full range, while the second-life packs occupy distinct, narrow deployment bands within it.

3.3.4 Cross-Dataset Comparability and Data Release

Fig. 3.3 compares the four SoH label distributions on a common axis. NASA RW spans nearly the whole range. The second-life packs occupy distinct, narrow deployment bands within it. The NCA pack sits high, well separated from the two LMO packs. Being units of one pack, the LMO packs share a lower band. This banding is a feature of the second-life deployment regime. Each retired pack arrives at its current SoH band before any in-house measurements begin. The lab targets thus probe narrow, non-overlapping (NCA vs. LMO) operating windows rather than the full degradation trajectory. Metrics are therefore reported per dataset in Section 3.5, since pooling the packs into one split would conceal this band structure. The three in-house datasets are released under the CC BY 4.0 license together with a sample loader (DOI reserved at submission).

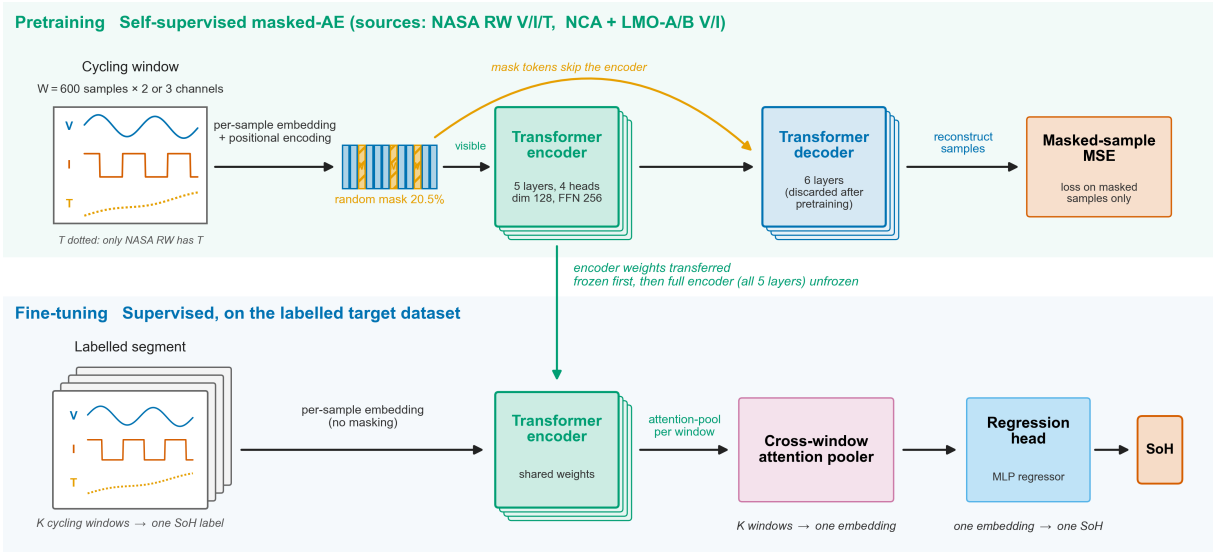


Figure 3.4. Two-stage framework. Pretraining reconstructs masked samples of a pretraining corpus with a Transformer encoder and decoder. Table 3.4 lists the pretraining conditions compared. Fine-tuning then transfers the encoder and trains it, frozen at first and unfrozen thereafter, together with a cross-window attention pooler and a regression head that predicts one SoH value per labeled cycling segment.

3.4 Proposed Framework

The overall two-stage methodology is illustrated in Fig. 3.4. The primary innovation of our approach lies in pretraining a shared masked-AE encoder on a multi-source corpus that spans both single-cell RW cycling and connected second-life battery packs. The same encoder is then fine-tuned with a cross-window attention pooler that aggregates the windows of each cycling segment into one embedding, which a regression head maps to a single SoH estimate. Pretraining uses masked-reconstruction self-supervision on V/I windows, and V/I/T for NASA RW. A fixed fraction of each window's timesteps is hidden, with the hidden positions drawn afresh on every pass, and the model reconstructs them. All sources share a common window length. At fine-tuning, only the encoder weights are transferred. The pooler and the head are trained from scratch on each dataset's own SoH-labeled cycling segments. We hypothesize that the multi-source pretraining corpus spans a wider range of V/I dynamics than any single dataset the encoder is later fine-tuned on. Pretraining therefore equips the encoder with representations it

could not learn from that target alone, which should lower SoH error on the target's held-out cells.

3.4.1 Cell-Level Stratified Data Preparation

Across all four datasets, a per-segment SoH label is assigned to the full window sequence of that segment, and all channels enter the model clipped and standardized with train-split statistics (Section 3.3.3). Each segment is partitioned into fixed-length windows of $W = 600$ samples, and each sample forms one timestep of the encoder input. At the 1 Hz sampling of NASA RW, a window therefore spans 10 min. The second-life segments are far longer. Their streams are subsampled by a factor of five when loaded. A window there spans 50 min. This keeps the median number of non-overlapping windows per segment comparable across the four datasets, between roughly 170 and 250. Consecutive windows overlap by 50 % (stride 300), which roughly doubles the number of windows drawn from each segment. The recorded cycling is fixed in both stages, so the overlap raises the training signal without new data. It helps most at fine-tuning time, because a segment carries a single SoH label however long it is. Fine-tuning keeps this stride when it evaluates, so that a segment is divided into windows the same way at training and at test time. Pretraining instead evaluates with stride 600. Its reconstruction windows then do not overlap, so the reported error weights every part of a held-out segment equally. Windows never cross segment boundaries. When a segment is shorter than W, the sequence is right-padded to one full window with a masking flag that excludes padded positions from both the reconstruction loss and the SoH prediction. When the sliding stride does not land on the segment end, an additional end-aligned window is appended so that the final samples of every segment are always covered.

For the fine-tuning stage, cell-level partitions are constructed by ranking all cells in a dataset by their minimum observed SoH in ascending order, with the most-degraded cell ranked first. The validation and test cells are then selected at evenly spaced positions along this ranking. NASA RW is split into 10 training, 2 validation, and 2 test cells. LMO-A and LMO-B are

each split into 12/2/2 pouches. NCA is split into 8/2/2 bricks. This minimum-SoH-stratified split guarantees that the train and test pools each span the full observed degradation spectrum, removing the trivial confound in which the test set occupies a narrower SoH band than the training set. The pretraining stage instead splits segments within each cell in a 70/15/15 ratio across pretrain-train/val/test pools. Its self-supervised objective does not require unseen-cell generalization. The data partitions are deterministic and bit-for-bit reproducible. Reported metrics carry the small seed-level run-to-run variation quantified in Section 3.5.

3.4.2 Masked-Autoencoder Pretraining

(1) Encoder Architecture

The encoder (Fig. 3.4, pretraining) is a standard Transformer encoder [94] adapted to multivariate battery time series. A window $\mathbf{x} \in \mathbb{R}^{W \times C}$ has $C \in \{2,3\}$ channels per dataset (V/I for the lab datasets, V/I/T for NASA RW). We treat each of the W timesteps as one token. A learnable linear projection $\mathbf{W}_{\text{in}} \in \mathbb{R}^{d_{\text{model}} \times C}$ maps each per-timestep feature vector into a higher-dimensional space, giving the encoder more capacity to represent the signal than the raw C input channels provide. The embedding width is fixed at $d_{\text{model}} = 128$ throughout, selected via hyperparameter optimization (HPO) on the pretraining task (Section 3.4.6). Sinusoidal positional encodings are then added, since self-attention is order-agnostic and the encodings supply each timestep's position within the window. A stack of $L_e = 5$ Transformer encoder blocks then processes the token sequence, each with $h = 4$ attention heads, a feed-forward width of $d_{\text{ff}} = 256$, layer normalization, and residual connections.

(2) Decoder Architecture

For pretraining, the encoder is paired with a Transformer decoder of $L_d = 6$ blocks of the same width and head count. Its depth was selected by the pretraining sweep (Section 3.4.6). A linear output projection $\mathbf{W}_{\text{out}} \in \mathbb{R}^{C \times d_{\text{model}}}$ then maps each decoded vector back to the C-channel input space, reconstructing the per-timestep feature vector at every position. Only the encoder is

retained for fine-tuning.

(3) Masking Strategy and Reconstruction Objective

A fraction $r = 0.205$ of the timesteps is masked uniformly at random. Masking acts on whole timesteps, so all C channels of a masked position are hidden together. We follow the asymmetric masked-AE design [73], in which the encoder processes only the unmasked timesteps. Dropping the masked positions keeps pretraining efficient. The masked values never reach the encoder, which must therefore infer them from the surrounding context. A shared learnable mask token, a single vector in $\mathbb{R}^{d_{\text{model}}}$, is then inserted at every masked position, and the decoder reconstructs both masked and unmasked timesteps. The reconstruction loss, computed only on the masked timesteps in the standardized input space, is

$$\mathscr{L}_{\text{rec}} = \frac{1}{C\,|\mathscr{M}|} \sum_{j \in \mathscr{M}} \left\| \hat{\mathbf{p}}_j - \mathbf{p}_j \right\|_2^2, \tag{3.2}$$

where $\mathscr{M}$ is the set of masked, non-padded timestep indices and $\mathbf{p}_j, \hat{\mathbf{p}}_j \in \mathbb{R}^C$ are the ground-truth and reconstructed per-timestep feature vectors. The normalization by $C\,|\mathscr{M}|$ makes the loss a per-element mean, so values are comparable between the $C = 2$ and $C = 3$ configurations. This loss is the training and model-selection objective. Reconstruction quality on held-out data is evaluated separately, with per-channel RMSE and mean-absolute-error (MAE) values in physical units (Section 3.5.1). The masked positions are resampled on every forward pass, so the model cannot rely on a fixed mask pattern and must learn representations that reconstruct any masked subset. Training uses the AdamW optimizer [95]. Its weight decay shrinks each parameter slightly toward zero at every update, which discourages overfitting. The learning rate and the decay strength are selected by HPO (Section 3.4.6).

3.4.3 Supervised Fine-Tuning

The fine-tuning stage attaches a cross-window attention pooler and a scalar regression head to the pretrained encoder (Fig. 3.4, fine-tuning).

(1) Cross-Window Attention Pooler

Each cycling segment carries a single scalar SoH label but is represented by a variable number of windows, so the model must aggregate evidence across all windows of the segment. Pooling proceeds in two levels. First, within each window, the encoder output is reduced to a single vector $\mathbf{e}_k \in \mathbb{R}^{d_{\text{model}}}$ by a learned attention pooling over its timesteps. Concretely, a linear layer scores each timestep, a softmax over these scores forms the pooling weights, and $\mathbf{e}_k$ is the weighted sum of the timestep outputs. Stacking the K window vectors of a segment gives $\mathbf{E} = (\mathbf{e}_1, \ldots, \mathbf{e}_K)$.

Second, the cross-window pooler aggregates $\mathbf{E}$ into one segment embedding. A small two-layer perceptron scores each window vector. Specifically, layer normalization first places all window vectors on a common scale. A Gaussian error linear unit (GELU) projection to hidden width $d_{\text{model}}/2 = 64$ and a final linear map then produce the scalar score. The GELU nonlinearity allows the score to reflect combinations of embedding features rather than a single linear direction. The halved hidden width keeps the scorer small, which guards against overfitting to the limited segment labels. A softmax over these scores across the K windows yields the window weights a_k. Supervision exists only at the segment level, whereas the SoH-relevant evidence is spread unevenly across the windows. We therefore adopt attention-based multiple-instance pooling [96]. Each window acts as an instance and the segment as the bag. The pooler then summarizes the K window vectors with three complementary statistics. These are the attention-weighted mean $\mathbf{z}_{\text{attn-mean}} = \sum_k a_k \mathbf{e}_k$, the attention-weighted standard deviation $\mathbf{z}_{\text{attn-std}}$, and the dimension-wise maximum $\mathbf{z}_{\max}$ over the K windows. The three statistics are concatenated and fused into the final

segment embedding

$$\mathbf{z} = \mathrm{Fuse}\big([\mathbf{z}_{\text{attn-mean}}; \mathbf{z}_{\text{attn-std}}; \mathbf{z}_{\max}]\big) + \mathbf{z}_{\text{attn-mean}}, \tag{3.3}$$

where Fuse : $\mathbb{R}^{3d_{\text{model}}} \to \mathbb{R}^{d_{\text{model}}}$ is a second two-layer perceptron, whose hidden width is d_{model} rather than the $d_{\text{model}}/2$ of the window scorer, with dropout inserted before its output layer at the same tuned rate as the regression head. Adding $\mathbf{z}_{\text{attn-mean}}$ outside Fuse forms a residual connection that preserves the attention-weighted mean for short segments, where the higher-order statistics are poorly conditioned. During training only, window dropout discards a random 0.240 fraction of each segment's windows. The pooler therefore sees a different subset of windows in every epoch and cannot depend on any single window.

(2) Regression Head and Objective

The regression head maps the segment embedding $\mathbf{z}$ to the scalar SoH estimate. The embedding is first layer-normalized. It then passes through a two-layer GELU multilayer perceptron (MLP) with a linear residual connection. Dropout follows each of the two GELU activations during training. A logistic (sigmoid) output unit produces the estimate,

$$\begin{aligned} \tilde{\mathbf{z}} &= \mathrm{LayerNorm}(\mathbf{z}), \\ \mathbf{h} &= \mathrm{MLP}(\tilde{\mathbf{z}}) + \mathbf{W}_{\mathrm{r}}\tilde{\mathbf{z}}, \\ \hat{y} &= \sigma\big(\mathbf{w}_o^{\top}\mathbf{h} + b_o\big), \end{aligned} \tag{3.4}$$

where the MLP hidden width is selected by HPO, $\mathbf{W}_{\mathrm{r}}$ is the linear residual projection, and $\mathbf{h}$ is the resulting hidden feature vector. The weight vector $\mathbf{w}_o$ and bias b_o form the linear output unit. The sigmoid σ constrains the estimate to $(0,1)$, which matches the SoH-fraction range and rules out physically impossible predictions. The training objective blends a Huber term and an MAE term. The Huber term is quadratic for small errors and thus follows the squared-error criterion behind the RMSE metric. The MAE term matches the MAE metric directly. The

Huber component also limits the gradient contributed by the rare, deeply degraded segments. Once the encoder is unfrozen in Stage 2 of the schedule below (Table 3.3), the objective adds an anchor-regularization term that penalizes drift of the encoder weights from their pretrained values,

$$\mathscr{L}_{\text{ft}} = \frac{1}{B}\Bigg(\sum_{i=1}^{B} \left[0.7\, H_{\delta}(\hat{y}_i - y_i) + 0.3\, |\hat{y}_i - y_i| \right] + \frac{\lambda_{\text{anc}}}{P} \left\| \theta_{\text{enc}} - \theta_{\text{enc}}^{\text{pre}} \right\|_2^2 \Bigg), \tag{3.5}$$

where $\hat{y}_i$ and y_i are the predicted and labeled SoH of the i-th segment in the batch, and H_{δ} is the Huber loss with $\delta = 0.05$, so errors beyond 5 %SoH are penalized linearly rather than quadratically. The batch size B counts segments, with $B = 2$ throughout. The anchor term measures the drift of the unfrozen encoder weights θ_{enc} from their pretrained values $\theta_{\text{enc}}^{\text{pre}}$. The average runs over the P entries of the encoder's weight matrices. Its coefficient is $\lambda_{\text{anc}} = 0.01$ in Stage 2 and $\lambda_{\text{anc}} = 0$ in Stage 1.

(3) Two-Stage Schedule

Fine-tuning proceeds in two stages (Table 3.3). Stage 1 trains only the cross-window pooler and regression head while the encoder stays frozen. Freezing shields the pretrained representations from the noisy early gradients of the two randomly initialized modules. Stage 2 then unfreezes the encoder and trains the full model, so the representations can adapt to the target dataset. The Stage 2 learning rate follows a rise-then-fall trajectory. A linear warm-up first raises the rate gradually, and a cosine decay then lowers it smoothly over the remaining epochs. This shape addresses the two risks of end-to-end fine-tuning, which sit at opposite ends of the stage. At the start, the just-unfrozen encoder receives task gradients for the first time and the rebuilt optimizer state is not yet calibrated, so the warm-up keeps these first updates small. Near the end, the risk is instead overshoot around the loss minimum, so the decay lets the weights settle rather than oscillate. On top of this trajectory, the encoder is trained at a fixed fraction of the learning rate used for the pooler and the head throughout Stage 2. The smaller

Table 3.3. Two-stage fine-tuning schedule.

	Stage 1	Stage 2
Epochs	12	60
Trainable	pooler + head	full encoder (all 5 layers) + pooler + head
LR schedule	cosine to $0.01\,lr_{\text{ft}}$	6-epoch linear warm-up to lr_{ft}, then cosine to $0.01\,lr_{\text{ft}}$, encoder at $0.226\times$ this rate
Encoder	frozen	anchor-regularization (Eq. 3.5)
EMA	over pooler + head	re-initialized at transition, over all trainable parameters
Early stopping	patience 10	patience 10, best score and counter reset, not counted during warm-up
Deployment	—	EMA weights of the best-scoring epoch

encoder steps limit drift away from the pretrained weights. Stage 1, by contrast, uses the cosine decay without warm-up, since its randomly initialized modules put no pretrained weights at risk. Both stages optimize with AdamW. Here the weight decay is applied only to the weight matrices, whose entries couple pairs of features. Bias and LayerNorm parameters are exempt. These one-dimensional parameters only shift and rescale activations, so shrinking them toward zero would distort calibration while providing little regularization.

