

THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Data-driven battery aging diagnostics and lifetime extension

YIZHOU ZHANG



CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Electrical Engineering
Chalmers University of Technology
Gothenburg, Sweden, 2024

Data-driven battery aging diagnostics and lifetime extension

YIZHOU ZHANG

ISBN 978-91-8103-120-1

To my family

Copyright © YIZHOU ZHANG 2024

All rights reserved.

Doktorsavhandlingar vid Chalmers tekniska högskola

Ny serie nr 5578

ISSN 0346-718X

This thesis has been prepared using L^AT_EX.

Department of Electrical Engineering

Chalmers University of Technology

SE-412 96 Gothenburg, Sweden

Phone: +46 (0)31 772 1000

www.chalmers.se

Printed by Chalmers Reproservice

Gothenburg, Sweden, Nov2024

To my family.

Data-driven battery aging diagnostics and lifetime extension
For automotive applications
Yizhou Zhang
Department of Electrical Engineering
Chalmers University of Technology

Abstract

Transportation electrification is critical to mitigating climate change, with lithium-ion (Li-ion) batteries playing a pivotal role in the shift to low-carbon energy sources. Given that batteries can account for up to 50% of an electric vehicle's cost, optimizing their lifespan and performance is critical for cost-effective operation. Batteries though, degrade in ways that are inhomogeneous, nonlinear, and dependent on multiple factors. This makes accurate aging diagnostics and prognostics essential for ensuring their safe, efficient use. Diverse operating conditions, complex aging mechanisms, unpredictable usage profiles, and cell-to-cell variations pose significant challenges. At the same time, battery performance, including energy and power, is influenced not only by health state but also by conditions such as temperature, State of Charge (SoC), and applied current.

This thesis presents a series of machine learning (ML) frameworks developed using field data from vehicles and laboratory cycling data. One proposed framework is a battery capacity estimation algorithm that integrates multiple ML models with a Kalman filter, accommodating the diverse usage profiles of electric vehicles (EVs) in real-world scenarios. To reduce warranty costs, a histogram-based usage-related ML framework is developed, combining offline global models with online cell-specific models to track and predict future aging. Additionally, a remaining useful life (RUL) prediction model improves accuracy by combining usage and time-series data is developed as well.

Beyond aging diagnostics, the thesis proposes a method to extract relationships between battery performance indicators (PIs) and various influencing factors like temperature, SoC, and aging, using a neural network-based

framework. Lastly, it introduces an online method to estimate battery plating potential, enabling faster charging while minimizing lithium plating risks to extend the lifetime of the battery. Collectively, these contributions provide practical tools for diagnostics, prognostics, and control, advancing safer, more efficient, and cost-effective use of Li-ion batteries in EVs.

Keywords: Lithium-ion batteries, battery management system, fast charging, state of health, remaining useful life, machine learning, neural network.

List of Publications

This thesis is based on the following publications:

[A] **Yizhou Zhang**, Torsten Wik, John Bergström, Changfu Zou, “State of health estimation for lithium-ion batteries under arbitrary usage using data-driven multi-model fusion”. Published in *IEEE Transactions on Transportation Electrification*, vol. 10, pp. 1494-1507, Apr. 2023.

[B] **Yizhou Zhang**, Torsten Wik, John Bergström, Michael Pecht, Changfu Zou, “A machine learning-based framework for online prediction of battery ageing trajectory and lifetime using histogram data”. Published in *Journal of Power Sources*, Apr. 526, p. 231110, Apr. 2022.

[C] **Yizhou Zhang**, Torsten Wik, Yicun Huang, John Bergström, Changfu Zou, “Data-driven battery life prediction considering both onsite measurement and usage information”. 22nd IFAC World Congress 2023, July 2023, Yokohama, Japan.

[D] **Yizhou Zhang**, Torsten Wik, John Bergström, Changfu Zou, “Machine learning-based lifelong estimation of lithium plating potential: A path to health-aware fastest battery charging”. Accepted for publication in *Energy Storage Materials*.

[E] **Yizhou Zhang**, Torsten Wik, John Bergström, Shafiq Urréhman, Changfu Zou, “Harmonizing performance indicators and unifying data for lifelong battery management”. Under review.

Other publications by the author, not included in this thesis, are:

[F] **Yizhou Zhang**, Torsten Wik, John Bergström, Changfu Zou, “Practical battery State of Health estimation using data-driven multi-model fusion”. 22nd IFAC World Congress 2023, July 2023, Yokohama, Japan.

[G] **Yizhou Zhang**, Torsten Wik, Changfu Zou, “A method for estimation state of health of a battery”. Patent application filed with European patent office with application No. EP22216134.1.

[H] **Yizhou Zhang**, Torsten Wik, Changfu Zou, “A method for real-time estimation of battery anode potential”. Patent application filed with European patent office with application No. EP24170319.8.

Acknowledgments

When I graduated with my master's from KTH in 2016, I was certain it would be the last time I walked into a university as a student. But life has a way of surprising you. In 2020, I found myself back in academia, starting a journey I never imagined—as an industrial PhD student. Now, in 2024, I once again believe that this is the end of my university days, ready to embrace the industry. But who knows if life has another surprise in store for me?

Looking back on my time as a PhD student, of course, there were moments of pressure, anxiety, and even frustration—especially after frustrating reviewer rejections. Overall, I have truly enjoyed the ride, and the whole journey has been much smoother than I could have hoped for, thanks to the incredible people I have had the pleasure of meeting and working with along the way.

First, my deepest gratitude goes to my supervisor, Prof. Changfu Zou, for his unwavering guidance, support, and mentorship. I have learned so much from our countless hours revising manuscripts together—the attention to every tiny word, notation, and detail in pursuit of perfection. Your dedication has helped me become a more rigorous researcher, constantly pushing me to aim higher, whether in research quality or aiming for more prestigious journals. Equally, I want to thank my other supervisor, Prof. Torsten Wik. Not only are you an excellent supervisor and researcher, but you are also a fantastic person. I have really enjoyed our discussions over the years—whether about research or just life beyond academia. I admire the way you treat your students as a central part of your career, always showing care and support. I would also like to express my gratitude to my industrial supervisors, Dr. John Bergström and Dr. Shafiq Urréhman, for providing invaluable insights from an industry perspective and helping to navigate the administrative side of things so I could focus on my research. Finally, a big thank you to my past and present team managers—Sven Sjöberg, Erik Preihs, Lars Nilsson, and Mark Saukkonen—for their support and guidance. Your encouragement has been invaluable. I also want to acknowledge all my colleagues at Zeekr Technology Europe and Chalmers for fostering an enjoyable, positive, and inspiring working environment.

To my wonderful son Aiwen, who is now three years old and has accompanied me through almost the entire PhD journey: your smile means the world to me, and you are the greatest gift I have ever received. Thank you for filling my life with so much joy. Sincerely, I could not have achieved all of this with-

out the full support at home from my beloved wife, Chenwen, who took on so many responsibilities during the countless hours and days that I was away. I also want to thank my parents for their unwavering trust and support, and I express my gratitude to all my friends and family—your support has meant so much to me. I am so grateful to have you all in my life.

Acronyms

GHG:	Greenhouse gas
Li-ion:	Lithium-ion
SaaS:	Software as a service
BMS:	Battery management system
SoH:	State of health
RUL:	Remaining useful life
EV:	Electric vehicle
NMC:	Nickel-manganese-cobalt
NCA:	Nickel-cobalt-aluminum
LCO:	Lithium cobalt oxide
LFP:	Lithium iron phosphate
SEI:	Solid electrolyte interphase
SoC:	State of charge
SoP:	State of power
SoE:	State of energy
BOL:	Beginning of life
EOL:	End of life
ML:	Machine learning
BRR:	Bayesian ridge regression
SVR:	Support vector regression
RFR:	Random forest regression

GPR:	Gaussian process regression
NN:	Neural network
RBF:	Radial basis function
ReLU:	Rectified linear unit function
ECM:	Equivalent circuit model
EM:	Electrochemical model
LSTM:	Long short-term memory
RPT:	Reference performance test

Contents

Abstract	i
List of Papers	iii
Acknowledgements	v
Acronyms	vii
I Overview	1
1 Introduction	3
1.1 Introduction	3
1.2 Research Questions	6
1.3 Contributions	6
SoH estimation	6
Aging trajectory prediction	7
Remaining useful life (RUL) prediction	7
Estimation of battery performance indicators	8
Health-conscious control	8
1.4 Thesis outline	8

2	Battery systems	9
2.1	Working principles	9
2.2	Battery aging	10
2.3	Battery states	12
2.4	Battery management system (BMS)	14
3	Machine learning	17
3.1	ML algorithms	18
	Bayesian ridge regression (BRR)	18
	Support vector regression (SVR)	18
	Random forest regression (RFR)	20
	Gaussian process regression (GPR)	20
	Neural network (NN)	22
4	Battery aging diagnostics	25
5	Battery aging prognostics	29
6	Battery fast charging optimization	33
7	Summary of included papers	37
7.1	Paper A	37
7.2	Paper B	38
7.3	Paper C	38
7.4	Paper D	39
7.5	Paper E	39
8	Concluding remarks and future work	41
	References	45
II	Papers	59
A	SoH estimation	A1
1	Introduction	A3
2	Dataset introduction	A7

3	Feature construction and engineering	A8
3.1	Charging scenarios	A9
3.2	Feature construction	A11
3.3	Feature engineering	A13
4	Estimator design	A13
4.1	Time-series data-based model for capacity estimation . .	A14
4.2	Histogram data-based model for capacity prediction . .	A16
4.3	Optimal model fusion	A18
5	Results and discussion	A19
5.1	Hyperparameter values	A19
5.2	Result of feature engineering	A20
5.3	SoH estimation results under lab tests	A22
5.4	SoH estimation results under vehicle usage	A26
5.5	SoH estimation results for NCA batteries	A29
6	Conclusion	A31
	References	A31

B Aging trajectory prediction

B1

1	Introduction	B3
2	Dataset	B6
2.1	Data collection and processing	B7
3	Methods	B10
3.1	Histogram data-based feature construction	B10
3.2	Feature engineering	B11
3.3	Global model and algorithm development	B13
3.4	Individualised model and algorithm development	B18
3.5	Evaluation metrics for model fidelity	B20
3.6	Machine learning software and libraries	B22
4	Results and discussion	B22
4.1	Results of feature engineering	B22
4.2	Results of battery ageing prognosis	B23
4.3	Requirements for computation and memory	B27
5	Conclusions	B29
6	Supplementary materials	B29
	References	B29

C	RUL prediction	C1
1	Introduction	C3
2	Battery dataset introduction	C5
3	Feature construction and engineering	C5
	3.1 Usage-related features	C5
	3.2 Time-series, measurement-related features	C7
4	Early prediction model	C7
5	Evaluation Matrices	C9
6	Results and discussions	C10
7	Conclusion	C16
	References	C16
D	Plating potential estimation	D1
1	Introduction	D3
2	Results and discussion	D7
	2.1 Three-electrode cell experiments	D7
	2.2 Synthetic data generation	D9
	2.3 Lifelong estimation framework for plating potential	D10
	2.4 Results of the three-electrode cell test	D12
	2.5 Estimation results for SoH indicators	D14
	2.6 Estimation results for lifelong plating potential	D16
	2.7 Computational efficiency	D19
	2.8 Discussion and outlook of future research	D20
3	Conclusions	D21
4	Experimental procedures	D21
	4.1 Electrochemical battery modeling and simulation	D21
	4.2 ML-based battery modeling and implementation	D23
	4.3 Evaluation matrices	D26
	4.4 Resource availability	D27
	4.5 Supplementary materials	D27
	References	D27
E	Field analysis	E1
1	Introduction	E3
2	Description and processing of field vehicle data	E7
3	Results and discussions	E10
	3.1 Data-driven battery aging diagnostics	E10

3.2	Resistance during acceleration and regenerative braking	E11
3.3	Capacity during vehicle charging	E18
3.4	Fleet level evaluation	E20
4	Conclusion	E22
5	Method	E23
5.1	Resource availability	E23
5.2	Data processing	E23
5.3	Real-time battery resistance estimation	E24
5.4	Real-time battery available capacity estimation	E25
5.5	Estimating battery internal parameter aging dependency using NN	E26
	References	E29

Part I

Overview

1.1 Introduction

Global warming and climate change represent some of the most pressing challenges of the 21st century, necessitating coordinated global efforts to mitigate their impacts[1]–[3]. Among the various sectors contributing to these environmental issues, the transportation sector stands out as a significant contributor to greenhouse gas emissions. As of 2022, this sector was responsible for 23% of global direct carbon dioxide (CO₂) emissions [4], with road vehicles alone accounting for 79% of these emissions as illustrated in Fig. 1.1. This trend underscores the urgent need for innovative solutions to reduce the environmental footprint of transportation, making it a critical area of focus in the fight against climate change [5], [6]. Consequently, improving the efficiency and sustainability of transportation technologies has become increasingly important.

The shift towards electromobility, powered by low-carbon electricity, is vital for reducing greenhouse gas emissions from transportation [8]. The electrification of the automotive industry, therefore, is not just beneficial but essential for meeting sustainability goals. Simultaneously, the industry is undergo-

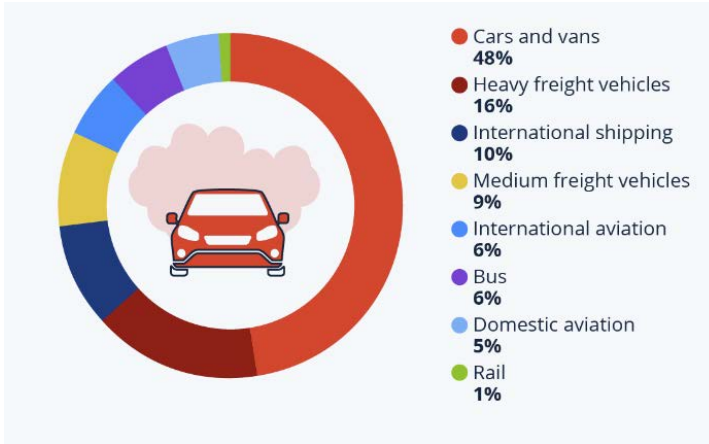


Figure 1.1: Estimated share of CO₂ emissions in the transportation sector worldwide by transport type (figure modified from [7]).

ing a profound transformation driven by advances in in-vehicle software, the rise of software-as-a-service (SaaS) business models, and the emergence of software-defined vehicles [9]. This evolution signifies a transition from producing mechanically-focused products to delivering software-centric mobility services.

Central to the transition to electrified transportation are lithium-ion (Li-ion) batteries, valued for their high energy density, cost-effectiveness, and relatively long lifespan [10], [11]. Despite their advantages, Li-ion batteries, being electrochemical devices, present complex and nonlinear aging characteristics [12], [13]. Their degradation is influenced by intrinsic factors such as pack design and manufacturing variability, as well as extrinsic factors like temperature and usage profiles [14], [15]. This complexity poses significant challenges for accurately estimating and predicting battery aging outside controlled laboratory environments. While understanding the aging process is critical, the ability to leverage this knowledge to extend battery lifespan is even more valuable. Developing strategies that not only diagnose but also mitigate aging effects can significantly enhance battery longevity and performance [16].

To ensure the safe and optimal use of these batteries, a robust battery management system (BMS) with accurate estimation and control algorithms

is crucial [17]. The growing focus on vehicle software functionality has made it feasible for automotive companies to implement advanced algorithms based on both model-based and data-driven methods. Model-based algorithms, particularly those using equivalent circuit models to represent battery dynamics, have been widely adopted in commercial products and have demonstrated promising results in tasks such as State of Charge (SoC) and impedance estimation [18]–[20]. However, these model-based methods tend to lose accuracy when operating conditions change or when the battery operates outside the calibrated parameter windows. This issue becomes more pronounced as the battery ages [21], [22]. In response, data-driven methods have emerged as a promising alternative for tracking battery aging and supporting continuous, adaptive battery control and optimization throughout the battery’s lifespan.

Amid increasing regulatory pressures and the growing emphasis on data, automakers are collecting operational usage data, including battery data, either on-board or via cloud systems [23]. However, managing and analyzing large volumes of data poses significant technical challenges, particularly in maintaining data quality and relevance. Storing all available data can lead to information overload, making it difficult to identify critical patterns and trends in battery aging and performance. Additionally, selective data collection is essential for real-time processing and decision-making, enabling automakers to efficiently perform aging diagnostics and usage optimization while minimizing computational load and latency. These considerations raise key questions about how to intelligently select and store only the most relevant data to ensure effective diagnostics and predictive insights without unnecessary complexity.

Furthermore, data collected from real-world applications often face inherent limitations, such as interrupted measurements and noisy signals, which fall short of the quality and consistency of lab-generated data. As a result, the robustness of battery aging diagnostic and prognostic algorithms is essential for effectively managing these real-world data challenges. Additionally, the vehicle development process, which includes extensive testing and verification before new models are launched, typically involves lengthy battery cycling campaigns. These campaigns generate substantial lab data that, when effectively integrated with real-world data, can significantly enhance the performance of aging estimation algorithms and improve the efficiency of usage optimization.

This research focuses on applying advanced data-driven modeling techniques to analyze field data collected during vehicle operation in conjunction with laboratory cycling data. The objective is to enhance battery aging diagnostics and prognostics while developing innovative control strategies that optimize battery usage, extend battery lifetime, and improve capacity utilization. The findings from this research contribute to the field of battery management systems, providing practical solutions for extending the operational lifespan of electric vehicle batteries.

1.2 Research Questions

- How can field data collected onboard vehicles be effectively used to robustly estimate battery capacity?
- How can large fleet data be efficiently leveraged to predict battery aging trajectory while adapting to individual vehicles?
- How can time-series measurement data and histogram usage data be combined to achieve more accurate and robust battery diagnostics and prognostics?
- How can key battery performance indicators, such as internal resistance and capacity, be effectively extracted directly from vehicle field data?
- How can insights from battery aging diagnostics be applied to optimize fast-charging strategies, enhance charging performance, and extend battery lifespan?

1.3 Contributions

Fig. 1.2 gives an overview of the problems that are investigated in this thesis. The contribution in each illustrated part can be summarized as follows, with more details introduced in subsequent chapters.

SoH estimation

Significant efforts have been made by both academia and industry to improve State of Health (SoH) estimation. However, relatively few methods

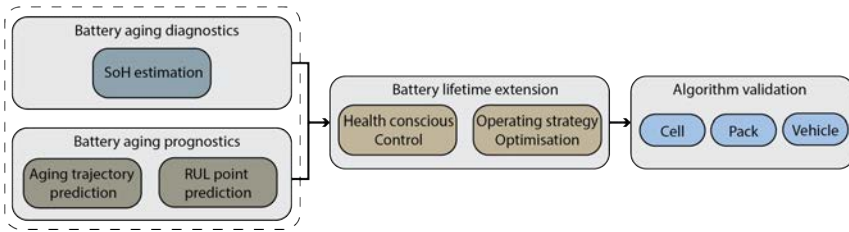


Figure 1.2: The overview of the thesis contribution.

have demonstrated the accuracy and robustness necessary for real-world applications. This thesis contributes to this area, as detailed in Paper A, by developing a practical battery capacity estimation method that works under various usage profiles for automotive applications, utilizing a data-driven multi-model fusion approach.

Aging trajectory prediction

Compared to SoH estimation, there is less literature available on aging trajectory prediction methods, despite their significant benefits, such as shortening the development cycle, enabling timely preventative maintenance, and reducing warranty and insurance costs. Paper B addresses this gap by proposing a systematic machine learning (ML) framework that uses histogram-based features to predict future battery aging trajectories.

Remaining useful life (RUL) prediction

Battery Remaining Useful Life (RUL) is a critical attribute from both technical and economic perspectives. Traditionally, time-series measurement data have been used for RUL prediction through machine learning methods. However, as shown in Paper C, histogram usage data can achieve comparable performance with much less data required. This work extends the study of aging trajectory prediction to RUL prediction, providing a comprehensive comparison of feature construction using both time-series data and usage-related histogram information.

Estimation of battery performance indicators

Battery performance is influenced not only by aging but also by operating conditions such as temperature, SoC levels, and current levels. Understanding and extracting these relationships is crucial for better battery control and utilization. In Paper E, a Neural Network (NN)-based machine learning framework is proposed to systematically extract battery performance indicators directly from real-world vehicle field data.

Health-conscious control

Unlike the stochastic and dynamic battery discharging profiles typical in electric vehicle (EV) applications, charging is more controllable, allowing engineers to predefine current levels and charging profiles. One significant aging mechanism, lithium plating, which negatively affects battery longevity, is more likely to occur during charging. To address this, Paper D presents an online method for lifelong estimation of battery plating potential, enabling faster charging while minimizing the risk of lithium plating.

1.4 Thesis outline

This thesis is structured into two parts. Part I provides the context and background for the research papers included in Part II. Part I consists of eight chapters. Chapter 1 briefly introduces the research project and describes the motivation of the work. Chapter 2 presents the battery system and its important components. This is followed by Chapter 3, which gives an overview of the used machine learning algorithms. Chapter 4 presents the methods for battery aging diagnostics, and Chapter 5 introduces the battery aging prognostics methods. Chapter 6 discusses the battery fast charging optimization. Chapter 7 summarizes the included papers. Lastly, Chapter 8 concludes the thesis and also highlights potential future research directions.

CHAPTER 2

Battery systems

Li-ion batteries have become the power source of choice for a wide range of applications, including portable electronic devices, uninterruptible power supplies, and energy storage systems. Moreover, they have emerged as the dominant battery technology for EVs due to their cost effectiveness, high energy density, relatively long cycle life, and low self-discharge rates. Given their widespread adoption, understanding the working principles and aging characteristics of Li-ion batteries is crucial for optimizing their performance and extending their lifespan. This chapter provides an overview of battery systems, with a particular emphasis on the factors influencing battery aging and the implications for long-term performance.

2.1 Working principles

Li-ion battery is composed of several key components: a negative electrode, a positive electrode, a separator, electrolyte, and two current collectors, as illustrated in Fig. 2.1. The negative electrode typically consists of graphite, or more recently, graphite enhanced with silica, while the positive electrode is generally made of a metal oxide, such as nickel-manganese-cobalt (NMC),

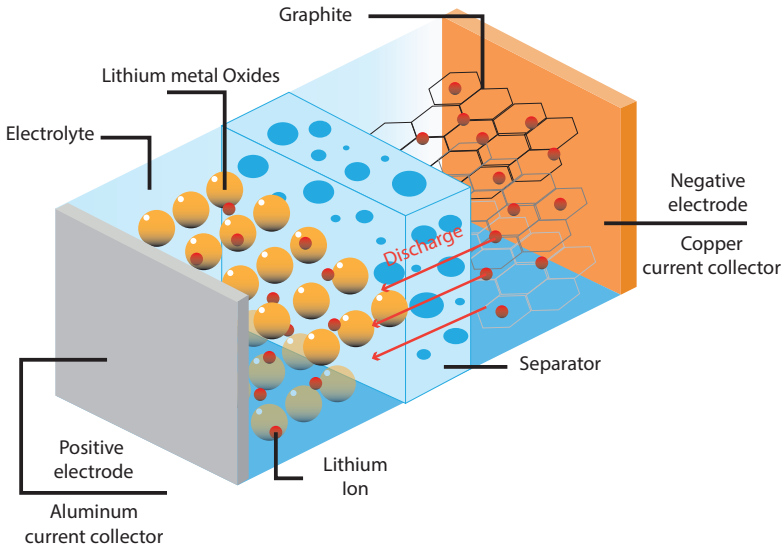


Figure 2.1: The overview of the working principle of the Li-ion battery.

nickel-cobalt-aluminum (NCA), or lithium cobalt oxide (LCO), or a phosphate-based material like lithium iron phosphate (LFP). The separator is a crucial component that permits ionic conductivity while ensuring electrical insulation between the electrodes.

During discharge, lithium ions move from the negative electrode to the positive electrode through the electrolyte, while electrons flow through an external circuit, generating electrical current. When the battery is charging, this process is reversed: lithium ions migrate back to the negative electrode, and the flow of electrons is directed in the opposite direction, restoring the battery's charge [17], [18].

2.2 Battery aging

As discussed in Section 1.1, battery performance gradually deteriorates over time due to a combination of internal factors, such as changes in cell mechanical and electrochemical properties, and external factors, such as operating

conditions [15]. Fig. 2.2 illustrates some of the common aging mechanisms that can occur within batteries. Among these, the formation of the solid electrolyte interphase (SEI) and lithium plating are considered by many researchers to be the most significant contributors to battery aging [12], [24], [25].

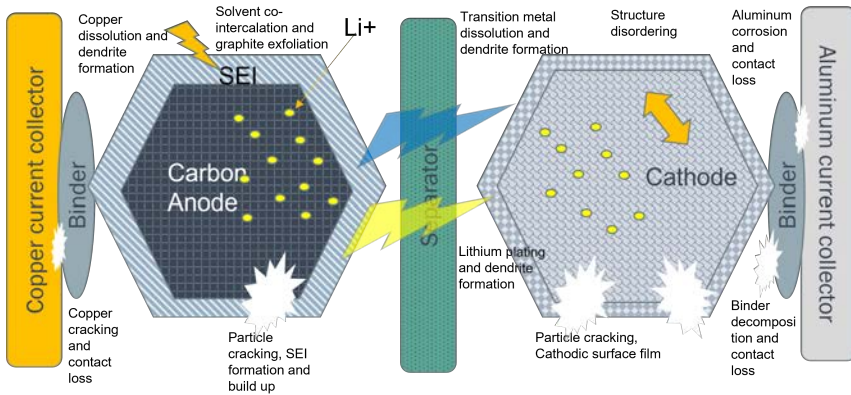


Figure 2.2: Different aging mechanisms that can potentially happen during the usage of the battery. Figure modified from [13].

The SEI layer is a passivation layer that forms on the surface of active electrode materials. Although the SEI can develop on both the positive and negative electrodes, it is more prevalent on the negative side. When the cell voltage exceeds the electrochemical stability window of the liquid electrolyte, reactions occur between the electrolyte, lithium ions, and electrons within the electrodes, leading to the formation of a solid electrolyte [24], [26]. The SEI layer begins to form during the first charge cycle, causing an initial capacity loss of roughly 10%. However, this layer serves to limit further reactions between the electrolyte and the electrodes [27].

Over time, the SEI layer continues to grow as the battery is used. New SEI formation can be triggered by factors such as newly exposed electrode surfaces after particle cracking, plated lithium, or the presence of transition metals [28]. This ongoing SEI growth leads to a gradual loss of capacity and an irreversible increase in resistance, further degrading the battery's performance [29]. The rate of SEI layer growth is strongly influenced by operating conditions, with higher temperatures during storage or cycling, accelerating

diffusion rates and potentially increasing SEI formation. Similarly, large SoC swings or higher currents can lead to more particle cracking, contributing to further SEI development [30], [31].

Lithium plating is another critical aging mechanism where lithium ions, instead of intercalating into the negative electrode, are deposited on its surface. This can lead to a reduction in battery capacity and, in severe cases, pose safety risks if the lithium penetrates the separator [32]–[34]. Lithium plating is particularly influenced by charging conditions, with factors such as high charging currents or low temperatures increasing the electrolyte potential, which favors lithium deposition over the normal intercalation process [35].

In addition to SEI formation and lithium plating, other aging processes may occur simultaneously, such as electrode structural changes, particle fractures, and electrolyte drying. These mechanisms often interact with each other, further complicating the already complex issue of battery degradation [12], [25], [36]. Understanding these interrelated processes is crucial for developing strategies to mitigate battery aging and extend the useful life of Li-ion batteries.

2.3 Battery states

Accurate estimation and prediction of a battery's states are crucial for ensuring its safe and optimal operation. In Li-ion batteries, understanding and managing these states is essential not only for maintaining performance but also for extending battery life and preventing potential safety hazards. Fig. 2.3 illustrates the flow of information related to different states of a battery.

The SoC represents the ratio between the currently available battery capacity and its capacity when fully charged. It essentially acts as a fuel gauge for electric vehicles, indicating how much charge remains. Extensive research has been dedicated to SoC estimation, with methods such as the Kalman filter, incremental capacity analysis, and machine learning being widely explored. Some of these methods have been successfully adopted by the industry and implemented in commercial products. For a comprehensive review of various SoC estimation techniques, readers are referred to [37]–[39].

Battery resistance is a critical parameter that influences the estimation of several key states, including SoC, State of Power (SoP), and SoH. Resistance varies with the battery's aging state and operating conditions, such as current,

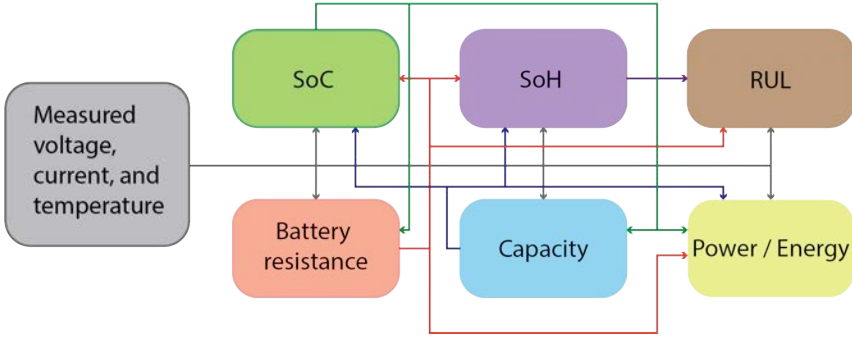


Figure 2.3: Information flow of different battery internal state estimations.