Throughout both stages, an exponential moving average (EMA) with decay $\alpha = 0.998$ maintains a running average of every currently trainable parameter. The averaged weights are swapped in for every validation pass, since evaluating the average rather than the latest raw weights suppresses epoch-to-epoch fluctuations in model selection. The running average is re-initialized at the stage transition because the trainable-parameter set changes. Model selection then reads this smoothed evaluation. Each epoch receives a selection score of $\frac{1}{2}\big(\text{RMSE}_{\text{val}} + \text{MAE}_{\text{val}}\big)$. Early stopping monitors this score. It restarts at the stage transition, so each stage is stopped on its own progress. The deployment checkpoint therefore stores the EMA weights of the best-scoring Stage 2 epoch.

3.4.4 Supervised Baselines

Before any transfer question is asked, the pretraining claim itself is tested on the NASA RW benchmark. The masked-AE-pretrained model is compared against four super-

vised baselines. All five models use the identical stratified-by-minimum-SoH cell assignment (Section 3.4.1) and the identical outlier-segment exclusions. The most closely matched baseline is the proposed model with the pretrained weights removed. The identical encoder, pooler, and head are trained from random initialization, so this control learns from the SoH labels alone and differs only in its initial weights. It keeps the pretraining-selected encoder structure and the tuned fine-tuning hyperparameters (Section 3.4.6), so any accuracy difference is attributable to the pretrained weights themselves. The schedule is the one deliberate exception. Because a frozen random encoder is uninformative, Stage 1 is shortened to one epoch and Stage 2 is extended to 70 epochs at early-stopping patience 14, close to the total budget of Table 3.3. The Stage-2 anchor regularization consequently references the random initial weights. The same run reappears as condition (a) of the transfer matrix (Section 3.4.5).

The three external baselines are chosen to represent three major families of data-driven SoH estimation. LightGBM [97] stands for the feature-engineering family, in which each segment is reduced to hand-designed features before regression. The BiLSTM stands for the recurrent family, a widely used deep architecture in battery SoH studies. PatchTST [98] stands for the modern Transformer family for long time series.

For LightGBM, each cycling segment is reduced to distributional statistics (mean, standard deviation, range, quartiles, slope, and endpoint) of V/I/T. These are complemented by first-difference (dV, dI, dT) dynamics, instantaneous power, a voltage/current resistance proxy, and log-duration, for 46 features in total. Nine LightGBM hyperparameters are tuned by the tree-structured Parzen estimator in Optuna [99] for 60 trials. The search runs under 5-fold GroupKFold grouped by cell on the train/validation pool, with the test cells held out throughout. The same grouped cross-validation scheme is used for the fine-tuning HPO of the proposed model (Section 3.4.6).

The BiLSTM and PatchTST operate directly on the raw V/I/T windows. Both take the proposed model's window size, training stride (Section 3.4.1), and blended Huber–MAE objective. Each produces one SoH per segment by mean-pooling over the segment's windows.

The BiLSTM stacks two bidirectional layers of hidden width 128 and adds a layer-normalized GELU regression head. PatchTST keeps the published patch length 16, $d_{\text{model}} = 128$, and three encoder layers, with dropout reduced and the learning rate raised to fit the small training pool. Unlike LightGBM and the proposed model, neither baseline receives a per-model Optuna search.

3.4.5 Cross-Corpus Transfer Experimental Design

To evaluate the contribution of masked-AE pretraining and the role of the source corpus, we construct the ablation matrix of Table 3.4. It opens with a random-initialization NASA run and then varies the pretraining source per fine-tune target, closing with a seen-target control. Condition (a) is the random-initialization NASA run of the baseline comparison (Section 3.4.4). Condition (b) applies random initialization at each lab target under the default schedule of Table 3.3, and provides the reference for the transfer conditions. The shortened Stage 1 of the NASA control is specific to that benchmark and is not used here. Condition (c) pretrains on NASA RW with temperature, so it tests whether NASA-only pretraining transfers despite a channel-format mismatch. The lab inputs are zero-padded to three channels at fine-tune time. Condition (d) drops the temperature channel and pretrains on V/I-only NASA windows. This isolates the temperature contribution and gives every target a channel-matched comparison. Condition (e) pretrains on the combined four-source pool, the union of NASA and the three lab datasets, and tests whether this broader pool improves transfer. The combined pool retains the target's own cells. Any advantage of (e) can therefore reflect prior exposure to the target itself rather than the breadth of the pool, which biases this comparison toward combined-pool success. Condition (f) is the leak-free complement, referred to below as the leave-one-out lab pool. For each lab target, the encoder is pretrained on the V/I union of the other two lab datasets, so the target contributes no pretraining data at all. Condition (g) is the seen-target control. For each lab target, the encoder is pretrained on NASA plus only that target's own pack, which is condition (e) with the other two lab sources removed. Conditions (e) and (g) both retain the target's own cells, so their contrast reveals whether a combined-pool advantage requires the other lab sources.

Table 3.4. Pretraining conditions evaluated per fine-tune target. Channels counts the pretraining input channels, and format match is stated for the V/I-only lab targets.

	Pretraining source	Channels	Format match
(a)	random weight initialization, NASA RW target	—	n/a
(b)	random weight initialization, lab targets	—	n/a
(c)	NASA (V/I/T)	3	no (zero-pad)
(d)	NASA (V/I)	2	yes
(e)	NASA + all three lab datasets (combined)	2	yes
(f)	other two lab datasets (leave-one-out)	2	yes
(g)	NASA + target's own pack	2	yes

All conditions, including the NASA benchmark fine-tune, are run at three seeds ($\{123, 42, 7\}$). Seed-paired statistics are reported in Section 3.5.

3.4.6 Hyperparameter Optimization

Architecture and training hyperparameters are optimized once on the three-channel NASA RW pool and reused across all other pretraining conditions for parity. The pretraining batch size is the one exception. The combined pool and the leave-one-out lab pools hold an order of magnitude more windows than NASA RW. At the selected batch size of 16 these pools would leave the GPU underused and would run past the available compute budget, so their final pretraining used 512 or 1024 instead. The learning rate was left unchanged, since AdamW sets each parameter's step from its own gradient statistics and therefore absorbs much of the batch-size change. The NASA-only pools and the seen-target control pools used the selected value. Table 3.5 lists the search range and selected value for every tuned hyperparameter, together with the fine-tuning batch size, which was fixed rather than searched. The pretraining sweep is scored by masked-reconstruction validation loss. The fine-tuning sweep uses the selection score of Section 3.4.3, computed without the EMA swap. The moving average is a final-training mechanism only, so a trial is scored on the weights as trained. The score is averaged over a 5-fold cross-validation grouped by cell on the NASA RW train/validation cells. The validation cells rotate across the folds, and the test cells are held out throughout. The fine-tuning sweep runs 30 trials. Each trial trains all five folds, 48 epochs per fold, of which the first 12 keep the encoder

Table 3.5. Hyperparameter search space and selected values.

Hyperparameter	Search range	Selected
Pretraining		
Model width d_{model}	$\{64, 128\}$	128
Attention heads h	$\{2, 4, 8\}$	4
Encoder layers L_e	4–8	5
Decoder layers L_d	3–6	6
Feed-forward width d_{ff}	256	256
Mask ratio r	0.15–0.25	0.205
Learning rate	2×10^{-4}–1×10^{-3}	2.6×10^{-4}
Weight decay	10^{-6}–10^{-4}	1.1×10^{-5}
Dropout	0–0.05	7.6×10^{-5}
Batch size	$\{16, 32, 64, 128\}$	16
Fine-tuning		
Learning rate lr_{ft}	6×10^{-4}–1.1×10^{-3}	6.5×10^{-4}
Weight decay	3×10^{-6}–1.5×10^{-5}	4.1×10^{-6}
Head/pooler dropout	0.15–0.25	0.229
Head width	$\{128, 192\}$	192
Unfrozen layers (Stage 2)	3–5	5
Stage-2 learning-rate scale	0.10–0.25	0.226
Window-dropout rate	0.15–0.30	0.240
Batch size B	fixed	2

frozen and the remaining 36 train it unfrozen. Both sweeps use Optuna with a tree-structured Parzen estimator and median pruning.

The pretraining sweep runs in two stages, so that the search budget is spent where it discriminates. A trial is one candidate configuration, trained from scratch for a fixed number of epochs and scored by its best validation loss. The first stage runs 60 trials of 10 epochs each, which ranks configurations coarsely at low cost. The second stage runs 40 trials of 20 epochs each. Its first eight trials re-run the eight best configurations of the first stage at this longer budget, which separates candidates that 10 epochs leaves tied. Those eight results then inform the sampler, which proposes the remaining 32 configurations. The winning configuration is finally trained for up to 150 epochs, stopping early once the validation loss has not improved for 20 consecutive epochs.

With the architecture and schedule held constant, performance differences between conditions trace to the pretraining corpus rather than to per-condition tuning. Separate per-

condition searches would blur this attribution, and they would repeat the expensive two-stage sweep for every pool. Such reuse is possible across channel counts because no tuned setting in Table 3.5 directly references the channel count C, so the shared configuration remains applicable to two-channel pools. Whether it remains effective after the channel change is verified empirically. The V/I-only NASA pretraining run reaches a masked-reconstruction mean squared error of the same order as the three-channel run (0.0019 versus 0.0014, Section 3.5.1). Reusing one configuration comes at the cost of possibly non-optimal per-target configurations. In particular, pools that differ most from NASA RW in size and composition may be fit less well.

The proposed model and its training loops are implemented in PyTorch, as are the BiLSTM and PatchTST baselines. LightGBM uses its own library. Optuna performs the hyperparameter searches, and scikit-learn supplies their grouped cross-validation splitter. Every deep-learning run used a single GPU. All pretraining runs, and all fine-tuning runs of the proposed model and of the randomly initialized control, used an NVIDIA RTX A5000 or an A100. The BiLSTM and PatchTST baselines used an NVIDIA GeForce RTX 3060 Ti. LightGBM was fitted on CPU.

3.4.7 Evaluation Metrics

Per-segment predictions are evaluated by

$$\mathrm{RMSE} = \sqrt{\frac{1}{N_s}\sum_{i=1}^{N_s}(\hat{y}_i - y_i)^2}, \tag{3.6}$$

$$\mathrm{MAE} = \frac{1}{N_s}\sum_{i=1}^{N_s}|\hat{y}_i - y_i|, \tag{3.7}$$

where N_s is the number of held-out cycling segments. RMSE is the headline metric for cross-condition comparison. Its quadratic penalty amplifies large errors, and these are the costly errors in screening because a single large SoH error can misroute a pack. RMSE is also the metric most widely reported in the battery SoH-estimation literature [65, 78, 90], so the results remain comparable with prior work. MAE complements it as a linear average that is less dominated by

rare large errors. For the lab transfer contrasts we additionally report 95 % confidence intervals (CIs). Each lab target has only two held-out pouches or bricks, so every contrast is estimated from the few segments they provide. A different set of held-out segments would give a different value, and the CI expresses the range of this sampling uncertainty. Concretely, each held-out segment contributes one error per seed run, and the errors of all three runs are concatenated into one pool per condition. The two pools are then resampled with replacement 10000 times, each independently of the other, the contrast is recomputed on every pair of resamples, and the central 95 % of the recomputed values forms the interval. Across-seed variation is tracked separately by a per-seed win count. The two conditions in a contrast are trained at the same three seeds, and we report at how many of the three seeds one condition beats the other, for example 3/3. The CIs should be read together with the win counts, because the intervals are descriptive rather than confirmatory.

The NASA RW benchmark is additionally evaluated under a leave-one-cell-out protocol, which measures how much the error depends on which cell is held out. Each of the 14 cells is held out once as the sole test cell, and the remaining 13 are split into 10 training and three validation cells, giving 14 folds. The validation cells are chosen to span the cycling regimes of the corpus. Hierarchical clustering on the cells' cycling statistics partitions the 14 cells into four protocol clusters, which are internal to this analysis and are not the dataset's release groups. One validation cell is then drawn from each cluster other than the test cell's, so the test cell's own cluster stays maximally represented in training. Results are reported per fold in Section 3.5.2.

3.5 Results

3.5.1 Pretraining: Masked Reconstruction

The pretrained model is first evaluated on its masked-reconstruction quality. On held-out NASA cycling segments the decoder recovers the randomly masked 20.5 % of timesteps in V/I/T, following the signal variation across many charge/discharge cycles and the full degradation

spectrum (Fig. 3.5). Table 3.6 quantifies this recovery. For each signal it reports the RMSE and MAE between the reconstructed and true masked values in the signal's own unit. The held-out windows are split into early-, mid-, and late-life thirds by their window-mean SoH, the average over the window of the per-sample SoH of Section 3.3.3. The voltage and current errors stay small at every stage relative to each signal's operating range. For both signals the MAE stays well below the RMSE, so most masked timesteps are recovered almost exactly and a few large misses dominate the RMSE. The large misses originate in aged cells. Higher internal resistance and reduced capacity drive the voltage to the 3.2–4.2 V RW switching limits almost instantly under the randomized high-current loads. The result is brief, large charge/discharge transients that grow markedly more frequent with degradation (Fig. 3.5(c)). Because these events are momentary, they are the hardest content to reconstruct, and they explain the mild end-of-life increase.

The temperature signal behaves differently. It varies slowly as a mild cycling-induced ripple on a baseline set by the ambient conditions. Neighboring timesteps therefore carry nearly the same value, so the unmasked context around each masked position almost determines it. The decoder consequently tracks the signal closely at every stage. The temperature error in Table 3.6 even decreases toward end of life. The fall traces to which cells populate each third, not to aging itself. Only five of the 14 cells reach the late-life third, and 91.3 % of its windows come from the three cells cycled continuously, without rest steps. These cells show the smallest within-window thermal swings in the dataset, so their temperature is the easiest to reconstruct at every stage. Restricted to these three cells, the late-life temperature error is in fact slightly higher than the early-life one (0.033 versus 0.027 °C). Because the ambient baseline also differs between the three stages of Fig. 3.5, each temperature panel is autoscaled. The panels therefore show reconstruction quality, and absolute temperatures are not comparable across stages. This fidelity establishes that the encoder has learned the raw-window V/I/T manifold that the fine-tuning stage subsequently exploits. Reconstruction fidelity alone, however, does not determine transfer. As the transfer experiments of Section 3.5.3 show, transfer is governed by the match between the

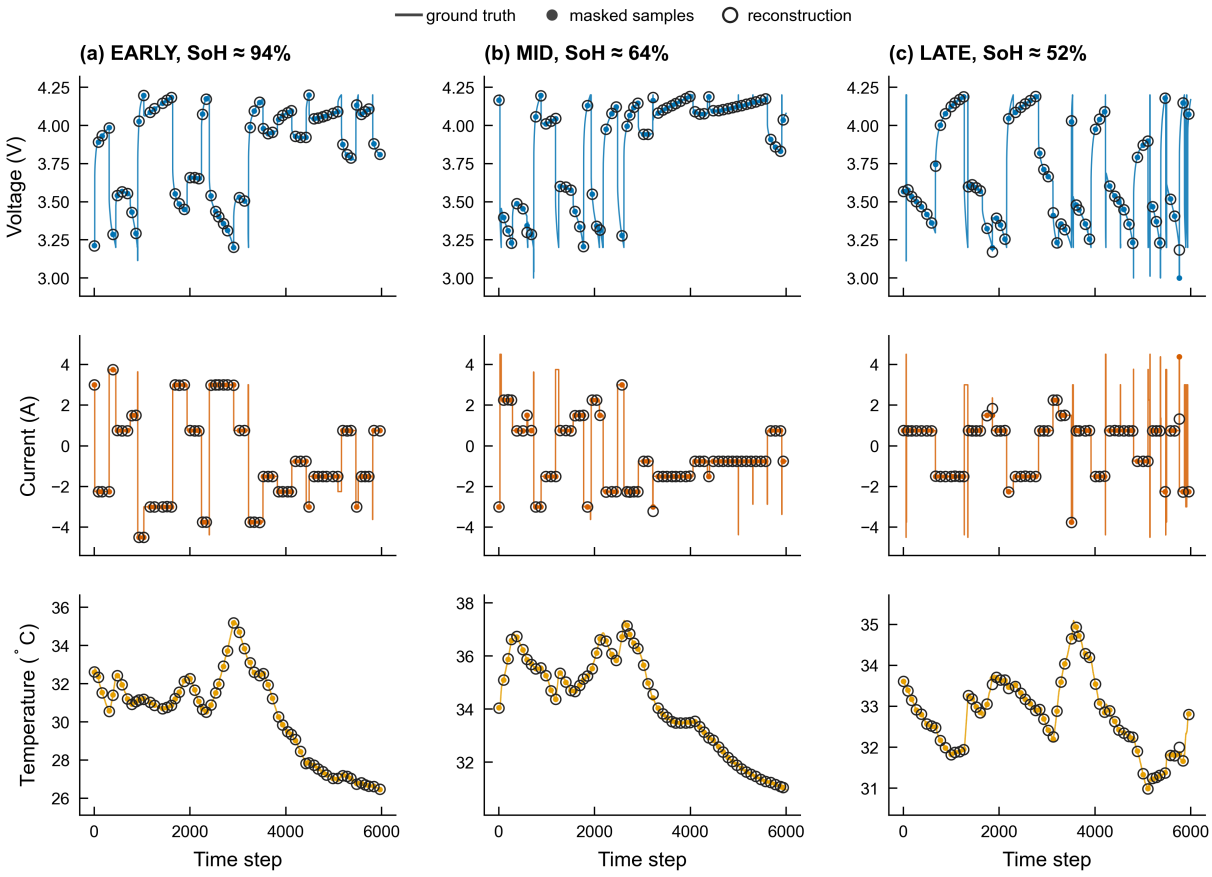


Figure 3.5. Masked reconstruction of the proposed model, following one cell's held-out segments across three degradation levels (V/I/T pretrain, seed 123; SoH $\approx$ 94/64/52 %). Each panel spans $\sim$6000 samples. The line is the ground truth; dots mark the masked ground truth and rings the reconstruction, downsampled to one masked timestep per $\sim$100 samples for legibility. A ring on its dot is a perfect recovery. Voltage and current share a y-axis across panels; each temperature panel is autoscaled.

pretraining corpus and the target.

3.5.2 Public-Benchmark Performance

Table 3.7 compares the SoH prediction performance of the masked-AE-pretrained SoHEncoder and four supervised baselines at three seeds. The masked-AE-pretrained model attains a test RMSE of 2.82 ± 0.28 %SoH, written here and throughout as the mean and standard deviation over the three seeds. The strongest baseline is the identical SoHEncoder architecture trained from random initialization, which reaches 2.72 ± 0.12 %SoH. Its advantage over the

Table 3.6. Masked-reconstruction error of the seed-123 V/I/T pretrain on held-out NASA RW segments, by channel and degradation stage. Each entry is RMSE / MAE over the masked timesteps in that stage.

Stage	Voltage (V)	Current (A)	Temperature (°C)
Early	0.008 / 0.001	0.080 / 0.011	0.068 / 0.025
Mid	0.008 / 0.001	0.082 / 0.008	0.047 / 0.018
Late	0.013 / 0.002	0.116 / 0.011	0.036 / 0.015
All	0.010 / 0.001	0.094 / 0.010	0.052 / 0.019

Table 3.7. SoH estimation on the NASA RW dataset. Test RMSE and MAE in %SoH, mean $\pm$ std over seeds $\{123, 42, 7\}$. "vs. proposed" is each method's RMSE divided by the proposed model's. Bold marks the results of the proposed model.

Method	Test RMSE	Test MAE	vs. proposed
Masked-AE-pretrained SoHEncoder	$\mathbf{2.82 \pm 0.28}$	$\mathbf{2.28 \pm 0.22}$	—
Randomly initialized SoHEncoder	2.72 ± 0.12	2.18 ± 0.06	1.0×
LightGBM	4.26 ± 0.69	3.41 ± 0.49	1.5×
BiLSTM	7.10 ± 2.31	5.48 ± 1.54	2.5×
PatchTST	9.05 ± 0.85	7.07 ± 0.90	3.2×

pretrained model is within the seed-to-seed variation of either model. With this benchmark's comparatively plentiful labels, the pretrained weights therefore bring no measurable in-domain gain. The control is nevertheless a product of the pretraining pipeline, because it inherits that pipeline's encoder structure and tuned training configuration, as specified in Section 3.4.4. The value of the pretrained weights appears instead on the label-scarce second-life targets, where random initialization loses to chemistry-matched pretraining at every seed on both LMO packs (Section 3.5.3). Parity between the two SoHEncoder variants does not mean the benchmark is easy. Tuned LightGBM [97] trails the proposed model at 1.5× its RMSE, the BiLSTM at 2.5×, and PatchTST [98] at 3.2×. The BiLSTM also shows the largest seed sensitivity of any method, a standard deviation of 2.31 %SoH. The advantage over these established methods therefore comes from the SoHEncoder architecture, which both variants share.

The across-seed spreads above hold the cell split fixed. Error variation across held-out cells is quantified by a leave-one-cell-out evaluation over 14 folds (Section 3.4.7, Table 3.8).

Table 3.8. Leave-one-cell-out generalization on NASA RW (14 folds, cluster-aware validation, seed 123). Per-cell test RMSE and MAE in %SoH with bootstrap 95% CIs (10,000 resamples).

Cell	Cluster	RMSE	RMSE 95% CI	MAE	MAE 95% CI
cell_1	A	4.58	[2.66, 6.18]	3.14	[1.88, 4.58]
cell_2	A	2.56	[1.84, 3.22]	2.07	[1.51, 2.68]
cell_3	A	2.50	[1.83, 3.08]	1.98	[1.38, 2.61]
cell_4	B	2.49	[1.95, 3.00]	1.97	[1.49, 2.49]
cell_5	B	2.12	[1.76, 2.47]	1.87	[1.56, 2.21]
cell_6	B	4.42	[3.72, 5.07]	3.88	[3.18, 4.59]
cell_7	C	1.72	[1.16, 2.16]	1.44	[0.84, 2.01]
cell_8	C	6.46	[4.65, 7.73]	5.95	[3.95, 7.68]
cell_9	C	4.03	[2.84, 5.01]	3.66	[2.57, 4.78]
cell_10	C	7.01	[5.73, 8.32]	6.74	[5.64, 8.02]
cell_11	D	6.63	[5.20, 7.86]	6.00	[4.35, 7.56]
cell_12	D	9.83	[8.60, 10.92]	9.55	[8.20, 10.77]
cell_13	D	3.79	[2.99, 4.46]	3.46	[2.49, 4.32]
cell_14	D	2.70	[1.70, 3.54]	2.24	[1.39, 3.16]
Median (14 folds)		3.91	[3.13, 4.30]	3.30	[2.57, 3.77]

The four protocol clusters that structure the folds are labeled A–D in Table 3.8. The median per-cell RMSE is 3.91 %SoH (bootstrap 95 % CI [3.13, 4.30]) and the median per-cell MAE is 3.30 %SoH (CI [2.57, 3.77]). The held-out cell's cluster sets most of the fold difficulty. Per-cluster RMSE medians range from 2.49 %SoH (cluster B) to 5.25 %SoH (cluster C). Cells cycled under low-current randomized profiles (clusters A and B) are estimated roughly twice as accurately as cells under high-current or skewed-random profiles (clusters C and D). This pattern is consistent with a coverage effect. The harsher C and D trajectories differ more from the rest of the 14-cell corpus. When one of these cells is held out, the remaining pool therefore offers fewer similar degradation trajectories to learn from.

The three-seed mean (2.82 ± 0.28 %SoH, canonical seed-123 2.49 %SoH) is in line with the strongest published cell-level result on this dataset, $\approx$2.45 %SoH averaged over all 256 possible 12/4 cell splits [78]. The present accuracy is reached with a raw-window input pipeline and no hand-engineered features. The two use different SoH conventions, cell counts, and split protocols (a single stratified 10/2/2 split versus a full-enumeration 12/4 mean), so the comparison is indicative rather than head-to-head. To our knowledge no prior masked-AE-pretrained method

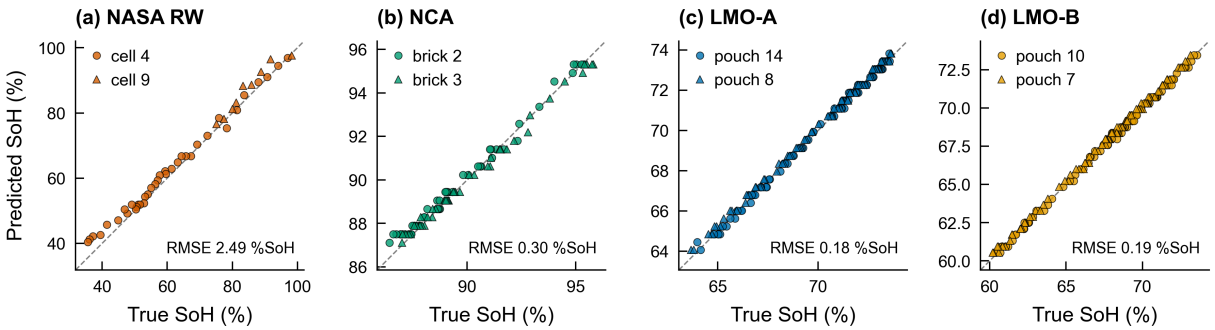


Figure 3.6. Predicted versus true SoH on held-out test cells, pouches, and bricks at seed 123. Panel (a) shows NASA RW with the proposed masked-AE-pretrained model. Panels (b)–(d) show each second-life target under its best-performing pretraining source, NASA V/I/T for the NCA pack and the combined pool for LMO-A/LMO-B. Marker shape distinguishes the two held-out cell-group units per panel, and the dashed line is identity. Axis ranges differ per panel because each target occupies a different SoH band.

has been benchmarked on NASA RW. None of the recent battery foundation models include it in their corpora [67–69].

The predicted-versus-true scatter on the held-out cells tracks the identity line across their full SoH span (Fig. 3.6(a)). That span runs from the corpus floor of 35.6 % up to near beginning of life, so the fit is exercised over essentially the whole degradation range the benchmark offers. The mild over-prediction in the scatter has two identifiable sources. The held-out pool contains the deepest-degraded cell (cell 4), whose lowest segments (35.6–39.4 %SoH) lie below the training-label floor of 40 %SoH. Their predictions saturate just above that floor. For the same cell's segments inside the training range, the average over-prediction is only $\sim$1 %SoH. This is a label-support boundary effect rather than a miscalibration of the learned dynamics-to-SoH mapping. The remaining over-prediction is specific to cell 9 and peaks in its early-to-mid life ($+2$ to $+5$ %SoH). Cell 9 was cycled under the same RW protocol group as training cells 7, 8, and 10. At matched cycle count its measured capacity runs up to 2.5 %SoH below those of group-mates 7 and 10 over exactly those cycles. It is a slightly weak cell of its protocol group, which the model maps toward the group-typical capacity. In deployment, a single initial reference capacity test on the incoming pack exposes and removes such an offset.

Table 3.9. Transfer-learning ablation on the three second-life battery systems. Each cell is test RMSE $\pm$ std / MAE $\pm$ std in %SoH (means over seeds $\{123, 42, 7\}$). Best RMSE per target in bold. The sixth source, the NASA+own-pack seen-target control, is reported in Table 3.10.

Pretraining source	LMO-A	LMO-B	NCA
NASA + all three lab datasets (combined)	$\mathbf{0.19 \pm 0.011}$ / 0.16 ± 0.009	$\mathbf{0.20 \pm 0.013}$ / 0.16 ± 0.010	0.50 ± 0.053 / 0.40 ± 0.043
Other two lab datasets (leave-one-out)	0.21 ± 0.009 / 0.17 ± 0.007	0.26 ± 0.014 / 0.22 ± 0.010	0.36 ± 0.189 / 0.29 ± 0.157
NASA (V/I/T)	0.21 ± 0.004 / 0.16 ± 0.003	0.27 ± 0.024 / 0.22 ± 0.017	$\mathbf{0.30 \pm 0.006}$ / 0.23 ± 0.003
NASA (V/I)	0.26 ± 0.026 / 0.21 ± 0.020	0.28 ± 0.029 / 0.23 ± 0.020	0.33 ± 0.017 / 0.26 ± 0.018
Random weight initialization	0.30 ± 0.029 / 0.25 ± 0.028	0.28 ± 0.024 / 0.23 ± 0.019	0.35 ± 0.061 / 0.28 ± 0.048

3.5.3 Transfer to Second-Life Systems

Table 3.9 presents the transfer ablation of five pretraining sources at three second-life targets, with every combination run at three seeds. Fig. 3.7 visualizes the mean and per-seed test RMSE of the same matrix, and Fig. 3.6(b)–(d) shows the predicted-versus-true scatter for each target under its best-performing pretraining source. Six findings emerge, which are discussed in the following sections. Together they substantiate the second-life transfer contributions of Section 3.1, namely composition, stability, and deployment-realistic transfer. Win counts written $x/3$ are the per-seed win counts of Section 3.4.7, computed on a cell-level split that is identical in every run. Two transfer regimes are distinguished, following the standard transfer-learning taxonomy [77]. In unseen-target (inductive) transfer the target system is absent from the pretraining corpus, the realistic case for a newly retired pack. In seen-target (transductive) transfer the target's own unlabeled data is already in the corpus, and only its held-out pouches or bricks are estimated. Of the five pretraining sources, only the combined four-source pool is seen-target. The two NASA sources and the leave-one-out lab pool are unseen-target, since each excludes the target by construction. Random initialization falls outside the taxonomy altogether, because it uses no pretraining corpus, and it therefore serves as the no-transfer reference for both regimes. A sixth pretraining source, NASA plus the target's own pack, is seen-target by construction and serves as a control in the second finding.

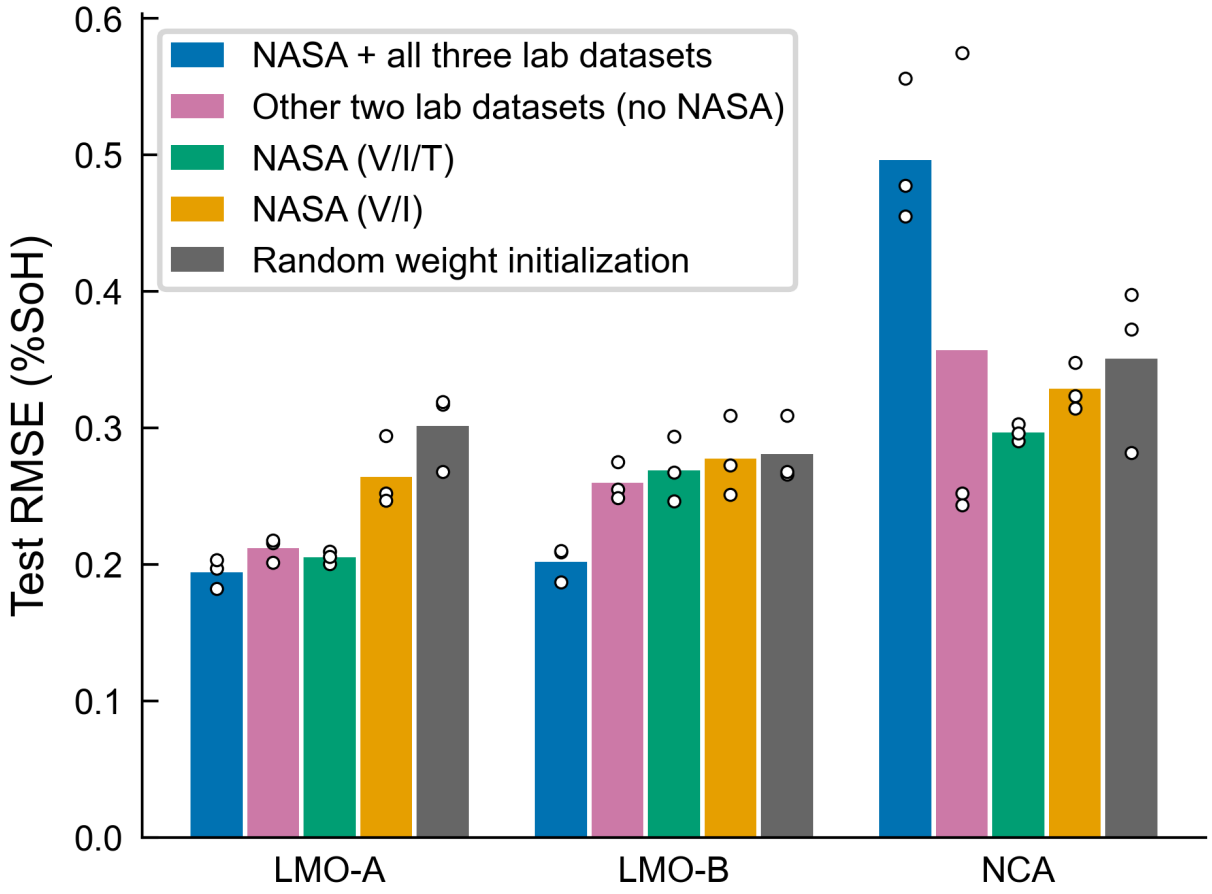


Figure 3.7. Transfer-learning ablation on the three second-life systems, five pretraining sources per target. Bars show mean test RMSE (%SoH) over seeds $\{123, 42, 7\}$ and circles show the individual seeds. The combined pool is NASA together with all three lab datasets, and the leave-one-out pool is the other two lab datasets without NASA. Source names match the row labels of Table 3.9.

(1) Pool Composition, Not Pool Size, Governs the Sign of Transfer

The combined four-source pool yields the best accuracy observed on both LMO packs, reaching 0.19 %SoH on LMO-A (-35.5 % versus random initialization, 3/3 seeds) and 0.20 %SoH on LMO-B (-28.1 %, 3/3). Yet the same checkpoint is the worst configuration on the NCA pack. There it reaches 0.50 %SoH, $+41.6$ % above random initialization with 0/3 wins, an increase robust to segment resampling ($+0.14$ %SoH [$+0.10$, $+0.19$]). This reversal follows from the composition of the pool. The two LMO packs numerically dominate it, supplying $\sim$72.9 % of its windows, and as a result $\sim$75.2 % of its SoH labels lie in the LMO

Table 3.10. Seen-target control (Section 3.5.3): pretraining on NASA plus the target's own unlabeled cycling data, without the other two lab sources. The first row gives the control's mean $\pm$ std test RMSE in %SoH over seeds $\{123, 42, 7\}$. The remaining rows give its RMSE gap to each comparator source of Table 3.9, with its 95% segment-level bootstrap CI.

	LMO-A	LMO-B	NCA
NASA+own-pack	0.22 ± 0.026	0.26 ± 0.021	0.40 ± 0.029
Control−Combined	$+0.03$ [+0.01, +0.04]	$+0.06$ [+0.04, +0.08]	-0.10 [-0.15, -0.05]
Control−NASA (V/I)	-0.04 [-0.06, -0.02]	-0.02 [-0.04, +0.00]	$+0.07$ [+0.03, +0.11]
Control−Random initialization	-0.08 [-0.10, -0.06]	-0.02 [-0.04, -0.00]	$+0.05$ [+0.00, +0.09]

band (Appendix A.1). The encoder is therefore tuned to LMO-like dynamics, while the NCA pack's own data, although included in the pool, is a swamped minority. Consequently, the same tuning that helps the LMO targets gives the NCA pack, whose SoH band the pool barely covers, a worse starting point than random initialization. Pool size still matters, but it is subordinate to composition. When the composition matches the target, scale helps. On both LMO packs the four-source pool accordingly leads all three smaller sources in mean error (Table 3.9). Scale thus amplifies a correctly signed transfer, but it cannot fix a mis-signed one.