SoC, and temperature. Estimating resistance can be achieved through techniques like electrochemical impedance spectroscopy, equivalent circuit models combined with Kalman filters, or electrochemical models using observers [39], [40].

The battery capacity refers to the battery’s ability to store and deliver charge at a given time. As the battery degrades, its capacity typically diminishes, leading to a reduced driving range and diminished vehicle performance. Therefore, accurate capacity estimation is essential. Both resistance and capacity represent one perspective of the battery SoH, which can be expressed as:

$$\text{SoH}_R = \frac{R}{R_{nom}} \times 100\%, \quad (2.1)$$

$$\text{SoH}_C = \frac{C}{C_{nom}} \times 100\%, \quad (2.2)$$

where R and C represent the current battery resistance and capacity, respectively, and R_{nom} and C_{nom} are the nominal battery resistance and capacity at the beginning of life (BOL) [41].

As batteries age, it becomes increasingly important to estimate how long or how many cycles the battery can continue to function before reaching its End of Life (EOL). This attribute, known as the RUL, is vital for predictive maintenance, maximizing the residual value of the battery, and reducing warranty claims [19], [42].

Battery state of power/energy (SoP/SoE) are two additional critical states

that impact both the drivability and longevity of electric vehicles. These states must be continuously monitored and estimated. SoP predicts the maximum available charge and discharge power of the battery without causing premature aging or compromising safety. SoE, on the other hand, indicates the amount of energy currently stored in the battery [39].

2.4 Battery management system (BMS)

An automotive BMS plays a crucial role in ensuring the safe and efficient operation of a vehicle's battery. Typically, all the battery's internal states, as discussed in the previous section, are calculated by a microcontroller onboard the vehicle, often referred to as the master of the BMS. This microcontroller not only handles core estimation tasks but also continuously monitors critical parameters such as cell voltage, current, and temperature. Additionally, the BMS is responsible for executing various protection functions, including controlling the battery disconnect unit, conducting isolation checks, and performing interlock checks, all of which are essential to maintaining safe operation and preventing hazardous situations [43]–[45]. Fig.2.4 provides an overview of the typical functions performed by a BMS.

Traditionally, the automotive industry has favored low-end microcontrollers for BMS applications due to cost constraints. These microcontrollers, however, come with limited computational power and memory, which restricts the implementation of advanced algorithms that require significant processing resources. Despite these limitations, the advancement of telecommunications technology and data storage capabilities is paving the way for next-generation cloud-based BMS solutions. These systems are becoming not only feasible but also potentially more economical [46]–[48]. Fig. 2.5 illustrates the structure of a cloud-based BMS.

In a cloud-based BMS, many data-intensive and computationally demanding algorithms can be executed in the cloud, with only the final results being sent back to the vehicle. This approach alleviates the processing burden on the vehicle's onboard systems. Moreover, the increasing recognition of the value of vehicle data has led to the logging and storage of vast amounts of operational information in data centers. To fully leverage this valuable resource, state-of-the-art data-driven methods are essential for more accurately assessing the battery's internal states and predicting future performance trends.

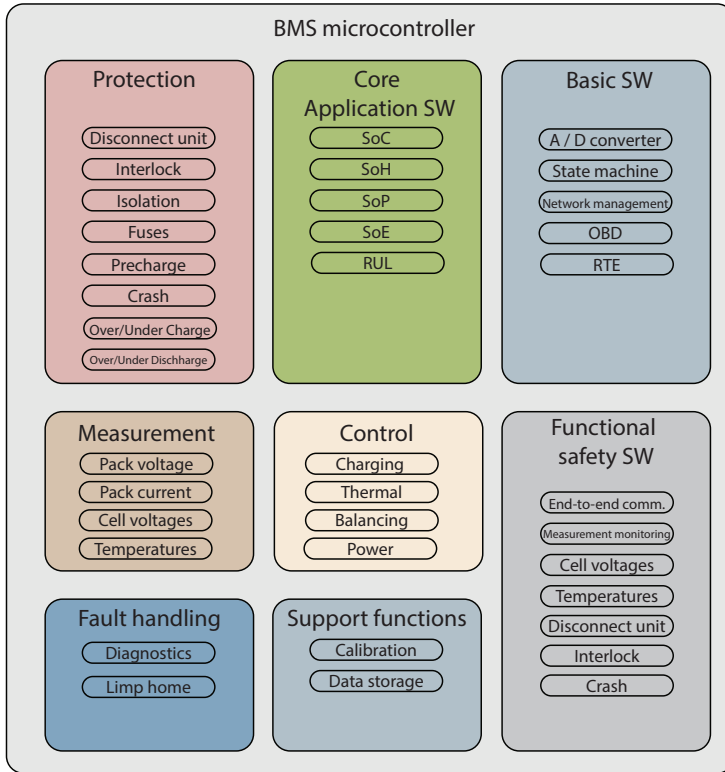


Figure 2.4: Overview of the functions of a typical BMS.

This shift towards cloud-based BMS not only enhances the capabilities of battery management but also represents a significant step forward in the ongoing evolution of electric vehicle technology, enabling more sophisticated monitoring, diagnostics, and optimization of battery systems.

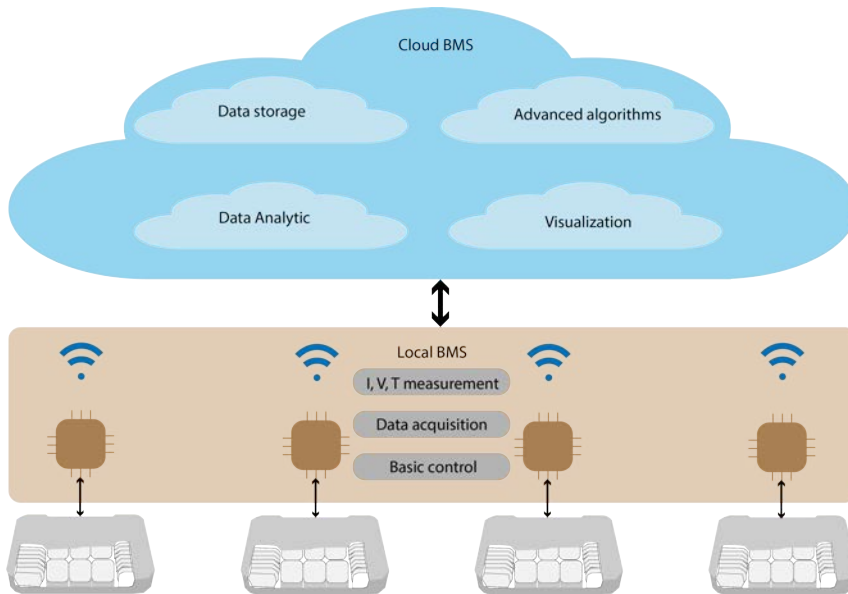


Figure 2.5: Illustration of a cloud-based BMS.

CHAPTER 3

Machine learning

Machine learning (ML) is a collection of methods that leverage data to enable learning, reasoning, and decision-making processes [49]. Through model training, ML algorithms develop models based on sampled data, which are then used to make predictions or execute actions. Generally, ML can be categorized into three types: supervised learning, unsupervised learning, and reinforcement learning.

Supervised learning involves training a model using labeled input-output data, where the desired outputs are predefined by domain experts. The primary objective in supervised learning is for the model to replicate the labeling process performed by these experts. Within the supervised learning framework, problems typically fall into two main categories: regression and classification [50], [51]. Regression focuses on estimating the relationship between one or more input features and a desired continuous output, while classification is concerned with identifying the category to which an observation belongs.

In this thesis, the focus is on battery aging diagnostics, prognostics, and internal state estimation, which are generally approached as supervised learning problems. Unless otherwise specified, the machine learning algorithms discussed in subsequent chapters pertain to the regression category.

3.1 ML algorithms

Among the large pool of ML algorithms, some are particularly interesting for solving battery aging diagnostics and prognostics tasks due to their characteristics. Here, several typical such algorithms, applied later in the appended papers, are briefly introduced.

Bayesian ridge regression (BRR)

Fitting a straight line or a hyperplane to data is probably the most straightforward way to form a linear model. The simple and neat model construction makes the computation extremely simple and the model output easy to interpret. The contracted model, usually called linear regression (for the scalar case a line) has the form $y = \theta^T \cdot x_p + c$, where θ and c are the model parameters that need to be found, and x_p is the training data. Applying a Bayesian approach to a linear regression model leads to the so-called BRR. Instead of treating the coefficient θ as a single variable, BRR assumes θ to be of a spherical Gaussian distribution, defined as $P(\theta) = \mathcal{N}(\theta; 0, \Sigma_0)$, having zero mean and a covariance Σ_0 . To simplify the calculation, Σ_0 is chosen to be $\mathcal{I}\alpha$, where \mathcal{I} is the identity matrix and α is a positive hyperparameter [52], [53]. For a detailed treatment of the algorithm, we refer to [54]. As with all linear models, the drawback is when the linear assumption does not hold. As a consequence, the model's performance will suffer when treating complex and nonlinear systems.

Support vector regression (SVR)

SVR is a kernel-based algorithm utilizing a nonlinear mapping function $\Phi(\cdot)$ to transform the data from low-dimensional space x_p to a high-dimensional space $\Phi(x_p)$ such that the estimated function f is linear in $\Phi(x_p)$, i.e., $f(x_p) = w^T \Phi(x_p) + b$, where both w and b are the model parameters that need to be found. It is similar to linear regression in the form but is solved as a quadratic programming problem. The optimization problem can be formulated as fol-

lows:

$$\begin{aligned} \min_{w,b,\xi,\xi^*} \quad & \frac{1}{2} \|w\|^2 + C \sum_p (\xi_p + \xi_p^*) \\ \text{subject to} \quad & \begin{cases} y_p - w^T \Phi(x_p) - b \leq \epsilon + \xi_p \\ w^T \Phi(x_p) + b - y_p \leq \epsilon + \xi_p^* \\ \xi_p, \xi_p^* \geq 0 \end{cases} \end{aligned} \quad (3.1)$$

where ξ_p and ξ_p^* are positive slack variables, C is a regularisation coefficient, and ϵ is the error tolerance coefficient. Both C and ϵ are hyperparameters to be determined using cross-validation. To avoid explicitly computing $\Phi(x_p)$, the kernel trick can be used [55]. Intuitively, SVR constructs a tube around a hyperplane with the aim of letting as many training points as possible fall into the tube and, at the same time, the hyperplane is kept as flat as possible, as depicted in Fig. 3.1.

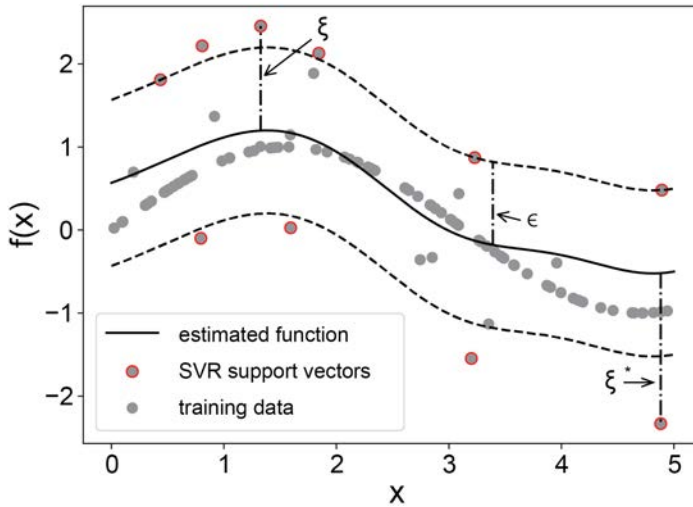


Figure 3.1: Illustration of SVR with one single feature.

Random forest regression (RFR)

RFR is an ensemble supervised learning method leveraging two core techniques called “bootstrap” and data aggregation. At the sampling collection stage, every individual tree randomly selects a subset of the training data and feature to conduct prediction, and this process is called “bootstrap”. After each tree has its own prediction result, RFR aggregates all the results by taking an average [56]. The total number of decision trees and the maximum depth of each individual tree are two important hyperparameters to make a trade-off between model performance, computational resources, and training time. The overall structure of the RFR algorithms is illustrated in Fig. 3.2. Comparatively, RFR has a better bias-variance trade-off than the single decision tree model. Moreover, the prediction results from RFR models are generally easier to interpret compared to other complex nonlinear ML algorithms.

Gaussian process regression (GPR)

GPR is a non-parametric supervised learning algorithm which means that a finite set of parameters cannot represent the model, which is rather given in a function format $f(\cdot)$ [57]. Often the function is assumed to be distributed according to a Gaussian process and represented as

$$f(x) \sim \text{GP}(m(x), \kappa(x, x')), \quad (3.2)$$

where x and x' are two arbitrary data samples, $m(x)$ represents the mean value of $f(x)$, and $\kappa(x, x')$ is the covariance of $f(\cdot)$ between the points x and x' . Notably, both the mean and the covariance functions can incorporate prior knowledge about the estimated function $f(x)$. The posterior probability of the predicted output $f(x^*)$ at any point x^* in the test set can be calculated by calculating the probability of conditioning f on the complete training dataset $(\mathbf{X}, f(\mathbf{X}))$, where \mathbf{X} is a vector of all the \mathcal{P} training samples, i.e., $\mathbf{X} = [x_1, \dots, x_{\mathcal{P}}]$. The predicted output probability distribution can be

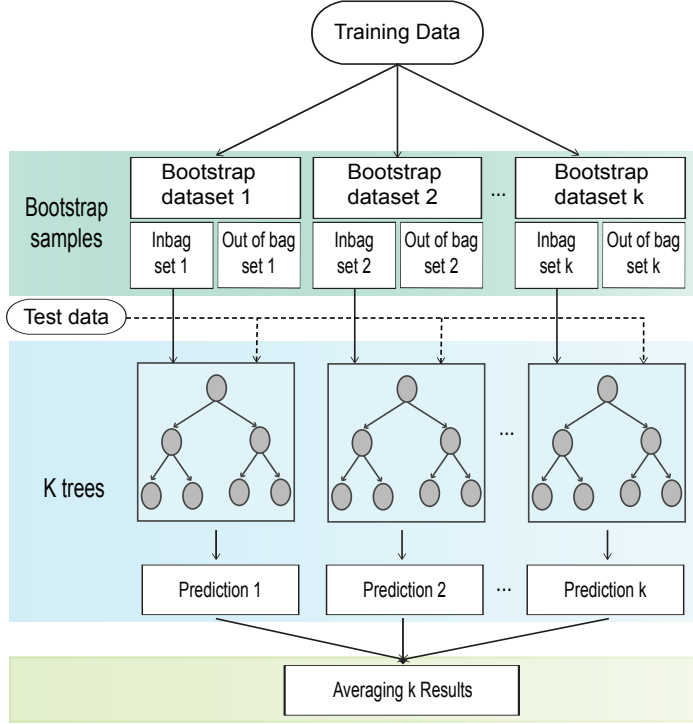


Figure 3.2: Illustration of RFR algorithm.

represented as:

$$p(f(x^*)|x^*, \mathbf{X}, f(\mathbf{X})) \sim \mathcal{N}(f(x^*); m^*, \Sigma^*) \quad (3.3)$$

$$m^* = \mathcal{K}(\mathbf{X}, x^*)^T \mathcal{K}(\mathbf{X}, x^*)^{-1} f(\mathbf{X}) \quad (3.4)$$

$$\Sigma^* = \mathcal{K}(\mathbf{X}, x^*) - \mathcal{K}(\mathbf{X}, x^*)^T \mathcal{K}(\mathbf{X}, x^*)^{-1} \mathcal{K}(\mathbf{X}, x^*) \quad (3.5)$$

where \mathcal{K} is the covariance kernel matrix having κ as elements. Usually, the kernel function $\kappa(x, x')$ can be seen as a hyperparameter that can be tuned during training using cross-validation. Typical kernel functions are radial basis function (RBF), Matérn kernel, and rational quadratic kernel [57]. Fig. 3.3 illustrates a typical GPR algorithm. Due to its Bayesian origin, GPR can naturally propagate its prediction uncertainty, which is seen as a merit for

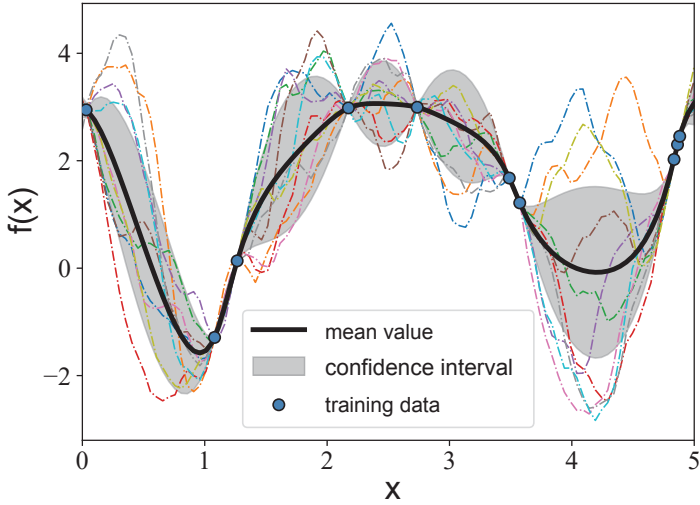


Figure 3.3: Illustration of GPR algorithm with one single feature.

many applications. However, when the data amount increases, the necessity of a matrix inversion during posterior probability population, as given in (3.5), becomes computationally expensive.

Neural network (NN)

Conceptually, the invention of the NN was inspired by the human brain's cognition process, where the goal is to learn from data and find the best model to map from input to output. NN usually consists of several layers, with each layer made of multiple units, which can also be referred to as neurons. The mathematical form for an arbitrary neuron $j \in \{1, \dots, J_l\}$ in a hidden layer $l \in \{0, 1, \dots, L\}$ can be represented as

$$z_{j,l} = \sum_{j=1}^{l_J} w_{j,l} h_{j,l-1} + b_{j,l} \quad (3.6)$$

$$h_{j,l} = a_l(z_{j,l}) \quad (3.7)$$

where w , b , and a are the weight of the unit, the bias factor, and the activation function, respectively. Depending on the preference of the out, the activation function can be chosen accordingly, or it can also be selected through cross-validation. Typical activation functions are the rectified linear unit (ReLU) function, the *tanh* function, and the *sigmoid* function. A loss function is chosen to measure the difference between the model's output and the ground truth. With all the weights and biases as decision variables, gradient descent-based algorithms are often adopted to solve the minimization of the loss functions [51]. A typical NN architecture and its mathematical representation are shown in Fig. 3.4.

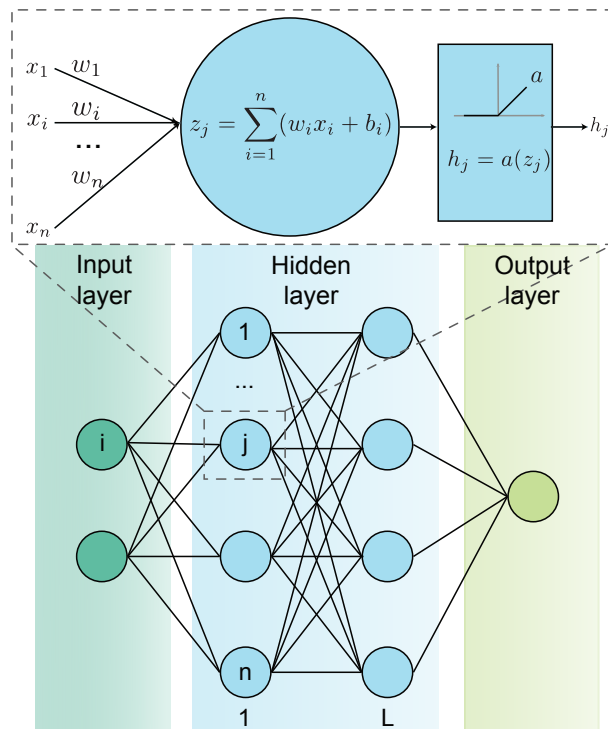


Figure 3.4: Illustration of NN.

The remarkable ability to detect patterns and identify trends from complicated or imprecise data of complex and nonlinear systems has made NN

an attractive method for many applications, such as image recognition and natural language processing. Furthermore, the capability of handling large datasets and the flexibility to accommodate parallel computation make NN a preferred choice in many real-world applications. On the other hand, NN also has its own pitfalls, e.g., a high required tuning effort, risk of overfitting, and relatively long training time.

CHAPTER 4

Battery aging diagnostics

During normal charging and discharging operations, batteries undergo not only the primary intended electrochemical reactions but also several complex and interacting side reactions, which contribute to battery degradation over time, as discussed in Chapter 3. To ensure the safe and optimal use of battery systems, accurate and reliable aging diagnostics are essential. Battery aging is often referred to as SoH estimation [58]. It is important to note that SoH can be defined in various ways, such as resistance growth or capacity decline. In this thesis, SoH is defined as the ratio between the current available capacity and the nominal capacity at the beginning of life (BOL).

In recent years, significant efforts have been made by both academia and industry to improve battery capacity estimation, as reviewed in [40], [59]–[63]. These methods can generally be categorized into three groups: empirical methods [64]–[68], model-based methods [69]–[72], and data-driven methods [73]–[75]. Empirical models, based on extensive laboratory cycling data, establish a mathematical relationship between cell capacity and common battery degradation features, such as Ah counting, equivalent cycle number, or elapsed time. These models were initially attractive due to their simplicity and niche applications. However, unlike the controlled conditions of labora-

tory testing, real-world applications involve much more dynamic and varying operating conditions. As a result, empirical models often fail to accurately extrapolate battery behavior in real-world environments [14].

Model-based methods offer better generalization and adaptability by continuously updating internal parameters based on real-time measurements, making them more suitable for specific applications [42]. These methods are typically divided into two categories: equivalent circuit models (ECMs) and electrochemical models (EMs). ECMs are constructed using electrical circuitry components and are favored for their straightforward implementation in battery aging diagnostics. However, their performance may degrade over time, particularly under the harsh and complex operating conditions encountered in EVs. Additionally, the accuracy of ECMs can decline over time due to the lack of regular re-calibration through Reference Performance Tests (RPTs) [63]. On the other hand, EMs are derived from the porous electrode and concentrated solution theory, capturing detailed internal characteristics of the battery [76], [77]. To model aging behaviors, aging models such as SEI formation [78], [79], lithium plating [33], and particle cracking [80] are coupled with the original EM. However, the high computational demand and challenges in parameterization make these methods difficult to apply in real-time in online applications [81].

In contrast to model-based approaches, data-driven methods do not rely on the underlying physical mechanisms but instead, capture the aging state by analyzing measured signals from battery operation. The growing trend of digitalization and the introduction of new technologies, such as digital twins and intelligent battery management systems [20], [47], [82], have generated significant interest in applying data-driven methods to battery diagnostics. Both feature-based machine learning algorithms and end-to-end deep learning methods have been proposed to estimate battery capacity [59]–[61], [73]–[75]. Feature-based methods involve manually selecting relevant features from raw data or estimated states using domain knowledge [83], [84], with common features extracted from voltage, current, temperature, or accumulated time measurements.

Filtered or calculated signals, such as incremental capacity (IC), differential voltage (DV), and differential temperature (DT), have also been used to estimate SoH. For example, Deng *et al.* [85] constructed features from the discharge capacity and voltage curve under pulse discharging conditions,

showing a clear correlation between these features and cell capacity. Sui *et al.* [86] successfully estimated battery SoH using the fuzzy entropy of the voltage response after pulse excitation. The signal measured during the charging relaxation period, which can indicate the aging state, has been investigated by Zhu *et al.* [87]. Differential signals that highlight material phase changes have been widely used as diagnostic tools [88]–[92], making them strong candidates as input features for machine learning algorithms. For example, She *et al.* [93] used IC-related features from real-world bus operating data to indicate battery aging status, while Wang *et al.* [94] and Li *et al.* [95] used features extracted from the DT curve to estimate battery SoH.

Deep learning algorithms, on the other hand, can bypass the feature engineering process by using raw measurement data as input. For instance, Li *et al.* [96] used voltage signals from repetitive CC-CV charging cycles to estimate cell capacity using Long Short-Term Memory (LSTM) neural networks. Tagade *et al.* [97] trained a deep Gaussian Process Regression (GPR) algorithm using raw discharging data to estimate battery capacity.

While many of these methods demonstrate strong estimation results, their practical applicability in real-world scenarios, particularly in automotive applications, has not been thoroughly investigated [14], [98], [99]. Real-world battery use involves more diverse cycling conditions, operating environments, and challenges, such as noisy measurements and varying sampling rates, which complicate the implementation of diagnostic algorithms. To address these challenges, a charging scenario-based feature extraction strategy is introduced in Paper A to ensure reliable SoH estimation under varying operating conditions. Additionally, a Kalman filter-based online model fusion algorithm is proposed to enhance estimation accuracy and robustness.

Battery aging prognostics

Battery aging prognostics focuses on predicting the future health and RUL of a battery. While diagnostics tell you the current state, prognostics attempt to forecast how the battery will age under different conditions. This helps in planning maintenance, replacements, or in optimizing battery usage to extend its life, as depicted in Fig. 5.1. Accurate aging prognostics are of great impor-

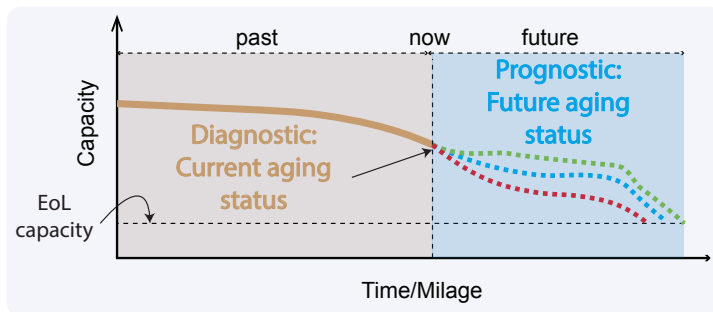


Figure 5.1: Illustration of battery aging diagnostics and prognostics, where the predicated capacity fades depends on future usage.

tance for predictive maintenance, health-conscious usage, and evaluation of residual value. According to the prediction target, the prognostic method can be categorized as future aging trajectory prediction and RUL point prediction [58]. The main difference is that the aging trajectory prediction aims to extract information about different aging states by predicting the aging states at selected points in time until the battery reaches its EOL. While RUL point prediction only focuses on how long the battery can still be used.

Due to similar characteristics, methods applied to battery aging diagnostics can often be adapted to conduct battery aging prognostics as well. Therefore, methods such as empirical methods and model-based methods can also be used for battery prognostic tasks as reviewed in [42], [60], [63]. However, the shortcomings of these methods for battery aging diagnostics apply also to prognostics, which will limit their effectiveness in a real-world deployment. Comparatively, data-driven methods have a better fidelity and complexity trade-off and may thus suit better for such a task.

For aging trajectory prediction, a growing number of data-driven methods have been developed. For instance, Li *et al.* [100] developed a recurrent neural network (NN)-based sequence-to-sequence model using the historic capacity information to predict the future capacity trajectory. Richardson *et al.* [101] applied a Gaussian process regression (GPR) to successfully predict both the short-term and long-term capacity trajectory and quantify the estimation confidence interval.

However, these methods often require repeatable cycling profiles, which are not comparable with the random usage pattern encountered in real-world applications. Additionally, extensively using time-series-based measurement as feature input may require large data storage space in the cloud, e.g., one vehicle with several complex cycles can easily reach gigabyte data size [14], [19]. Therefore, from both economic and energy-saving points of view, they impose challenges. Several pilot works explored the usage of histogram-like data to formulate feature input to minimize the input data size. Richardson *et al.* [102] adopted capacity throughput and time duration as feature inputs, while Greenbank *et al.* [103] mainly used calendar time and time spent in specific voltage windows as inputs. However, all data used in these works are still laboratory repetitive cycling data; thus, the practical capability of such a model may be limited. In Paper B, a machine learning framework is proposed leveraging histogram-based usage data to formulate features to predict the

future aging trajectory of the battery. Additionally, to tackle the problem of cell-to-cell variations, an online adaptive correction model is also developed.

Similar to aging trajectory prediction, data-driven methods can also be applied to predict the RUL point prediction. Severson *et al.* [21] innovatively discovered a feature based on the variance of the time series voltage signal and used an elastic net model to successfully predict the RUL of the battery only using the first 100 cycles information. Jones *et al.* [104] applied electrochemical impedance spectroscopy measurements as feature input and trained probabilistic ML models to predict the future discharge capacity. However, to the best of the author's knowledge, there is no work using usage-related histogram information to forecast battery lifetime and examine its relationship with the methods using time-series measurements. In Paper C, a comprehensive comparison study was conducted to investigate the prediction performance between using time series measurement data and adopting histogram usage information.

Battery fast charging optimization

While accurately estimating a battery's health and predicting its remaining lifetime is crucial, applying this knowledge to optimally control and adapt battery usage for lifetime extension is even more significant. Due to varying usage profiles and customer preferences, battery discharging tends to be stochastic and difficult to control. However, during vehicle charging, both the charging rate and profiles can be deliberately set by engineers. As discussed in Chapter 2, one of the primary aging mechanisms—lithium plating—is particularly prone to occur during charging. Therefore, optimizing and carefully controlling battery charging, especially during fast charging, is both necessary and logical.

Fast charging aims to reduce the charging time of lithium-ion batteries to a level comparable with refueling conventional vehicles, ideally within 15 minutes [105], [106]. However, simply increasing the charging current can lead to accelerated battery aging, capacity fade, and safety hazards as the battery operates near its physical and chemical limits. To prevent harmful side effects like lithium plating, which occurs when lithium ions deposit on the anode instead of intercalating, careful control of the charging process is essential, particularly under conditions of high charging currents, low temperatures,

and high SoC [107], [108].

Enabling fast charging requires a multifaceted approach. Improvements in battery materials, such as advanced electrode and electrolyte compositions, are key to enhancing performance and safety during fast charging [109], [110]. Additionally, system-level strategies, including efficient thermal management and optimized battery design, are critical to support faster charging rates while minimizing risks [111]–[115]. This chapter focuses on the battery management system perspective, specifically how to design a health-conscious fast charging strategy for a given battery system.

Fast charging control algorithms optimize the charging process by dynamically adjusting the charging current based on real-time battery conditions. These algorithms can be broadly categorized into heuristic-based and model-based approaches, each with its own strengths and limitations [116], [117].

Heuristic-based strategies are derived from empirical observations and testing. For example, Boost Charging involves starting with a high initial current to rapidly increase the battery’s charge, then reducing the current as the SOC increases, where the risk of lithium plating and other degradation mechanisms becomes more significant [118], [119]. Pulse Charging alternates between high and low current pulses; the high current accelerates charging, while the low current allows the battery to relax, mitigating risks such as lithium plating and excessive heat buildup. Studies have shown that this approach may effectively maintain battery health during fast charging by reducing the adverse effects of sustained high currents [120], [121]. Another method, the Multistage Constant Current (MSCC) strategy, applies different current levels at various stages of the charging process—starting with a high current, then moving to moderate, and finally low current as the battery nears full charge [122], [123].

In contrast, model-based strategies use mathematical models to simulate battery behavior and predict the impact of fast charging on the battery’s internal components. These models offer more precise control over the charging process. Electrochemical models, for instance, simulate the plating potential during battery usage. When combined with advanced control algorithms like model predictive control, they can regulate the charging current in real-time to prevent plating [124]–[126]. However, running these models for every cell in large battery packs remains computationally challenging due to the high number of cells involved.

Equivalent Circuit Models (ECMs) provide a more simplified approach by

using electrical circuit analogs to represent the battery’s behavior. Enhanced versions of these models incorporate thermal dynamics and aging processes, allowing them to predict temperature rise and capacity fade during charging. This capability is particularly useful for real-time adjustments of the charging current, helping to prevent overheating and shortened battery life [127], [128]. However, parameterizing these models over the battery’s entire lifetime is challenging and may limit their effectiveness in real-world applications.

With the increasing collection of battery operation data and the use of cloud-based BMS that provide greater computational power, data-driven methods have gained popularity for battery diagnosis and prognosis. With large, high-quality datasets, these methods demonstrate strong performance in estimating battery SoC, SoH, and RUL [48], [129]–[131]. While data-driven methods are promising, there have been limited attempts to estimate battery plating potential and prevent lithium plating during charging. Lin [132] employed a long short-term memory neural network (LSTM) to estimate plating potential using synthetic data generated from a physics-based battery model. However, this approach considered only limited operating conditions, which may reduce its applicability. Hamar *et al.* [133] extended this work by incorporating more realistic conditions, including varying temperatures, initial SoC levels, and multi-step constant charging. Nevertheless, the plating potential under repeated charging cycles can be significantly influenced by the battery’s aging state [134], [135]. Older batteries are more susceptible to lithium plating, making it crucial to account for aging effects when estimating plating potential. To date, there is no established method capable of accurately and continuously estimating plating potential throughout the battery’s entire lifespan. In Paper D, we developed an ML framework capable of estimating lithium plating potential throughout the battery’s lifecycle, integrating diverse operating conditions and aging states to support the fastest lithium plating-free charging of Li-ion batteries.

Summary of included papers

This chapter provides a summary of the included papers.

7.1 Paper A

Yizhou Zhang, Torsten Wik, John Bergström, Changfu Zou
State of health estimation for lithium-ion batteries under arbitrary usage
using data-driven multi-model fusion
Published in IEEE Transactions on Transportation Electrification,
vol. 10, pp. 1494-1507, Apr. 2023.
©<https://doi.org/10.1109/TTE.2023.3267124> .

This paper investigates the problem of real-world battery aging diagnostics. More specifically, estimating the battery state of health under the conditions of dynamic vehicle operating conditions, stochastic user behaviors, and cell-to-cell variations. First, all possible operating conditions are categorized into six scenarios which are complementary to each other. Second, a comprehensive feature pool is extracted for each scenario to indicate the battery aging state. Then four machine learning algorithms are applied, of which two are

probabilistic-based, and the other two are frequentist-based to estimate the aging state individually using time-series data. A histogram-based online adaptive model is trained to predict the one-step-ahead capacity of the battery. In the end, a Kalman filter is applied to systematically fuse all estimation and prediction results to increase accuracy and robustness.

7.2 Paper B

Yizhou Zhang, Torsten Wik, John Bergström, Michael Pecht, Changfu Zou

A machine learning-based framework for online prediction of battery ageing trajectory and lifetime using histogram data

Published in Journal of Power Sources,

vol. 526, p. 231110, Apr. 2022.

©<https://doi.org/10.1016/j.jpowsour.2022.231110> .

This paper developed a data-driven method to systematically predict the future aging trajectory for batteries used in real-world applications. The wide range of operating conditions, highly nonlinear aging trends, and cell-to-cell variations pose challenges for battery health prediction. To tackle this, a histogram-based usage-related feature pool was first constructed regardless of raw data format and usage profiles. The proposed prediction framework has been used for training and testing of both laboratory cycling data and real-world vehicle fleet data, all showing promising results. Moreover, an online adaptation algorithm was developed to address the cell-to-cell variation problems and further reduce the errors by up to 13%.

7.3 Paper C

Yizhou Zhang, Torsten Wik, Yicun Huang, John Bergström, Changfu Zou

Data-driven battery life prediction considering both onsite measurement and usage information

22nd IFAC World Congress 2023, July 2023, Yokohama, Japan .

This work developed a data-driven battery early-life prediction algorithm using both time-series, measurement-related features, and usage-related features. A comprehensive comparison study using these two different feature

sources is also conducted. The comparison indicates these two feature sources are, in principle, complementary to each other. Therefore, if possible, combining these two feature sources can increase the prediction accuracy and enhance the robustness. Additionally, four commonly used machine learning algorithms are applied to compare the effectiveness and performance of these algorithms on battery early-life prediction problems. Finally, batteries with different cell chemistry are used to verify the applicability and generality of the developed methods.

7.4 Paper D

Yizhou Zhang, Torsten Wik, John Bergström, Changfu Zou

Machine learning-based lifelong estimation of lithium plating potential:
A path to health-aware fastest battery charging

Accepted for publication in *Energy Storage Materials* .

This paper introduces an innovative approach that combines experimental validation with machine learning to address the challenge of balancing fast charging and prolonged battery lifespan in lithium-ion batteries. The proposed method utilizes lithium plating potential control during charging, as demonstrated through three-electrode cell experiments, allowing batteries to charge at least 30% faster while nearly doubling their lifespan. To enable practical implementation in commercial applications, a machine learning framework is developed to estimate plating potential over the battery's lifecycle using real-time data, eliminating the need for costly and complex reference electrodes. This framework integrates multiple predictive models to ensure accurate and reliable predictions across various operating conditions and aging stages, advancing the field of battery management by enabling health-conscious fast charging strategies.

7.5 Paper E

Yizhou Zhang, Torsten Wik, John Bergström, Shafiq Urréhman, Changfu Zou

Harmonizing performance indicators and unifying data for lifelong battery management

Under review .

This work presents a novel machine learning-based framework for harmonizing and evaluating battery performance indicators using extensive field data from electric vehicles. This research addresses the challenges posed by the stochastic and uncontrolled conditions of real-world battery usage, which make it difficult to assess battery performance accurately. The proposed framework projects key battery performance indicators, such as internal resistance and capacity, onto standardized reference conditions, enabling accurate real-time monitoring and prognosis of battery performance throughout its lifecycle. The framework is trained and validated using one of the largest datasets known from a fleet of plug-in hybrid electric vehicles, ensuring robustness and reliability. This approach allows for systematic assessment and comparison of battery performance indicators across different operating conditions and aging levels, as well as between different battery packs. Moreover, the study highlights the potential of this unified framework to standardize data storage formats and reduce data volume, thus significantly improving the efficiency of data transmission and storage. By harmonizing noisy and drifted performance indicators, the framework enhances the accuracy of battery diagnostics and facilitates predictive maintenance, ultimately contributing to extending battery life.

Concluding remarks and future work

Li-ion batteries play a pivotal role in the deployment of EVs, integration of renewable energy sources, and reliable operation of electric grids, thereby contributing substantially to global efforts in addressing climate change. As essential components of these energy systems, batteries require safe and efficient management grounded in precise information about their degradation state, projected aging trajectories, and RUL. This thesis has developed a suite of machine learning-driven methodologies to advance the accuracy of battery aging estimation and prediction, leveraging both laboratory cycling data and real-world field data. These approaches have been validated across multiple battery chemistries, underscoring their robustness and adaptability.

Additionally, a significant contribution of this thesis is the establishment of techniques for deriving general battery performance indicators directly from vehicle field data. This advancement enables continuous, real-time monitoring and assessment of critical parameters, such as internal resistance and capacity, under authentic operating conditions. By capitalizing on field data, this research provides valuable insights into battery behavior, paving the way for more informed and dynamic battery management strategies that can extend the operational lifespan of batteries.

The thesis also addresses a critical challenge in fast-charging technology: lithium plating during high-current charging events. By introducing a novel methodology for estimating the long-term plating potential of Li-ion batteries, this work supports the design of charging protocols that mitigate lithium plating risks, even at elevated charging rates. This innovation is particularly impactful, as it enables faster charging without compromising battery safety or durability, which is essential for the widespread adoption and sustainability of EVs.

Together, these contributions advance the field of battery management systems by providing robust tools for diagnosing, predicting, and managing battery aging. The methodologies developed in this thesis not only enhance the accuracy of battery health assessments but also open new pathways for optimizing battery performance and prolonging battery life. In doing so, they contribute meaningfully to the overarching goals of sustainability and efficiency in modern energy systems.

Although significant advancements have been made in battery technology, including the contributions of this thesis, many research questions remain unanswered. Highlighted here are several research prospects that could lead to more fruitful results in the future.

Current battery state estimation methods primarily rely on sensing data from battery voltage, current, and temperature. However, to enhance the accuracy and reliability of these estimations—whether for SoC, SoH, or state of safety—integrating additional types of sensors could provide deeper insights into the internal states of batteries. For example, pressure sensors can offer data related to volume changes during charging and discharging cycles, ultrasound sensors may assist in monitoring internal structural shifts, and optical sensors could be used to detect gas evolution or other internal reactions. Such multi-sensor approaches could significantly improve the precision and robustness of battery diagnostics.

The adoption of vehicle field data for battery aging diagnostics and the extraction of performance indicators offers new opportunities, but it also presents significant challenges. One of the most pressing challenges is the reliance on field data that often lacks ground truth validation. Without controlled environments to conduct comprehensive RPTs, the accuracy and reliability of diagnostic and prognostic models can be compromised. Future research could focus on developing quick or integrated RPTs that can be

seamlessly incorporated into regular battery usage scenarios, such as during routine vehicle charging cycles. Implementing such tests would provide more reliable validation data, thereby enhancing the precision and robustness of predictive models.

Another compelling avenue for future research is the efficient and optimal integration of physics-based models with data-driven approaches for battery aging diagnostics and prognostics. Physics-based models offer detailed insights into the aging mechanisms and internal state changes of batteries, providing a more interpretable framework for understanding battery health. However, these models are often difficult to parameterize and computationally intensive. In contrast, data-driven models, while computationally efficient and capable of handling large datasets, operate as black boxes, making it challenging to interpret their estimation or prediction results. By combining the strengths of both approaches—leveraging the detailed mechanistic insights from physics-based models and the computational efficiency and flexibility of data-driven models—future research could develop hybrid models that are both accurate and interpretable, offering a more comprehensive solution for battery management.

Finally, advancing fast-charging technologies that minimize lithium plating is crucial, especially for applications in automotive battery packs. This thesis proposes a method for lithium plating-free fast charging and demonstrates it over the full lifecycle of a laboratory-scale coin cell. Extending this method to commercial-scale cells presents further challenges and opportunities. A natural next step is to adapt this approach for use in commercial cells. One potential strategy is to insert a reference electrode into a commercial cell to measure the lithium plating potential directly and then construct a closed-loop fast charging control system based on this measurement. This approach could ensure the maintenance of positive plating potential, enhancing safety and longevity. However, challenges persist, as reference electrodes may fail to capture critical plating regions within the cell due to spatial inhomogeneity, potentially causing side effects or failing to represent the unmodified cell's behavior accurately. Alternatively, a more extensive parameterization of the commercial battery cell across its entire lifecycle could be conducted to develop a full-scale electrochemical model for real-time plating potential estimation during charging. Such a model, though valuable, would be complex to build due to the intricacies of battery aging processes and remains an active area of

research. Addressing these challenges through future work can help bridge the gap between laboratory-scale findings and practical applications in commercial battery systems.

References

- [1] L. Zaval, E. A. Keenan, E. J. Johnson, and E. U. Weber, “How warm days increase belief in global warming,” *Nat. Clim. Change.*, vol. 4, no. 2, pp. 143–147, 2014.
- [2] T. Geng, F. Jia, W. Cai, *et al.*, “Increased occurrences of consecutive la niña events under global warming,” *Nature*, vol. 619, no. 7971, pp. 774–781, 2023.
- [3] J. K. Balch, J. T. Abatzoglou, M. B. Joseph, *et al.*, “Warming weakens the night-time barrier to global fire,” *Nature*, vol. 602, no. 7897, pp. 442–448, 2022.
- [4] R. Shaw, Y. Luo, T. Cheong, *et al.*, “Transportation,” in *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK and New York, USA: Cambridge University Press, 2022, pp. 1457–1579.
- [5] X. Chen, H. Zhang, Z. Xu, C. P. Nielsen, M. B. McElroy, and J. Lv, “Impacts of fleet types and charging modes for electric vehicles on emissions under different penetrations of wind power,” *Nat. Energy*, pp. 413–421, Apr. 2018.
- [6] A. Jenn, “Emissions benefits of electric vehicles in Uber and Lyft ride-hailing services,” *Nat. Energy*, pp. 520–525, Jun. 2020.

- [7] *Cars cause biggest share of transportation co2 emissions*, <http://https://tex.stackexchange.com/questions/3587/how-can-i-use-bibtex-to-cite-a-web-page>, Accessed: 2024-08-12.
- [8] K. Pietrzak and O. Pietrzak, "Environmental effects of electromobility in a sustainable urban public transport," *Sustainability*, vol. 12, no. 3, p. 1052, 2020.
- [9] C. Llopis-Albert, F. Rubio, and F. Valero, "Impact of digital transformation on the automotive industry," *Technological Forecasting and Social Change*, vol. 162, p. 120 343, 2021.
- [10] B. Nykvist and M. Nilsson, "Rapidly falling costs of battery packs for electric vehicles," *Nat. Clim. Change*, pp. 329–332, Apr. 2015.
- [11] R. Schmuch, R. Wagner, G. Hörpel, T. Placke, and M. Winter, "Performance and cost of materials for lithium-based rechargeable automotive batteries," *Nat. Energy*, pp. 267–278, Apr. 2018.
- [12] J. M. Reniers, G. Mulder, and D. A. Howey, "Review and Performance Comparison of Mechanical-Chemical Degradation Models for Lithium-Ion Batteries," *J. Electrochem. Soc.*, A3189–A3200, 2019.
- [13] C. R. Birkl, M. R. Roberts, E. McTurk, P. G. Bruce, and D. A. Howey, "Degradation diagnostics for lithium ion cells," *J. Power Sources*, pp. 373–386, Feb. 2017.
- [14] V. Sulzer, P. Mohtat, A. Aitio, *et al.*, "The challenge and opportunity of battery lifetime prediction from field data," *Joule*, Jul. 2021.
- [15] M. Woody, M. Arbabzadeh, G. M. Lewis, G. A. Keoleian, and A. Stefanopoulou, "Strategies to limit degradation and maximize Li-ion battery service lifetime - Critical review and guidance for stakeholders," *J. Energy Storage*, p. 101 231, Apr. 2020.
- [16] H. Ruan, J. V. Barreras, T. Engstrom, Y. Merla, R. Millar, and B. Wu, "Lithium-ion battery lifetime extension: A review of derating methods," *J. Power Sources*, vol. 563, p. 232 805, 2023.
- [17] G. L. Plett, *Battery management systems, Volume I: Battery modeling*. Artech House, 2015, vol. 1.
- [18] C. D. Rahn and C.-Y. Wang, *Battery systems engineering*. John Wiley & Sons, 2013.

-
- [19] X. Hu, Z. Deng, X. Lin, Y. Xie, and R. Teodorescu, “Research directions for next-generation battery management solutions in automotive applications,” *Renewable and Sustainable Energy Reviews*, vol. 152, p. 111 695, 2021.
- [20] H. Dai, B. Jiang, X. Hu, X. Lin, X. Wei, and M. Pecht, “Advanced battery management strategies for a sustainable energy future: Multi-layer design concepts and research trends,” *Renewable and Sustainable Energy Reviews*, vol. 138, p. 110 480, 2021.
- [21] K. A. Severson, P. M. Attia, N. Jin, *et al.*, “Data-driven prediction of battery cycle life before capacity degradation,” *Nat. Energy*, pp. 383–391, May 2019.
- [22] P. M. Attia, A. Grover, N. Jin, *et al.*, “Closed-loop optimization of fast-charging protocols for batteries with machine learning,” *Nature*, vol. 578, no. 7795, pp. 397–402, Feb. 2020.
- [23] *A new eu regulatory framework for batteries*, <https://www.europarl.europa.eu/news/en/press-room/20221205IPR60614/batteries-deal-on-new-eu-rules-for-design-production-and-waste-treatment>, Accessed: 09-Feb-2023, 2022.
- [24] J. Vetter, P. Novák, M. Wagner, *et al.*, “Ageing mechanisms in lithium-ion batteries,” *J. Power Sources*, pp. 269–281, Sep. 2005.
- [25] J. S. Edge, S. O’Kane, R. Prosser, *et al.*, “Lithium ion battery degradation: What you need to know,” *Phys. Chem. Chem. Phys.*, pp. 8200–8221, 2021.
- [26] C. Lin, A. Tang, H. Mu, W. Wang, and C. Wang, “Aging mechanisms of electrode materials in lithium-ion batteries for electric vehicles,” *Journal of Chemistry*, vol. 2015, 2015.
- [27] P. Bai, J. Li, F. R. Brushett, and M. Z. Bazant, “Transition of lithium growth mechanisms in liquid electrolytes,” *Energy & Environmental Science*, vol. 9, no. 10, pp. 3221–3229, 2016.
- [28] G. Liu and W. Lu, “A model of concurrent lithium dendrite growth, sei growth, sei penetration and regrowth,” *Journal of The Electrochemical Society*, vol. 164, no. 9, A1826, 2017.

- [29] M. Safari, M. Morcrette, A. Teyssot, and C. Delacourt, “Multimodal physics-based aging model for life prediction of li-ion batteries,” *Journal of The Electrochemical Society*, vol. 156, no. 3, A145, 2008.
- [30] P. M. Attia, S. Das, S. J. Harris, M. Z. Bazant, and W. C. Chueh, “Electrochemical kinetics of sei growth on carbon black: Part i. experiments,” *Journal of The Electrochemical Society*, vol. 166, no. 4, E97–E106, 2019.
- [31] S. Das, P. M. Attia, W. C. Chueh, and M. Z. Bazant, “Electrochemical kinetics of sei growth on carbon black: Part ii. modeling,” *Journal of The Electrochemical Society*, vol. 166, no. 4, E107–E118, 2019.
- [32] M. Dubarry and D. Beck, “Big data training data for artificial intelligence-based li-ion diagnosis and prognosis,” *Journal of Power Sources*, vol. 479, p. 228 806, 2020.
- [33] X.-G. Yang, Y. Leng, G. Zhang, S. Ge, and C.-Y. Wang, “Modeling of lithium plating induced aging of lithium-ion batteries: Transition from linear to nonlinear aging,” *J. Power Sources*, pp. 28–40, Aug. 2017.
- [34] I. D. Campbell, M. Marzook, M. Marinescu, and G. J. Offer, “How observable is lithium plating? differential voltage analysis to identify and quantify lithium plating following fast charging of cold lithium-ion batteries,” *Journal of the electrochemical society*, vol. 166, no. 4, A725–A739, 2019.
- [35] S. E. O’Kane, I. D. Campbell, M. W. Marzook, G. J. Offer, and M. Marinescu, “Physical origin of the differential voltage minimum associated with lithium plating in li-ion batteries,” *Journal of The Electrochemical Society*, vol. 167, no. 9, p. 090 540, 2020.
- [36] B. M., B. Ph, B. F., *et al.*, “Main aging mechanisms in Li ion batteries,” *J. Power Sources*, pp. 90–96, Aug. 2005.
- [37] Y. Wang, J. Tian, Z. Sun, *et al.*, “A comprehensive review of battery modeling and state estimation approaches for advanced battery management systems,” *Renewable and Sustainable Energy Reviews*, p. 110 015, Oct. 2020.
- [38] S. Piller, M. Perrin, and A. Jossen, “Methods for state-of-charge determination and their applications,” *J. Power Sources*, pp. 113–120, Jun. 2001.

-
- [39] W. Waag, C. Fleischer, and D. U. Sauer, “Critical review of the methods for monitoring of lithium-ion batteries in electric and hybrid vehicles,” *J. Power Sources*, pp. 321–339, Jul. 2014.
- [40] M. Lucu, E. Martinez-Laserna, I. Gandiaga, and H. Camblong, “A critical review on self-adaptive Li-ion battery ageing models,” *J. Power Sources*, pp. 85–101, Oct. 2018.
- [41] A. Farmann, W. Waag, A. Marongiu, and D. U. Sauer, “Critical review of on-board capacity estimation techniques for lithium-ion batteries in electric and hybrid electric vehicles,” *J. Power Sources*, pp. 114–130, May 2015.
- [42] R. Xiong, Y. Pan, W. Shen, H. Li, and F. Sun, “Lithium-ion battery aging mechanisms and diagnosis method for automotive applications: Recent advances and perspectives,” *Renewable and Sustainable Energy Reviews*, p. 110 048, Oct. 2020.
- [43] D. Marcos, M. Garmendia, J. Crego, and J. A. Cortajarena, “Functional Safety BMS Design Methodology for Automotive Lithium-Based Batteries,” *Energies*, p. 6942, Oct. 2021.
- [44] H. J. Bergveld, W. S. Kruijt, and P. H. L. Notten, *Battery Management Systems*. Dordrecht: Springer Netherlands, 2002.
- [45] M. Lelie, T. Braun, M. Knips, *et al.*, “Battery Management System Hardware Concepts: An Overview,” *Applied Sciences*, p. 534, Mar. 2018.
- [46] R. Teodorescu, X. Sui, S. B. Vilsen, P. Bharadwaj, A. Kulkarni, and D.-I. Stroe, “Smart Battery Technology for Lifetime Improvement,” *Batteries*, p. 169, Oct. 2022.
- [47] B. Wu, W. D. Widanage, S. Yang, and X. Liu, “Battery digital twins: Perspectives on the fusion of models, data and artificial intelligence for smart battery management systems,” *Energy and AI*, vol. 1, p. 100 016, 2020.
- [48] W. Li, M. Rentemeister, J. Badedo, D. Jöst, D. Schulte, and D. U. Sauer, “Digital twin for battery systems: Cloud battery management system with online state-of-charge and state-of-health estimation,” *J. energy storage*, vol. 30, p. 101 557, 2020.

- [49] A. Lindholm, N. Wahlström, F. Lindsten, and T. B. Schön, *Machine Learning - A First Course for Engineers and Scientists*. Cambridge University Press, 2022.
- [50] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006, vol. 4.
- [51] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [52] M. E. Tipping, “Sparse bayesian learning and the relevance vector machine,” *J. machine learning research*, vol. 1, no. Jun, pp. 211–244, 2001.
- [53] D. J. MacKay, “Bayesian interpolation,” *Neural computation*, vol. 4, no. 3, pp. 415–447, 1992.
- [54] C. M. Bishop, *Pattern recognition and machine learning* (Information science and statistics). New York: Springer, 2006.
- [55] A. J. Smola and B. Schölkopf, “A tutorial on support vector regression,” *Stat. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004.
- [56] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, pp. 5–32, 2001.
- [57] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning* (Adaptive computation and machine learning). Cambridge, Mass: MIT Press, 2006.
- [58] Y. Che, X. Hu, X. Lin, J. Guo, and R. Teodorescu, “Health prognostics for lithium-ion batteries: Mechanisms, methods, and prospects,” *Energy Environ. Sci.*, 2023.
- [59] M.-F. Ng, J. Zhao, Q. Yan, G. J. Conduit, and Z. W. Seh, “Predicting the state of charge and health of batteries using data-driven machine learning,” *Nat Mach Intell*, pp. 161–170, Mar. 2020.
- [60] Y. Li, K. Liu, A. M. Foley, *et al.*, “Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review,” *Renewable and Sustainable Energy Reviews*, p. 109254, Oct. 2019.
- [61] H. Tian, P. Qin, K. Li, and Z. Zhao, “A review of the state of health for lithium-ion batteries: Research status and suggestions,” *Journal of Cleaner Production*, p. 120813, Jul. 2020.

-
- [62] X. Han, L. Lu, Y. Zheng, *et al.*, “A review on the key issues of the lithium ion battery degradation among the whole life cycle,” *eTransportation*, p. 100 005, Aug. 2019.
- [63] X. Hu, L. Xu, X. Lin, and M. Pecht, “Battery Lifetime Prognostics,” *Joule*, pp. 310–346, Feb. 2020.
- [64] J. Wang, P. Liu, J. Hicks-Garner, *et al.*, “Cycle-life model for graphite-LiFePO₄ cells,” *J. Power Sources*, vol. 196, no. 8, pp. 3942–3948, 2011.
- [65] L. Lam and P. Bauer, “Practical capacity fading model for Li-ion battery cells in electric vehicles,” *IEEE Trans. Power Electron.*, vol. 28, no. 12, pp. 5910–5918, 2012.
- [66] W. He, N. Williard, M. Osterman, and M. Pecht, “Prognostics of lithium-ion batteries based on Dempster-Shafer theory and the Bayesian Monte Carlo method,” *J. Power Sources*, vol. 196, no. 23, pp. 10 314–10 321, 2011.
- [67] X. Tang, C. Zou, K. Yao, J. Lu, Y. Xia, and F. Gao, “Aging trajectory prediction for lithium-ion batteries via model migration and Bayesian Monte Carlo method,” *Appl. Energy*, vol. 254, p. 113 591, 2019.
- [68] I. Bloom, B. Cole, J. Sohn, *et al.*, “An accelerated calendar and cycle life study of Li-ion cells,” *J. Power Sources*, vol. 101, no. 2, pp. 238–247, 2001.
- [69] P. Arora, M. Doyle, and R. E. White, “Mathematical modeling of the lithium deposition overcharge reaction in lithium-ion batteries using carbon-based negative electrodes,” *J. Electrochem. Soc.*, vol. 146, no. 10, p. 3543, 1999.
- [70] J. Christensen and J. Newman, “A mathematical model for the lithium-ion negative electrode solid electrolyte interphase,” *J. Electrochem. Soc.*, vol. 151, no. 11, A1977, 2004.
- [71] P. Ramadass, B. Haran, P. M. Gomadam, R. White, and B. N. Popov, “Development of first principles capacity fade model for li-ion cells,” *J. Electrochem. Soc.*, vol. 151, no. 2, A196, 2004.
- [72] X.-G. Yang, Y. Leng, G. Zhang, S. Ge, and C.-Y. Wang, “Modeling of lithium plating induced aging of lithium-ion batteries: Transition from linear to nonlinear aging,” *J. Power Sources*, vol. 360, pp. 28–40, 2017.

- [73] X. Shu, S. Shen, J. Shen, *et al.*, “State of health prediction of lithium-ion batteries based on machine learning: Advances and perspectives,” *iScience*, p. 103 265, Nov. 2021.
- [74] X. Sui, S. He, S. B. Vilsen, J. Meng, R. Teodorescu, and D.-I. Stroe, “A review of non-probabilistic machine learning-based state of health estimation techniques for Lithium-ion battery,” *Appl. Energy*, p. 117 346, Oct. 2021.
- [75] H. Rauf, M. Khalid, and N. Arshad, “Machine learning in state of health and remaining useful life estimation: Theoretical and technological development in battery degradation modelling,” *Renewable Sustainable Energy Rev.*, p. 111 903, Mar. 2022.
- [76] M. Doyle, T. F. Fuller, and J. Newman, “Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell,” *J. Electrochem. Soc.*, vol. 140, no. 6, p. 1526, 1993.
- [77] V. Ramadesigan, K. Chen, N. A. Burns, V. Boovaragavan, R. D. Braatz, and V. R. Subramanian, “Parameter Estimation and Capacity Fade Analysis of Lithium-Ion Batteries Using Reformulated Models,” *J. Electrochem. Soc.*, p. 8, 2011.
- [78] A. V. Randall, R. D. Perkins, X. Zhang, and G. L. Plett, “Controls oriented reduced order modeling of solid-electrolyte interphase layer growth,” *J. Power Sources*, pp. 282–288, Jul. 2012.
- [79] T. R. Tanim and C. D. Rahn, “Aging formula for lithium ion batteries with solid electrolyte interphase layer growth,” *J. Power Sources*, pp. 239–247, Oct. 2015.
- [80] R. Deshpande, M. Verbrugge, Y.-T. Cheng, J. Wang, and P. Liu, “Battery cycle life prediction with coupled chemical degradation and fatigue mechanics,” *J. The Electrochemical Society*, vol. 159, no. 10, A1730, Aug. 2012.
- [81] J. C. Forman, S. J. Moura, J. L. Stein, and H. K. Fathy, “Genetic identification and Fisher identifiability analysis of the Doyle–Fuller–Newman model from experimental cycling of a LiFePO₄ cell,” *J. Power Sources*, vol. 210, pp. 263–275, 2012.