(2) A Seen-Target Control: Target Presence Is Not the Active Ingredient

Condition (g) of Section 3.4.5 tests whether the combined pool's behavior is driven by the presence of the target's own data rather than by the rest of the corpus. For each lab target the encoder is re-pretrained on NASA plus only that target's own pack, which is the combined pool with the other two lab sources removed. The temperature-free setting, fine-tune cells, hyperparameters, and protocol are identical, so the contrast is controlled. Table 3.10 reports the control and its gaps to the comparator sources, whose accuracies appear in Table 3.9. Because removing the other sources also shrinks the pool (Table A.1), the control isolates the target's mere presence but leaves corpus diversity and corpus scale entangled.

The results reject target presence as the active ingredient on the failing and the succeeding targets alike. On the NCA pack the target's own data supplies $\sim$65 % of the NASA+NCA control corpus. Even so, the control sits $+14.1$ % above random initialization at all three seeds, so a

corpus can consist mostly of target data and still transfer negatively. The control further shows where the combined pool's larger failure on this target comes from. The two corpora differ only by the two LMO packs, and dropping them improves NCA accuracy by 19.4 %. The combined pool's harm therefore comes mostly from the two LMO packs that dominate it, not from the target's presence in the corpus. On the two LMO packs the comparison runs the other way, since there the combined pool helps. The control retains the target's own data yet trails the combined pool at all three seeds, by +14.1 % on LMO-A and +29.3 % on LMO-B. Most of the combined pool's gain therefore comes from the sources beyond the target. All three Control−Combined gaps are robust to segment resampling, since each of their 95 % CIs in Table 3.10 excludes zero. With three seeds and two held-out pouches or bricks per target, however, these comparisons indicate consistent directions rather than statistically conclusive differences.

A final comparison isolates what the target's own pack adds to NASA (V/I) pretraining. The control differs from NASA (V/I) only by the target's own pack, so the Control−NASA (V/I) row of Table 3.10 measures the pack's contribution directly. The addition helps modestly on LMO-A (−16.1 %), is within bootstrap noise on LMO-B (−5.9 %), and hurts the NCA pack (+21.8 %). The target's own data is thus neither sufficient for the combined pool's gains nor consistently helpful when added to NASA (V/I). Whether it is necessary at all is settled by the next finding, where target-free sources transfer positively.

(3) Cross-System Transfer Is Real but Chemistry-Bounded

Transfer holds along two independent lines, neither of which requires the target system in the pretraining corpus. The first line is transfer from NASA (V/I/T), which uses a different cathode chemistry (LCO), topology (individual 18650 cells), and cycle regime (the RW protocol) from any lab pack. It nonetheless lowers LMO-A error to 0.21 %SoH (−32.0 % versus random initialization, 3/3 seeds). Among all sources it also has the lowest seed-to-seed standard deviation (0.004 %SoH). By mean error, the best source on LMO-A is the combined pool (Table 3.9). Its edge over NASA (V/I/T), however, is within bootstrap noise (−0.01 %SoH [−0.03, 0.00]), so

NASA (V/I/T) matches the best source on LMO-A. The gain thus arises despite fully mismatched battery properties, so it cannot be an artifact of pretrain–target similarity.

The second line is transfer from the leave-one-out lab pools, which exclude the target but share its V/I-only connected-pack sensing and its dispatch cycling protocol. They beat random initialization at all three seeds on both LMO packs ($-29.8\,\%$ on LMO-A, $-7.6\,\%$ on LMO-B). Each leave-one-out pool still contains the target's sister pack, since LMO-A and LMO-B are two packs of one model. The gain could therefore reflect copying from a near-identical pack rather than transfer. Comparing the leave-one-out pool with the combined pool cannot settle this question. The sister pack belongs to both pools, so the contrast is blind to what the sister contributes. The two pools also differ by the NASA corpus, not only by the target's own data. The objection's premise, however, does not survive inspection. The two packs are not near-copies. They entered testing at different states of degradation and aged at different rates under the same protocol. They also differ in SoH band and segment count (Section 3.3.2). Moreover, they respond differently to identical pretraining. Concretely, relative to random initialization, NASA (V/I/T) lowers LMO-A error by $32.0\,\%$ yet barely moves LMO-B ($-4.2\,\%$). The two packs share cathode chemistry and pack model yet diverge in their response, so the leave-one-out gain is genuine pack-to-pack transfer rather than accuracy inherited from a near-copy.

(4) Estimates on the NCA Pack Are Seed-Sensitive, and NASA Pretraining Stabilizes Them

On the NCA pack, accuracy varies markedly with the random seed unless the encoder is pretrained on NASA data alone. The variation is quantified by the across-seed coefficient of variation, the standard deviation of test RMSE over the three seeds divided by their mean. On the NCA pack the coefficient is $10.7\,\%$ for the combined pool, $52.9\,\%$ for the leave-one-out pool, and $17.4\,\%$ for random initialization, which involves no pretraining corpus at all (Table A.2). The two NASA sources cut it to $2.1\,\%$ for V/I/T and $5.3\,\%$ for V/I. All three non-NASA configurations are also more seed-sensitive on the NCA pack than on either LMO pack, so the susceptibility

is specific to this target. Three properties of the pack plausibly underlie it. It has the narrowest operating SoH band and the fewest labeled cycling segments of the three packs. It is also the only target whose voltage channels are read at the brick level, where 64 parallel-connected cells share one terminal voltage (Table 3.1). This parallel aggregation smooths the very V/I features the estimator relies on. Band width and parallel-brick aggregation are thus confounded as causes of the pack's hardness.

What sets the NASA data apart is the property the NCA pack lacks, wide SoH coverage. Its records span 35.6–100.0 %SoH (Table 3.1), while the lab pools are saturated by the narrow LMO band. NASA (V/I/T) is accordingly at once the most accurate and the most stable configuration on the NCA pack. It reaches the lowest mean error of any source (0.30 %SoH, -15.4 % versus random initialization, Table 3.9), an accuracy gain robust to segment resampling (-0.06 %SoH [-0.10, -0.02]). It also holds the smallest coefficient of variation of the five sources. Against random initialization the per-seed win count is 2/3. At the remaining seed, random initialization lands on its own best result, and even there NASA (V/I/T) stays within random initialization's seed spread (Table A.2). This seed stability is not specific to the NCA pack. NASA (V/I/T) is also the lowest-variance source on LMO-A (coefficient of variation 2.2 %, Table A.2). A wide-SoH corpus thus appears to regularize fine-tuning across targets.

(5) A Single Seed Can Misrank Pretraining Sources

The seed sensitivity of the NCA pack carries a practical lesson for evaluating pretraining sources. The leave-one-out pool delivers the lowest test error of any source at 2/3 seeds, yet at the remaining seed it delivers the highest, worse than random initialization (Table A.2). A single-seed comparison at a lucky seed would therefore crown the least reliable corpus as the best source on this pack. Pretraining corpora must instead be judged across seeds.

(6) Temperature-Informed Features Transfer to Temperature-Blind Targets

No lab target senses temperature during fine-tuning, yet removing the temperature channel from NASA pretraining still degrades transfer (Table 3.9). NASA (V/I) trails NASA (V/I/T) in mean error by $+28.9\,\%$ on LMO-A and $+10.8\,\%$ on the NCA pack (3/3 seeds each). On LMO-B the mean gap is only $+3.2\,\%$ (2/3 seeds), smaller than either source's seed-to-seed spread. The temperature gaps on LMO-A and the NCA pack arise without any temperature input at fine-tune time, so the advantage must be carried by the pretrained representations. The encoder therefore appears to use temperature during pretraining to learn better-calibrated V/I representations. This interpretation predicts a temperature benefit only where NASA pretraining itself transfers well (error reduction versus random initialization), and the results follow this pattern. The reduction from NASA (V/I/T) is largest on LMO-A ($-32.0\,\%$) and the NCA pack ($-15.4\,\%$), the two targets with the widest temperature gaps. On LMO-B the reduction is near zero ($-4.2\,\%$), and the temperature gap vanishes with it.

3.5.4 Summary of Evidence

On the label-rich public benchmark, the pretrained and randomly initialized variants of the proposed architecture differ within seed-to-seed variation (Section 3.5.2). Both outperform the established supervised baselines by a wide margin, so the benchmark advantage comes from the shared SoHEncoder architecture rather than from the pretrained weights. The case for the pretrained weights rests on the three second-life packs, where labels are scarce.

On these packs the evidence divides into consistent gains and diagnosable failures. The combined and leave-one-out pools, which contain LMO data, improve on random initialization on both LMO packs (3/3 seeds each). On LMO-A every pretraining source is at least as accurate as random initialization (3/3 seeds, Table A.2). The gains take two forms. On the LMO packs the combined pool lowers mean error the most. On the narrow-band NCA pack the wide-SoH NASA corpus delivers a smaller mean gain but collapses the seed-to-seed variability. No other

source is both accurate and stable there. Neither the NASA sources nor the leave-one-out pools contain any target data, so their gains are unseen-target transfer, the setting a newly retired pack presents. The failures are equally systematic. Every configuration that fails on the NCA pack draws on a corpus saturated by one narrow SoH band. The LMO-dominated combined pool and the NCA-majority seen-target control stay less accurate than random initialization (3/3 seeds), and the pure-LMO leave-one-out pool swings on that pack from the best source to the worst with the seed. Failure is therefore diagnosable in advance from corpus composition.

3.6 Discussion

The second-life campaign reveals that self-supervised transfer in battery SoH estimation depends on the composition of the pretraining pool relative to the target, not on the volume of pretraining data. Three results follow. The first two establish the claim directly, one for the sign of transfer and one for its stability. The third establishes its relevance to deployment. First, the combined four-source pool is the largest corpus in the study. On both LMO packs it yields the lowest mean error of any source and beats random initialization (3/3 seeds). On the NCA pack, however, the same checkpoint has the highest mean error of any configuration (Section 3.5.3). It trails random initialization (3/3 seeds). The reversal follows from the pool's composition. The two LMO packs numerically dominate the pool. Concretely, $\sim$75.2 % of the pool's SoH labels lie in the LMO band against only $\sim$19.4 % in the NCA band. The NCA pack's own data is included but swamped. The pool that helps most where the target is well-represented harms most where it is not. This qualifies the prevailing scale-first narrative of battery foundation models [67–69, 83, 84]. Indiscriminately enlarging a pretraining corpus can actively damage downstream accuracy on a mismatched target. Moreover, the damage is predictable from corpus composition alone.

The finding is observable only because the lab packs serve double duty in the experimental design. Each pack is both a fine-tune target and a member of several pretraining pools. Corpus

composition could therefore be varied while the target was held fixed. The same design rules out two competing readings, one for the gains on the two LMO packs and one for the harm on the NCA pack. The combined pool might win on the two LMO packs simply because it contains their own cycling data. That reading predicts that the seen-target control, NASA plus the target's own pack, matches the combined pool. Instead the control recovers only part of the gain. It trails the combined pool by 14.1–29.3 % (Table 3.10). Even positive transfer therefore depends on the rest of the corpus, not the target's mere presence in the pool. The mirror-image reading attributes the harm on the NCA pack to the small share of its own data in the pool. That reading predicts that the pack's seen-target control, in which its own data supplies the majority of the windows ($\sim$65 %), at least matches random initialization. Instead this control also trails random initialization (3/3 seeds). NASA (V/I/T) contains no NCA data at all and spans a wide SoH range. On the NCA pack it nonetheless beats random initialization on the three-seed mean (-15.4 %). Every configuration that fails on this pack therefore draws on a corpus saturated by one narrow SoH band. In the combined pool that band is the LMO band. In the NCA pack's seen-target control it is the pack's own high-SoH band. Saturation by itself does not predict harm, since the same LMO-saturated combined pool is the best source on both LMO packs. Harm follows when the saturating band also sits far from the target's own operating band.

Second, composition also decides the stability of transfer, a dimension that single-seed evaluation cannot see. The clearest case is the NCA pack. With the combined pool, the leave-one-out pool, or random initialization, the across-seed coefficient of variation on this pack runs from 10.7 % to 52.9 %. The two NASA sources, however, hold it at 2.1 % for V/I/T and 5.3 % for V/I (Table A.2). NASA (V/I/T) is at once the most accurate and the most stable configuration there. NASA (V/I) is likewise stable and likewise more accurate than random initialization, so temperature sharpens an advantage that the wide SoH range already provides. What separates the stable sources from the unstable ones is the SoH coverage of the pretraining corpus, not its distance from the target. Concretely, the two NASA sources are far from the target in chemistry and format but cover a wide SoH range. The combined and leave-one-out pools are closer in

pack-level dynamics but concentrate in the narrow LMO band (Section 3.5.3). Corpus volume does not separate them either. The combined pool contains the entire NASA (V/I) corpus plus all three lab packs. It nonetheless sits in the unstable group, and NASA (V/I) sits in the stable one. This advantage of coverage over closeness and volume runs against the transfer literature. The usual findings are that a closer in-domain corpus transfers better [100], that fine-tuning robustness is driven by pretraining-data quantity rather than diversity (the variety of the pretraining data) [101], and that diversity helps only when the source and target distributions are close [102]. The tension, however, is specific to this setting. The NCA target occupies so narrow an operating band that closeness and wide coverage are mutually exclusive. A corpus close to this target is necessarily narrow, and a wide corpus is necessarily out of distribution. The stability comparison above settles the trade-off in favor of coverage. For deployment on hard, narrow-band targets, a wide-SoH public corpus can thus be the more dependable choice. Why wide coverage stabilizes fine-tuning remains open. Two readings are consistent with the data. One is a regularization effect, in which features learned from the wide-SoH corpus constrain the fine-tuning optimization on the narrow-band target. The other is exposure, in which the wide-SoH corpus has already encountered the degradation states, and their voltage signatures, that the narrow target never visits. The experimental design neither separates the two readings nor tests either directly. What the data establish is the stability gain itself, not its mechanism.

Third, transfer needs no data from the target pack. The first result showed that such data is not sufficient. This result shows it is not necessary either. The unseen-target regime is also the one a fleet operator faces, since cycling data for an incoming retired pack's exact design is rarely in hand. Two lines of evidence establish it. NASA (V/I/T) shares no chemistry, topology, or cycle regime with the lab packs. It nonetheless transfers well to LMO-A ($-32.0\,\%$ versus random initialization, 3/3 seeds). The gain therefore cannot be an artifact of pretrain–target similarity. The leave-one-out pools exclude the target by construction. They beat random initialization on both LMO packs (3/3 seeds each). Each such pool contains the held-out pack's sister. The gain could thus reflect copying from a near-identical pack rather than transfer. The

two packs, though one model, behave as distinct systems in entry degradation state, in aging rate, and in their response to identical pretraining (Section 3.5.3). The leave-one-out gain is therefore genuine pack-to-pack transfer, not accuracy inherited from a near-copy. This freedom is nonetheless chemistry-bounded. For the NCA pack, the leave-one-out pool is the two LMO packs alone. This pure-LMO pool is unreliable toward the chemistry-distinct target. Concretely, its mean error is nominally neutral versus random initialization. Across seeds, however, its worst-seed error reaches 2.4$\times$ its best-seed error (Table A.2). The choice of pretraining corpus is therefore unconstrained by any need to include the target, even for an unseen pack. It remains consequential, however, since outcomes across target-free pools range from strong to unreliable. The first two results establish what should guide it.

The per-pouch breakdown tempers the first result's gains. On LMO-B the combined pool's aggregate advantage is carried almost entirely by one of the two held-out pouches. That pouch is the one random initialization fits worst. There the combined pool cuts error by 36.3 %. The other held-out pouch is essentially unchanged (+1.3 %). Both pouches span the same SoH range. The gain is therefore a cell-specific correction, not a systematic improvement across the degradation band. Genuine deep-degradation difficulty does, however, surface on the NASA benchmark, where the parity deviations trace to a label-support boundary and one weak cell (Section 3.5.2). Panel (b) of Fig. 3.6 appears to scatter more widely than panel (a). The impression is an artifact of axis range. The NCA pack occupies a $\sim$10 %SoH deployment band. Its test RMSE (0.30 %SoH) is far below panel (a)'s (2.49 %SoH). The same small deviations nonetheless fill much more of the plotted axis.

Finally, composition includes not only which SoH bands the corpus covers but which channels it carries. Removing the temperature channel from NASA pretraining raises mean transfer error by up to +28.9 % (Section 3.5.3). No lab target, however, observes temperature during fine-tuning. A temperature-informed corpus thus aids even temperature-blind targets. For practitioners, V/I-only sensing is sufficient at fine-tuning time and matches the channel availability of many production BMSs. At pretraining time, corpora should include temperature

when available.

Four limitations bound the scope of these conclusions. First, the seed ensemble is small ($n = 3$). The paper therefore reports per-seed win counts, relative deltas, and segment-level bootstrap intervals rather than parametric significance tests. The stability claims rest on coefficient-of-variation contrasts that are large (2–25$\times$) relative to the ensemble size. Second, the NCA pack contributes only 49 cycling segments over a 9.9-percentage-point SoH band. Its results should be read as a case study of a hard, narrow-band target rather than a chemistry-wide conclusion about NCA. Third, label efficiency is not re-measured. The reduced-labeling advantage of self-supervised pretraining is established by prior work [71, 72]. It is inherited here as motivation, not demonstrated with a label-budget sweep on these packs. Fourth, the pretraining batch size was not held constant across conditions. The lab pools are far larger than NASA RW and were pretrained at 512 or 1024 rather than the selected 16 (Section 3.4.6), so a contrast that crosses that boundary carries a second difference besides corpus composition. The contrasts within the lab pools, and the seen-target controls, are free of it.