-
- [82] M. H. Lipu, M. Hannan, T. F. Karim, *et al.*, “Intelligent algorithms and control strategies for battery management system in electric vehicles: Progress, challenges and future outlook,” *J. Cleaner Production*, vol. 292, p. 126 044, 2021.
- [83] Y. Li, D.-I. Stroe, Y. Cheng, H. Sheng, X. Sui, and R. Teodorescu, “On the feature selection for battery state of health estimation based on charging–discharging profiles,” *J. Energy Storage*, vol. 33, p. 102 122, 2021.
- [84] X. Hu, Y. Che, X. Lin, and S. Onori, “Battery health prediction using fusion-based feature selection and machine learning,” *IEEE Trans. Transp. Electrification.*, vol. 7, no. 2, pp. 382–398, 2020.
- [85] 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 Trans. Mechatron.*, vol. 26, pp. 1295–1306, Jun. 2021.
- [86] X. Sui, S. He, J. Meng, R. Teodorescu, and D.-I. Stroe, “Fuzzy entropy-based state of health estimation for li-ion batteries,” *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 9, no. 4, pp. 5125–5137, 2021.
- [87] J. Zhu, Y. Wang, Y. Huang, *et al.*, “Data-driven capacity estimation of commercial lithium-ion batteries from voltage relaxation,” *Nat. Commun.*, p. 2261, Dec. 2022.
- [88] M. Dubarry and B. Y. Liaw, “Identify capacity fading mechanism in a commercial LiFePO₄ cell,” *J. Power Sources*, vol. 194, pp. 541–549, Oct. 2009.
- [89] M. Dubarry, B. Y. Liaw, M.-S. Chen, *et al.*, “Identifying battery aging mechanisms in large format Li ion cells,” *J. Power Sources*, vol. 196, pp. 3420–3425, Apr. 2011.
- [90] I. Bloom, L. K. Walker, J. K. Basco, D. P. Abraham, J. P. Christopher, and C. D. Ho, “Differential voltage analyses of high-power lithium-ion cells. 4. Cells containing NMC,” *J. Power Sources*, vol. 195, pp. 877–882, Feb. 2010.

- [91] Y. Merla, B. Wu, V. Yufit, N. P. Brandon, R. F. Martinez-Botas, and G. J. Offer, “Novel application of differential thermal voltammetry as an in-depth state-of-health diagnosis method for lithium-ion batteries,” *J. Power Sources*, vol. 307, pp. 308–319, Mar. 2016.
- [92] T. Shibagaki, Y. Merla, and G. J. Offer, “Tracking degradation in lithium iron phosphate batteries using differential thermal voltammetry,” *J. Power Sources*, vol. 374, pp. 188–195, Jan. 2018.
- [93] C. She, Y. Li, C. Zou, T. Wik, Z. Wang, and F. Sun, “Offline and online blended machine learning for lithium-ion battery health state estimation,” *IEEE Trans. Transp. Electrification.*, vol. 8, no. 2, pp. 1604–1618, 2022.
- [94] Z. Wang, C. Yuan, and X. Li, “Lithium battery state-of-health estimation via differential thermal voltammetry with gaussian process regression,” *IEEE Trans. Transp. Electrification.*, vol. 7, no. 1, pp. 16–25, 2020.
- [95] X. Li, C. Yuan, Z. Wang, and J. Xie, “A data-fusion framework for lithium battery health condition estimation based on differential thermal voltammetry,” *Energy*, vol. 239, p. 122 206, 2022.
- [96] 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,” *J. power sources*, vol. 482, p. 228 863, 2021.
- [97] “Deep gaussian process regression for lithium-ion battery health prognosis and degradation mode diagnosis,” *J. Power Sources*, vol. 445, p. 227 281, 2020, ISSN: 0378-7753.
- [98] F. von Bülow and T. Meisen, “A review on methods for state of health forecasting of lithium-ion batteries applicable in real-world operational conditions,” *J. Energy Storage*, p. 105 978, Jan. 2023.
- [99] A. Aitio and D. A. Howey, “Predicting battery end of life from solar off-grid system field data using machine learning,” *Joule*, vol. 5, no. 12, pp. 3204–3220, 2021.

-
- [100] W. Li, N. Sengupta, P. Dechent, D. Howey, A. Annaswamy, and D. U. Sauer, “One-shot battery degradation trajectory prediction with deep learning,” *Journal of Power Sources*, vol. 506, p. 230 024, 2021, ISSN: 0378-7753.
- [101] R. R. Richardson, M. A. Osborne, and D. A. Howey, “Gaussian process regression for forecasting battery state of health,” *Journal of Power Sources*, pp. 209–219, Jul. 2017.
- [102] R. R. Richardson, M. A. Osborne, and D. A. Howey, “Battery health prediction under generalized conditions using a Gaussian process transition model,” *J. Energy Storage*, pp. 320–328, Jun. 2019.
- [103] S. Greenbank and D. Howey, “Automated feature extraction and selection for data-driven models of rapid battery capacity fade and end of life,” *IEEE Trans. Ind. Inf.*, pp. 1–1, 2021.
- [104] P. K. Jones, U. Stimming, and A. A. Lee, “Impedance-based forecasting of lithium-ion battery performance amid uneven usage,” *Nat. Commun.*, p. 4806, Aug. 2022.
- [105] *Strategic research & innovation agenda*, <https://www.horizon-europe.gov.fr/sites/default/files/2024-02/t-1-charger-le-sria-batt4eu-f-vrier-2024--6449.pdf>, Accessed: 16-Aug-2024, 2024.
- [106] *Batteries, charging, and electric vehicles*, <https://www.energy.gov/eere/vehicles/batteries-charging-and-electric-vehicles>, Accessed: 16-Aug-2024, 2019.
- [107] T. R. Tanim, P. P. Paul, V. Thampy, *et al.*, “Heterogeneous behavior of lithium plating during extreme fast charging,” *Cell Reports Physical Science*, vol. 1, no. 7, 2020.
- [108] T. Waldmann, B.-I. Hogg, and M. Wohlfahrt-Mehrens, “Li plating as unwanted side reaction in commercial li-ion cells—a review,” *J. Power Sources*, vol. 384, pp. 107–124, 2018.
- [109] H. Zhao, L. Wang, Z. Chen, and X. He, “Challenges of fast charging for electric vehicles and the role of red phosphorous as anode material,” *Energies*, vol. 12, no. 20, p. 3897, 2019.

- [110] W. Xie, X. Liu, R. He, *et al.*, “Challenges and opportunities toward fast-charging of lithium-ion batteries,” *Journal of Energy Storage*, vol. 32, p. 101 837, 2020.
- [111] S. Ahmed, I. Bloom, A. N. Jansen, *et al.*, “Enabling fast charging—a battery technology gap assessment,” *Journal of Power Sources*, vol. 367, pp. 250–262, 2017.
- [112] M. Keyser, A. Pesaran, Q. Li, *et al.*, “Enabling fast charging—battery thermal considerations,” *Journal of Power Sources*, vol. 367, pp. 228–236, 2017.
- [113] A. Meintz, J. Zhang, R. Vijayagopal, *et al.*, “Enabling fast charging—vehicle considerations,” *Journal of Power Sources*, vol. 367, pp. 216–227, 2017.
- [114] C. Michelbacher, S. Ahmed, I. Bloom, *et al.*, *Enabling fast charging—introduction and overview*, 2017.
- [115] A. Burnham, E. J. Dufek, T. Stephens, *et al.*, “Enabling fast charging—infrastructure and economic considerations,” *Journal of Power Sources*, vol. 367, pp. 237–249, 2017.
- [116] N. Wassiliadis, J. Schneider, A. Frank, *et al.*, “Review of fast charging strategies for lithium-ion battery systems and their applicability for battery electric vehicles,” *Journal of energy storage*, vol. 44, p. 103 306, 2021.
- [117] A. Tomaszewska, Z. Chu, X. Feng, *et al.*, “Lithium-ion battery fast charging: A review,” *ETransportation*, vol. 1, p. 100 011, 2019.
- [118] P. H. Notten, J. O. het Veld, and J. Van Beek, “Boostcharging li-ion batteries: A challenging new charging concept,” *Journal of Power Sources*, vol. 145, no. 1, pp. 89–94, 2005.
- [119] S. S. Zhang, “The effect of the charging protocol on the cycle life of a li-ion battery,” *Journal of power sources*, vol. 161, no. 2, pp. 1385–1391, 2006.
- [120] J. Li, E. Murphy, J. Winnick, and P. A. Kohl, “The effects of pulse charging on cycling characteristics of commercial lithium-ion batteries,” *Journal of Power Sources*, vol. 102, no. 1-2, pp. 302–309, 2001.

-
- [121] B. Purushothaman and U. Landau, "Rapid charging of lithium-ion batteries using pulsed currents: A theoretical analysis," *Journal of The Electrochemical Society*, vol. 153, no. 3, A533, 2006.
- [122] M. A. Monem, K. Trad, N. Omar, *et al.*, "Lithium-ion batteries: Evaluation study of different charging methodologies based on aging process," *Applied Energy*, vol. 152, pp. 143–155, 2015.
- [123] M. Abdel-Monem, K. Trad, N. Omar, O. Hegazy, P. Van den Bossche, and J. Van Mierlo, "Influence analysis of static and dynamic fast-charging current profiles on ageing performance of commercial lithium-ion batteries," *Energy*, vol. 120, pp. 179–191, 2017.
- [124] F. Ringbeck, M. Garbade, and D. U. Sauer, "Uncertainty-aware state estimation for electrochemical model-based fast charging control of lithium-ion batteries," *J. Power Sources*, vol. 470, p. 228 221, 2020.
- [125] C. Zou, X. Hu, Z. Wei, T. Wik, and B. Egardt, "Electrochemical estimation and control for lithium-ion battery health-aware fast charging," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 8, pp. 6635–6645, 2017.
- [126] Y. Li, T. Wik, Y. Huang, and C. Zou, "Nonlinear model inversion-based output tracking control for battery fast charging," *IEEE Transactions on Control Systems Technology*, 2023.
- [127] Y. Lu, X. Han, Z. Chu, *et al.*, "A decomposed electrode model for real-time anode potential observation of lithium-ion batteries," *J. Power Sources*, vol. 513, p. 230 529, 2021.
- [128] H. E. Perez, X. Hu, S. Dey, and S. J. Moura, "Optimal charging of li-ion batteries with coupled electro-thermal-aging dynamics," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 9, pp. 7761–7770, 2017.
- [129] Y. Zhang, T. Wik, J. Bergström, and C. Zou, "State of health estimation for lithium-ion batteries under arbitrary usage using data-driven multi-model fusion," *IEEE Trans. Transp. Electrif.*, 2023.
- [130] Y. Zhang, T. Wik, J. Bergström, M. Pecht, and C. Zou, "A machine learning-based framework for online prediction of battery ageing trajectory and lifetime using histogram data," *J. Power Sources*, vol. 526, p. 231 110, 2022.

- [131] Z. Wei, X. Yang, Y. Li, H. He, W. Li, and D. U. Sauer, “Machine learning-based fast charging of lithium-ion battery by perceiving and regulating internal microscopic states,” *Energy Stor. Mater.*, vol. 56, pp. 62–75, 2023.
- [132] X. Lin, “Real-time prediction of anode potential in li-ion batteries using long short-term neural networks for lithium plating prevention,” *J. Electrochem. Soc.*, vol. 166, no. 10, A1893, 2019.
- [133] J. C. Hamar, S. V. Erhard, C. Zoerr, and A. Jossen, “Anode potential estimation in lithium-ion batteries using data-driven models for online applications,” *J. Electrochem. Soc.*, vol. 168, no. 3, p. 030 535, 2021.
- [134] N. Wassiliadis, J. Kriegler, K. A. Gamra, and M. Lienkamp, “Model-based health-aware fast charging to mitigate the risk of lithium plating and prolong the cycle life of lithium-ion batteries in electric vehicles,” *J. Power Sources*, vol. 561, p. 232 586, 2023.
- [135] J. Sieg, A. U. Schmid, L. Rau, *et al.*, “Fast-charging capability of lithium-ion cells: Influence of electrode aging and electrolyte consumption,” *Appl. Energy*, vol. 305, p. 117 747, 2022.

tive for a range of different types of machine learning methods, represented by SVR, RFR, GPR, and ANN. Specifically, RFR and ANN outperform their alternatives in terms of MAPE and achieve an error of 0.93% and 1.13%, respectively. Such smallness of prediction errors and strong resilience over the whole capacity range make the proposed models very competitive to the prevalent models developed directly upon time series data, e.g., [9], [70].

When we look into the details, the predicted points in Fig. 7(b) and (d) are located in the bottom triangle slightly more often than in the top triangle, and the orange and red lines in Fig. 7(e) show lower values in the right-hand side of the positive half-plane. This implies that ANN and RFR tend to less often overpredict the true capacity, which is beneficial to traction battery applications in which safety and reliability are of paramount importance. One advantage of GPR is the provided confidence intervals for its predictions. The two-sigma confidence interval has been highlighted as grey areas in Supplementary Fig. 1(c). However, the covered areas are too large to guide our prediction task. This is potentially due to high variations among different battery cells and because the uncertainties are quite large when applying a global model to predict a specific cell online. One example is cell ‘b2c47’ in the Stanford dataset that was aged extremely slowly compared to other cells. Its capacity trajectory has been emphasised as dark grey in Fig. 7(a)–(d). Not surprisingly, all the developed global models fail to capture its behaviour and result in an unacceptably large prediction error. This echoes our earlier statement and emphasises the importance of online adaptive prediction.

The prediction performance of the individualised models, with RFR as an example, is depicted in Fig. 7(f). It can be calculated that the individualised model is capable of reducing the prediction error by 13.7% in the best case and by 8.6% on average. These improvements are significant in battery ageing and lifetime prognosis, particularly considering their subsequent applications to health optimisation and lifetime extension of a large number of battery cells. Aside from the case of handling all the testing cells in a batch, we also examine the performance of online adaptation for individual cells. It can be seen from Fig. 7(f)–(h) that for both the abnormal cell ‘b2c47’ and a normal cell ‘b2c35’, the individualised model effectively learns from the historical ageing information and continuously adjusts the global predictions along the ageing trajectories to approach the ground truth, leading to more accurate and robust predictions.

Table 1: MAPE and RMSPE of the predicted capacity change by four machine learning methods.

Algorithms	MAPE (%)			RMSPE (%)		
	Stanford	NASA	Vehicle fleet	Stanford	NASA	Vehicle fleet
SVR	1.61	4.41	1.46	3.22	6.33	2.18
RFR	0.93	4.21	1.43	2.12	5.76	2.16
GPR	1.35	3.23	1.44	2.58	4.45	2.11
ANN	1.13	3.64	1.41	1.92	5.08	2.12

Vehicle driving schedule test With proven effectiveness against the Stanford dataset, we further assess the designed algorithms for battery cells covered in the NASA and vehicle fleet datasets. The calibrated capacity profiles for the NASA dataset occasionally have some local peaks where the measurements may have drifted from the actual capacity. For the fleet data, the capacity measurements were generated by onboard ECUs, and their accuracy is unknown. In such cases, we assume all the measurements as the ground truth, though this will to some degree affect the numerical results, which is a common situation for studies on real-world battery data. With this in mind, the obtained results are presented in Table 1 and Fig. 8.

For the fleet dataset, the global models based on different machine learning methods all have nearly the same prediction errors, which are around 1.45% in MAPE and 2.15% in RMSPE. For the NASA dataset, although the prediction performance cannot match those obtained with the other two datasets, the MAPE of 3.23% can still satisfy many industrial applications that typically require the error to be less than 5%. These results further corroborate the effectiveness and practicability of the histogram-based models for battery ageing prognosis. For the NASA dataset having the fewest data, GPR outperforms all its alternatives in both MAPE and RMSPE, which aligns well with GPR’s advantages discussed in Section 3.3. On the other hand, for each method, the prediction accuracy for battery cells installed in the vehicle fleet is much higher than that for the NASA ones. This is mainly attributed to the substantially increased number of features and data samples in the vehicle fleet dataset. The numerical relationship between model accuracy and data size is referred to the learning curve presented in Supplementary Fig. 2. These results confirm the importance of having a sufficiently large dataset.

The online adaptation algorithm is also tested on the NASA dataset. Its

efficacy in enhancing model accuracy and robustness in the presence of cell variations is again verified. Specifically, as demonstrated in Fig. 8, the individualised model is able to decrease the prediction errors at almost every prediction start index, with a maximum reduction of 7.5% in RMSPE. Similar to the results of Fig. 7(g)–(h), the individualised model follows the measured capacity profile more closely than the global model. Moreover, this is achieved no matter whether the global model overpredicts or underpredicts the measured values, and the performance is maintained even though battery degradation exceeds 50% of the nominal capacity. Additional results are demonstrated in Supplementary Figs. 1–6.

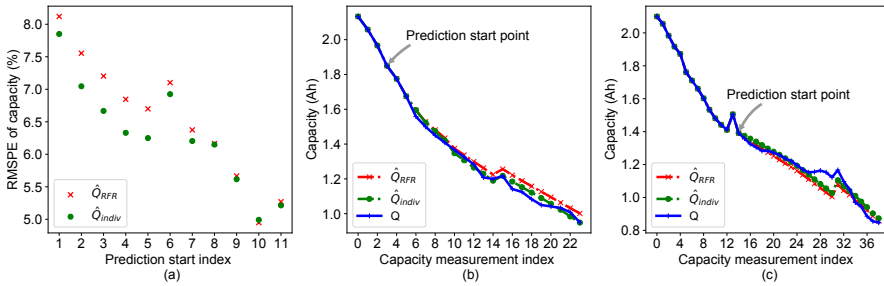


Figure 8: Comparison of the RFR-based global model and the individual model for predicting battery capacity of the NASA dataset. (a) presents RMSPE_Q^k for each prediction start index k . (b) and (c) show the predicted and measured capacity trajectory of cell No. 15 and No. 10, respectively.

4.3 Requirements for computation and memory

The online prediction algorithm associated with different machine learning methods was sequentially implemented on a typical laptop computer with specifications, software, and machine learning libraries introduced in Section 3.6. The hyperparameters for the employed machine learning methods are provided in Supplementary Tables 10–12. To minimise the stochasticity in quantifying the computational time, with each machine learning method the algorithm has been tested ten times under the same settings, and then the average time has been recorded. Table 2 summarises the time spent in training and testing. It can be first observed that the time to train the most demanding machine learning method associated with the largest dataset and

the most features was less than 1150 s. In this regard, we can conclude that the time required for offline model development is very low and thus can be readily satisfied. Second, the online prediction algorithm embedded with any of the developed global models accomplished the task at each sample for all cells in the corresponding dataset within 3.2 s. By using the most suitable machine learning method for online usage, the computational efficiency can be improved by 15–200 times, depending on the specified dataset, feature number, and hyperparameters. Along with the batteries’ age, the prediction horizon decreases, and the computational time can be further shortened. In either case, the required time is negligibly short compared to the timescale of battery degradation and the time interval between two adjacent data samples, which are several weeks or even months for real-world vehicle applications. It is notable that the high prediction performance presented in Section 4.2 was achieved at very small data. With up to 10,000 data samples for training, the needed memory resources were extremely limited and could be easily covered by today’s cloud system or even a desktop computer. This is a great advantage of the proposed histogram-based algorithm compared to its time series counterparts that are either oversized for onboard ECUs to store or plagued by a range of problems in data quality, security, cost, and energy consumption for transmission to data centres.

Table 2: Computational time and data size to train and implement the battery ageing predictor.

Methods	Training time (s) ^a			Testing time (μ s) ^b		
	NASA	Stanford	Vehicle fleet	NASA	Stanford	Vehicle fleet
SVR	0.01	2.4	6.9	18.9	144.5	593.0
RFR	0.2	1.1	29.5	70.9	10.5	108.3
GPR	1.4	291.7	1140.1	53.5	1933.5	1073.0
ANN	2.3	15.1	11.1	282.2	37.0	40.0

^aThe average time in seconds (s) to run the predictor on the complete training set.

^bThe average time in microseconds (μ s) to run the predictor on one data sample in the test set.

5 Conclusions

Data-driven modelling has been widely cited as the most promising approach for battery ageing prediction, but existing models commonly require specific operating profiles, a substantial amount of time series data and lack online adaptability. To address these problems, we have introduced a novel machine learning-based prediction framework. The novelties first come from using features based on data in histograms instead of time series, thereby significantly saving computational power and memory and allowing predictions under generalised operating conditions. Second, the framework is equipped with an online model adaptation algorithm that systematically handles cell variations, measurement noise, and disturbances.

This paper explored four widely adopted machine learning methods for the offline development of global models and evaluated the framework against three large datasets measured from batteries of different sizes, types of chemistry, and usage profiles. This framework has been verified to be effective for each method and dataset, and the computational time to predict the ageing trajectory of all batteries in the corresponding dataset is less than 3.2 s. Specifically, the best global models achieved 0.93% test error on laboratory data and 1.41% test error on the real-world fleet data, and the online algorithm further reduced the errors by up to 13.7%. Overall, this work proves the feasibility and benefits of using histogram data and highlights the importance of online adaptation for data-driven modelling of battery prognostics.

6 Supplementary materials

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.jpowsour.2022.231110>.

References

- [1] *Road transport: Reducing CO₂ emissions from vehicles*, https://ec.europa.eu/clima/policies/transport/vehicles_en, Accessed: 14-April-2021, 2017.
- [2] G. J. Offer, “Automated vehicles and electrification of transport,” *Energy Environ. Sci.*, vol. 8, no. 1, pp. 26–30, 2015.

- [3] B. Nykvist and M. Nilsson, “Rapidly falling costs of battery packs for electric vehicles,” *Nat. Clim. Change*, pp. 329–332, Apr. 2015.
- [4] R. Schmuch, R. Wagner, G. Hörpel, T. Placke, and M. Winter, “Performance and cost of materials for lithium-based rechargeable automotive batteries,” *Nat. Energy*, pp. 267–278, Apr. 2018.
- [5] Y. Wang, J. Tian, Z. Sun, *et al.*, “A comprehensive review of battery modeling and state estimation approaches for advanced battery management systems,” *Renewable and Sustainable Energy Reviews*, p. 110 015, Oct. 2020.
- [6] J. F. Peters, M. Baumann, B. Zimmermann, J. Braun, and M. Weil, “The environmental impact of li-ion batteries and the role of key parameters – a review,” *Renew. Sustain. Energy Rev.*, vol. 67, pp. 491–506, 2017.
- [7] G. Zubi, R. Dufo-López, M. Carvalho, and G. Pasaoglu, “The lithium-ion battery: State of the art and future perspectives,” *Renew. Sustain. Energy Rev.*, vol. 89, pp. 292–308, 2018.
- [8] M.-F. Ng, J. Zhao, Q. Yan, G. J. Conduit, and Z. W. Seh, “Predicting the state of charge and health of batteries using data-driven machine learning,” *Nat Mach Intell*, pp. 161–170, Mar. 2020.
- [9] K. A. Severson, P. M. Attia, N. Jin, *et al.*, “Data-driven prediction of battery cycle life before capacity degradation,” *Nat. Energy*, pp. 383–391, May 2019.
- [10] X. Hu, L. Xu, X. Lin, and M. Pecht, “Battery Lifetime Prognostics,” *Joule*, pp. 310–346, Feb. 2020.
- [11] V. Sulzer, P. Mohtat, A. Aitio, *et al.*, “The challenge and opportunity of battery lifetime prediction from field data,” *Joule*, Jul. 2021.
- [12] Y. Li, D. Karunathilake, D. M. Vilathgamuwa, *et al.*, “Model order reduction techniques for physics-based lithium-ion battery management: A survey,” *IEEE Ind. Electron. Mag.*, 2021.
- [13] X. Hu, H. Yuan, C. Zou, Z. Li, and L. Zhang, “Co-estimation of state of charge and state of health for lithium-ion batteries based on fractional-order calculus,” *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 10 319–10 329, 2018.

-
- [14] J. Wang, P. Liu, J. Hicks-Garner, *et al.*, “Cycle-life model for graphite-LiFePO₄ cells,” *J. Power Sources*, vol. 196, no. 8, pp. 3942–3948, 2011.
- [15] L. Lam and P. Bauer, “Practical capacity fading model for Li-ion battery cells in electric vehicles,” *IEEE Trans. Power Electron.*, vol. 28, no. 12, pp. 5910–5918, 2012.
- [16] W. He, N. Williard, M. Osterman, and M. Pecht, “Prognostics of lithium-ion batteries based on Dempster-Shafer theory and the Bayesian Monte Carlo method,” *J. Power Sources*, vol. 196, no. 23, pp. 10 314–10 321, 2011.
- [17] X. Tang, C. Zou, K. Yao, J. Lu, Y. Xia, and F. Gao, “Aging trajectory prediction for lithium-ion batteries via model migration and Bayesian Monte Carlo method,” *Appl. Energy*, vol. 254, p. 113 591, 2019.
- [18] I. Bloom, B. Cole, J. Sohn, *et al.*, “An accelerated calendar and cycle life study of Li-ion cells,” *J. Power Sources*, vol. 101, no. 2, pp. 238–247, 2001.
- [19] P. Arora, M. Doyle, and R. E. White, “Mathematical modeling of the lithium deposition overcharge reaction in lithium-ion batteries using carbon-based negative electrodes,” *J. Electrochem. Soc.*, vol. 146, no. 10, p. 3543, 1999.
- [20] J. Christensen and J. Newman, “A mathematical model for the lithium-ion negative electrode solid electrolyte interphase,” *J. Electrochem. Soc.*, vol. 151, no. 11, A1977, 2004.
- [21] P. Ramadass, B. Haran, P. M. Gomadam, R. White, and B. N. Popov, “Development of first principles capacity fade model for li-ion cells,” *J. Electrochem. Soc.*, vol. 151, no. 2, A196, 2004.
- [22] X.-G. Yang, Y. Leng, G. Zhang, S. Ge, and C.-Y. Wang, “Modeling of lithium plating induced aging of lithium-ion batteries: Transition from linear to nonlinear aging,” *J. Power Sources*, vol. 360, pp. 28–40, 2017.
- [23] R. Xiong, Y. Pan, W. Shen, H. Li, and F. Sun, “Lithium-ion battery aging mechanisms and diagnosis method for automotive applications: Recent advances and perspectives,” *Renewable and Sustainable Energy Reviews*, p. 110 048, Oct. 2020.

- [24] Y. Li, K. Liu, A. M. Foley, *et al.*, “Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review,” *Renewable and Sustainable Energy Reviews*, p. 109254, Oct. 2019.
- [25] M. Lucu, E. Martinez-Laserna, I. Gandiaga, and H. Camblong, “A critical review on self-adaptive Li-ion battery ageing models,” *J. Power Sources*, pp. 85–101, Oct. 2018.
- [26] M. Doyle, T. F. Fuller, and J. Newman, “Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell,” *J. Electrochem. Soc.*, vol. 140, no. 6, p. 1526, 1993.
- [27] V. Ramadesigan, K. Chen, N. A. Burns, V. Boovaragavan, R. D. Braatz, and V. R. Subramanian, “Parameter Estimation and Capacity Fade Analysis of Lithium-Ion Batteries Using Reformulated Models,” *J. Electrochem. Soc.*, p. 8, 2011.
- [28] K. Khodadadi Sadabadi, X. Jin, and G. Rizzoni, “Prediction of remaining useful life for a composite electrode lithium ion battery cell using an electrochemical model to estimate the state of health,” *J. Power Sources*, p. 228861, Jan. 2021.
- [29] J. M. Reniers, G. Mulder, and D. A. Howey, “Review and Performance Comparison of Mechanical-Chemical Degradation Models for Lithium-Ion Batteries,” *J. Electrochem. Soc.*, A3189–A3200, 2019.
- [30] B. M., B. Ph, B. F., *et al.*, “Main aging mechanisms in Li ion batteries,” *J. Power Sources*, pp. 90–96, Aug. 2005.
- [31] J. Vetter, P. Novák, M. Wagner, *et al.*, “Ageing mechanisms in lithium-ion batteries,” *J. Power Sources*, pp. 269–281, Sep. 2005.
- [32] J. C. Forman, S. J. Moura, J. L. Stein, and H. K. Fathy, “Genetic identification and Fisher identifiability analysis of the Doyle–Fuller–Newman model from experimental cycling of a LiFePO₄ cell,” *J. Power Sources*, vol. 210, pp. 263–275, 2012.
- [33] I. Bloom, J. Christophersen, and K. Gering, “Differential voltage analyses of high-power lithium-ion cells 2. Applications,” *J. Power Sources*, vol. 139, no. 1-2, pp. 304–313, Jan. 2005.
- [34] M. Dubarry and B. Y. Liaw, “Identify capacity fading mechanism in a commercial LiFePO₄ cell,” *J. Power Sources*, vol. 194, pp. 541–549, Oct. 2009.

-
- [35] T. Shibagaki, Y. Merla, and G. J. Offer, "Tracking degradation in lithium iron phosphate batteries using differential thermal voltammetry," *J. Power Sources*, vol. 374, pp. 188–195, Jan. 2018.
- [36] 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," *J. Power Sources*, vol. 482, p. 228 863, Jan. 2021.
- [37] X. Hu, Y. Che, X. Lin, and Z. Deng, "Health prognosis for electric vehicle battery packs: A data-driven approach," *IEEE/ASME Trans. Mechatronics*, pp. 1–1, 2020.
- [38] "Deep gaussian process regression for lithium-ion battery health prognosis and degradation mode diagnosis," *J. Power Sources*, vol. 445, p. 227 281, 2020, ISSN: 0378-7753.
- [39] X. Li, C. Yuan, and Z. Wang, "Multi-time-scale framework for prognostic health condition of lithium battery using modified Gaussian process regression and nonlinear regression," *J. Power Sources*, vol. 467, p. 228 358, Aug. 2020.
- [40] W. Li, M. Rentemeister, J. Badedo, D. Jöst, D. Schulte, and D. U. Sauer, "Digital twin for battery systems: Cloud battery management system with online state-of-charge and state-of-health estimation," *J. energy storage*, vol. 30, p. 101 557, 2020.
- [41] K. Li, P. Zhou, Y. Lu, X. Han, X. Li, and Y. Zheng, "Battery life estimation based on cloud data for electric vehicles," *J. Power Sources*, vol. 468, p. 228 192, 2020.
- [42] N. Subramanian and A. Jeyaraj, "Recent security challenges in cloud computing," *Comput. Electr. Eng.*, vol. 71, pp. 28–42, 2018.
- [43] R. R. Richardson, M. A. Osborne, and D. A. Howey, "Battery health prediction under generalized conditions using a Gaussian process transition model," *J. Energy Storage*, pp. 320–328, Jun. 2019.
- [44] S. Greenbank and D. Howey, "Automated feature extraction and selection for data-driven models of rapid battery capacity fade and end of life," *IEEE Trans. Ind. Inf.*, pp. 1–1, 2021.

- [45] S. J. Harris, D. J. Harris, and C. Li, "Failure statistics for commercial lithium ion batteries: A study of 24 pouch cells," *J. Power Sources*, vol. 342, pp. 589–597, Feb. 2017.
- [46] T. Baumhöfer, M. Brühl, S. Rothgang, and D. U. Sauer, "Production caused variation in capacity aging trend and correlation to initial cell performance," *J. Power Sources*, vol. 247, pp. 332–338, Feb. 2014.
- [47] X. Lin, H. E. Perez, J. B. Siegel, and A. G. Stefanopoulou, "Robust estimation of battery system temperature distribution under sparse sensing and uncertainty," *IEEE Trans. Contr. Syst. Technol.*, vol. 28, no. 3, pp. 753–765, 2019.
- [48] A. Allam, E. Catenaro, and S. Onori, "Pushing the envelope in battery estimation algorithms," *Iscience*, vol. 23, no. 12, p. 101 847, 2020.
- [49] W. Li, N. Sengupta, P. Dechent, D. Howey, A. Annaswamy, and D. U. Sauer, "One-shot battery degradation trajectory prediction with deep learning," *Journal of Power Sources*, vol. 506, p. 230 024, 2021, ISSN: 0378-7753.
- [50] G. dos Reis, C. Strange, M. Yadav, and S. Li, "Lithium-ion battery data and where to find it," *Energy and AI*, p. 100 081, Apr. 2021.
- [51] P. M. Attia, A. Grover, N. Jin, *et al.*, "Closed-loop optimization of fast-charging protocols for batteries with machine learning," *Nature*, vol. 578, no. 7795, pp. 397–402, Feb. 2020.
- [52] B. Bole, C. Kulkarni, and M. Daigle, "Adaptation of an Electrochemistry-based Li-Ion Battery Model to Account for Deterioration Observed Under Randomized Use," in *PHM 2014 - Proceedings of the Annual Conference of the Prognostics and Health Management Society 2014*, 2014.
- [53] M. Woody, M. Arbabzadeh, G. M. Lewis, G. A. Keoleian, and A. Stefanopoulou, "Strategies to limit degradation and maximize Li-ion battery service lifetime - Critical review and guidance for stakeholders," *J. Energy Storage*, p. 101 231, Apr. 2020.
- [54] C. R. Birkel, M. R. Roberts, E. McTurk, P. G. Bruce, and D. A. Howey, "Degradation diagnostics for lithium ion cells," *J. Power Sources*, pp. 373–386, Feb. 2017.

-
- [55] G. Ning, B. Haran, and B. N. Popov, “Capacity fade study of lithium-ion batteries cycled at high discharge rates,” *J. Power Sources*, vol. 117, pp. 160–169, May 2003.
- [56] T. Guan, S. Sun, F. Yu, *et al.*, “The degradation of LiCoO₂/graphite batteries at different rates,” *Electrochim. Acta*, vol. 279, pp. 204–212, Jul. 2018.
- [57] T. Guan, P. Zuo, S. Sun, *et al.*, “Degradation mechanism of LiCoO₂/mesocarbon microbeads battery based on accelerated aging tests,” *J. Power Sources*, vol. 268, pp. 816–823, Dec. 2014.
- [58] D. Wong, B. Shrestha, D. A. Wetz, and J. M. Heinzel, “Impact of high rate discharge on the aging of lithium nickel cobalt aluminum oxide batteries,” *J. Power Sources*, vol. 280, pp. 363–372, Apr. 2015.
- [59] T. Waldmann, M. Wilka, M. Kasper, M. Fleischhammer, and M. Wohlfahrt-Mehrens, “Temperature dependent ageing mechanisms in Lithium-ion batteries – A Post-Mortem study,” *J. Power Sources*, vol. 262, pp. 129–135, Sep. 2014.
- [60] “Spearman rank correlation coefficient,” in *The Concise Encyclopedia of Statistics*. New York, NY: Springer New York, 2008, pp. 502–505, ISBN: 978-0-387-32833-1.
- [61] M. Kuhn and K. Johnson. “Applied predictive modeling.” (2013).
- [62] A. J. Smola and B. Schölkopf, “A tutorial on support vector regression,” *Stat. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004.
- [63] A. Lindholm, N. Wahlström, F. Lindsten, and T. B. Schön, *Machine Learning - A First Course for Engineers and Scientists*. Cambridge University Press, 2022.
- [64] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, pp. 5–32, 2001.
- [65] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning* (Adaptive computation and machine learning). Cambridge, Mass: MIT Press, 2006.
- [66] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [67] L. Ljung, *System Identification (2nd Ed.): Theory for the User*. USA: Prentice Hall PTR, 1999, ISBN: 0136566952.

- [68] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, “Scikit-learn: Machine learning in python,” *J. Mach. Learn. Res*, vol. 12, pp. 2825–2830, 2011.
- [69] F. Chollet *et al.*, *Keras*, <https://keras.io>, Accessed: 02-October-2020, 2015.
- [70] D. Roman, S. Saxena, V. Robu, M. Pecht, and D. Flynn, “Machine learning pipeline for battery state-of-health estimation,” *Nat. Mach. Intell.*, Apr. 2021.

PAPER C

**Data-driven battery life prediction considering both onsite
measurement and usage information**

Yizhou Zhang, Torsten Wik, Yicun Huang, John Bergström, Changfu Zou

22nd IFAC World Congress 2023, July 2023, Yokohama, Japan

The layout has been revised.

Abstract

Due to dynamic operating conditions, random user behaviors, and cell-to-cell variations, accurately predicting battery life is challenging, especially using information from only a few early cycles. This work proposes a data-driven battery early-life prediction pipeline using both time-series, measurement-related features and usage-related features. By comparing the prediction performance of using these two feature sources individually, we show that they are effectively interchangeable and complementary to one another. Then two methods of systematically combining these two feature sources are devised, both achieving excellent prediction results with a root mean square error of around 150 cycles using only the first 100 cycle's data. Additionally, four machine learning algorithms with different characteristics are applied to compare their performances on battery prognostic problems. Finally, experimental data of different cell types and cycling conditions are used to verify the effectiveness and the generality of the developed method.

1 Introduction

For a more sustainable future, many efforts have been put into different sectors, such as transportation and electricity production, to replace fossil fuel usage with sustainable solutions. Lithium-ion (Li-ion) batteries play an important role in this transition by serving as traction batteries for electric vehicles or energy storage devices for the power grid, due to their lower cost, higher reliability, and longer lifetime compared to other alternative solutions. However, as electrochemical devices, Li-ion batteries are also plagued by their nonlinear, complex, and path-dependent aging characteristics, imposing significant challenges for their broad adoption by end customers [1]. Therefore, optimally controlling battery usage to extend their lifetime and enhance performance becomes indispensable. To this end, an accurate and robust early life prediction algorithm is pivotal to pave the way for a later health-conscious control strategy.

Existing battery lifetime prediction approaches can be divided into empirical, physics-based, and data-driven methods [2]. The simplicity and low computational need of the empirical battery aging models made it an attractive way in the early days. However, the dynamic operating profiles and cell-to-cell variations in real-life applications make such methods lose accuracy and reliability. Aging models can also be derived from first principles of the internal battery mechanical and electrochemical processes for use in the physics-based method. Despite their profound, detailed insights into the aging mechanisms, the complicated parameterization and high computation requirement make such methods unpractical for online applications. Data-driven methods though are flexible, mechanism-agnostic, and can recognize trends and patterns under complex and dynamic situations, thus being a promising way to solve battery life prediction problems.

The choice of inputs to a data-driven model, often referred to as features, is of great importance for such a method. Many efforts have been made to extract features from the charge or discharge time-series measurement data to predict battery capacity, lifetime, and knee point [2]. Additionally, different features based on electrochemical impedance spectroscopy (EIS), cell pressure change, or acoustic analysis have also been investigated [3]. However, the cost of instrumentation necessary to collect these measurements may hinder the applicability of such methods. As most of the data-driven methods are developed based on a repetitive cycling of the battery in a laboratory, deploying such a model in real-world applications undergoing diverse cycling profiles and an ever-changing cycling environment is very challenging [4]. Some early works have used usage-related features, where only information of how the battery has been used is recorded, to successfully predict the aging trajectory of traction batteries used in the vehicle [5]. However, to the best of our knowledge, there is no work using such information to forecast battery life and investigating its relationship to methods using time-series measurements.

In this work, we aim to fill the identified research gap by proposing a data-driven battery early-life prediction pipeline using both time-series measurement data and usage-related information. The contributions arise from three perspectives. First, we demonstrate that using usage-related features alone can achieve the same prediction performance as using the time-series data, indicating that these two feature sources are effectively interchangeable and also complementary. Second, we developed two methods to systematically combine

the information from these two feature sources, both methods achieving excellent performance with increased accuracy and tightened confidence intervals compared to using them individually. Last, the generality of the developed pipeline is demonstrated on two different data sources of different cell chemistry.

2 Battery dataset introduction

Two battery datasets are investigated in this study, one originating from Stanford University, which will be referred to as the Stanford dataset, and another from Sandia National Laboratories (SNL), which will be called the SNL dataset hereafter [6], [7]. For the Stanford dataset, the cells used in the test campaign are of Lithium-iron phosphate (LFP) type, whereas for the SNL dataset, nickel magnesium cobalt (NMC) type batteries were used. Moreover, the cycling profile in the Stanford dataset is various fast charging policies with identical 4C constant discharge profiles, while in the SNL dataset more diverse cycling profiles are used to investigate their impact on cell aging performance. The detailed cell specifications are listed in Table 1.

Table 1: Battery cell specifications

Dataset	Stanford	SNL
Battery type	LFP	NMC
Manufacture	A123	LG Chem
Nominal capacity (Ah)	1.1	3
Voltage range (V)	2 to 3.6	2 to 4.2
Max discharge current (A)	30	20
Operating Temperature (°C)	-30 to 60	-5 to 50

3 Feature construction and engineering

3.1 Usage-related features

The usage profiles have a profound impact on the battery life. Therefore, we presume that incorporating both the historic usage recordings and predicted future usage patterns will considerably improve the prediction accuracy and

robustness. As per Woody *et al.*, the following usage-related features, also known as stress factors, are of interest: depth of discharge (DoD), charge current rate, discharge current rate, cycling/calendar temperature, voltage, SoC and cycling/calendar time. These stress factors form the foundation for the usage-related feature pool that we will construct for later machine learning (ML) algorithm training. It is worth noting that not all features mentioned are available for extraction due to limitations of the datasets. For example, in the Stanford dataset only the charge current and charge time can be used due to other features being identical for all the batteries across the whole dataset.

For the Stanford dataset, we construct our features based on the fact that most cells experience a unique charging policy. First, the complete SoC range is divided into twenty intervals of 5%. Then, the corresponding average charge current in C-rate is assigned to each interval to form a feature vector of 20 elements. The statistical property of this vector is then calculated to reduce the feature dimension with the hope of not affecting the prediction performance. Apart from using the charge current information, the charging time of the first cycle is also used as one of the features to indicate how the battery is being charged. As for the SNL dataset, the initial SoC, DoD, discharge current, mean and variance of the ambient cycling temperature, and mean and variance of the cell temperature are selected, all as scalar values.

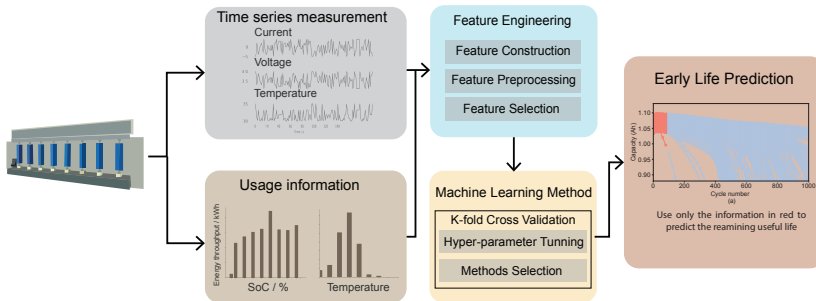


Figure 1: The overall battery early life prediction pipeline.

3.2 Time-series, measurement-related features

With a fixed charge or discharge policy, both the current and voltage curves will gradually change when the battery degrades. Considering the hidden aging mechanisms that can potentially be manifested by the current and voltage profiles, the on-purpose feature construction based on these time-series measurement data can be quite useful to predict the lifetime of the battery. Here we first investigate the cycle-to-cycle evolution of the accumulated charge/discharge capacity as a function of the voltage curve ($dQ(V)$), which has been applied to successfully predict the battery lifetime [6]. Second, we study the effectiveness of using the features constructed from the incremental capacity (IC) curves. IC is the ratio between the difference of the accumulated charge capacity and the difference of the voltage over a given time interval, which is a powerful tool for battery aging diagnostics. Thus, we hypothesize that the rich aging information that the IC curve contains can also make IC-related features a good health indicator for battery early life prediction. In practice, for the Stanford dataset, we adopt the $dQ(V)$ curve and the IC curve during discharge as the baseline features because the batteries experience the same constant discharge profile across the whole dataset. Conversely, with the same reasoning, we use the charge part of the profile to form the baseline features for the SNL dataset.

4 Early prediction model

The problem of battery early lifetime prediction is formulated as a regression problem with the aim of minimizing the true measured lifetime and the model predictions for all batteries in the test set. Four ML algorithms are considered in the prediction pipeline to evaluate their performance in such cases. Among them, two are linear models, i.e., elastic net (EN) and Bayesian ridge regression (BRR), and the other two are nonlinear support vector regression (SVR) and random forest regression (RFR) models. Fig. 1 illustrates the overall prediction pipeline for the battery early lifetime prediction.

Linear models have advantages, such as efficient training and relatively low computational cost during online deployment, making them the first choice when the accuracy of the model output can meet the requirement of the desired applications. Additionally, the simple and neat model structure makes the final prediction result highly interpretable. EN consists of an ordinary least

square formulation together with two regularization terms, e.g., L1 and L2 regularization, with the goal of solving the following optimization problem:

$$\hat{\theta} = \arg \min_{\theta} \|y - X\theta\|_2^2 + \lambda \frac{1-\alpha}{2} \|\theta\|_2^2 + \alpha \|\theta\|_1, \quad (\text{C.1})$$

where y is the prediction target, in this case the measured lifetime of the battery, X is the feature matrix, θ is the model coefficient vector, and both λ and α are hyperparameters that are tuned using cross-validation. Similarly, applying a Bayesian approach to the ordinary least square problem leads to the BRR. The algorithm can incorporate the regularization parameters in the estimation procedure and naturally and systematically propagate its prediction uncertainty.

A nonlinear model, such as SVR and RFR, may require a complex calculation procedure for both the training and testing but may potentially increase the prediction accuracy and robustness for certain applications. Mathematically, an SVR algorithm tries to minimize the ϵ -insensitive loss:

$$L(y, \hat{y}) = \begin{cases} 0 & \text{in } |y - \hat{y}| < \epsilon \\ |y - \hat{y}| - \epsilon & \text{otherwise,} \end{cases} \quad (\text{C.2})$$

where \hat{y} is the predicted lifetime value. By introducing the slack variables and constructing the Lagrangian of the primal formulation, we can find the solution to such an optimization problem. The built-in sparsity characteristics and outstanding fitting performance for nonlinear systems make SVR a popular algorithm for battery prognostics. RFR is an ensemble learning method, which is used through bootstrapping and data aggregation to achieve a good bias and variance trade-off. Due to its ensemble characteristics, RFR is a robust and accurate supervised learning algorithm for highly nonlinear systems with complicated data and features. To quantify the estimation uncertainty, we adopt the method from [9], using the bootstrap replicates during the training process to calculate the confidence interval of the prediction result for RFR.

Combining the features from different sources during the training process is one way to systematically incorporate information from both the time-series measurement information and the battery usage information. Alternatively, fusing the prediction result from two independent models trained with different feature inputs is also a plausible candidate. One of the motivations

for considering both is to have the flexibility to choose which model to use depending on the availability of the feature input, as practical data collection challenges (communication delays, data corruption, or memory shortage) may hinder the applicability of a certain type of the feature input. Herein, we adopt an inverse-variance weighting method to aggregate the prediction results, for which the fused prediction result is given by:

$$\hat{y}_f = \frac{\sum_m \hat{y}_m / \sigma_m^2}{\sum_m 1 / \sigma_m^2}, \quad (\text{C.3})$$

where \hat{y}_f is the fused result, \hat{y}_m is the individual prediction result using model m , and σ_m is the standard deviation of the prediction result from model m . Correspondingly, the prediction variance of the fused result can be calculated as follows:

$$\text{Var}(\hat{y}_f) = \frac{1}{\sum_m 1 / \sigma_m^2}. \quad (\text{C.4})$$

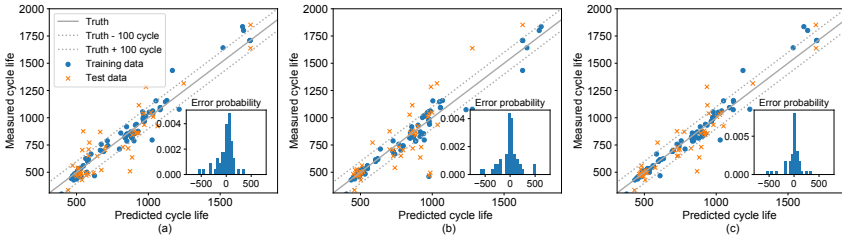


Figure 2: Prediction results using different feature inputs for Stanford dataset. (a) Using time-series, measurement-related features only. (b) Using usage-related features only. (c) Using both feature sources.

5 Evaluation Matrices

For the train-test set split, a ratio of 8/2 is adopted, where 80% of the batteries are randomly selected as the train set, and the remaining unseen data are assigned to the test set. Three error evaluation variables are chosen to quantify the performance of the developed method, i.e., coefficient of determination (R^2), root mean squared error (RMSE), and mean absolute percentage

error (MAPE), which are defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (\text{C.5})$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (\text{C.6})$$

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{y_i} \times 100\%, \quad (\text{C.7})$$

where \bar{y} is the mean value of the measured lifetime and N is the number of data points in the test set.

6 Results and discussions

We first discuss the results obtained from the Stanford dataset, followed by the SNL dataset. To compare the prediction performance using different feature sources, we applied the same ML algorithm, RFR in this case, but trained with different feature inputs. The results are summarized in Table 2. It can be seen that the prediction performance of using time-series, measurement-related features, or usage-related features alone are similar in terms of the MAPE, with usage-related features slightly outperforming the time-series, measurement-related features. Based on this, we could conclude that the two feature sources are effectively interchangeable for battery life prediction. However, it is worth highlighting that when both feature sources are used, significant performance improvement has been achieved, showing that the two feature sources are complementary and should be used together when possible. By examining the detailed prediction result shown in Fig. 2, the model trained with combined features tends to violate less the ± 100 cycles prediction boundary of the true measured lifetime when compared to using individual feature sources. Additionally, the zoom-in figures show the error histogram of the prediction results using different feature sources. This result again shows the superiority of combining both feature sources with less frequent extreme prediction and narrower distribution shapes. The superiority stems from the fact that the usage-related feature may indicate how the cycling profiles affect the battery life in an average sense, while the time-series, measurement-related

features are able to identify the cell-to-cell variations. Fig. 3 illustrates the

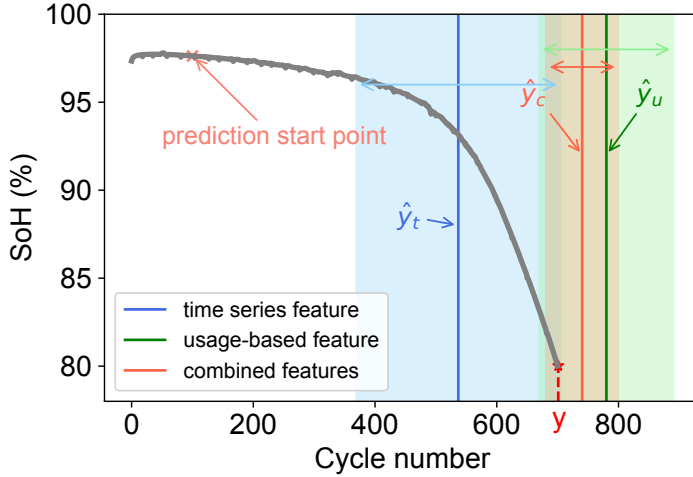


Figure 3: The prediction result and its confidence interval using different feature sources were demonstrated on a randomly selected cell, where the shaded color illustrates the 95% prediction confidence interval. \hat{y}_c , \hat{y}_t , and \hat{y}_u indicate the predicted cycle life and y is the measured true cycle life.

prediction result on a randomly selected cell in the test set using different feature sources. The prediction result of using combined feature sources is not only better in terms of prediction accuracy but it also has a narrower prediction confidence interval compared to the other two predictions using only one feature source. As stated in Section 4, instead of combining the fea-

Table 2: Results of different feature input and two combination methods for battery early lifetime prediction using Stanford dataset

Feature input	R^2	RMSE (cycles)	MAPE (%)
Time-series	0.78	197	15.41
Usage related	0.85	162	14.31
combined features	0.87	149	10.51
combined models	0.87	151	10.11

ture sources and then training one ML algorithm, an alternative option is to

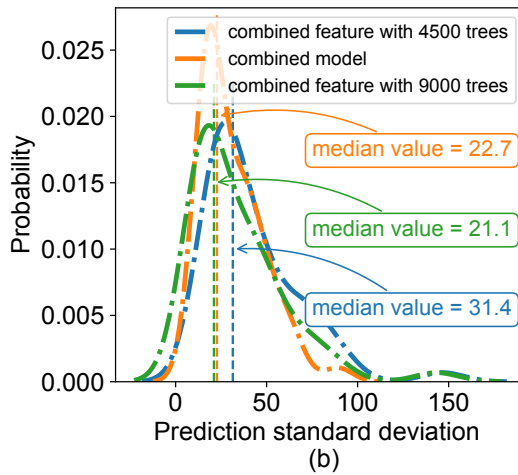
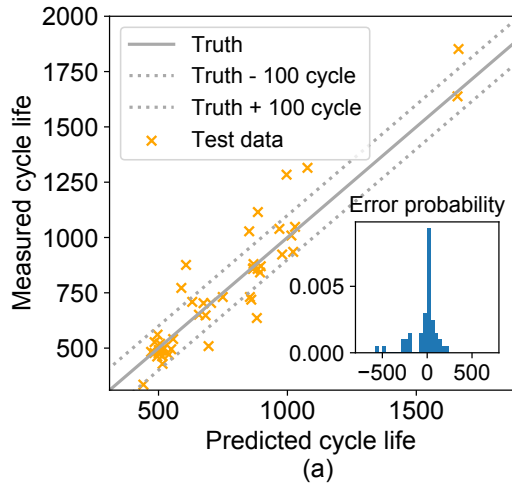


Figure 4: (a) Detailed prediction result by systematically fusing the individual prediction output trained with different feature sources using RFR. (b) Comparison result of the prediction standard deviation for different combination methods.

train two ML algorithms using different feature inputs and fuse the prediction result. The last row in Table 2 shows the numerical result of training two

RFR prediction models and then fusing the prediction results of the two using (C.3). Not surprisingly, the obtained results are similar to the ones using combined features to train one ML algorithm. The detailed prediction results in Fig. 2(c) and Fig. 4(a) also verify the similar prediction performance of the two methods. We would like to attribute this result to the fact that these two feature sources are largely complementary. Nevertheless, when using the same number of trees in the RFR algorithm to propagate prediction uncertainty, one of the advantages of training two RFR models and fusing the final results is that the confidence interval of the outputs is considerably tightened. Additionally, to achieve a similar confidence interval level, the number of trees needs to be doubled for the model combining the feature sources. Therefore, combining the results can save computational resources. The histogram of the standard deviation of the prediction uncertainty and their median value for different methods are shown in Fig. 4(b).

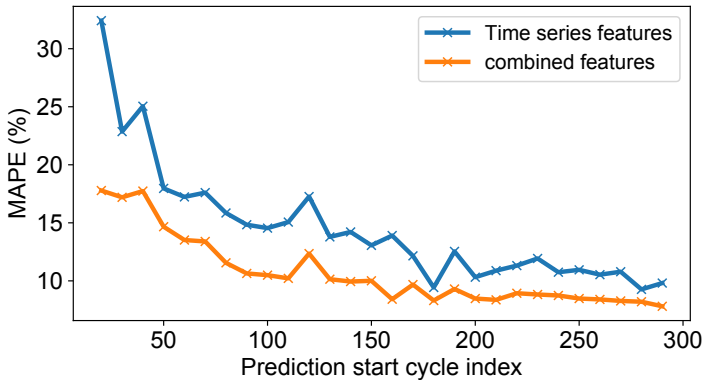


Figure 5: Prediction error as a function of prediction start cycle number.

It is worth mentioning that all the prediction results acquired so far only use the cycling information from the first 100 cycles. Therefore, to demonstrate the robustness of the developed method, we evaluate the prediction result from as early as the 20th cycle to the 300th cycle with the prediction MAPE shown in Fig. 5. It can be noted that incorporating usage-related features can dramatically increase the prediction accuracy, especially when performing early cycle prediction. For example, when we start from the 20th cycle, the prediction accuracy can be improved by around 45%. Furthermore, the

superiority of using combined features is consistent over the whole examined prediction start range, verifying the importance of having such information included in the feature construction step. Finally, the prediction results of

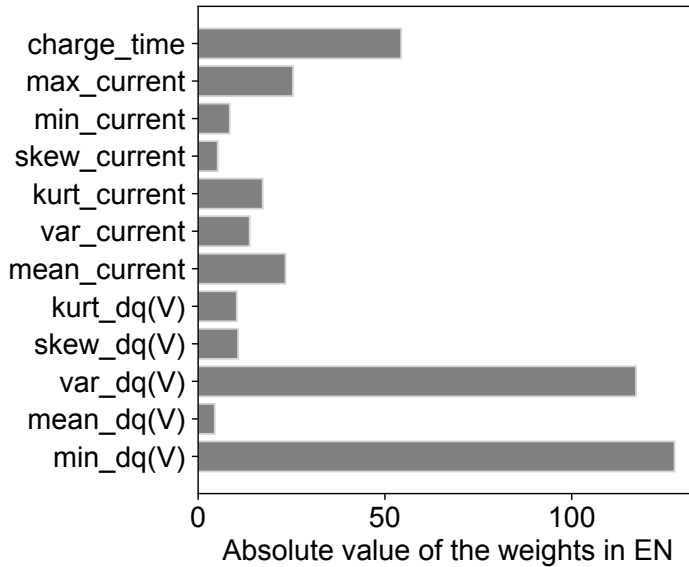


Figure 6: Feature importance analysis.

different ML algorithms trained with combined feature input are reported in Table 3. The best-performing algorithm, in this case, is RFR with a MAPE of around 10%. However, we want to emphasize the importance of having several ML algorithms of different characteristics in the prediction pipeline, as which algorithm that performs the best may depend on the dataset.