3.7 Conclusion

This paper asked whether masked-AE pretraining, central to recent battery foundation models, remains effective under the measurement constraints of second-life screening. On the public NASA RW benchmark the answer separates the weights from the design. The pretrained model reaches 2.82 ± 0.28 %SoH with no hand-engineered features, in line with the strongest published cell-level result. The identical architecture trained from random initialization matches it. With plentiful in-domain labels the pretrained weights are therefore dispensable. Both variants outperform LightGBM, BiLSTM, and PatchTST, which trail at 1.5–3.2$\times$ the pretrained model's RMSE. On three connected second-life packs, where labels are scarce, the answer becomes conditional on the composition of the pretraining corpus. On both LMO packs the combined pool has the lowest mean error of any source. On the NCA pack it has the highest of any

configuration. Wide-SoH NASA (V/I/T) pretraining is instead the most accurate and the most stable configuration on that narrow-band target. Its temperature-free counterpart, NASA (V/I), is likewise stable there and likewise more accurate than random initialization. Across these three packs, pool composition governs the sign of transfer. Pool size does not. The seen-target controls show that the target's own data is not sufficient. The leave-one-out pools show it is not necessary. Every configuration that fails on the NCA pack draws on a corpus saturated by one narrow SoH band. Failure is therefore diagnosable in advance from corpus composition. Seed-level stability emerges as a second axis of pretraining value. Wide SoH coverage of the corpus, not closeness to the target, delivers it. Temperature-informed features transfer to targets that never sense temperature.

Estimating SoH where labels are scarce and early-life data are unavailable, as in second-life screening, needs both accuracy and stability. The findings translate into four suggestions. First, a corpus matched to the target's chemistry class is the default choice. Second, when the target is narrow-band, a broad corpus spanning a wide SoH range can be the more dependable choice. Third, temperature channels are worth including at pretraining time even when deployment sensing is V/I-only. Fourth, amassing more cycling data from the packs to be screened is not the main lever. What is collected matters more than whether it covers the target. Following these suggestions still presumes a known chemistry class and reference capacity tests per target to supply fine-tuning labels. Pretraining on a public corpus does not remove that labeling need. Natural next steps are extending the transfer matrix to additional chemistries and curating battery foundation-model corpora by SoH coverage rather than volume.

Together with the previous chapter, this chapter completes the estimation half of the dissertation. The capacity-based SoH can now be estimated in both data regimes, from engineered features where cells are fully instrumented, and from learned representations where packs expose only voltage and current. What remains is the question of worth. Chapter 4 converts the estimated capacity of a storage system into economic value on the grid.

Acknowledgments

Chapter 3 is based on the following manuscript, currently under review: Long Ling, Shaojie Yang, Shijie Tong, and Jan Kleissl, "Masked-Autoencoder Pretraining for State-of-Health Estimation of Second-Life Battery Packs," submitted to *IEEE Transactions on Transportation Electrification*. The dissertation author was the primary researcher and author of this material.

The author thanks the UC San Diego Data Science / Machine Learning Platform for the GPU computing resources used in the fine-tuning campaign.

The three second-life datasets (NCA, LMO-A, LMO-B) are released on Zenodo (DOI: 10.5281/zenodo.21422753). The cached hyperparameter configurations, the training code, and the trained checkpoints are released on Zenodo (DOI: 10.5281/zenodo.21423164). The processed NASA RW benchmark used here ships with the released code and credits the NASA Ames Prognostics Center of Excellence as the data source.

Chapter 4

An Optimal Dispatch Model for Economic Benefit Valuation of Grid Energy Storage

4.1 Introduction

The two preceding chapters established how the capacity-based state of health (SoH) of a battery system can be estimated from its operating data, through engineered features on instrumented cells (Chapter 2) and through learned representations on retired packs (Chapter 3). Health, however, is a measure of capability, not of value. This chapter closes the dissertation's chain by asking what a storage system's capability earns when dispatched against electricity market prices, and which battery attributes drive that earning.

Innovative battery technologies—such as sodium-ion, vanadium redox flow, and all-iron redox flow batteries—are being developed for potential applications in large-scale energy storage systems essential to the electric grid. These emerging technologies are competing with lithium-ion batteries, the current market leader primarily driven by the rapidly expanding electric vehicle (EV) sector [2]. The constantly evolving grid energy storage market drives increasingly diverse applications, ranging from short-term power regulation services, such as frequency regulation and voltage support, to longer-term energy management strategies, such as energy shifting and load leveling. This evolution enhances the grid's flexibility, reliability, and efficiency but also imposes varied cost and performance targets for new battery technologies. Consequently, optimal dispatch models that target specific energy storage applications and consider key battery

performance characteristics are needed to evaluate the economic benefits of these technologies. Such optimization results can align battery candidates with suitable application segments and help innovators identify application-specific challenges, enabling targeted design improvements to maximize economic returns.

Optimal economic dispatch models have been proposed specifically for energy time shifting (ETS) [103], energy arbitrage [104], and renewable integration [105]. These models all consider operational limits such as energy capacity, power and voltage ranges, and state of charge (SoC) iterations. However, these parameters may not be sufficient to fully capture battery characteristics such as ohmic voltage losses and the nonlinear dependence of battery voltage on SoC. Introducing a battery model into the constraints can simulate battery behavior, thus enabling the optimization to generate more realistic outputs. A single-particle model and simplified pseudo-two-dimensional electrochemical models were considered in energy dispatch models for grid-connected lithium-ion and EV batteries, respectively [106, 107]. Although electrochemical models can precisely depict battery electrochemical dynamics, many battery mechanism-related experiments need to be performed to calibrate the model parameters. Additionally, the differential equations within these models make the optimization problem nonlinear and difficult to solve. The first-order resistance–capacitance (RC) equivalent circuit model (ECM) has also been applied in this field, which can be regarded as a balance between optimization effort and realistic output [108]. However, the differential equations involved in this ECM have not yet been effectively relaxed.

To address both battery characterization and optimization solvability, this study proposes an economic dispatch model that represents battery behavior using the simplest ECM, known as the Rint model [109]. The associated nonlinear constraints are relaxed into linear representations, formulating a mixed-integer linear programming (MILP) problem that enhances computational efficiency. A case study was conducted to demonstrate that the proposed model effectively evaluates the economic benefits of three battery technologies—vanadium redox flow, sodium-ion, and lithium-ion—in an ETS application within the real-time market.

The main contributions of this chapter are: 1) An optimal dispatch model for economic benefit valuation of different grid energy storage solutions is proposed. This model is applicable to ETS, energy arbitrage, and potentially renewable integration applications that share price-based revenue mechanisms. 2) The analysis reveals that higher C-rates and stronger power capabilities enable battery energy storage systems (BESSs) to cycle more frequently and better capitalize on price differences, thereby generating greater profit in energy dispatch-based applications.

The remainder of this chapter is organized as follows: Section 4.2 introduces the problem formulation of the proposed economic dispatch model. Section 4.3 describes the model's implementation through a case study. Section 4.4 analyzes the optimization results and highlights key findings. Finally, the conclusions are presented in Section 4.5.

4.2 Problem Formulation

4.2.1 Objective Function

The objective function aims to minimize the operational costs of a BESS in a grid energy storage application, as mathematically defined in (4.1).

$$\min \sum_{i=1}^{n} \left(p_{\mathrm{c}}^{i} / \eta_{\mathrm{ACDC}} - p_{\mathrm{d}}^{i} \cdot \eta_{\mathrm{DCAC}} \right) \cdot p_{\mathrm{e}}^{i} \cdot \Delta t \tag{4.1}$$

where p_{c}^{i} and $p_{\mathrm{d}}^{i} \in \mathbb{R}$ represent the charging and discharging power in kW at the i-th time step, respectively; η_{ACDC} and η_{DCAC} denote the efficiencies of alternating current (AC) to direct current (DC) and DC to AC conversion, respectively; $p_{\mathrm{e}}^{i} \in \mathbb{R}$ is the electricity price in \$/kWh at the i-th time step; Δt is the duration of one time step in hours; $i \in [1, n]$ is the current time step; and n is the total number of time steps. Each term thus multiplies an energy in kWh by a price in \$/kWh. The objective value is therefore in dollars.

This study assumes that the market electricity auction process is complete and prices are settled. Given the rapid advancements in battery technologies and the volatility in supply and demand for battery energy and raw materials, costs related to battery degradation, production, and

maintenance fluctuate significantly. Consequently, this study focuses on the ETS revenue alone, deferring cost considerations to future research due to the challenges of accurate quantification. Demand charges, the fees a tariff levies on a customer's peak power draw, are also outside the model. A battery that shaves its peak draw would earn additional demand charge savings, and their inclusion is deferred alongside the cost considerations.

4.2.2 Constraints

To enhance the realism of the economic dispatch, the Rint model, illustrated in Fig. 4.1, is selected to simulate battery behavior. Its simplicity ensures computational efficiency, and its absence of RC networks—typically used to represent diffusion-related processes in lithium-ion batteries—enables broad applicability to diverse chemistries, such as vanadium redox flow batteries, where such processes are absent. The model also represents the entire storage system as one lumped cell. Cell-level interactions, such as the natural self-balancing among parallel-connected cells and the balancing a battery management system performs, are therefore not taken into account. $U_{\mathrm{OCV}} \in \mathbb{R}^n$ represents the open circuit voltage (OCV), $U_{\mathrm{t}} \in \mathbb{R}^n$ denotes the terminal voltage, $R \in \mathbb{R}$ is the ohmic resistance, and $I_{\mathrm{L}} \in \mathbb{R}^n$ refers to the load current. We use I_{c}^i and $I_{\mathrm{d}}^i \in \mathbb{R}$ to differentiate I_{L} during charging and discharging. These parameters follow Ohm's law, as expressed in (4.2). In addition, U_{OCV} exhibits a nonlinear relationship with $SoC \in \mathbb{R}^n$, driven by thermodynamic properties and, depending on the battery type, phase transitions, electrode material characteristics, or electrolyte concentration dynamics.

$$\begin{cases} U_{\mathrm{t}}^i = U_{\mathrm{OCV}}^i - (I_{\mathrm{d}}^i - I_{\mathrm{c}}^i) \cdot R, \\ U_{\mathrm{OCV}}^i = f(SoC^i), \\ U_{\mathrm{t}}^{\min} \leq U_{\mathrm{t}}^i \leq U_{\mathrm{t}}^{\max}. \end{cases} \tag{4.2}$$

The initial SoC value of a BESS is assumed to be 50%, and the terminal SoC value is expected to return to this initial state after all time steps. Additional SoC-related constraints are summarized in (4.3). The current-related and power-related constraints are outlined in (4.4) and

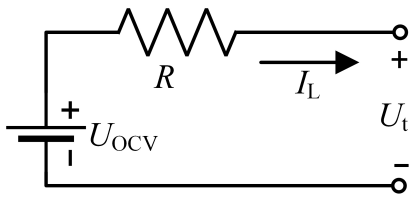


Figure 4.1. Rint battery model.

(4.5), respectively. The binary variable $h \in \{0,1\}^n$ ensures that either charging, discharging, or neither occurs at any given time step.

$$\begin{cases} SoC^0 = 0.5, SoC^n = 0.5, \\ SoC^{\min} \leq SoC^i \leq SoC^{\max}, \\ SoC^j = SoC^{j-1} + (I_c^j \cdot \eta_C - I_d^j) \cdot \Delta t/Q,\ j = 1,2,\ldots,n. \end{cases} \tag{4.3}$$

where η_C denotes the Coulombic efficiency and Q represents the ampere-hour capacity of a BESS.

$$\begin{cases} I_c^{\min} \leq I_c^i \leq I_c^{\max}, I_d^{\min} \leq I_d^i \leq I_d^{\max}, \\ I_c^{j+1} - I_d^{j+1} - I_c^j + I_d^j \leq I_r, \\ I_c^j - I_d^j - I_c^{j+1} + I_d^{j+1} \leq I_r, \quad j = 1,2,\ldots,n-1. \end{cases} \tag{4.4}$$

$$\begin{cases} p_c^{\min} \leq p_c^i \leq p_c^{\max} \cdot h^i, p_d^{\min} \leq p_d^i \leq p_d^{\max} \cdot (1-h^i), \\ p_c^i = U_t^i \cdot I_c^i, p_d^i = U_t^i \cdot I_d^i. \end{cases} \tag{4.5}$$

The products $U_t^i \cdot I_c^i$ and $U_t^i \cdot I_d^i$ are expressed in kW, consistent with the units of (4.1).

4.2.3 Relaxation Techniques

The *SoC*-OCV relationship $f(SoC)$ and the equation constraints in (4.5) are nonlinear, making the optimization problem prone to local minima and challenging to solve. To address this, relaxation techniques are applied to linearize these constraints, converting the problem into a MILP problem. The relaxations preserve the variation of U_{OCV} with *SoC* along $f(SoC)$, while the

battery parameters $f(SoC)$, R, η_{C}, and Q stay at their characterized values over the optimization horizon. Battery degradation is therefore not modeled, either as a drift in these parameters or as a cost in the objective. Its primary effects, a lower Q and a higher R, would reduce both the energy a system can shift and the revenue it earns. They would also change the optimal dispatch schedule itself, because a smaller usable capacity and larger resistive losses shift when, and at what power, the system charges and discharges. Chemistry, by contrast, does enter the model. Each battery technology is characterized by its own tests in Section 4.3, and the chemistry sensitivity of cycling behavior enters through these characterized parameters. Degradation also differs across chemistries in rate and mechanism. The dispatch changes described above would therefore differ across technologies as well, and the comparison among them would shift as the systems age. This chemistry dependence is excluded along with degradation itself. Dispatch that updates the capacity Q with the estimated SoH is outlined as a future direction in Chapter 5.

(1) Piecewise Linear Approximation

Assume the SoC-OCV relationship $f(SoC)$ in (4.2) can be approximated using m evenly distributed sampling points $SoC^1, SoC^2, \cdots, SoC^m$. The function values of $f(SoC)$ can then be evaluated by interpolating between neighboring pairs of consecutive break points. $f(SoC)$ is reformulated as shown in (4.6).

$$\begin{cases} \sum_{k=0}^{m} \alpha^k = 1, \sum_{k=1}^{m} \beta^k = 1, \\ \beta^k \leq \alpha^{k-1} + \alpha^k, \quad k = 1, \ldots, m, \\ \mathrm{SoC}^x = \sum_{k=1}^{m} \beta^k \, \mathrm{SoC}^k, U^x_{\mathrm{OCV}} = \sum_{k=1}^{m} \beta^k U^k_{\mathrm{OCV}}. \end{cases} \tag{4.6}$$

where SoC^x and U^x_{OCV} denote any SoC and U_{OCV} values, respectively.

(2) McCormick Envelopes

A bilinear constraint, such as $p^i = U^i_{\mathrm{t}} \cdot I^i$ can be relaxed into a linear form by constructing its concave and convex envelopes. The resulting linear constraints are provided in (4.7).

$$\begin{cases} p^i \leq U_{\mathrm{t}}^{\max} \cdot I^i + U_{\mathrm{t}}^i \cdot I^{\min} - U_{\mathrm{t}}^{\max} \cdot I^{\min}, \\ p^i \leq U_{\mathrm{t}}^i \cdot I^{\max} + U_{\mathrm{t}}^{\min} \cdot I^i - U_{\mathrm{t}}^{\min} \cdot I^{\max}, \\ p^i \geq U_{\mathrm{t}}^{\min} \cdot I^i + U_{\mathrm{t}}^i \cdot I^{\min} - U_{\mathrm{t}}^{\min} \cdot I^{\min}, \\ p^i \geq U_{\mathrm{t}}^{\max} \cdot I^i + U_{\mathrm{t}}^i \cdot I^{\max} - U_{\mathrm{t}}^{\max} \cdot I^{\max}. \end{cases} \tag{4.7}$$

where the subscripts c and d for p^i and I^i are omitted to facilitate a generalized representation applicable to both charging and discharging.

4.3 Case Study

A case study is conducted to demonstrate the implementation of the proposed economic dispatch model. This work evaluates three advanced battery technologies—lithium-ion, sodium-ion, and vanadium redox flow batteries—in an ETS application within a real-time market.

4.3.1 Price Time Series Data

The ETS application leverages market price fluctuations by charging a BESS during off-peak periods and discharging during peak periods. Fig. 4.2(a)-(d) presents the hourly locational marginal prices at the La Jolla node, sourced from the California Independent System Operator (CAISO) real-time market throughout 2022. The figure reports the prices in \$/MWh. Daily optimization is employed for two key reasons: 1) it is computationally efficient compared to long-term optimizations; and 2) it captures typical short-term price patterns and market operational dynamics that may be less evident in long-term optimizations, facilitating the analysis of general daily performance. Each daily optimization spans $n = 24$ hourly steps with Δt equal to one hour.

With a price data size of 365×24, daily optimizations for each BESS would require 365 runs, making it computationally intensive. To reduce the computational burden, we treated each day's hourly price data as a feature vector and applied Principal Component Analysis for dimensionality reduction. As a result, only 8 principal components out of the original 24 were retained. Subsequently, K-means clustering was employed to partition the 365 observations

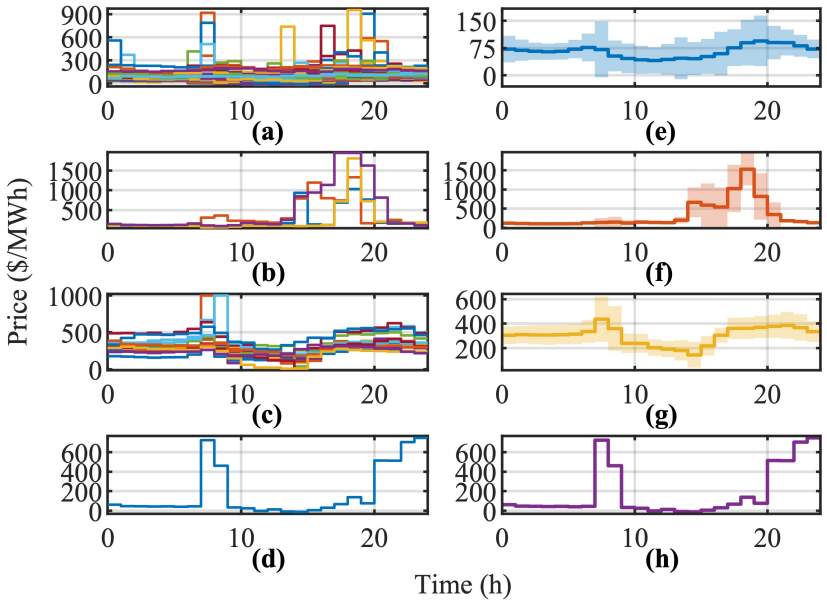


Figure 4.2. Hourly electricity prices in the real-time market at the La Jolla node in 2022, sourced from CAISO: (a)–(d) show the raw price data, categorized into four clusters; (e)–(h) present the cluster centers for Clusters 1 to 4 (solid lines) with standard deviations (shaded areas).

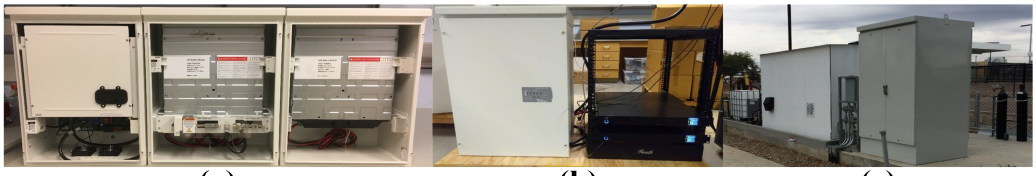
into four clusters. As shown in Fig. 4.2(e)-(h), each cluster center represents a typical day's price profile with dimensions of 1×24 and is used for optimization. This approach reduced the required optimizations per BESS to four, significantly decreasing computational effort while still providing representative results of the entire year's price data. The number of days assigned to each cluster is 342, 4, 18, and 1, respectively.

4.3.2 Battery Modeling

(1) Battery Test

Battery characteristics are sensitive to specific application conditions. To ensure accurate battery modeling, three BESSs were deployed at laboratory and campus sites at the University of California San Diego (UCSD), as shown in Fig. 4.3. Constant current charging and discharging tests were then conducted to determine the ampere-hour capacity and Coulombic efficiency. Additionally, the *SoC*-OCV relationship was established through low-current OCV tests [110]. An optimal dispatch algorithm was developed to generate duty cycle power profiles for the

BESSs in ETS applications within the real-time market [111]. Utilizing MILP, the algorithm efficiently manages energy storage resources to maximize revenue, considering fundamental battery parameters and operational limits specified by manufacturers and validated through initial tests, as summarized in Table 4.1. These power profiles were further applied in dynamic tests to identify ohmic resistances.



(a) (b) (c)

Figure 4.3. BESSs deployed at UCSD: (a) lithium-ion battery; (b) sodium-ion battery; (c) vanadium redox flow battery.

Table 4.1. Battery parameters and operating limits.

Battery Chemistry	Voltage Range (V)	Power Range (W)	Current Range (A)	Current Ramp Limit (A)	Energy Capacity (Wh)	Energy-to-Power Ratio (h)
Lithium-ion	48.4-55.7	0-3000	0-53.9	65	13000	4.3
Sodium-ion	41.7-57.4	0-200	0-3.5	5	370	1.5
Vanadium redox flow	25.0-32.7	0-3360	0-102.9	135	2000	0.6

(2) Battery Model Parameter Identification

The identified battery model parameters are outlined in Table 4.2. Ampere-hour capacity was computed by integrating the current during discharging. Coulombic efficiency was determined by dividing the discharge capacity by the charge capacity. As shown in Fig. 4.4, the OCV curves for charging and discharging were averaged to mitigate certain hysteresis effects on voltage response, primarily for lithium-ion and sodium-ion batteries, where such effects are influenced by factors such as internal resistance, electrode kinetics, and ion diffusion. To balance accuracy with computational efficiency, 21 evenly distributed *SoC* data points ranging from 0 to

1 were selected to compute the average and define the *SoC*-OCV relationship. This averaged relationship was subsequently utilized to identify ohmic resistances from collected dynamic test data by solving the first equation in (4.2). The resistance values calculated at each time step were averaged to obtain the final values.

Table 4.2. Identified battery model parameters.

Battery Chemistry	Ampere-hour Capacity (Ah)	Coulombic Efficiency (%)	Ohmic Resistance (Ω)
Lithium-ion	181.37	97.58	1.39E-02
Sodium-ion	6.91	96.44	9.82E-02
Vanadium redox flow	101.89	92.20	9.94E-03

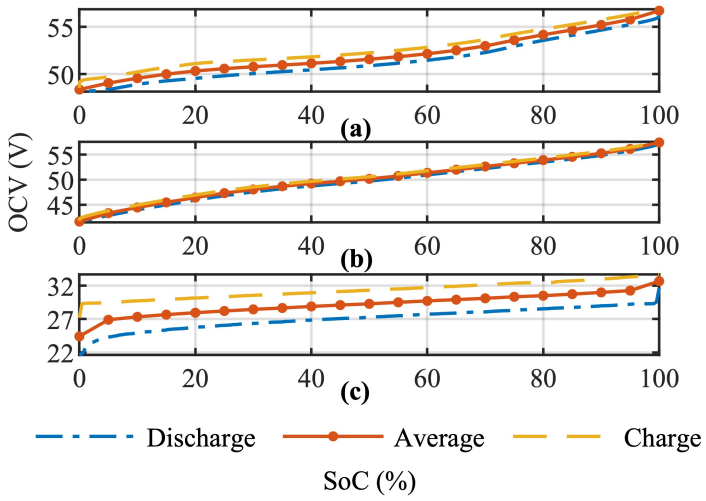


Figure 4.4. *SoC*-OCV curves identified from battery characteristic tests: (a) lithium-ion battery; (b) sodium-ion battery; (c) vanadium redox flow battery.

4.3.3 Optimization Setup

Gurobi 11.0.3 was employed to solve the MILP problem described in Section 4.2. Specifically, the *ConcurrentMIP* parameter was set to 2, enabling the solver to execute two MILP solves in parallel and thereby leveraging multicore processing to enhance computational

efficiency. Additionally, the *MIPFocus* parameter was set to 1 to prioritize the rapid identification of high-quality feasible solutions, accelerating convergence toward optimality. All other solver parameters were maintained at their default settings.

The AC-DC (η_{ACDC}) and DC-AC (η_{DCAC}) efficiencies were set to a constant value of 0.9, as this is a commonly used value that accommodates varying power levels and simplifies computations [112]. To ensure a fair comparison given the differing operational limits of the three BESSs, their maximum terminal voltages were first scaled to 1000 V, followed by standardizing their energy capacities to 1 MWh. The resulting scale factors were then applied to adjust the remaining parameters accordingly.

4.4 Results and Discussion

Fig. 4.5 presents a comparison of revenue and cycle numbers across different battery chemistries and price clusters. A rainflow counting algorithm was utilized to calculate the cycles experienced by the BESSs [113]. Overall, all optimizations successfully captured price fluctuations and generated revenue through energy trading. As shown in Fig. 4.2(e)-(h) and Fig. 4.5(a), a higher price range is associated with increased revenue. Moreover, Clusters 3 and 4 demonstrate that similar price ranges can result in significantly different revenues. Specifically, the standard deviations of price data in Cluster 3 (0.26 \$/kWh) and Cluster 4 (0.34 \$/kWh) suggest that BESSs can achieve higher profitability with greater price dispersion.

Optimization results for the three BESSs across four price clusters are aggregated in Table 4.3 using weighted summation. The results highlight that higher daily revenue is primarily attributed to both increased cycles and higher revenue per cycle. The C-rate, defined as the charge or discharge rate relative to a BESS's ampere-hour capacity, plays a key role. A larger C-rate means a BESS can be fully charged or discharged in less time, allowing it to run more cycles within the same time period. For example, in the most representative price cluster (Cluster 1), the maximum C-rates achieved by lithium-ion, sodium-ion, and vanadium redox flow BESSs

are 0.2, 0.4, and 0.7, respectively, resulting in 1.0, 1.8, and 2.2 cycles, respectively, as shown in Fig. 4.5(b). Fig. 4.6(a)-(b) graphically depicts this finding, illustrating that BESSs with higher C-rates exhibit more frequent *SoC* and voltage cycling. These curves also show that, due to the constraints imposed by the *SoC*-OCV relationship (i.e., maximum $U_t <$ maximum U_{OCV}) and internal resistance, the maximum *SoC* of the lithium-ion BESS remains below 100%, even when its terminal voltage reaches 1000 V. This indicates that the integrated battery model effectively constrains the system, ensuring the generation of reasonable and realistic results.

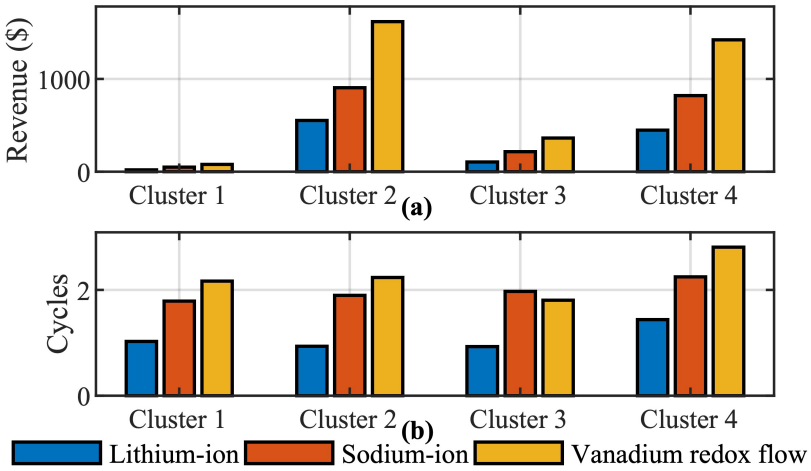


Figure 4.5. Economic dispatch results for four electricity price clusters across three BESSs: (a) daily revenue; (b) daily cycles.

Table 4.3. Aggregated daily optimization results.

Battery Chemistry	Revenue ($)	Cycles	Revenue per Cycle ($)
Lithium-ion	31.17	1.02	30.62
Sodium-ion	67.63	1.80	37.61
Vanadium redox flow	111.08	2.15	51.56

Furthermore, higher power capabilities enable BESSs to generate more profit by capitalizing on price differences, resulting in higher revenue per cycle. As shown in Fig. 4.6(c), the vanadium redox flow BESS leverages its superior power capacity to discharge during price peaks and charge during price valleys at more than twice the power rate of the other two BESSs,

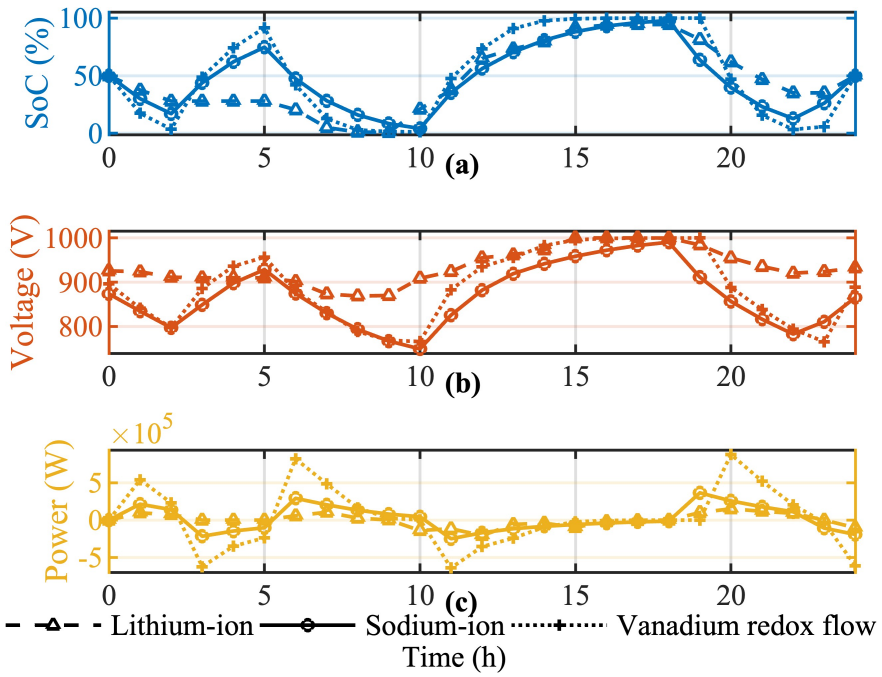


Figure 4.6. Economic dispatch results for the most representative electricity price cluster across three BESSs: (a) SoC; (b) terminal voltage; (c) power. Positive power indicates discharging, while negative power indicates charging.

achieving the highest revenue per cycle. Similarly, under the given system configurations, the sodium-ion BESS demonstrates stronger power capabilities than the lithium-ion BESS in energy arbitrage. While scaling adjustments can further enhance the power limits of vanadium redox flow and sodium-ion BESSs, high power capability is an intrinsic feature of these technologies.

It is important to note that the battery technologies examined in this research are at significantly different stages of engineering development and commercialization. Consequently, the analysis primarily showcases the methodology, while the technologies themselves may vary in terms of readiness for commercial application.

4.5 Conclusion

This study proposed a MILP model for scheduling energy in BESSs comprising different battery technologies. The case study on an ETS application confirmed that the applied relaxation techniques effectively enabled problem-solving across different parameter configurations. The

optimization results demonstrated that the introduced battery model appropriately constrained battery operations, producing outcomes that align with realistic operational behavior. Further analysis revealed that C-rate and power capability significantly influence battery profitability in an ETS application, validating the model's feasibility in assessing economic performance differences among the three technologies. Thus, the proposed model shows strong potential for economic evaluations of emerging BESSs in applications such as ETS, energy arbitrage, and renewable integration, where revenue depends on price-based gains. Future work will extend the model to include broader cost analysis, additional energy storage applications, and more representative battery metrics for comprehensive BESS evaluation.

This chapter completes the chain the dissertation set out to build. Health estimated under realistic cycling and deployment constraints supplies the capacity input, and optimal dispatch converts the system's capability into revenue. Chapter 5 draws the three studies together and outlines the step this chain motivates next, dispatch that adapts to the estimated and evolving health of the system it schedules.

Acknowledgments

Chapter 4 is based on the following publication: Long Ling, Cristian Cortes-Aguirre, Shaojie Yang, and Shijie Tong, "An Optimal Dispatch Model for Economic Benefit Valuation of Grid Energy Storage Solutions," in *Proceedings of the IEEE Green Technologies Conference (GreenTech)*, Wichita, KS, USA, 2025, pp. 113–117. The dissertation author was the primary investigator and author of this paper.

Chapter 5

Conclusions

This dissertation developed the chain from battery health estimation to the economic value of grid energy storage in three steps. Manually engineered features estimate the capacity-based state of health (SoH) under randomized cycling. Learned representations carry SoH estimation to second-life battery packs with limited sensing and scarce labels. An optimal dispatch model converts the estimated capacity into economic benefit on the grid.

5.1 Summary of Contributions

Each of the three studies delivered one of the four contributions stated in the introduction. Their synthesis into one workflow for a storage operator, presented in Section 5.2, forms the fourth.

Chapter 2 developed an SoH estimation framework for single cells cycled under randomized loads, where no repeating charge or discharge template exists for classical features to align against. The framework summarizes each stretch of cycling by the probability density functions (PDFs) of voltage and temperature, identifies the underlying bimodal and normal distribution parameters with the expectation–maximization algorithm, and feeds the identified parameters and statistical descriptors most correlated with SoH to a support vector regression (SVR). On the NASA random walk benchmark the framework reached a mean absolute error (MAE) of 2.63 % and a root mean square error (RMSE) of 3.31 %. Its maximum error was the lowest of

all forty algorithm-input combinations. Its accuracy held across four test groups with different load profiles and ambient temperatures. Moreover, extending the sampling interval from 1 s to 5 s raised the error indicators by less than 5 % and cut the average storage requirement from 3.18 MB to 0.64 MB. This tolerance makes the framework practical for an embedded battery management system (BMS).

Chapter 3 developed a two-stage masked-autoencoder framework for SoH estimation on second-life battery packs. Pretraining learns a Transformer encoder on unlabeled cycling data. Fine-tuning then transfers that encoder and trains it together with a cross-window attention pooler and a regression head that predicts one SoH value per labeled cycling segment. On the NASA random walk benchmark the pretrained model reached a test RMSE of 2.82 ± 0.28 %SoH with no manually engineered features. LightGBM, BiLSTM, and PatchTST baselines trailed at 1.5–3.2$\times$ its error. The chapter's core result came from a transfer ablation across five pretraining sources, three genuinely retired packs, and three seeds. Pool composition, not pool size, decides whether pretraining helps or hurts. On both lithium-manganese-oxide (LMO) packs the combined four-source pool was the best source (-35.5 % and -28.1 % versus random initialization, 3/3 seeds each). On the chemistry-distinct nickel-cobalt-aluminum-oxide (NCA) pack, however, the same checkpoint was the worst configuration ($+41.6$ % versus random initialization). Wide-SoH NASA voltage, current, and temperature (V/I/T) pretraining was instead the most accurate and the most stable configuration on that pack. The target pack's own cycling data proved neither sufficient nor necessary for transfer. Moreover, removing the temperature channel from NASA pretraining raised mean transfer error by up to $+28.9$ %, although no target pack senses temperature during fine-tuning. Together these findings make transfer failure diagnosable in advance from corpus composition alone.