The model parameter θ of the EN model can be used to evaluate how much each feature contribute to the final prediction result. Fig. 6 shows the absolute value of the weighting coefficient for the selected features. Among these, the variance and the minimum value of the $dQ(V)$ curve are the most important features with the highest contribution to the final prediction result, which is in line with the conclusion drawn in [6]. Other than that, the next most important features are extracted from the charging time and the mean charge current, which demonstrate the effectiveness of including the usage-

related information in the feature pool to increase prediction accuracy and robustness.

Table 3: Results of different ML algorithms for battery life prediction using Stanford dataset

ML algorithm	R^2	RMSE (cycles)	MAPE (%)
EN	0.83	175	17.91
BR	0.81	181	14.73
SVR	0.61	263	20.62
RFR	0.87	149	10.51

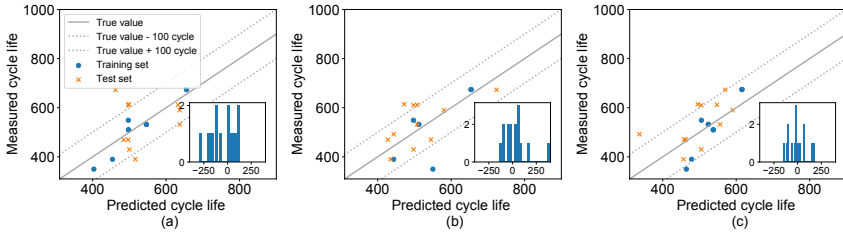


Figure 7: Prediction results using different feature input for SNL dataset. (a) Using time-series, measurement-related features only. (b) Using usage-related features only. (c) Using both feature sources.

To verify the generalizability of the proposed method, we applied our approach to the SNL dataset with NMC-type chemistry. The numerical prediction results of various kinds of feature inputs are shown in Table 4, and the detailed prediction results are illustrated in Fig. 7. Generally, the conclusions that we drew previously on the Stanford dataset are also valid here. Comparatively, the prediction performance discrepancy between using the time-series, measurement-related features, and using usage-related features is larger in the SNL dataset. The reason could be that the cells in this dataset experience more diverse cycling profiles than the ones in the Stanford dataset. Moreover, the charging part of the $dQ(V)$ curve does not manifest as much aging information as the discharging part of the $dQ(V)$ curve, which has also been noticed by [10]. Another interesting finding is that cell consistency, which is how much the battery lifetime varies if the cells experience exactly the same cycling profile, will considerably affect the prediction accuracy using usage-

related features. We have noted here that the lifetime variance of the cells that conduct the same cycling profiles inside the Stanford dataset is 78 cycles, whereas for the SNL dataset it is 42 cycles. Consequently, the prediction accuracy of the SNL dataset adopting usage-related features is also higher than the one for the Stanford dataset.

Table 4: Results of different feature input for battery life prediction using SNL dataset

Feature input	R^2	RMSE (cycles)	MAPE (%)
Time-series	0.95	136	16.38
Usage related	0.96	131	11.32
combined features	0.98	94	10.52

7 Conclusion

In this work, we developed a battery early lifetime prediction pipeline that is accurate and robust under various cycling profiles with different cell chemistry, highlighting the importance and benefits of simultaneously using both time-series, measurement-related features and usage-related features. We foresee such a method should be considered as a standard way to formulate feature pools for battery prognostic problems.

In the future, the efficacy of such a prediction model under more dynamic usage protocols, e.g., the NEDC cycle or even using field data directly, should be investigated. More broadly, further development of the optimal control strategy to extend the lifetime and enhance the battery’s performance based on such an early lifetime prediction model can bring more value to the battery communities.

References

- [1] C. R. Birkl, M. R. Roberts, E. McTurk, P. G. Bruce, and D. A. Howey, “Degradation diagnostics for lithium ion cells,” *J. Power Sources*, pp. 373–386, Feb. 2017.

-
- [2] Y. Li, K. Liu, A. M. Foley, *et al.*, “Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review,” *Renewable and Sustainable Energy Reviews*, p. 109 254, Oct. 2019.
 - [3] S. Yang, C. Zhang, J. Jiang, W. Zhang, L. Zhang, and Y. Wang, “Review on state-of-health of lithium-ion batteries: Characterizations, estimations and applications,” *J. Cleaner Production*, vol. 314, p. 128 015, 2021, ISSN: 0959-6526.
 - [4] V. Sulzer, P. Mohtat, A. Aitio, *et al.*, “The challenge and opportunity of battery lifetime prediction from field data,” *Joule*, Jul. 2021.
 - [5] Y. Zhang, T. Wik, J. Bergström, M. Pecht, and C. Zou, “A machine learning-based framework for online prediction of battery ageing trajectory and lifetime using histogram data,” *J. Power Sources*, Apr. 2022.
 - [6] K. A. Severson, P. M. Attia, N. Jin, *et al.*, “Data-driven prediction of battery cycle life before capacity degradation,” *Nat. Energy*, pp. 383–391, May 2019.
 - [7] Y. Preger, H. M. Barkholtz, A. Fresquez, *et al.*, “Degradation of Commercial Lithium-Ion Cells as a Function of Chemistry and Cycling Conditions,” *J. Electrochem. Soc.*, vol. 167, Sep. 2020.
 - [8] M. Woody, M. Arbabzadeh, G. M. Lewis, G. A. Keoleian, and A. Stefanopoulou, “Strategies to limit degradation and maximize Li-ion battery service lifetime - Critical review and guidance for stakeholders,” *J. Energy Storage*, p. 101 231, Apr. 2020.
 - [9] S. Wager, T. Hastie, and B. Efron, “Confidence intervals for random forests: The jackknife and the infinitesimal jackknife,” *J. Machine Learning Research*, vol. 15, no. 1, pp. 1625–1651, 2014.
 - [10] V. Sulzer, P. Mohtat, S. Lee, J. B. Siegel, and A. G. Stefanopoulou, “Promise and challenges of a data-driven approach for battery lifetime prognostics,” in *2021 American Control Conference (ACC)*, 2021, pp. 4427–4433.

PAPER **D**

Machine learning-based lifelong estimation of lithium plating potential: A path to health-aware fastest battery charging

Yizhou Zhang, Torsten Wik, John Bergström, Changfu Zou

Accepted for publication in Energy Storage Materials

The layout has been revised.

Abstract

To enable a shift from fossil fuels to renewable and sustainable transport, batteries must allow fast charging and exhibit extended lifetimes—objectives that traditionally conflict. Current charging technologies often compromise one attribute for the other, leading to either inconvenience or diminished resource efficiency in battery-powered vehicles. For lithium-ion batteries, the way to meet both objectives is for the lithium plating potential at the anode surface to remain positive. In this study, we address this challenge by introducing a novel method that involves real-time monitoring and control of the plating potential in lithium-ion battery cells throughout their lifespan. Our experimental results on three-electrode cells reveal that our approach can enable batteries to charge at least 30% faster while almost doubling their lifetime. To facilitate the adoption of these findings in commercial applications, we propose a machine learning-based framework for lifelong plating potential estimation, utilizing readily available battery data from electric vehicles. The resulting model demonstrates high fidelity and robustness under diverse operating conditions, achieving a mean absolute error of merely 3.37 mV. This research outlines a practical methodology to prevent lithium plating and enable the fastest health-conscious battery charging.

1 Introduction

At the forefront of a transition to sustainable mobility are lithium-ion (Li-ion) batteries, because of their high energy density, cost-effectiveness, and durability, establishing themselves as the primary power source for electric vehicles (EVs) [1]–[4]. Yet, as the industry strives to match the convenience of traditional fueling methods, it encounters significant hurdles in fast-charging technology that do not compromise battery lifespan [5].

Rapid charging introduces a dilemma; on the one hand, it necessitates high current levels, leading to excessive heat generation that, if not adequately

dissipated through advanced thermal management systems, can significantly accelerate several battery aging mechanisms, such as solid electrolyte interface (SEI) growth [6], [7]. On the other, even when thermal issues are addressed, the escalated current can trigger intricate mechanical and electrochemical reactions within the battery, further exacerbating its degradation [8]. These challenges have catalyzed a plethora of research aimed at developing fast-charging strategies while protecting the health of the battery [9], [10].

Among the myriad of factors influencing battery degradation during fast charging, lithium plating emerges as a critical concern [10]–[12]. This phenomenon—characterized by the deposition of metallic lithium on the anode’s surface—directly undermines the battery’s capacity and efficiency by reducing the cyclable lithium and impeding the normal intercalation process. The consensus among researchers is that lithium plating needs to be prevented for prolonged battery life and maintained performance during rapid charging [13]–[16]. Recent studies have demonstrated that the microstructural inhomogeneities in graphite electrodes significantly influence the onset and severity of lithium plating, necessitating the refinement of both electrode design and charging protocols [17]. Despite a well-established theoretical foundation for regulating the plating potential to mitigate this issue [5], [13], [18], empirical validation and practical application of such strategies, especially over the battery’s lifespan, remain limited.

Methods to extract the information of plating potential can be categorized into three groups, including direct/indirect measurements, model-based methods, and data-driven methods [5], [8]. The most straightforward approach involves inserting a lithium metal reference electrode between the negative electrode and separator of a battery cell to enable the measurement of the anode voltage during battery operation [18], [19]. However, the insertion of such reference electrodes can be practically challenging due to the incurred additional costs and space requirements, making it unfeasible for any cost- and energy-density-sensitive applications. Moreover, the normal electrochemical reaction may be affected by the inserted reference electrode, for example, by blocking certain areas for Li-ion intercalation and deintercalation [20], [21]. Last but not least, if lithium metal was used as the reference electrode, the extra lithium added to the cell may participate in the charge and discharge reactions, potentially altering the cell’s behavior compared to a normal cell without reference electrodes.

Using a high-fidelity model to estimate the detailed internal battery states is another way to capture the plating potential information during battery usage. Lu *et al.* [22] constructed an equivalent circuit model (ECM) and used an extended Kalman filter to estimate the plating potential. However, due to the coupled effect of the cathode and anode, the plating potential may become unidentifiable. Another commonly employed class of model for such an estimation task is physics-based models, such as the pseudo-two-dimensional (P2D) model [23]. In contrast to the ECM, which only captures the battery's lumped electric behavior, the P2D model uses the porous electrode and concentrated solution theory to simulate the internal electrochemical reactions, which can predict the plating potential's evaluation in response to external operating conditions. Based on the P2D model coupled with thermal dynamics, Ringback *et al.* [24] designed an observer to estimate the plating potential together with its lower bound in the presence of model uncertainties. Compared to employing the full-order electrochemical model, which requires an excessive amount of computational power and a complex solver to solve the partial differential-algebraic equations (PDAE), control-oriented models that require light computation while capturing key battery dynamics are desired. Following this trend, Li *et al.* [25] designed an observer for the plating potential based on the single particle model (SPM). The SPM can achieve reasonable accuracy at low to moderate current rates but lose fidelity at high currents. Additionally, to successfully use physics-based models, it is essential to have accurate parameters. However, identifying a large set of parameters over various battery operation conditions, e.g., different state of charge (SoC) levels, current profiles, temperatures, and different health levels, remains an open research question [26], [27]. Moreover, even with a reliable and simplified electrochemical model, running it for every battery cell is still computationally challenging, considering the targeted applications often contain hundreds or even thousands of cells.

Due to the ever-increasing awareness of collecting battery operation data and employing the concept of cloud battery management systems (BMS) to access much higher computational power, data-driven methods have gained popularity in dealing with battery diagnosis and prognosis. With sufficiently large and high-quality datasets, data-driven methods often demonstrate excellent performance in estimating or predicting battery SoC, state of health (SoH), and remaining useful life (RUL) [28]–[31]. In this regard, data-driven

methods can be a good candidate, whereas there are very limited attempts to estimate battery plating potential and avoid the onset of lithium plating during charging. Lin [32] adopted a long short-term memory neural network (LSTM) to estimate the plating potential using synthetic data generated from a physics-based battery model. However, only limited operating conditions were considered, e.g., the constant current charging, standard driving profiles, and short time horizon, which may hinder the model’s applicability. Hamar *et al.* [33] further extended the work to incorporate more realistic operating conditions with different cycling temperatures, various initial SoC levels, and multi-step constant charging. Nonetheless, under the same charging conditions repeated over different cycles, the plating potential can be significantly affected by how the battery has aged [14], [18]. In fact, old batteries are more prone to trigger lithium plating. Therefore, it is crucial to consider the impact of battery aging on the plating potential. To the best of our knowledge, there is no established work capable of accurately and online estimating the plating potential over the battery’s entire life.

Our study aims to combine theoretical insights with practical applications in the fast charging of batteries, driven by two principal innovations. First, we validate the effectiveness of a plating potential controlled charging method using three-electrode battery cells, demonstrating marked improvements in both battery lifetime and charging rate over conventional methods. Building upon this foundation, we then develop an innovative ML-based framework for the lifelong estimation of battery plating potential, thus avoiding the three-electrode setup. This framework utilizes real-time measurements to learn electrochemical behaviors on a fast timescale alongside online estimated SoH indicators for which the effect of aging on plating potential has been learned. Comprehensive validation of this framework through extensive battery cycling data under diverse operating conditions and aging levels demonstrates its robustness and accuracy. These innovations collectively represent a significant stride towards health-aware, fast-charging solutions for EV batteries.

2 Results and discussion

2.1 Three-electrode cell experiments

In the experimental setup, the EL-CELL PAT-Cell was used for testing three-electrode cells, as illustrated in Fig. 1a and b. For the three-electrode cells, Nickle Manganese Cobalt (NMC) oxides of type 811 were used as the positive electrode, and artificial graphite was used as the negative electrode, both supplied by Customcells Itzehoe GmbH. Detailed specifications of the electrode materials are available in the *Supplemental Information* (i.e., Note 4). All the cell assembling operations were carried out in an argon-filled glove box. First, the electrode sheets were die-cut to a coin size with a diameter of 18 mm. Then, the positive electrode, the negative electrode, and the separator were assembled into the PAT-Cell stack. Borosilicate glass fiber separators (260 μm , Whatman FG/A) were used, in which a lithium metal ring served as a reference electrode to measure the electrode potential during battery operation. The electrolyte used was a 90 μL lithium hexafluorophosphate solution in a mixture of ethylene carbonate and dimethyl carbonate (LiPF₆ LP50). The nominal capacity of the assembled three-electrode cells is 5 mAh, which is used as the basis when defining C-rates in subsequent test instructions.

The experimental procedures were carried out in a controlled temperature chamber (ESPEC LHU-124), utilizing the PAT-Stand-1 cell holder, and subjected to cycling using a NEWARE coin cell cycler. The temperature chamber was maintained at a constant temperature of 25°C. After the assembly of the cells, a formation cycle was executed following the guidelines provided by Customcells. The formation cycle profile was first, according to Profile 1, repeated twice as listed in Table 1, then following Profile 2, repeated twice. To assess the aging status of the battery cells, reference performance tests (RPTs) were scheduled every 25 normal cycles. During these tests, Profile 3 was executed twice, and the average value of the two discharge capacities was defined as the cell's reference capacity.

In this study, a total of four cells were assembled and used in the experimental campaign. Two of the cells served as a reference group, undergoing constant current-constant voltage (CC-CV) charging according to Profile 4 in Table 1. The charging process for these cells was terminated when the current met 0.1C or the accumulated charged capacity reached 4 mAh, which is equivalent to 80% of the nominal capacity (see Fig. 1d). The remaining two

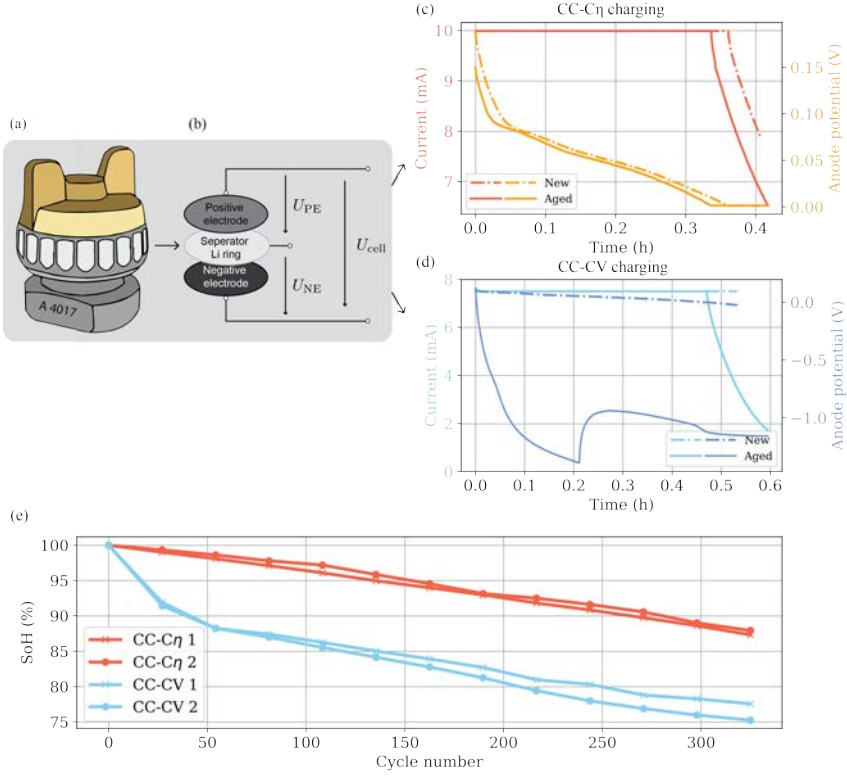


Figure 1: The three-electrode cell test setup used for experiments. (a) shows the EL-cell three-electrode cell holder. (b) illustrates the internal structure of the cell and the measured signals, including the cell voltage U_{cell} and the anode voltage U_{NE} . (c) illustrates the current and anode potential profiles of a studied cell at different aging levels and under CC-C η charging. (d) illustrates the current and anode potential profiles of another studied cell under CC-CV charging. (e) presents the aging curves of four cells experiencing the two different charging strategies.

cells were operated under a lithium-plating-free charging strategy, wherein the charging current was controlled to maintain non-negativity of the plating potential (η), i.e. ensuring $\eta > 0$. A safety margin of 1 mV is applied to account for the potential variability of the plating potential measurement. Designated as Profile 5 and referred to as CC-C η charging, this approach constituted the

Table 1: CC-CV charging and CC discharging profiles for the designed three-electrode cell experiment.

		Charge phase			Discharge phase	
		Current	Voltage limit	Cut-off condition	Current	Voltage limit
CC-CV	Profile 1	0.1C	4.2V	1/25C	0.1C	3V
	Profile 2	0.2C	4.2V	1/25C	0.2C	3V
	Profile 3	1C	4.2V	1/10C	1C	3V
	Profile 4	1.5C	4.2V	1/10C or 4 mAh	1C	3V
CC-C η	Profile 5	2C	4.5V	4 mAh	1C	3V

experimental group for advanced comparative analysis (depicted in Fig. 1c). Here, a higher current value was adopted in the CC stage for CC-C η charging compared to CC-CV, ensuring that the two strategies have comparable charging durations. This arrangement facilitates a direct comparison of their effects on battery degradation. For all the defined tests, we measured and collected data on current, voltage, temperature, and plating potential at a sampling frequency of 1 Hz.

The rationale for maintaining a positive plating potential ($\eta > 0$) to prevent the initiation of lithium plating is grounded in equilibrium thermodynamics principles [15], [34]–[36], which are used to explain when lithium plating begins. According to this theoretical framework, lithium plating becomes thermodynamically favorable when the potential of the graphite electrode falls below 0 V versus Li/Li⁺. This situation may arise if the intercalation reaction experiences significant kinetic limitations, leading to a large overpotential. This overpotential can surpass the equilibrium potential of the lithium-graphite phase diagram, consequently causing the graphite potential to drop below 0 V versus Li/Li⁺. While this voltage criterion alone is not sufficient to guarantee lithium plating, it is a necessary condition for its occurrence [37], [38]. Thus, from a practical application perspective, using this criterion is both feasible and advantageous for mitigating lithium plating.

2.2 Synthetic data generation

To cover wide operating conditions and diverse aging trajectories that batteries may encounter in real-world applications, synthetic data was generated by simulating a single particle model, including electrolyte dynamics (SPMe), and coupled with several aging mechanisms. The modeling of battery aging

is an active research area. So far, there is no universally accepted model that covers all aging mechanisms possibly triggered in typical Li-ion batteries. In this regard, three major aging mechanisms are considered in the simulations, including SEI layer growth in the anode and loss of active materials in the anode and cathode [39]. For a more detailed introduction to the employed SPMe-aging model, readers are referred to the literature works [39]–[41]. It is worth mentioning that apart from these three aging mechanisms, lithium plating is also known as one of the major aging mechanisms during the usage of Li-ion batteries, particularly at high charging currents, low temperatures, high SoC levels, and in aged cells [42], [43]. However, since this study eventually aims to develop a fast-charging strategy capable of keeping the plating potential positive to prevent lithium plating, we do not consider it in the simulation model.

By varying aging-related parameters sequentially, the above battery model was simulated extensively on the platform PyBaMM [44] under a wide range of operating conditions (as detailed in Section 4.1). The simulations that were conducted resulted in a large synthetic dataset consisting of two parts. One part of the dataset includes time-series data of current, voltage, temperature, and plating potential recorded at a sampling rate of 1 Hz. The other part contains four SoH indicators, i.e., the battery capacity, lithium inventory loss (LLI), loss of active materials in the negative electrode (LAM_{NE}), and loss of active material in the positive electrode (LAM_{PE}), recorded at the end of each RPT cycle. To partially visualize the dataset, the generated 1,392 aging trajectories are depicted in Fig. S1 (see *Supplemental Information*). With the end of life defined at 80% of the nominal capacity, it can be seen that the battery lifetimes corresponding to the different parameter settings span from 20 to 2000 cycles. Further details for the employed battery model and its implementation are introduced in Section 4.1.

2.3 Lifelong estimation framework for plating potential

In this study, we introduce an innovative framework designed for the life-long estimation of plating potential, as illustrated in Fig. 2. Central to our approach are two pivotal tasks executed through the integration of three ML models. Specifically, the framework employs two slow timescale models (Model 1 and Model 2) for estimating the battery’s capacity and the three aging modes both on a slow timescale and Model 3 for estimating the plat-

ing potential on a fast timescale. This structured framework leverages the strengths of each model to deliver an accurate, robust, and real-time estimation of the plating potential over the battery’s lifespan.

For Model 1, the inputs, also called features, are manually constructed from time-series and histogram-based raw data, e.g., the slope of the voltage curves and the accumulated energy throughput. The output of this model is battery capacity. For Model 2, the inputs include time-series current, voltage, and temperature measurements, while the outputs are the three aging modes (namely LLI, LAM_{PE}, and LAM_{NE}). Considering the nature of slow aging dynamics, we develop Model 1 and Model 2 from the data collected during RPT tests, which, as specified in Section 2.1, were conducted every 25 normal cycles. For real-world vehicle applications, the data samples for training these two ML models can be collected during destination charging, in which the charging current is low (often lower than 1/8C). This provides beneficial conditions, such as wide SoC ranges, relatively stable temperatures, and mild lithium diffusion processes, for extracting high-quality data samples. To label the four SoH indicators as the model output, the battery capacity can be calibrated according to Profile 3 of Section 2.1, and the three aging modes are calculated from the open-circuit voltage (OCV), as introduced in Note 1 of the *Supplemental Information*.

With the plating potential as the unique output, Model 3 has two types of inputs. The first type of input comprises time-series current, voltage, and temperature measurements, from which we aim to learn the fast battery dynamics. As the ultimate target is to optimize battery charging performance in real-time, the input data could be extracted from the corresponding fast charging profile. The second type is the four outputs from Model 1 and Model 2, by which we inform Model 3 of the battery health characteristics. Essentially, Model 3 combines time-series and non-temporal features to adaptively estimate the plating potential over the battery’s entire lifetime. To pave the way for later conservative fast-charging control implementation, an underestimated plating potential result is preferred compared to an overestimate. We therefore propose an underestimate-enhanced long short-term memory (UE-LSTM) to refine the precision of plating potential estimations while inherently biasing towards underestimations. The idea is to ensure prediction accuracies but with a preference for conservative estimates.

In the proposed framework, the three ML models operate in parallel. At

low current rates, such as those encountered during destination charging, the risk of lithium plating diminishes, reducing the necessity for real-time plating potential estimation. In such circumstances, it may be unnecessary to activate Model 3. Conversely, during fast charging, precise estimation of the plating potential becomes critical to prevent lithium plating, highlighting the critical role of Model 3. It is worth mentioning that this framework can be seamlessly integrated with existing vehicle BMS, utilizing readily available measurements to enable onsite quantification of all the SoH indicators and real-time estimation of the plating potential.

2.4 Results of the three-electrode cell test

By implementing the experiments designed in Section 2.1, the aging trajectories of all the studied battery cells can be obtained, as depicted in Fig. 1e. To the best of our knowledge, these are the first published experimental results that directly measure the plating potential and use it to dynamically control the charging current in a closed-loop fashion during the battery’s lifelong operation. It can be seen that at the end of the test, corresponding to 325 normal cycles, both cells in the CC-CV group had passed their end of life, with a SoH below 80%. Specifically, these cells suffered from a sharp capacity loss, i.e., 12% of the nominal capacity, within the first 50 cycles, attributed to the onset of lithium plating, as evidenced by the negative plating potential shown in Fig. 1d. In contrast, the two cells in the CC-C η group were able to retain an SoH above 87% beyond the 325th cycle, suggesting a markedly extended lifespan compared to their counterparts. This improvement is derived from the CC-C η charging strategy continuously monitoring the plating potential and ensuring its non-negativity.

Beyond extending battery lifetime, the CC-C η charging strategy significantly surpassed CC-CV charging in terms of efficiency. It also achieves a 30% initial reduction in charging time and a 40% reduction by the 200th cycle, as illustrated in Fig. 1c-d.

The above experimental results demonstrate that real-time control of the charging profile based on the plating potential (η) can significantly extend battery lifetime while enabling faster charging. These findings not only confirm the viability of health-aware fastest charging but also point out how to achieve it. It strongly motivates the lifelong estimation of η in Section 2.3, given it is not economically feasible to incorporate reference electrodes into

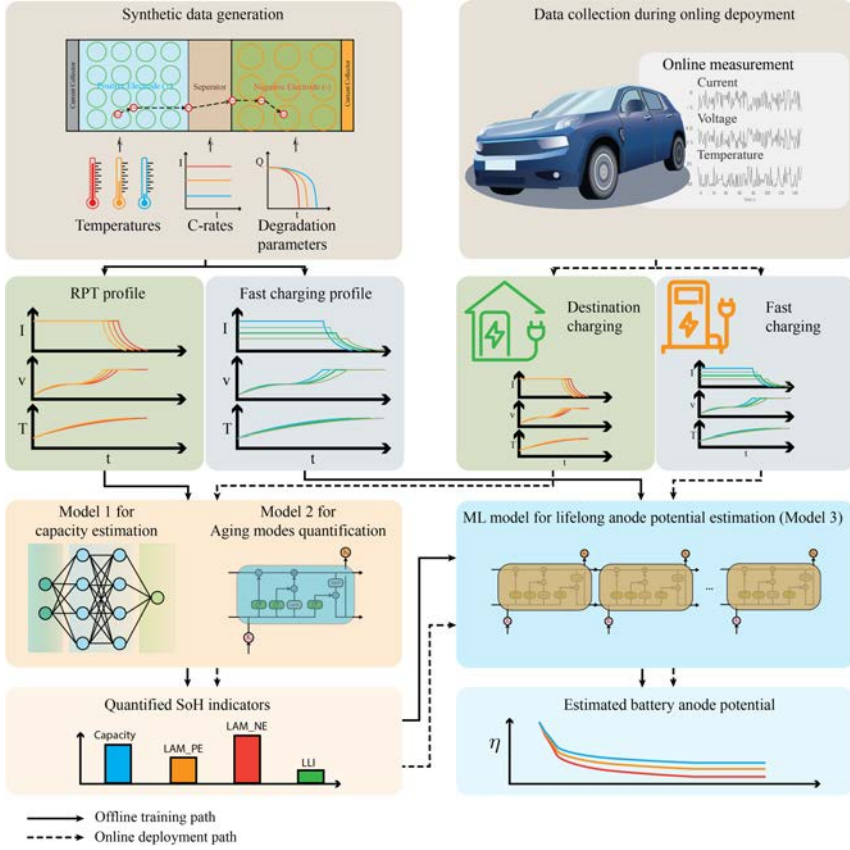


Figure 2: The proposed framework for lifelong estimation of battery plating potential.

each commercial battery.

Conventionally, the evaluation of a given charging strategy with respect to battery degradation and lifetime relies heavily on extensive experiments over thousands of cycles, especially considering diverse operating conditions. As per the results in Fig. 1, the long-term effects of different charging strategies on battery health can be largely revealed by the profile of η in the early cycling stage, thus significantly mitigating the need for extremely time-consuming and expensive experiments. This benefit will, in turn, accelerate the improvement

and optimization process of existing battery materials and design for faster charging and a longer lifetime.