Chapter 4 developed an optimal dispatch model for economic benefit valuation of battery energy storage systems in price-based grid applications. The model represents battery behavior inside the constraints with the Rint equivalent circuit model. Piecewise linear approximation and McCormick envelopes then relax the nonlinear constraints into a mixed-integer linear

programming (MILP) problem. A case study dispatched three battery technologies, each characterized by in-house tests, against clustered 2022 real-time prices from the California Independent System Operator in an energy time shifting application. With all three systems scaled to the same maximum terminal voltage and energy capacity, the vanadium redox flow, sodium-ion, and lithium-ion systems earned aggregated daily revenues of $111.08, $67.63, and $31.17, respectively. The reason is that higher C-rates and stronger power capability let a system cycle more often and earn more per cycle. These are therefore the battery attributes that matter most in price-based grid applications.

5.2 Synthesis: From Estimated Health to Dispatched Value

The two estimation chapters cover complementary data regimes. Chapter 2 assumes instrumented cells. Voltage and temperature channels are both available. Labeled reference tests are plentiful across a wide SoH range. In that regime manually engineered features are accurate, cheap to compute, and robust to downsampling. Chapter 3 assumes deployed second-life packs. Sensing is limited to group-level voltage and one shared string current. Labels are scarce and narrow-band. In that regime the representation, the features an estimator extracts from raw cycling data, must be learned rather than manually engineered. The framework therefore pretrains it on abundant unlabeled cycling data so that the pack's scarce labels are spent only on fine-tuning. The choice between the two approaches is therefore set by the data regime. Moreover, both estimators predict SoH from operating data without taking the battery out of service.

That SoH estimate is one of the physical inputs that Chapter 4 consumes. The dispatch model schedules a storage system subject to its capacity, resistance, and power limits. SoH determines the capacity directly. The three chapters therefore combine into one workflow for a storage operator. Where a system exposes cell-level voltage and temperature, for example an instrumented first-life application, the manually engineered features of Chapter 2 track SoH

even under randomized usage. For second-life packs, the operator instead builds the estimator of Chapter 3 once, on a small set of sample packs. Pretraining consumes abundant unlabeled cycling data. The composition rule points to a corpus that spans a wide SoH range as the dependable choice. Fine-tuning consumes reference-labeled segments from the sample packs. An incoming retired pack is then screened for remaining capacity and tracked in service from its operating data alone, without a reference test of its own. The tracked capacity enters the dispatch model of Chapter 4, which converts it into daily revenue under the prices the system actually faces.

The integration presented here is conceptual rather than operational. The three studies were validated separately. Chapter 2 used single cells from the NASA random walk benchmark, Chapter 3 used the same benchmark together with three retired second-life packs, and Chapter 4 used three battery technologies. Moreover, the packs that supplied fine-tuning labels in Chapter 3 were also the packs evaluated, so label-free screening of a new pack remains to be demonstrated. No experiment in this dissertation yet closes the loop by feeding an estimated, uncertain SoH trajectory into the dispatch model. Closing that loop is the natural next step. It motivates the directions below.

5.3 Future Directions

(1) Health-aware dispatch.

The dispatch model of Chapter 4 treats capacity as a fixed parameter identified once. Coupling it to the estimators of Chapters 2–3 would let the ampere-hour capacity Q evolve with the estimated SoH. The dispatch schedule and the resulting revenue would then track the capacity the system still holds as it ages. This coupling would close the estimation–valuation loop that the present chapters connect only conceptually.

(2) Online deployment.

Both estimators operate on data that production systems already log. The downsampling robustness of Chapter 2 and the ability of Chapter 3 to operate on voltage and current (V/I) alone

are the enabling properties. Streaming SoH estimation inside a BMS still raises questions the laboratory studies did not face. One is real-time response. The Transformer encoder of Chapter 3 would have to run within the memory and compute budget of production BMS hardware. Another is data accumulation. Both frameworks estimate from a stretch of cycling data rather than from an instantaneous reading. A deployed system must therefore decide when enough data has accumulated for a dependable estimate.

(3) Broader chemistries and corpus curation.

The transfer ablation of Chapter 3 covered one NCA pack and two LMO packs. Extending the ablation to additional chemistries and pack topologies would test how far the composition rule generalizes. The composition rule also suggests curating battery foundation-model corpora by SoH coverage rather than volume, a criterion that can be evaluated before any training run.

(4) End-to-end uncertainty quantification.

An SoH estimate carries uncertainty. Chapter 3 already quantifies it, reporting every result across three training seeds and showing that estimates on the NCA pack are seed-sensitive. The dispatch model of Chapter 4, however, consumes one fixed capacity value. Solving the dispatch at the low and high ends of an estimated SoH band would turn the estimation uncertainty into a revenue range. An investor would then see the best and worst revenue a system supports at its current SoH instead of a single number.

Appendix A

Supplementary Material for Chapter 3

A.1 Pretraining Corpus Composition

Table A.1 breaks down the combined four-source pretraining pool by source, quantifying the corpus imbalance discussed in Section 3.6. The two LMO packs supply ∼72.9 % of the pretraining windows, against 17.7 % for the NCA pack and 9.3 % for NASA RW. The imbalance arises because the LMO packs contribute both more monitored pouches and far longer cycling records. In SoH terms the imbalance is sharper still. Concretely, ∼75.2 % of the pool's SoH labels fall in the LMO band (60–74 %) and only ∼19.4 % in the NCA band (86–96 %). The LMO packs contribute nothing to that band.

Table A.1. Composition of the combined four-source pretraining pool. Pretraining windows are the 600-sample windows (training stride 300) the masked AE optimizes on; share is their fraction of the pool.

Source	Cells	Samples (M)	Windows	Share (%)
NASA RW	14	24.8	82,531	9.3
NCA	12	47.2	157,168	17.7
LMO-A	16	89.3	297,094	33.5
LMO-B	16	104.9	349,158	39.4
Total	58	266.2	885,951	100.0

A.2 Per-Seed Results

Table A.2 lists the per-seed test RMSE behind the aggregate statistics of Table 3.9. It also lists the three NASA+own-pack seen-target controls summarized in Table 3.10.

Table A.2. Per-seed test RMSE in %SoH for all 18 transfer configurations, the 15 source–target pairs of Table 3.9 plus one NASA+own-pack seen-target control per target (Table 3.10). Combined stands for NASA + all three lab datasets, and Leave-one-out for the other two lab datasets.

Target	Pretrain	seed 123	seed 42	seed 7	mean	std
LMO-A	Combined	0.18	0.20	0.20	0.19	0.011
	Leave-one-out	0.22	0.22	0.20	0.21	0.009
	NASA (V/I/T)	0.21	0.20	0.21	0.21	0.004
	NASA (V/I)	0.29	0.25	0.25	0.26	0.026
	Random initialization	0.32	0.27	0.32	0.30	0.029
	NASA+LMO-A	0.20	0.21	0.25	0.22	0.026
LMO-B	Combined	0.19	0.21	0.21	0.20	0.013
	Leave-one-out	0.26	0.27	0.25	0.26	0.014
	NASA (V/I/T)	0.25	0.27	0.29	0.27	0.024
	NASA (V/I)	0.27	0.31	0.25	0.28	0.029
	Random initialization	0.27	0.31	0.27	0.28	0.024
	NASA+LMO-B	0.25	0.29	0.25	0.26	0.021
NCA	Combined	0.56	0.48	0.46	0.50	0.053
	Leave-one-out	0.57	0.25	0.24	0.36	0.189
	NASA (V/I/T)	0.30	0.29	0.30	0.30	0.006
	NASA (V/I)	0.31	0.32	0.35	0.33	0.017
	Random initialization	0.40	0.37	0.28	0.35	0.061
	NASA+NCA	0.41	0.42	0.37	0.40	0.029

Bibliography

[1] X. Hu, C. Zou, C. Zhang, and Y. Li, "Technological developments in batteries: A survey of principal roles, types, and management needs," *IEEE Power and Energy Magazine*, vol. 15, pp. 20–31, 2017.

[2] F. Duffner, N. Kronemeyer, J. Tübke, J. Leker, M. Winter, and R. Schmuch, "Post-lithium-ion battery cell production and its compatibility with lithium-ion cell production infrastructure," *Nature Energy*, vol. 6, pp. 123–134, 2021.

[3] Y. Che, X. Hu, X. Lin, J. Guo, and R. Teodorescu, "Health prognostics for lithium-ion batteries: Mechanisms, methods, and prospects," *Energy & Environmental Science*, vol. 16, pp. 338–371, 2023.

[4] V. S., H. S. Che, J. Selvaraj, K. S. Tey, J. W. Lee, H. Shareef, and R. Errouissi, "State of health (SoH) estimation methods for second life lithium-ion battery—review and challenges," *Applied Energy*, vol. 369, p. 123542, 2024.

[5] F. Guo, X. Wu, L. Liu, J. Ye, T. Wang, L. Fu, and Y. Wu, "Prediction of remaining useful life and state of health of lithium batteries based on time series feature and Savitzky-Golay filter combined with gated recurrent unit neural network," *Energy*, vol. 270, p. 126880, 2023.

[6] L. Ling, D. Sun, X. Yu, and R. Huang, "State of charge estimation of lithium-ion batteries based on the probabilistic fusion of two kinds of cubature kalman filters," *Journal of Energy Storage*, vol. 43, p. 103070, 2021.

[7] L. Ling, C. Cortes-Aguirre, S. Yang, and S. Tong, "An optimal dispatch model for economic benefit valuation of grid energy storage solutions," in *Proc. IEEE Green Technologies Conf. (GreenTech)*, Wichita, KS, USA, 2025, pp. 113–117.

[8] X. Bian, Z. G. Wei, W. Li, J. Pou, D. U. Sauer, and L. Liu, "State-of-health estimation of lithium-ion batteries by fusing an open-circuit-voltage model and incremental capacity analysis," *IEEE Transactions on Power Electronics*, p. 1, 2021.

[9] S. Khaleghi, D. Karimi, S. H. Beheshti, M. S. Hosen, H. Behi, M. Berecibar, and J. van

Mierlo, "Online health diagnosis of lithium-ion batteries based on nonlinear autoregressive neural network," *Applied Energy*, vol. 282, p. 116159, 2021.

[10] H. Obeid, R. Petrone, H. Chaoui, and H. Gualous, "Higher order sliding-mode observers for state-of-charge and state-of-health estimation of lithium-ion batteries," *IEEE Transactions on Vehicular Technology*, vol. 72, pp. 4482–4492, 2023.

[11] S. Wang, S. Zhang, S. Wen, and C. Fernandez, "An accurate state-of-charge estimation of lithium-ion batteries based on improved particle swarm optimization-adaptive square root cubature kalman filter," *Journal of Power Sources*, vol. 624, p. 235594, 2024.

[12] X. Li, C. Yuan, Z. Wang, J. He, and S. Yu, "Lithium battery state-of-health estimation and remaining useful lifetime prediction based on non-parametric aging model and particle filter algorithm," *eTransportation*, vol. 11, p. 100156, 2022.

[13] C. Lin, L. Wu, X. Tuo, C. Liu, W. Zhang, Z. Huang, and G. Zhang, "A lightweight two-stage physics-informed neural network for SOH estimation of lithium-ion batteries with different chemistries," *Journal of Energy Chemistry*, vol. 105, pp. 261–279, 2025.

[14] S. Wang, Y. Fan, S. Jin, P. Takyi-Aninakwa, and C. Fernandez, "Improved anti-noise adaptive long short-term memory neural network modeling for the robust remaining useful life prediction of lithium-ion batteries," *Reliability Engineering & System Safety*, vol. 230, p. 108920, 2023.

[15] X. Xia, Y. Chen, J. Shen, Y. Liu, Y. Zhang, Z. Chen, and F. Wei, "State of health estimation for lithium-ion batteries based on impedance feature selection and improved support vector regression," *Energy*, vol. 326, p. 136135, 2025.

[16] S. Wang, F. Wu, P. Takyi-Aninakwa, C. Fernandez, D.-I. Stroe, and Q. Huang, "Improved singular filtering-gaussian process regression-long short-term memory model for whole-life-cycle remaining capacity estimation of lithium-ion batteries adaptive to fast aging and multi-current variations," *Energy*, vol. 284, p. 128677, 2023.

[17] Y. Wang, K. Li, P. Peng, and Z. Chen, "Health diagnosis for lithium-ion battery by combining partial incremental capacity and deep belief network during insufficient discharge profile," *IEEE Transactions on Industrial Electronics*, vol. 70, pp. 11 242–11 250, 2023.

[18] X. Chen, Y. Hu, S. Li, Y. Wang, D. Li, C. Luo, X. Xue, F. Xu, Z. Zhang, Z. Gong, Y. Li, and Y. Yang, "State of health (SoH) estimation and degradation modes analysis of pouch NMC532/graphite li-ion battery," *Journal of Power Sources*, vol. 498, p. 229884, 2021.

[19] J. Yang, Y. Cai, and C. Mi, "Lithium-ion battery capacity estimation based on battery surface temperature change under constant-current charge scenario," *Energy*, vol. 241, p. 122879, 2022.

[20] H. Liu, I. H. Naqvi, F. Li, C. Liu, N. Shafiei, Y. Li, and M. Pecht, "An analytical model for the CC-CV charge of li-ion batteries with application to degradation analysis," *Journal of Energy Storage*, vol. 29, p. 101342, 2020.

[21] K. A. Severson, P. M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M. H. Chen, M. Aykol, P. K. Herring, D. Fraggedakis, M. Z. Bazant, S. J. Harris, W. C. Chueh, and R. D. Braatz, "Data-driven prediction of battery cycle life before capacity degradation," *Nature Energy*, vol. 4, pp. 383–391, 2019.

[22] Z. Lin, Y. Cai, W. Liu, C. Bao, J. Shen, and Q. Liao, "Estimating the state of health of lithium-ion batteries based on a probability density function," *International Journal of Electrochemical Science*, vol. 18, p. 100137, 2023.

[23] M. Lin, X. Zeng, and J. Wu, "State of health estimation of lithium-ion battery based on an adaptive tunable hybrid radial basis function network," *Journal of Power Sources*, vol. 504, p. 230063, 2021.

[24] H. Dong, J. Fan, Y. Zou, L. Xiao, B. Deng, Y. Ou, Y. Yang, and J. Guan, "Robustness evaluation of probability density function based features on state-of-health estimation used in electric vehicles," *IOP Conference Series: Earth and Environmental Science*, vol. 585, p. 012141, 2020.

[25] A. Maheshwari, N. G. Paterakis, M. Santarelli, and M. Gibescu, "Optimizing the operation of energy storage using a non-linear lithium-ion battery degradation model," *Applied Energy*, vol. 261, p. 114360, 2020.

[26] J. Wu, H. Su, J. Meng, and M. Lin, "State of health estimation for lithium-ion battery via recursive feature elimination on partial charging curves," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 11, pp. 131–142, 2023.

[27] Z. Wei, H. Ruan, Y. Li, J. Li, C. Zhang, and H. He, "Multistage state of health estimation of lithium-ion battery with high tolerance to heavily partial charging," *IEEE Transactions on Power Electronics*, vol. 37, pp. 7432–7442, 2022.

[28] Z. Deng, X. Hu, P. Li, X. Lin, and X. Bian, "Data-driven battery state of health estimation based on random partial charging data," *IEEE Transactions on Power Electronics*, vol. 37, pp. 5021–5031, 2022.

[29] W. Li, N. Sengupta, P. Dechent, D. Howey, A. Annaswamy, and D. U. Sauer, "Online capacity estimation of lithium-ion batteries with deep long short-term memory networks," *Journal of Power Sources*, vol. 482, p. 228863, 2021.

[30] E. Schaltz, D.-I. Stroe, K. Nørregaard, B. Johnsen, and A. Christensen, "Partial charging method for lithium-ion battery state-of-health estimation," in *2019 Fourteenth Interna-*

tional Conference on Ecological Vehicles and Renewable Energies (EVER), Monte-Carlo, Monaco, 2019, pp. 1–5.

[31] E. Braco, I. San Martín, P. Sanchis, A. Ursúa, and D.-I. Stroe, "State of health estimation of second-life lithium-ion batteries under real profile operation," *Applied Energy*, vol. 326, p. 119992, 2022.

[32] J. He, S. Meng, X. Li, and F. Yan, "Partial charging-based health feature extraction and state of health estimation of lithium-ion batteries," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 11, pp. 166–174, 2023.

[33] S. Kohtz, Y. Xu, Z. Zheng, and P. Wang, "Physics-informed machine learning model for battery state of health prognostics using partial charging segments," *Mechanical Systems and Signal Processing*, vol. 172, p. 109002, 2022.

[34] Z. Deng, X. Hu, X. Lin, L. Xu, Y. Che, and L. Hu, "General discharge voltage information enabled health evaluation for lithium-ion batteries," *IEEE/ASME Transactions on Mechatronics*, vol. 26, pp. 1295–1306, 2021.

[35] C. Zhao, P. B. Andersen, C. Træholt, and S. Hashemi, "Data-driven battery health prognosis with partial-discharge information," *Journal of Energy Storage*, vol. 65, p. 107151, 2023.

[36] J. Yang, B. Xia, W. Huang, Y. Fu, and C. Mi, "Online state-of-health estimation for lithium-ion batteries using constant-voltage charging current analysis," *Applied Energy*, vol. 212, pp. 1589–1600, 2018.

[37] S. Park, H. Lee, Z. K. Scott-Nevros, D. Lim, D.-H. Seo, Y. Choi, H. Lim, and D. Kim, "Deep-learning based spatio-temporal generative model on assessing state-of-health for li-ion batteries with partially-cycled profiles," *Materials Horizons*, vol. 10, pp. 1274–1281, 2023.

[38] Z. Zhou, Y. Liu, Z. Zhao, H. Xia, Z. Chen, and Y. Zhang, "Automatic feature extraction-enabled lithium-ion battery capacity estimation using random fragmented charging data," *IEEE Transactions on Transportation Electrification*, vol. 10, pp. 8845–8856, 2024.

[39] Y. Jiang, Y. Chen, F. Yang, and W. Peng, "State of health estimation of lithium-ion battery with automatic feature extraction and self-attention learning mechanism," *Journal of Power Sources*, vol. 556, p. 232466, 2023.

[40] M. Wang, Y. Chen, D. Guo, and Z. Xu, "A micromesh multi-scaled features extraction network for li-ion batteries SOH estimation," *IEEE Transactions on Vehicular Technology*, pp. 1–11, 2025.

[41] L. Tian, Y. Gao, H. Yang, and R. Wang, "Multi-scenario long-term degradation prediction of PEMFC based on generative inference informer model," *Applied Energy*, vol. 377, p. 124398, 2025.

[42] W. Tang, J. Chen, and D. Chen, "Predicting EV battery state of health using long short term degradation feature extraction and FEA TimeMixer," *Scientific Reports*, vol. 15, p. 2200, 2025.

[43] Y. Liu, G. Fan, B. Zhou, S. Chen, Z. Sun, Y. Wang, and X. Zhang, "Rapid and flexible battery capacity estimation using random short-time charging segments based on residual convolutional networks," *Applied Energy*, vol. 351, p. 121925, 2023.

[44] X. Feng, C. Weng, X. He, L. Wang, D. Ren, L. Lu, X. Han, and M. Ouyang, "Incremental capacity analysis on commercial lithium-ion batteries using support vector regression: a parametric study," *Energies*, vol. 11, p. 2323, 2018.

[45] B. Bole, S. Kulkarni, and M. Daigle, "Adaptation of an electrochemistry-based li-ion battery model to account for deterioration observed under randomized use," in *Annual Conference of the PHM Society*, vol. 6, 2014.

[46] X. Feng, J. Li, M. Ouyang, L. Lu, J. Li, and X. He, "Using probability density function to evaluate the state of health of lithium-ion batteries," *Journal of Power Sources*, vol. 232, pp. 209–218, 2013.

[47] R. Qaddoura, W. A. Manaseer, M. A. M. Abushariah, and M. A. Alshraideh, "Dental radiography segmentation using expectation-maximization clustering and grasshopper optimizer," *Multimedia Tools and Applications*, vol. 79, pp. 22 027–22 045, 2020.

[48] D. M. Sundaram and A. Loganathan, "A new supervised clustering framework using multi discriminative parts and expectation–maximization approach for a fine-grained animal breed classification (SC-MPEM)," *Neural Processing Letters*, vol. 52, pp. 727–766, 2020.

[49] B. E. Olivares, M. A. Cerda Muñoz, M. E. Orchard, and J. F. Silva, "Particle-filtering-based prognosis framework for energy storage devices with a statistical characterization of state-of-health regeneration phenomena," *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 2, pp. 364–376, 2013.

[50] B. Epding, B. Rumberg, H. Jahnke, I. Stradtmann, and A. Kwade, "Investigation of significant capacity recovery effects due to long rest periods during high current cyclic aging tests in automotive lithium ion cells and their influence on lifetime," *Journal of Energy Storage*, vol. 22, pp. 249–256, 2019.

[51] M. Lewerenz, J. Münnix, J. Schmalstieg, S. Käbitz, M. Knips, and D. U. Sauer, "Systematic aging of commercial $LiFePO_4$|Graphite cylindrical cells including a theory explaining

rise of capacity during aging," *Journal of Power Sources*, vol. 345, pp. 254–263, 2017.

[52] Q. Li, D. Li, K. Zhao, L. Wang, and K. Wang, "State of health estimation of lithium-ion battery based on improved ant lion optimization and support vector regression," *Journal of Energy Storage*, vol. 50, p. 104215, 2022.

[53] Y. Zhang, Y. Liu, J. Wang, and T. Zhang, "State-of-health estimation for lithium-ion batteries by combining model-based incremental capacity analysis with support vector regression," *Energy*, vol. 239, p. 121986, 2022.

[54] X. Li, C. Yuan, and Z. Wang, "State of health estimation for li-ion battery via partial incremental capacity analysis based on support vector regression," *Energy*, vol. 203, p. 117852, 2020.

[55] T. Liu, D. Li, K. Wang, and Q. Lu, "Estimation of battery capacity degeneration based on an improved neural fuzzy inference system under dynamic operating conditions," *Journal of Energy Storage*, vol. 102, p. 113988, 2024.

[56] M. El-Dalahmeh, M. Al-Greer, M. El-Dalahmeh, and I. Bashir, "Capacity estimation of lithium-ion batteries based on adaptive empirical wavelet transform and long short-term memory neural network," *Journal of Energy Storage*, vol. 70, p. 108046, 2023.

[57] M. Chen, G. Ma, W. Liu, N. Zeng, and X. Luo, "An overview of data-driven battery health estimation technology for battery management system," *Neurocomputing*, vol. 532, pp. 152–169, 2023.

[58] P. Eleftheriadis, M. Gangi, S. Leva, A. V. Rey, E. Groppo, and L. Grande, "Comparative study of machine learning techniques for the state of health estimation of li-ion batteries," *Electric Power Systems Research*, vol. 235, p. 110889, 2024.

[59] J. Zhao, D. Li, Y. Li, D. Shi, J. Nan, and A. F. Burke, "Battery state of health estimation under fast charging via deep transfer learning," *iScience*, vol. 28, p. 112235, 2025.

[60] Z. Zhang, R. Zhang, X. Liu, C. Zhang, G. Sun, Y. Zhou, Z. Yang, X. Liu, S. Chen, X. Dong, P. Jiang, and Z. Sun, "Advanced state-of-health estimation for lithium-ion batteries using multi-feature fusion and KAN-LSTM hybrid model," *Batteries*, vol. 10, p. 433, 2024.

[61] E. Martinez-Laserna, I. Gandiaga, E. Sarasketa-Zabala, J. Badeda, D.-I. Stroe, M. Swierczynski, and A. Goikoetxea, "Battery second life: Hype, hope or reality? A critical review of the state of the art," *Renewable and Sustainable Energy Reviews*, vol. 93, pp. 701–718, 2018.

[62] C. Xu, P. Behrens, P. Gasper, K. Smith, M. Hu, A. Tukker, and B. Steubing, "Electric vehicle batteries alone could satisfy short-term grid storage demand by as early as 2030,"

Nature Communications, vol. 14, p. 119, 2023.

[63] J. Zhu, I. Mathews, D. Ren, W. Li, D. Cogswell, B. Xing, T. Sedlatschek, S. N. R. Kantareddy, M. Yi, T. Gao, Y. Xia, Q. Zhou, T. Wierzbicki, and M. Z. Bazant, "End-of-life or second-life options for retired electric vehicle batteries," *Cell Reports Physical Science*, vol. 2, p. 100537, 2021.

[64] R. R. Richardson, M. A. Osborne, and D. A. Howey, "Battery health prediction under generalized conditions using a Gaussian process transition model," *Journal of Energy Storage*, vol. 23, pp. 320–328, 2019.

[65] L. Ling, S. Yang, and S. Tong, "Lithium-ion battery state of health estimation based on statistical features derived from voltage and temperature probability density functions under realistic randomized cycling conditions," *Journal of Energy Storage*, vol. 137, p. 118496, 2025.

[66] A. Aliberti, F. Boni, A. Perol, M. Zampolli, R. Jaboeuf, P. Tosco, E. Macii, and E. Patti, "Comparative analysis of neural networks techniques for lithium-ion battery SOH estimation," in *Proc. IEEE 46th Annu. Computers, Software, and Applications Conf. (COMPSAC)*, 2022.

[67] H. Lu, J. Chen, J. Zhang, G. He, X. Han, and M. Ouyang, "Multitask battery management with flexible pretraining," arXiv:2509.01323, 2025, flexible Masked Autoencoder (FMAE); verified: corpus excludes NASA Randomized Battery.

[68] R. Tan, W. Hong, J. Li, J. Huang, and T.-Y. Zhang, "Pretrained battery transformer (PBT): A foundation model for universal battery life prediction," arXiv:2512.16334, 2025.

[69] J. Chan, H. Wang, H. Pan, W. Wu, Z. Wang, Z. Chen, E. Pan, M. Xie, and L. Xi, "Universal battery degradation forecasting driven by foundation model across diverse chemistries and conditions," arXiv:2601.00862, 2026.

[70] Y. Che, Y. Zheng, X. Sui, and R. Teodorescu, "Boosting battery state of health estimation based on self-supervised learning," *Journal of Energy Chemistry*, vol. 84, pp. 335–346, 2023.

[71] T. Wang, Z. Ma, S. Zou, Z. Chen, and P. Wang, "Lithium-ion battery state-of-health estimation: A self-supervised framework incorporating weak labels," *Applied Energy*, vol. 355, p. 122332, 2024.

[72] T. Chen, S. Kornblith, K. Swersky, M. Norouzi, and G. Hinton, "Big self-supervised models are strong semi-supervised learners," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, 2020, arXiv:2006.10029.

[73] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 16 000–16 009.

[74] B. Bole, C. S. Kulkarni, and M. Daigle, "Randomized battery usage data set," NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA, USA, 2014.

[75] S. Tao, R. Guo, J. Lee, S. Moura, L. C. Casals, S. Jiang, J. Shi, S. Harris, T. Zhang, C. Y. Chung, G. Zhou, J. Tian, and X. Zhang, "Immediate remaining capacity estimation of heterogeneous second-life lithium-ion batteries via deep generative transfer learning," *Energy & Environmental Science*, vol. 18, pp. 7413–7426, 2025.

[76] T. Rahman and N. Deb, "Deep learning-driven cell-to-vehicle transfer of differential voltage and incremental capacity features for electric vehicle battery state-of-health diagnostics," *Energy Conversion and Management: X*, vol. 30, p. 101678, 2026.

[77] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[78] B. Ospina Agudelo, W. Zamboni, F. Postiglione, and E. Monmasson, "Battery state-of-health estimation based on multiple charge and discharge features," *Energy*, vol. 263, p. 125637, 2023.

[79] Y. Shen, D. Guo, Y. Wang, J. Chen, X. Liu, X. Han, Y. Zheng, and M. Ouyang, "Enhanced few-shot state-of-health estimation for lithium-ion batteries via masked autoencoder," *Energy*, vol. 335, p. 138263, 2025.

[80] Z. Wang, S. Yuan, S. Cai, R. Huang, Y. Liu, M. Wei, P. Zhang, and J. Hu, "Masked self-supervised transformer framework for state-of-health estimation of electric vehicle lithium batteries," *Energy & Fuels*, vol. 40, pp. 4844–4863, 2026.

[81] Z. Yuan, Z. Deng, Y. He, Z. Ning, and J. Liu, "Multi-step prediction of battery state of health based on self-supervised pre-training and transfer learning using the xPatch model," *Energy*, vol. 341, p. 139410, 2025.

[82] J.-C. Chen, "Degradation self-supervised learning for lithium-ion battery health diagnostics," arXiv:2503.08083, 2025.

[83] J. Li, Y. Yang, H. Su, J. Liu, Y. Chen, J. Zhang, and L. Pan, "LiPM: Foundation model for lithium-ion battery analysis," in *Proc. 31st ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, 2025, pp. 1412–1423.

[84] J. Zhao, Z. Wang, Y. Wu, and A. F. Burke, "Predictive pretrained transformer (PPT) for

real-time battery health diagnostics," *Applied Energy*, vol. 377, p. 124746, 2025.

[85] S. Yang, L. Ling, X. Li, J. Han, and S. Tong, "Industrial battery state-of-health estimation with incomplete limited data toward second-life applications," *Journal of Dynamics, Monitoring and Diagnostics*, vol. 3, pp. 246–257, 2024.

[86] J. Lu, R. Xiong, J. Tian, C. Wang, and F. Sun, "Deep learning to estimate lithium-ion battery state of health without additional degradation experiments," *Nature Communications*, vol. 14, p. 2760, 2023.

[87] Y. Feng, G. Hu, and Z. Zhang, "GPT4Battery: An LLM-driven framework for adaptive state of health estimation of raw Li-ion batteries," arXiv:2402.00068, 2024.

[88] K. Liu, Q. Peng, Y. Che, Y. Zheng, K. Li, R. Teodorescu, D. Widanage, and A. Barai, "Transfer learning for battery smarter state estimation and ageing prognostics: Recent progress, challenges, and prospects," *Advances in Applied Energy*, vol. 9, p. 100117, 2023.

[89] B. Bole, C. S. Kulkarni, and M. Daigle, "Adaptation of an electrochemistry-based Li-ion battery model to account for deterioration observed under randomized use," in *Proc. Annu. Conf. Prognostics and Health Management Society (PHM)*, 2014.

[90] M. Berecibar, I. Gandiaga, I. Villarreal, N. Omar, J. Van Mierlo, and P. Van Den Bossche, "Critical review of state of health estimation methods of Li-ion batteries for real applications," *Renewable and Sustainable Energy Reviews*, vol. 56, pp. 572–587, 2016.

[91] A. Devie, G. Baure, and M. Dubarry, "Intrinsic variability in the degradation of a batch of commercial 18650 lithium-ion cells," *Energies*, vol. 11, no. 5, p. 1031, 2018.

[92] S. Lv, X. Wang, W. Lu, J. Zhang, and H. Ni, "The influence of temperature on the capacity of lithium ion batteries with different anodes," *Energies*, vol. 15, no. 1, p. 60, 2022.

[93] S. Tong, H. Yang, and W. Torre, "Energy storage system dispatching optimization in stacked applications for utility grid," in *Proc. Electrical Energy Storage Applications and Technologies (EESAT)*, 2017, mILP dispatch over five CAISO/SDG&E market applications; source of the 7-day aging schedule.

[94] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017, pp. 5998–6008.

[95] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2019.

[96] M. Ilse, J. M. Tomczak, and M. Welling, "Attention-based deep multiple instance learning,"

in *Proc. 35th Int. Conf. Machine Learning (ICML)*, ser. PMLR, vol. 80, 2018, pp. 2127–2136.

[97] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

[98] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, "A time series is worth 64 words: Long-term forecasting with transformers," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2023.

[99] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proc. 25th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, 2019, pp. 2623–2631.

[100] T. Isazawa and J. M. Cole, "How beneficial is pretraining on a narrow domain-specific corpus for information extraction about photocatalytic water splitting?" *Journal of Chemical Information and Modeling*, vol. 64, no. 8, pp. 3205–3212, 2024.

[101] V. Ramanujan, T. Nguyen, S. Oh, L. Schmidt, and A. Farhadi, "On the connection between pre-training data diversity and fine-tuning robustness," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2023, arXiv:2307.12532.

[102] H. A. A. K. Hammoud, T. Das, F. Pizzati, P. H. S. Torr, A. Bibi, and B. Ghanem, "On pretraining data diversity for self-supervised learning," in *Proc. European Conf. Computer Vision (ECCV)*, ser. Lecture Notes in Computer Science, vol. 15114, 2024, arXiv:2403.13808.

[103] R. Karandeh, T. Lawanson, and V. Cecchi, "A multi-application energy management of behind-the-meter battery energy storage system," in *Proceedings of the 2020 52nd North American Power Symposium*, Tempe, AZ, USA, 2021, pp. 1–6.

[104] A. Grimaldi, F. D. Minuto, J. Brouwer, and A. Lanzini, "Profitability of energy arbitrage net profit for grid-scale battery energy storage considering dynamic efficiency and degradation using a linear, mixed-integer linear, and mixed-integer non-linear optimization approach," *Journal of Energy Storage*, vol. 95, p. 112380, 2024.

[105] D. Yang, C. He, T. Wang, and M. Zhang, "Optimal economic dispatch of microgrid with battery energy storage considering high penetration of renewable generation," in *Proceedings of the 2022 IEEE/IAS Industrial and Commercial Power System Asia*, Shanghai, China, 2022, pp. 353–358.

[106] J. M. Reniers, G. Mulder, S. Ober-Blöbaum, and D. A. Howey, "Improving optimal control of grid-connected lithium-ion batteries through more accurate battery and degradation

modelling," *Journal of Power Sources*, vol. 379, pp. 91–102, 2018.

[107] Y. Chen, K. Zheng, Y. Gu, J. Wang, and Q. Chen, "Optimal energy dispatch of grid-connected electric vehicle considering lithium battery electrochemical model," *IEEE Transactions on Smart Grid*, vol. 15, no. 3, pp. 3000–3015, May 2024.

[108] J. Gonzalez-Saenz and V. Becerra, "Optimal battery energy storage dispatch for the day-ahead electricity market," *Batteries*, vol. 10, no. 7, p. 228, 2024.

[109] M. Tekin and M. İ. Karamangil, "Comparative analysis of equivalent circuit battery models for electric vehicle battery management systems," *Journal of Energy Storage*, vol. 86, Part B, p. 111327, 2024.

[110] F. Zheng, Y. Xing, J. Jiang, B. Sun, J. Kim, and M. Pecht, "Influence of different open circuit voltage tests on state of charge online estimation for lithium-ion batteries," *Applied Energy*, vol. 183, pp. 513–525, 2016.

[111] S. Tong, H. Yang, and W. Torre, "Energy storage system dispatching optimization in stacked applications for utility grid," in *2017 Electrical Energy Storage Applications and Technologies Conference*, 2018, [Online]. Available: https://api.semanticscholar.org/CorpusID:198977269.

[112] C. Lange, A. Rueß, A. Nuß, R. Öchsner, and M. März, "Dimensioning battery energy storage systems for peak shaving based on a real-time control algorithm," *Applied Energy*, vol. 280, p. 115993, 2020.

[113] S. Upadhya and M. J. Wagner, "A dispatch optimization model for hybrid renewable and battery systems incorporating a battery degradation model," *ASME Journal of Energy Resources Technology*, vol. 144, no. 7, p. 070907, Jul. 2022.