2.5 Estimation results for SoH indicators

The results for the estimation of the four SoH indicators using Model 1 and Model 2 applied to the full SoC operating window are presented in Fig. 3a and the two first rows of Table 2. The models exhibit high estimation fidelity for all the SoH indicators, achieving a mean absolute error (MAE) below 0.4% and a root-mean-square error (RMSE) below 0.55%. The estimation accuracy is slightly lower in LAM_{NE} compared to LAM_{PE}. However, although the estimation of LAM_{NE} results in the highest average error among all the indicators, the absolute error remains within 1.07% for 95% of the test dataset, underscoring the model’s high accuracy and robustness across a broad spectrum of operating conditions. The relatively reduced accuracy in LAM_{NE} estimations, as compared to LAM_{PE}, may be attributed to the more flat open-circuit potential (OCP) of the anode compared to that of the cathode, which complicates the discernment of LAM_{NE} variations via the cell’s current and voltage. This observation aligns well with previous findings in [45].

For the training of our ML algorithms, we extensively varied aging-related parameters in the simulation to generate a large and diverse synthetic dataset. For further examination, we randomly select a simulated battery cell from the dataset. The estimation trajectories for each SoH indicator over the cell’s entire lifetime are shown in Fig. 3b. As can be seen, the estimates closely follow the calibrated values, with maximum estimation errors staying well within the $\pm 2.5\%$ error bounds. This verifies the prediction capability of our developed Models 1 and 2. It is important to note that the randomly selected simulated cell, exhibiting nearly linear SoH indicator trajectories, is a special case. Among the 1,392 simulated aging trajectories, the aging behavior of other cells varies significantly. As illustrated in Fig. S1 (see *Supplemental Information*), many cells demonstrate distinct patterns of capacity degradation, with clear non-linear characteristics. Similarly, the SoH indicators exhibit pronounced non-linearity over the cycle life, as visualized in Fig. S2. To underscore the necessity of employing the proposed non-linear model, we benchmark it against two linear models for estimating the SoH indicators. The results from these comparisons strongly indicate that a non-linear approach is crucial for accurately capturing the complex aging behavior of batteries. A detailed

comparison of the results is provided in Tables S3 and S4 (see *Supplemental Information*).

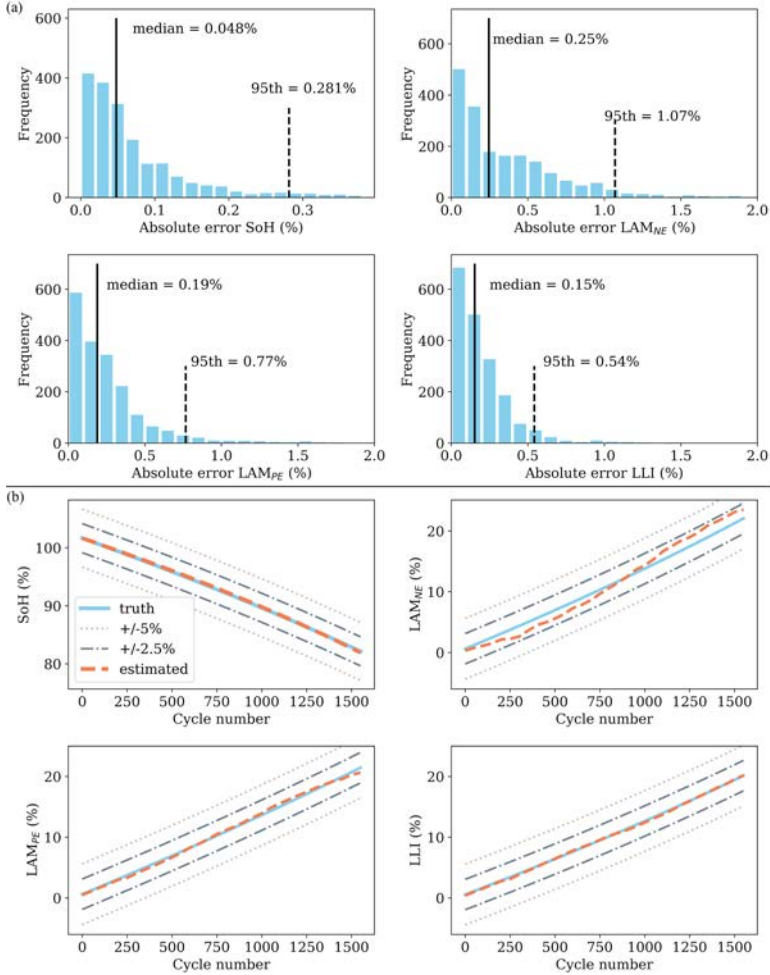


Figure 3: Estimation results of the four SoH indicators for batteries operated under the full SoC window. (a) displays the histogram results for all data samples in the test set. For a randomly selected ‘cell’ from the simulation pool, (b) presents the estimated trajectories against their ground truth along the equivalent full cycles.

In real-world vehicular applications, battery discharge and charge cycles rarely span the full 0–100% SoC window. Instead, batteries typically undergo various partial charging cycles, including the aforementioned destination charging for EVs. Such a context necessitates a comprehensive evaluation of the developed ML models for SoH indicator quantification under realistic charging conditions. Accordingly, we assess the performance of Model 1 and Model 2 across six distinct usage scenarios, each characterized by a unique SoC window. Table 2 presents the estimation errors for the four SoH indicators across these scenarios. Despite a slight reduction in accuracy relative to the full SoC window scenario, the obtained estimates maintain high reliability, with average errors remaining below 1% for all the scenarios and SoH indicators. These results point to the real-world applicability of Model 1 and Model 2 as used in the proposed estimation framework. Note 6 of the *Supplemental Information* provides more results regarding the overall estimation errors under different usage conditions.

Table 2: Estimation results for battery SoH indicators under different charging scenarios.

SoC window (%)		SoH capacity (%)	LAM _{NE} (%)	LAM _{PE} (%)	LLI (%)
0–100	MAE	0.077	0.375	0.269	0.200
	RMSE	0.131	0.546	0.439	0.289
20–50	MAE	0.335	0.458	0.268	0.232
	RMSE	0.429	0.702	0.421	0.370
30–60	MAE	0.357	0.603	0.344	0.241
	RMSE	0.554	0.777	0.887	0.872
40–70	MAE	0.431	0.470	0.451	0.312
	RMSE	0.719	0.789	0.641	0.432
50–80	MAE	0.389	0.580	0.633	0.398
	RMSE	0.610	0.880	0.982	0.586
60–90	MAE	0.335	0.557	0.393	0.255
	RMSE	0.429	0.859	0.673	0.408
70–100	MAE	0.099	0.544	0.385	0.161
	RMSE	0.165	0.956	0.578	0.245

2.6 Estimation results for lifelong plating potential

1) **Noise specification for model training.** In the development of Model 3, we have meticulously accounted for uncertainties inherent to the estimation of SoH indicators and the measurement of current, voltage, and temperature

encountered in real-world applications. This is achieved by augmenting the training dataset with Gaussian white noise, thereby simulating the influence of estimation and measurement inaccuracies on model training. For all four SoH indicators, the noise is calibrated based on the most challenging real-world conditions, using the largest RMSE observed across the six partial charging scenarios (as highlighted in Table 2) to define the standard deviation of the Gaussian noise. The choice of measurement noise for current, voltage, and temperature is based on the specifications and experience of widely utilized battery sensor technologies. Specifically, we used the sensor noise with standard deviations corresponding to Level 4 defined in Table 4 as the default.

2) The overall performance of Model 3. Our examination of the plating potential estimation, as delineated in Fig. 4, reveals that Model 3 is capable of predicting the battery’s plating potential with a high degree of accuracy over the entire lifespan, limiting the maximum error to less than 20 mV. This precision is further highlighted in the zoomed-in section of Fig.4a, where the cumulative error histogram demonstrates that 95% of the estimations maintain an absolute error below 10 mV. Further quantitative analysis, represented in the last row of Table 3, showcases that the model has an MAE of 3.37 mV and an RMSE of 4.77 mV. These values affirm the model’s high fidelity and robustness in estimating the plating potential across diverse operating conditions.

A detailed case study, illustrated in Fig. 4c, tracks the plating potential estimation for a randomly selected cell throughout its lifespan. The performance in estimating plating potentials near 0 V, essential for optimizing battery charging performance, is highlighted in the zoomed-in figure. This in-depth analysis reveals the model’s capacity to not only closely follow the cell’s true trajectory of plating potential but also adapt dynamically to its aging status.

3) The effect of different SoH indicators. Four established SoH indicators have been integrated into the feature set of Model 3. To study the importance of the individual SoH indicators for plating potential estimation. We change the number of SoH indicators utilized inside the model, resulting in four alternative models characterized by their inputs, where Model 3a utilizes only the directly measured current, voltage, and temperature (see Table 3 for details).

It is observed from Table 3 and Fig. 4b that with an increasing number of SoH indicators added to the feature set, the model accuracy is consistently

improved, regardless of the number of data samples tested. Furthermore, leveraging the full array of SoH indicators, Model 3 outperforms all its counterparts that lack equally comprehensive aging information. In comparison to Model 3a, Model 3 achieves a reduction in the MAE and RMSE by 22% and 24%, respectively. This enhancement in predictive accuracy demonstrates the importance of including detailed information on battery capacity and aging modes within the estimation framework.

4) The effect of different ML models. To evaluate the effectiveness of our proposed UE-LSTM in the development of Model 3, we implemented three alternative ML algorithms for comparison: the gated recurrent unit (GRU) neural network, an artificial neural network (ANN), and Transformer (see Section 4.2 for details).

The comparative analysis between LSTM and UE-LSTM is illustrated in Fig. 4d–e. Although the overall performance in estimation accuracy of these two algorithms is similar, a notable distinction is observed in the nature of the estimation errors. Specifically, 56% of the estimation errors generated by the UE-LSTM are attributable to underestimation, marking a 7% increase compared to the conventional LSTM. This underestimation characteristic of UE-LSTM implies a potential for providing larger safety margins in the design of control strategies based on the estimated plating potential. Additional comparative results, particularly against GRU, ANN, and Transformer models, are available in Note 3 of the *Supplemental Information*. These results further attest to the UE-LSTM’s superior estimation accuracy.

Table 3: Estimation results of the plating potential using different feature sets.

	Feature input	MAE (mV)	RMSE (mV)
Model 3a	Current (I), Voltage (U), Temperature (T)	4.30	6.28
Model 3b	I , U , T , Capacity	3.91	5.71
Model 3c	I , U , T , Capacity, LLI	3.82	5.39
Model 3d	I , U , T , Capacity, LLI, LAM_{NE}	3.44	4.88
Model 3	I , U , T , Capacity, LLI, LAM_{PE} , LAM_{NE}	3.37	4.77

5) The effect of different levels of measurement noise. For battery applications such as EVs, the accuracy of signals measured from batteries is often reduced relative to laboratory experiments due to the need to balance sensor costs against accuracy. It is therefore crucial for the developed estima-

tion model to possess robustness against sensor noise. For a comprehensive evaluation, we introduce ten levels of Gaussian white noise to current, voltage, and temperature measurements, with their standard deviations defined in Table 4.

The results in Fig. 4f demonstrate an almost linear increase in the estimation error with escalating levels of sensor noise. Notably, even under the most severe conditions (noise level 10, which substantially surpasses typical industry standards), our developed estimation model maintains an MAE of less than 16 mV. This finding underscores that the proposed lifelong estimation framework and its embedded Model 3 are capable of delivering accurate and robust performance in the presence of large sensor noise.

Table 4: Ten levels of Gaussian white noise introduced to measurements.

Sensor noise level	0	1	2	3	4	5	6	7	8	9	10
Current (mA)	0	5	10	15	20	25	30	35	40	45	50
Voltage (mV)	0	2.5	5	7.5	10	12.5	15	17.5	20	22.5	25
Temperature (°C)	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5

2.7 Computational efficiency

The computational efficiency of the designed estimation algorithm, which integrates Models 1–3, is critical for enabling optimal charging control in a vehicle BMS. We have therefore assessed the computational load of our algorithm using a Nvidia Tesla A100 GPU in a Python environment. To reduce randomness in the evaluation, the algorithm is tested over ten times under identical conditions, with the mean computation times reported in Table 5.

Despite the extensive dataset used, the training process for all three developed models only requires 1060 seconds, while the testing process exhibits remarkable computational efficiency, completing in 5.27 seconds. The execution time for estimating the plating potential over a single real-world charging cycle is merely 0.16 seconds. Note that such calculations are conservative, as the three models can be executed in parallel to reduce the total required computation time. These results underscore the algorithm’s high applicability for real-world battery management.

Table 5: Computation time required by the proposed estimation algorithm.

	Computation time (s)		
	Training	Testing	Single charge
Model 1	13	0.28	0.01
Model 2	366	2.12	0.09
Model 3	681	2.87	0.06
Total	1060	5.27	0.16
	Data size		
	Training	Testing	Single charge
Model 1	7552	1888	1
Model 2	2.75×10^7	6.88×10^6	3645
Model 3	8.30×10^6	2.08×10^6	250

2.8 Discussion and outlook of future research

This study demonstrates that the plating potential-controlled charging strategy can substantially extend battery lifetime and enhance charging speed. Nonetheless, uncertainties linger regarding the implications of applying high currents at the CC stage of each charge. Despite careful control to maintain a positive plating potential, the possibility for high currents to expedite aging mechanisms, aside from lithium plating, has not been fully investigated. Additionally, while using a constant control threshold to limit the onset of lithium plating has shown practical benefits in our closed-loop control system, it may not fully capture the variability of the actual plating potential in the negative electrodes. Factors such as thermal gradients, uneven particle distribution, and inhomogeneous electrodes contribute to this variability. Therefore, a slightly higher threshold of 1 mV is selected in this work to account for such variability, albeit with a small performance trade-off. Furthermore, the design of laboratory-scale three-electrode cells, with their limited surface area and metal enclosure, inherently minimizes thermal fluctuations, neglecting the impact of thermal dynamics on fast charging. These areas are critical to our understanding of battery degradation under fast charging and will be a future research direction. With the estimated plating potential, another crucial next step is to explore its usage for the development of safety- and health-conscious BMS.

3 Conclusions

This study challenges the conventional compromise between the rapid charging of Li-ion batteries and their longevity. Through experimental validation on three-electrode battery cells, we have developed a plating potential-controlled charging strategy. This strategy not only enables over 30% faster charging rates but also significantly extends battery lifespan by mitigating lithium plating, a notorious catalyst for premature battery degradation.

To adapt these laboratory findings to commercial EV batteries without a reference electrode, we introduced a novel machine learning (ML)-based framework. This framework seamlessly integrates three distinct ML models to estimate the plating potential throughout the battery's life. These models are activated selectively and intelligently based on usage scenarios, timescales of the involved system dynamics, and available data sources of the battery during real-world conditions. Specifically, Model 1 and Model 2 are designed to estimate the battery's capacity and three aging modes on a slow timescale and at low current rates, such as during destination charging. Model 3, in contrast, operates on a fast timescale to provide real-time estimates of the plating potential during fast charging. By leveraging our newly developed UE-LSTM algorithm, time-series features from field data collected during fast charging, and non-temporal features outputted from Model 1 and Model 2, Model 3 is able to provide lifelong robust and accurate predictions under a variety of operating conditions and measurement uncertainties.

This new framework can be seamlessly integrated with existing vehicle BMS, utilizing readily available measurements. Our results represent a significant step forward in the development of more efficient and durable EV battery utilization.

4 Experimental procedures

4.1 Electrochemical battery modeling and simulation

SEI layer growth model

The liquid electrolyte used in Li-ion battery cells becomes unstable and reacts with lithium ions and electrons within the electrodes to form a passive solid-electrolyte interface (SEI) during charging in the first cycle, leading to around

10% capacity loss [46]–[48]. The SEI layer will continue to grow during battery usage, leading to capacity loss and resistance increase [49]. There are several ways to model SEI growth, as reviewed in [50]. To describe the dynamics of the SEI growth, a kinetics-limited model developed in [51] is adopted here, in which the exchange current density of the SEI-based side reactions is expressed in the Tafel equation

$$i_{\text{SEI}} = i_{0,\text{SEI}} \exp\left(\frac{\alpha n F}{RT} \eta_{\text{SEI}}\right), \quad (\text{D.1})$$

$$\eta_{\text{SEI}} = U_{\text{NE}}^{\text{ref}} + \eta_{\text{NE}} - U_{\text{SEI}}^{\text{ref}}. \quad (\text{D.2})$$

In (D.1), $i_{0,\text{SEI}}$ is the reference exchange current density, α is a transfer coefficient, n is the number of electrons participating in the reaction, F is the Faraday’s constant, R is the ideal gas constant, and T represents the battery temperature. In (D.2), η_{SEI} is the overpotential of the SEI-based side reaction, $U_{\text{NE}}^{\text{ref}}$ is the equilibrium anode potential which is evaluated as a function of the solid phase concentration at the surface of the particle, η_{NE} is the anode overpotential, and $U_{\text{SEI}}^{\text{ref}}$ is the equilibrium potential of the SEI growth reaction, respectively.

LAM model

The continuous cycling of Li-ion batteries may cause electrode dissolution [41], particle cracking [52], and detachment from the binding materials [53], resulting in LAM in both the negative and positive electrodes. The volume fraction of active materials in each electrode, denoted by ϵ , can be modeled by [41], [54]

$$\frac{\partial \epsilon}{\partial t} = \beta \left(\frac{\sigma_{h,\text{max}} - \sigma_{h,\text{min}}}{\sigma_{\text{yield}}} \right)^m, \quad (\text{D.3})$$

where $\sigma_{h,\text{max}}$ and $\sigma_{h,\text{min}}$ are the minimum and maximum hydraulic stress, respectively, depending on the particle radius and the average lithium concentration within the particle. The values of the parameters β , m , and σ_{yield} are often fitted from experimental data.

Physical parameters and simulation conditions

The SPMe-aging battery model employed in the simulation was previously parameterized against LG M50T cylindrical battery cells in [41], [50], [55], [56]. We take the parameter values from the referred references and gradually vary a set of aging-rated parameters, such as $i_{0,\text{SEI}}$ in (D.1) and β in (D.3), among different simulations. The detailed parameter values and simulation setup are given in Note 5 of the *Supplemental Information*.

Each cycle of the cell operation includes a CC-CV charge with a cut-off voltage of 4.2 V at the CC stage and a cut-off current of 50 mA at the CV stage, followed by a CC discharge with a cut-off voltage of 2.5 V. Between adjacent charge and discharge phases, a one-hour relaxation period is introduced to mimic battery behavior under calendar aging. For comprehensive coverage of diverse operating conditions encountered in real-world scenarios, the simulation incorporates various charging current rates and ambient temperatures. The simulation process is terminated if the cell has reached its end of life. Every 50 cycles, an RPT is performed to get the pseudo-OCV measurement, in which the cell is charged and discharged at 1/10C until the cut-off voltage limit of 4.2 V is reached.

4.2 ML-based battery modeling and implementation

The methods for developing the three ML models in the proposed lifelong estimation framework are introduced in this subsection.

KF-based fusion of ML models for capacity estimation

The methods for online capacity estimation have been well-established in the literature. A data-driven multi-model fusion method was recently developed and demonstrated high accuracy and practicability under arbitrary vehicle usage conditions [29]. We adopt it in Model 1 to estimate battery capacity continuously over the lifespan. While the details are referred to [29], we introduce the key idea here for completeness. First, a tailored set of features is meticulously curated from time-series current, voltage, and temperature measurements during battery charging. Based on these features, four distinct ML algorithms, including two probabilistic and two frequentist approaches, are applied to estimate the battery capacity. Second, leveraging usage-related histogram data, a one-step-ahead capacity prediction model is developed. In

the end, a Kalman filter (KF) is utilized to systematically fuse the results of all the estimation and prediction models. The comprehensive features derived from the partial slow charging process for Model 1 are provided in Note 2 of the *Supplemental Information*.

LSTM for aging mode quantification

The cell OCV curve or electrochemical impedance spectroscopy (EIS) can potentially be obtained at each RPT test and used to achieve the targets of Model 2, as per [57]–[59]. However, for real-world EV applications, it is practically challenging to receive either reliable EIS measurements over a wide frequency range or OCV measurements over the full SoC window. In contrast, extracting a pseudo-OCV curve, e.g., during destination charging or overnight home charging, is comparatively straightforward. Thus, we formulate features from the partial charging events within an LTSM model to quantify the three aging modes as the output of Model 2.

Other than traditional neural networks (NNs) that take a single data point as the input and require manual construction of features, LSTM takes a sequence of data and performs the regression or classification without the need for feature selection and engineering. This characteristic makes LSTM well-suited for speech recognition, language translation, and robot control tasks. Considering battery applications, the collected data are usually in time series by nature. Consequently, it is suitable to use LSTM to build an estimation model. Compared to recurrent neural networks (RNNs), also designed to process time-series data, LSTM can effectively avoid vanishing gradient during training [60].

A typical LSTM unit consists of two internal cells, i.e., a memory cell c_t and a hidden cell h_t , where t represents the current time step. The memory cells can be viewed as the state that saves information over a long period of time. The hidden cell, together with the gating mechanism composed of three main gates (i.e., the input, forget, and output gates), controls the information flowing within the neural network. The detailed calculation process can be

represented by the following equations [61]:

$$\begin{aligned} f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f), & i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i), \\ o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o), & \tilde{c}_t &= \tanh(W_c x_t + U_c h_{t-1} + b_c), \\ c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t, & h_t &= o_t \odot \tanh(c_t), \end{aligned}$$

where x_t is the input vector, i_t , f_t , and o_t are the activation vectors of the input, forget, and output gates, respectively, and \tilde{c}_t is the memory cell's input activation vector. W and U represent the weight matrices for the input and hidden state, respectively, b is the bias term, and \odot denotes the element-wise product.

Alternative methods for SoH indicator estimation

Alternatively to the above approaches, two other types of approaches can be utilized to estimate the four SoH indicators. The first type is to extend the LSTM model in Section 4.2 to include all four SoH indicators as outputs. The second type is to extract a relevant and appropriate feature set for each SoH indicator and use them for training point-to-point ML models, such as Gaussian process regression (GPR), Bayesian regression (BR), random forest regression (RFR), and NN. A systematic comparison between the proposed approach and its two benchmarks is conducted. Illustrative results in Note 2 of the *Supplemental Information* demonstrate that the proposed approach in Section 4.2–4.2 can provide the most accurate estimation results for all the indicators, albeit with some exceptions.

UE-LSTM for plating potential estimation

For fast charging control based on plating potential, it is crucial to have a small safety margin so that the plating potential will not enter the negative region. Consequently, it becomes imperative to prioritize models exhibiting a tendency towards underestimation rather than overestimation, particularly as the plating potential approaches zero. To achieve this, we propose a novel algorithm, termed UE-LSTM, which employs a unique loss function J defined as follows:

$$J = \frac{1}{N} \sum_i^N J_1(i), \quad (\text{D.4})$$

$$J_1(i) = \begin{cases} w_1 J_2(i), & \text{if } y_i \geq \hat{y}_i \\ J_2(i), & \text{otherwise} \end{cases} \quad (\text{D.5})$$

$$J_2(i) = \begin{cases} w_2 (y_i - \hat{y}_i)^2, & \text{if } |y_i| \leq 0.02 \\ (y_i - \hat{y}_i)^2, & \text{otherwise} \end{cases} \quad (\text{D.6})$$

where y_i and \hat{y}_i are the true and estimated plating potential values, respectively. (D.5) is designed with a weighting factor $w_1 = 10$ to favor scenarios of underestimation. (D.6) incorporates a large weighting factor $w_2 = 50$ to improve estimation accuracy within the plating potential range of $[-0.02, 0.02]$ V.

Training data labeling

The capacity of the battery cell can either be directly measured during RPT tests or estimated using an accurate and reliable estimator, e.g., the one developed in [29]. As per [62], there are established methods to experimentally extract LLI information through an advanced lab setting. How to measure the LAM operando in each electrode is still an active research question. Instead, we calculate the three aging modes continuously over the battery's life using pseudo-OCV measurements and the diagnostic algorithm of [63] (see Note 1 of the *Supplemental Information* for details).

4.3 Evaluation matrices

The MAE and RMSE are used to quantify the estimation performance for the four considered SoH indicators. These two evaluation matrices are defined as

$$MAE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i) \quad (\text{D.7})$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (\text{D.8})$$

where N represents the total number of samples in the test set, y_i and \hat{y}_i are the true value and the estimated value, respectively, and i is the time step. For the plating potential estimation, in addition to the MAE and RMSE, the coefficient of determination is used to evaluate how well the model fits the test samples, as defined by

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad (\text{D.9})$$

where the average estimated value $\bar{y} = \frac{1}{N} \sum_{i=1}^N \hat{y}_i$.

4.4 Resource availability

Data availability

The raw cycling data for the three-electrode battery cells can be found in the online version of the paper.

4.5 Supplementary materials

Supplementary material related to this article can be found in the online version of the paper.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT4 to polish the language. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

References

- [1] *Zero emission vehicles: First fit for 55 deal will end the sale of new co2 emitting cars in europe by 2035*, https://ec.europa.eu/commission/presscorner/detail/en/ip_22_6462, Accessed: 05-July-2023, 2022.

- [2] B. Nykvist and M. Nilsson, “Rapidly falling costs of battery packs for electric vehicles,” *Nat. Clim. Change*, vol. 5, no. 4, pp. 329–332, 2015.
- [3] M. S. Ziegler and J. E. Trancik, “Re-examining rates of lithium-ion battery technology improvement and cost decline,” *Energy Environ. Sci.*, vol. 14, no. 4, pp. 1635–1651, 2021.
- [4] N. Wassiliadis, M. Steinsträter, M. Schreiber, *et al.*, “Quantifying the state of the art of electric powertrains in battery electric vehicles: Range, efficiency, and lifetime from component to system level of the volkswagen id. 3,” *Etransportation*, vol. 12, p. 100 167, 2022.
- [5] N. Wassiliadis, J. Schneider, A. Frank, *et al.*, “Review of fast charging strategies for lithium-ion battery systems and their applicability for battery electric vehicles,” *Journal of energy storage*, vol. 44, p. 103 306, 2021.
- [6] C.-Y. Wang, T. Liu, X.-G. Yang, *et al.*, “Fast charging of energy-dense lithium-ion batteries,” *Nature*, vol. 611, no. 7936, pp. 485–490, 2022.
- [7] G. K. Pephrah, T. Wik, Y. Huang, F. Altaf, and C. Zou, “Control-oriented 2D thermal modelling of cylindrical battery cells for optimal tab and surface cooling,” in *American Control Conference*, IEEE, 2024, pp. 1–6.
- [8] A. Tomaszewska, Z. Chu, X. Feng, *et al.*, “Lithium-ion battery fast charging: A review,” *ETransportation*, vol. 1, p. 100 011, 2019.
- [9] I. Aghabali, J. Bauman, P. J. Kollmeyer, Y. Wang, B. Bilgin, and A. Emadi, “800-v electric vehicle powertrains: Review and analysis of benefits, challenges, and future trends,” *IEEE Trans. Transp. Electrification*, vol. 7, no. 3, pp. 927–948, 2020.
- [10] E. J. Dufek, D. P. Abraham, I. Bloom, *et al.*, “Developing extreme fast charge battery protocols—a review spanning materials to systems,” *J. Power Sources*, vol. 526, p. 231 129, 2022.
- [11] M. Li, M. Feng, D. Luo, and Z. Chen, “Fast charging li-ion batteries for a new era of electric vehicles,” *Cell Rep. Phys. Sci.*, vol. 1, no. 10, 2020.
- [12] W. Xie, X. Liu, R. He, *et al.*, “Challenges and opportunities toward fast-charging of lithium-ion batteries,” *Journal of Energy Storage*, vol. 32, p. 101 837, 2020.

-
- [13] J. Sieg, J. Bandlow, T. Mitsch, *et al.*, “Fast charging of an electric vehicle lithium-ion battery at the limit of the lithium deposition process,” *J. Power Sources*, vol. 427, pp. 260–270, 2019.
- [14] N. Wassiliadis, J. Krieglner, K. A. Gamra, and M. Lienkamp, “Model-based health-aware fast charging to mitigate the risk of lithium plating and prolong the cycle life of lithium-ion batteries in electric vehicles,” *J. Power Sources*, vol. 561, p. 232 586, 2023.
- [15] T. Waldmann, B.-I. Hogg, and M. Wohlfahrt-Mehrens, “Li plating as unwanted side reaction in commercial li-ion cells—a review,” *J. Power Sources*, vol. 384, pp. 107–124, 2018.
- [16] T. R. Tanim, P. P. Paul, V. Thampy, *et al.*, “Heterogeneous behavior of lithium plating during extreme fast charging,” *Cell Reports Physical Science*, vol. 1, no. 7, 2020.
- [17] X. Lu, M. Lagnoni, A. Bertei, *et al.*, “Multiscale dynamics of charging and plating in graphite electrodes coupling operando microscopy and phase-field modelling,” *Nat. Commun.*, vol. 14, no. 1, p. 5127, 2023.
- [18] J. Sieg, A. U. Schmid, L. Rau, *et al.*, “Fast-charging capability of lithium-ion cells: Influence of electrode aging and electrolyte consumption,” *Appl. Energy*, vol. 305, p. 117 747, 2022.
- [19] T. Amietszajew, E. McTurk, J. Fleming, and R. Bhagat, “Understanding the limits of rapid charging using instrumented commercial 18650 high-energy Li-ion cells,” *Electrochim. Acta*, vol. 263, pp. 346–352, 2018.
- [20] Y. Hoshi, Y. Narita, K. Honda, T. Ohtaki, I. Shitanda, and M. Itagaki, “Optimization of reference electrode position in a three-electrode cell for impedance measurements in lithium-ion rechargeable battery by finite element method,” *J. Power Sources*, vol. 288, pp. 168–175, 2015.
- [21] S. J. An, J. Li, C. Daniel, S. Kalnaus, and D. L. Wood, “Design and demonstration of three-electrode pouch cells for lithium-ion batteries,” *J. Electrochem. Soc.*, vol. 164, no. 7, A1755, 2017.
- [22] Y. Lu, X. Han, Z. Chu, *et al.*, “A decomposed electrode model for real-time anode potential observation of lithium-ion batteries,” *J. Power Sources*, vol. 513, p. 230 529, 2021.

- [23] M. Doyle, T. F. Fuller, and J. Newman, “Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell,” *J. Electrochem. Soc.*, vol. 140, no. 6, p. 1526, 1993.
- [24] F. Ringbeck, M. Garbade, and D. U. Sauer, “Uncertainty-aware state estimation for electrochemical model-based fast charging control of lithium-ion batteries,” *J. Power Sources*, vol. 470, p. 228 221, 2020.
- [25] L. Li, Y. Ren, K. O’Regan, *et al.*, “Lithium-ion battery cathode and anode potential observer based on reduced-order electrochemical single particle model,” *J. Energy Storage*, vol. 44, p. 103 324, 2021.
- [26] M. Andersson, M. Streb, J. Y. Ko, *et al.*, “Parametrization of physics-based battery models from input–output data: A review of methodology and current research,” *J. Power Sources*, vol. 521, p. 230 859, 2022.
- [27] A. M. Bizeray, J.-.-H. Kim, S. R. Duncan, and D. A. Howey, “Identifiability and parameter estimation of the single particle lithium-ion battery model,” *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 5, pp. 1862–1877, 2018.
- [28] W. Li, M. Rentemeister, J. Badedo, D. Jöst, D. Schulte, and D. U. Sauer, “Digital twin for battery systems: Cloud battery management system with online state-of-charge and state-of-health estimation,” *J. energy storage*, vol. 30, p. 101 557, 2020.
- [29] Y. Zhang, T. Wik, J. Bergström, and C. Zou, “State of health estimation for lithium-ion batteries under arbitrary usage using data-driven multi-model fusion,” *IEEE Trans. Transp. Electrif.*, 2023.
- [30] Y. Zhang, T. Wik, J. Bergström, M. Pecht, and C. Zou, “A machine learning-based framework for online prediction of battery ageing trajectory and lifetime using histogram data,” *J. Power Sources*, vol. 526, p. 231 110, 2022.
- [31] Z. Wei, X. Yang, Y. Li, H. He, W. Li, and D. U. Sauer, “Machine learning-based fast charging of lithium-ion battery by perceiving and regulating internal microscopic states,” *Energy Stor. Mater.*, vol. 56, pp. 62–75, 2023.

-
- [32] X. Lin, “Real-time prediction of anode potential in li-ion batteries using long short-term neural networks for lithium plating prevention,” *J. Electrochem. Soc.*, vol. 166, no. 10, A1893, 2019.
- [33] J. C. Hamar, S. V. Erhard, C. Zoerr, and A. Jossen, “Anode potential estimation in lithium-ion batteries using data-driven models for online applications,” *J. Electrochem. Soc.*, vol. 168, no. 3, p. 030535, 2021.
- [34] S. Ahmed, I. Bloom, A. N. Jansen, *et al.*, “Enabling fast charging—a battery technology gap assessment,” *Journal of Power Sources*, vol. 367, pp. 250–262, 2017.
- [35] P. Arora, M. Doyle, and R. E. White, “Mathematical modeling of the lithium deposition overcharge reaction in lithium-ion batteries using carbon-based negative electrodes,” *J. Electrochem. Soc.*, vol. 146, no. 10, p. 3543, 1999.
- [36] C. Uhlmann, J. Illig, M. Ender, R. Schuster, and E. Ivers-Tiffée, “In situ detection of lithium metal plating on graphite in experimental cells,” *J. Power Sources*, vol. 279, pp. 428–438, 2015.
- [37] X. Lin, K. Khosravinia, X. Hu, J. Li, and W. Lu, “Lithium plating mechanism, detection, and mitigation in lithium-ion batteries,” *Prog. Energy Combust. Sci.*, vol. 87, p. 100953, 2021.
- [38] T. Gao, Y. Han, D. Fraggedakis, *et al.*, “Interplay of lithium intercalation and plating on a single graphite particle,” *Joule*, vol. 5, no. 2, pp. 393–414, 2021.
- [39] J. S. Edge, S. O’Kane, R. Prosser, *et al.*, “Lithium ion battery degradation: What you need to know,” *Phys. Chem. Chem. Phys.*, vol. 23, no. 14, pp. 8200–8221, 2021.
- [40] S. J. Moura, F. B. Argomedeo, R. Klein, A. Mirtabatabaei, and M. Krstic, “Battery state estimation for a single particle model with electrolyte dynamics,” *IEEE Trans. Control Syst. Technol.*, vol. 25, no. 2, pp. 453–468, 2016.
- [41] S. E. O’Kane, W. Ai, G. Madabattula, *et al.*, “Lithium-ion battery degradation: How to model it,” *Phys. Chem. Chem. Phys.*, vol. 24, no. 13, pp. 7909–7922, 2022.

- [42] X.-G. Yang and C.-Y. Wang, "Understanding the trilemma of fast charging, energy density and cycle life of lithium-ion batteries," *J. Power Sources*, vol. 402, pp. 489–498, 2018.
- [43] Y. Zhang, X. Li, L. Su, Z. Li, B. Y. Liaw, and J. Zhang, "Lithium plating detection and quantification in Li-ion cells from degradation behaviors," *ECS Trans.*, vol. 75, no. 23, p. 37, 2017.
- [44] V. Sulzer, S. G. Marquis, R. Timms, M. Robinson, and S. J. Chapman, "Python Battery Mathematical Modelling (PyBaMM)," *J. Open Res. Softw.*, vol. 9, no. 1, p. 14, 2021.
- [45] H. Ruan, J. Chen, W. Ai, and B. Wu, "Generalised diagnostic framework for rapid battery degradation quantification with deep learning," *Energy AI*, vol. 9, p. 100 158, 2022.
- [46] J. Vetter, P. Novák, M. Wagner, *et al.*, "Ageing mechanisms in lithium-ion batteries," *J. Power Sources*, vol. 147, no. 1, pp. 269–281, 2005.
- [47] M. Broussely, P. Biensan, F. Bonhomme, *et al.*, "Main aging mechanisms in Li ion batteries," *J. Power Sources*, vol. 146, no. 1-2, pp. 90–96, 2005.
- [48] Y. Gao, J. Jiang, C. Zhang, W. Zhang, Z. Ma, and Y. Jiang, "Lithium-ion battery aging mechanisms and life model under different charging stresses," *J. Power Sources*, vol. 356, pp. 103–114, 2017.
- [49] P. Bai, J. Li, F. R. Brushett, and M. Z. Bazant, "Transition of lithium growth mechanisms in liquid electrolytes," *Energy & Environmental Science*, vol. 9, no. 10, pp. 3221–3229, 2016.
- [50] J. M. Reniers, G. Mulder, and D. A. Howey, "Review and performance comparison of mechanical-chemical degradation models for lithium-ion batteries," *J. Electrochem. Soc.*, vol. 166, no. 14, A3189–A3200, 2019.
- [51] X.-G. Yang, Y. Leng, G. Zhang, S. Ge, and C.-Y. Wang, "Modeling of lithium plating induced aging of lithium-ion batteries: Transition from linear to nonlinear aging," *J. Power Sources*, vol. 360, pp. 28–40, 2017.
- [52] F. M. Kindermann, J. Keil, A. Frank, and A. Jossen, "A sei modeling approach distinguishing between capacity and power fade," *J. Electrochem. Soc.*, vol. 164, no. 12, E287, 2017.

-
- [53] X. Zhang, W. Shyy, and A. M. Sastry, “Numerical simulation of intercalation-induced stress in li-ion battery electrode particles,” *J. Electrochem. Soc.*, vol. 154, no. 10, A910, 2007.
- [54] I. Laresgoiti, S. Käbitz, M. Ecker, and D. U. Sauer, “Modeling mechanical degradation in lithium ion batteries during cycling: Solid electrolyte interphase fracture,” *J. Power Sources*, vol. 300, pp. 112–122, 2015.
- [55] C.-H. Chen, F. B. Planella, K. O’Regan, D. Gastol, W. D. Widanage, and E. Kendrick, “Development of experimental techniques for parameterization of multi-scale lithium-ion battery models,” *J. Electrochem. Soc.*, vol. 167, no. 8, 2020.
- [56] J. Sturm, A. Rheinfeld, I. Zilberman, *et al.*, “Modeling and simulation of inhomogeneities in a 18650 nickel-rich, silicon-graphite lithium-ion cell during fast charging,” *J. Power Sources*, vol. 412, pp. 204–223, 2019.
- [57] C. Pastor-Fernández, K. Uddin, G. H. Chouchelamane, W. D. Widanage, and J. Marco, “A comparison between electrochemical impedance spectroscopy and incremental capacity-differential voltage as Li-ion diagnostic techniques to identify and quantify the effects of degradation modes within battery management systems,” *J. Power Sources*, vol. 360, pp. 301–318, 2017.
- [58] C. Pastor-Fernández, T. F. Yu, W. D. Widanage, and J. Marco, “Critical review of non-invasive diagnosis techniques for quantification of degradation modes in lithium-ion batteries,” *Renewable Sustainable Energy Rev.*, vol. 109, pp. 138–159, 2019.
- [59] R. Xiong, Y. Pan, W. Shen, H. Li, and F. Sun, “Lithium-ion battery aging mechanisms and diagnosis method for automotive applications: Recent advances and perspectives,” *Renewable Sustainable Energy Rev.*, vol. 131, p. 110 048, 2020.
- [60] R. Pascanu, T. Mikolov, and Y. Bengio, “On the difficulty of training recurrent neural networks,” in *International conference on machine learning*, Pmlr, 2013, pp. 1310–1318.
- [61] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

- [62] Y. Fang, A. J. Smith, R. W. Lindström, G. Lindbergh, and I. Furó, “Quantifying lithium lost to plating and formation of the solid-electrolyte interphase in graphite and commercial battery components,” *Appl. Mater. Today*, vol. 28, p. 101 527, 2022.
- [63] C. R. Birkel, M. R. Roberts, E. McTurk, P. G. Bruce, and D. A. Howey, “Degradation diagnostics for lithium ion cells,” *J. Power Sources*, vol. 341, pp. 373–386, 2017.

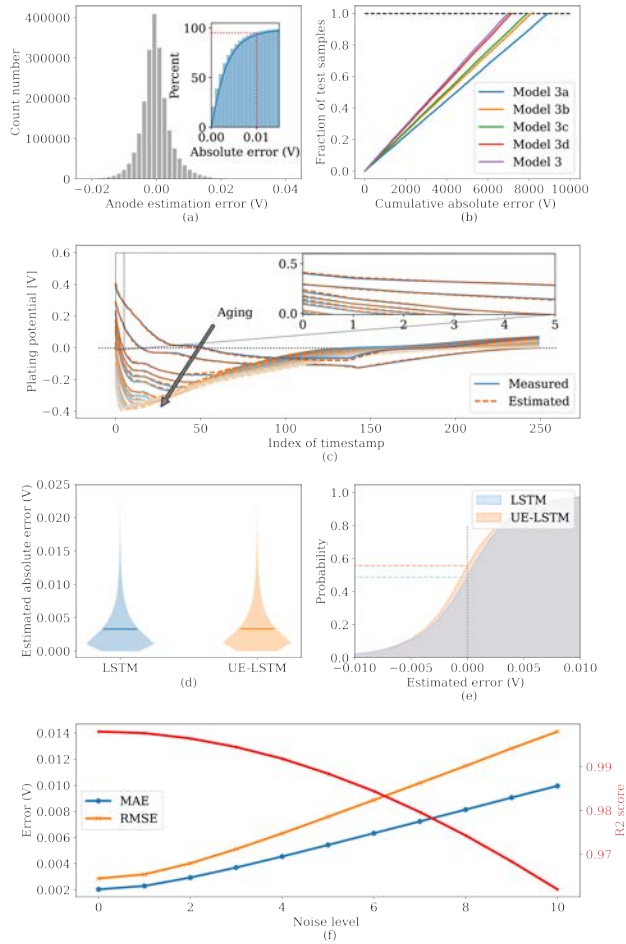


Figure 4: Estimation results for the plating potential. (a) presents the distribution of estimation errors for Model 3. (b) shows the cumulative absolute errors of Model 3 in comparison with four alternative models, each characterized by distinct input variables. (c) contrasts the plating potential estimated by Model 3 against the simulation profile for a randomly selected cell throughout its lifespan. (d) shows the violin plots of the estimation errors on the test dataset using the proposed UE-LSTM and conventional LSTM, where the horizontal lines represent the mean errors. (e) depicts the probability distribution of accumulative estimation errors using UE-LSTM and LSTM models. (f) shows the estimation errors and R_2 score for plating potential when using measurements polluted by different levels of noise.

PAPER **E**

**Harmonizing performance indicators and unifying data for lifelong
battery management**

Yizhou Zhang, Torsten Wik, John Bergström, Shafiq Urréhman,
Changfu Zou

Under review

The layout has been revised.

Abstract

Lithium-ion batteries represent an enabling technology for transportation electrification and renewable energy integration, yet their production requires substantial energy and resources. To optimize their usage, continuous monitoring and evaluation of battery performance indicators are essential. However, real-world vehicle applications face challenges due to stochastic usage patterns, data quality issues, uncontrolled environments, and insufficient validation data. To address these issues, we develop a neural network-based framework to harmonize battery performance indicators. This framework is able to project any performance indicators onto standardized reference conditions, thereby harmonizing traditionally noisy, drifted, and non-comparable indicators. With these harmonized indicators, we offer accurate real-time monitoring and prognosis of real-world batteries and enable systematic assessment and comparison under varying conditions and aging levels, as well as across different batteries. Furthermore, this unified approach can standardize data storage formats and sampling rates, significantly reducing data volume for storage and transmission. This work also reveals the impact of operational environments on battery performance and provides a potential standard for lifelong battery monitoring, predictive maintenance, and extended battery service life.

1 Introduction

The escalating challenge of global warming poses a profound threat to the environment, economy, and public health. Urgent and decisive actions are imperative to curtail greenhouse gas emissions and mitigate the far-reaching consequences of climate change [1]. Amidst this urgency, lithium-ion (Li-ion) batteries have emerged as a key player in the pursuit of sustainable solutions. Renowned for their low cost, high energy density, rapid response capabilities, and comparatively long lifespan, lithium-ion batteries are pivotal in driving the electrification of transportation and facilitating the increased penetration

of renewable energy sources [2], [3]. However, as the popularity of Li-ion batteries grows at an extraordinary pace, several significant challenges emerge. These include the high energy consumption during manufacturing, supply chain issues particularly related to cobalt, and the nascent state of battery recycling processes [4]–[6]. Consequently, it is of critical importance to extract the full potential of these batteries once installed. To achieve this, the key is to design a sophisticated battery management system (BMS) that accurately monitors battery states online and provides precise predictive informatics for future safety, reliability, and aging.

Batteries, as electrochemical devices, experience variations in energy and power capabilities due to factors such as operating temperature, cycling current/voltage profiles, manufacturing variability, and pack design. Additionally, battery performance is significantly affected by degradation, which is driven by complex and interactive mechanisms, and results in heterogeneous aging both between cells and within individual cells [7]–[10]. The intricate nature of battery degradation mechanisms and the variations across battery components introduce significant challenges in accurately assessing and monitoring battery performance changes, particularly for real-world applications where only field data are available. The direct application of field data to estimate battery performance indicators is further complicated by the stochastic and uncontrolled usage patterns of batteries in real-world scenarios, along with practical issues related to data quality and the absence of standardized validation tests [11]–[13].

Several indicators are widely used in the industry to assess battery performance, including but not limited to the state of energy, capacity fade, battery resistance, state of health (SoH), and state of safety. These indicators directly or indirectly impact the overall performance of end products, such as fast charging speed, total driving range, energy efficiency, and acceleration performance of electric vehicles (EVs). Quantifying these indicators and predicting their future characteristics require a comprehensive understanding of battery behavior in response to various external stress factors. For a given cell chemistry, these factors could include temperature profiles, current rates, depth of discharge, depth of charge, and other conditions over the battery’s entire lifespan. To address this problem, the current landscape of battery management algorithms encompasses various approaches categorized into model-based and data-driven methodologies [14], [15]. Model-based techniques typically involve

a representative model and a parameter estimation algorithm to monitor internal battery states. Empirical models, characterized by their simplicity, rely on predefined parametric functions derived from laboratory cycling data to describe battery input-output relationships. Despite their widespread use in the industry, these models exhibit limitations when extrapolating results beyond the specific cycling conditions employed in the laboratory, potentially leading to significant errors [16]–[18]. Another prevalent model-based approach involves equivalent circuit models, often expressed in a state-space format. Leveraging techniques derived from feedback control theory, these models facilitate the estimation of states or parameters, such as the internal resistance and capacity, using Bayesian recursive filtering or adaptive observers. However, applying such models in highly dynamic real-world scenarios can pose challenges, as their accuracy may diminish across a broad range of operating conditions due to parameter drifts [19]–[21]. Finally, electrochemical models, often based on porous electrode and liquid thermodynamics theories, can potentially capture the battery’s internal state dynamics. Aside from the high computational requirements, this type of model entails accurate parameterization suitable for different operating conditions and SoH levels. This complex process makes electrochemical models less favorable in real-world applications [22]–[24].

Compared to their model-based counterparts, data-driven methods have drawn much attention for their high flexibility and good capability to recognize complex patterns and mechanism-agnostic characteristics [25]–[27]. Leveraging only measurable signal data to formulate feature inputs, these data-driven approaches exhibit commendable capability in mapping battery indicators and time-varying parameters as the output. For such estimation tasks, methods such as random forest regression, Gaussian process regression (GPR), neural network (NN), and support vector regression have been adopted [28]–[31]. However, a predominant focus in the literature has been on the investigation of laboratory battery cycling data conducted under relatively controlled conditions. Despite some pioneering efforts involving deliberate cycling with random profiles to emulate real-world battery usage scenarios, the datasets generated from such endeavors remain notably limited in size compared to the datasets derived from batteries in active service [11], [13]. This discrepancy underscores a gap in existing research, calling for advanced data-driven methods to effectively track battery performance indicators under authentic

and dynamic real-world operating conditions.

Owing to constrained access to battery field data resources, only a few published studies have explored this domain, and most of them are focused on battery capacity and power fade. A comprehensive survey of the literature reveals notable investigations. Zhang et al. [32] examined the field data of 7,296 plug-in hybrid EVs, employing four distinct machine learning (ML) models and an online adaptation method to track battery state trajectories under arbitrary operating conditions. Deng et al. [33] focused on 20 EVs over 29 months, constructing two residual models to predict remaining capacity sequences. However, the labeled output in these studies relied on onboard BMS-reported capacity estimates or post-process methods, such as Ampere hour counting, which inadequately considers the impact of operating conditions on available battery capacity. Huo et al. [34] investigated two years of operational data from 16 EVs, employing a Bayesian network to estimate the SoH of battery packs. The ML output target was obtained through periodic battery capacity calibration during a standardized charging event under controlled temperature conditions. However, this calibration, intended for in-service vehicles, inevitably disrupts normal operations and incurs additional costs. By She et al. [35], electric bus operational data was utilized together with a radial basis function NN to monitor incremental capacity (IC) peak value changes during battery degradation. A support vector regression was applied to mitigate the impact of operational conditions on IC peak values by smoothing the acquired raw data. Similarly, Zhou et al. [36] studied resistance changes over the battery’s lifetime using similar data sources. They employed an empirical method to address the substantial temperature impact on resistance, though potential oversimplification arose due to other factors like state of charge (SoC) and current rates also affecting battery resistance. To address these challenges, for lead-acid batteries, Aitio et al. [37] proposed a recursive GPR framework to model the in-field charging resistance development. However, due to dataset limitations, the battery capacity was not considered, and the use of GPR over extensive data still requires significant training time despite a recursive framework. A recent work by Pozzato et al. [12] analyzed one year of lithium-ion battery field data from an Audi e-tron, presenting a data processing pipeline to assess battery health indicators (HIs) during vehicle operations. The results underscored the seasonal-dependent temperature’s impact on the studied HIs, yet the more intricate degradation

behavior of these HIs was not explored, possibly due to the short data collection period.

To fundamentally address the challenges identified above in both model-based and data-driven methods, this work proposes an ML-based framework to harmonize battery performance indicators and unify real-world field data. This framework projects key battery performance indicators, such as internal resistance and capacity, onto standardized reference conditions. Trained and tested comprehensively using the largest known dataset from a fleet of EVs, this approach harmonizes traditionally noisy, drifted, and non-comparable performance indicators. As a result, it offers accurate real-time monitoring and prediction of battery performance. Additionally, we enable systematic assessment and comparison of battery performance indicators under varying conditions and aging levels, as well as across different batteries. Furthermore, our established standard process for indicator extraction unifies battery data and significantly reduces the required data storage and transmission. The harmonized indicators also enhance model-based methods by systematically accommodating physical parameters, thereby maintaining high model fidelity across a wide range of applications.

2 Description and processing of field vehicle data

The dataset utilized in this study originates from onboard BMS of LYNK&Co 01 cars, compact plug-in hybrid electric vehicles (PHEVs) equipped with a campus area network (CAN) for data transmission. These vehicles, predominantly operated in Gothenburg, Sweden, have been in service for about two and a half years, with the highest mileage vehicle reaching 140,000 km. The traction battery pack of these vehicles comprises eight series-connected modules, each with 12 series-connected cells, giving a total rated voltage of 345.6V and a nominal capacity of 50Ah for both individual cells and the entire pack. Detailed schematics and topologies of the battery pack and its modules are illustrated in Fig. 1a-c.

The BMS architecture follows a master-slave configuration, employing a differential daisy chain for efficient communication between the master and slave boards. Each module features a slave board responsible for measuring cell voltages and the surface temperatures of two cells. These slave boards also manage the operation of passive balancing circuits. The master board, on

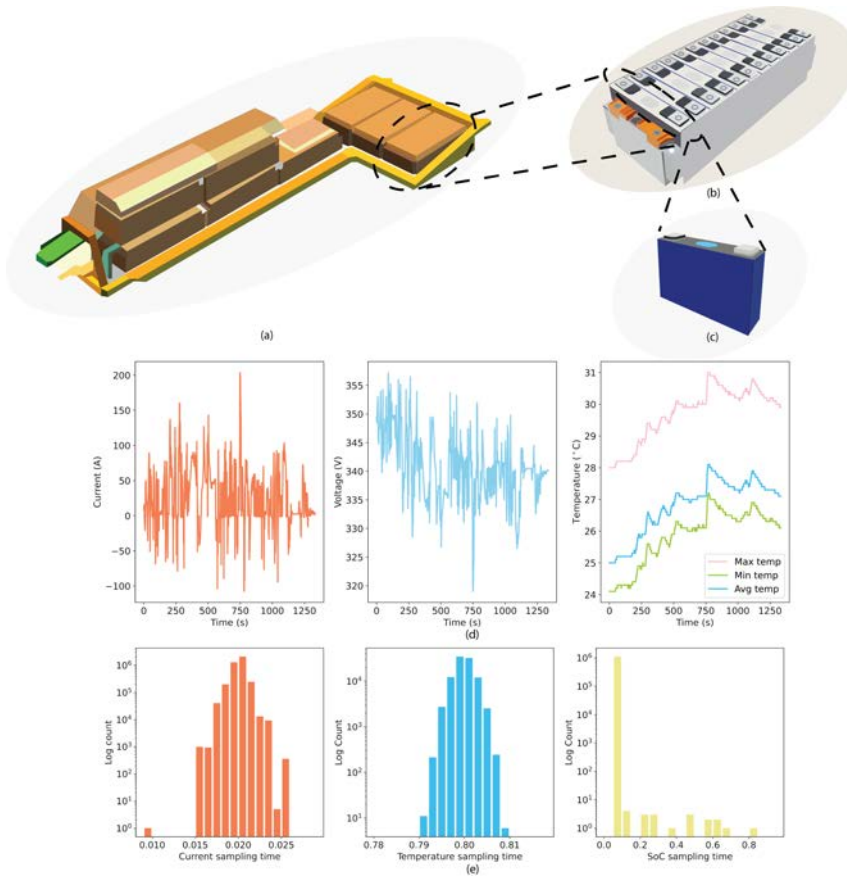


Figure 1: Overview of the Lynk & Co 17.6 kWh battery pack and its components, as well as the measurement data from its BMS. (a) Schematic of the Lynk & Co 17.6 kWh battery pack, consisting of 8 modules, each with 12 cells in an all-series topology. (b) Detailed schematic of a battery module. (c) Representation of an individual 50 Ah prismatic cell. (d) Driving cycle measurements including current, voltage, and temperature data from the pack’s sensors. (e) Histograms depicting the sampling times for the current, temperature, and estimated pack SoC by CAN signals.

the other hand, plays a pivotal role in the battery’s internal state estimations,

thermal management, fault diagnostics, control of the battery disconnect unit (BDU), and communication of critical battery information with external electronic control units (ECUs) within the vehicle. The pack-level current and voltage measurements, executed in the BDU, are transmitted to the master board via an internal CAN, ensuring a timely and synchronized data relay.

Our battery dataset contains comprehensive driving and charging data from six PHEVs, accumulated from March 2021 to October 2023, resulting in a big dataset of over one terabyte. To the best of the authors' knowledge, this is the largest and most detailed dataset publicly shared for field data-driven studies of battery aging. All battery-related signals on the vehicle propulsion CAN are stored together with critical signals from other propulsion system components. Since the data points were allocated to different CAN frames, time-series measurements for the same signal demonstrated distinct sampling frequencies. The pack current and voltage are supposed to be measured every 20 ms, and battery temperatures are nominally measured every 800 ms. Additionally, technical limitations, such as the CAN communication priority transmission scheme and inherent communication delays, resulted in non-uniform measurement frequencies across signals. Fig. 1e presents histograms depicting the distribution of battery pack current, average temperature measurements, and estimated pack SoC alongside their respective sampling times throughout ten consecutive driving cycles. The data was logged during the vehicle's daily usage, including both driving and charging segments. During parking, however, the vehicle's ECUs entered a sleep mode, suspending data collection.

For data-driven battery modeling and management, the first and fundamental step is to develop a systematic and robust data processing pipeline. This pipeline must extract all critical information while reducing the dataset to a manageable size, facilitating subsequent training and deployment of ML models. Moreover, it should be scalable and universally applicable across the entire fleet rather than being limited to specific vehicles. This task faces a set of practical challenges caused by sparse, unsynchronized, and noisy signal measurements. A comprehensive description of our constructed pipeline is provided in Section 5.

3 Results and discussions

3.1 Data-driven battery aging diagnostics

In principle, vehicle field measurement data contains rich information to indicate the overall performance of the battery. However, systematically extracting and evaluating the performance indicators under standardized conditions along the lifetime of a vehicle, and between different vehicles in a fleet, remains challenging. Conventionally, the battery resistance and capacity are used as key parameters to indicate the overall performance of the battery. In laboratory cycling cases, reference performance tests (RPTs) are often conducted periodically between dynamic cycling sessions to assess changes in battery capacity and resistance. Usually, several constant-current constant-voltage (CC-CV) charge and discharge cycles at fixed and low C-rates, where C-rate is defined as the current divided by the nominal cell capacity, are performed to determine the cell capacity, and a hybrid pulse power characterization (HPPC) test is often used to calculate the cell resistance. Additionally, all these tests are conducted at a controlled temperature, usually 25°C. However, for batteries deployed in EVs or serving as a stationary energy storage system (ESS), finding such specific cycling conditions is unfeasible.

Both the resistance and capacity of battery cells highly depend on their operating conditions. Specifically, the resistance is influenced by the operating temperature, current level, SoC level, and aging status. Similarly, the cell capacity is affected by the current level, cycling temperature, and aging status. Therefore, to accurately probe and evaluate the performance indicators of the battery, it is critical to assess and compare the resistance and capacity values under standardized conditions. In other words, developing effective methods to compensate for all these influencing factors is essential for accurate monitoring of battery resistance and capacity. To achieve this, we leverage the big battery dataset from EV fleets and ML techniques to propose a robust and efficient framework for systematically harmonizing and assessing battery performance indicators.

Conventional BMSs often apply recursive estimation methods, such as Kalman filters (KF) and recursive least squares (RLS), for battery internal resistance estimation. For capacity estimation, ampere-hour counting and its variants remain the most dominant methods. However, as shown in blue in Fig. 2a for resistance estimation using RLS and Fig. 5a for capacity estimation using

ampere-hour counting, the estimation results are severely polluted by noise, making it difficult to draw constructive conclusions about battery performance and its evolution with aging. Similar observations and challenges have been highlighted in [12]. Nevertheless, such results can still be good representations of the actual battery resistance and capacity under the specific measurement conditions under which the estimation was conducted. Hence, from a supervised learning perspective, these estimation results can be treated as the output or target of the ML model within our framework. Correspondingly, the operating conditions during the estimation can be parameterized and utilized as the inputs or features of the designed ML model. For the detailed implementation of real-time resistance and capacity estimation, including feature construction and ML model training, see Section 5. With the ML model obtained, all battery performance indicators and parameters can be projected and harmonized to a standardized reference space or point, such as the conditions for conducting RPTs. These harmonized values can then be analyzed and compared across different cells or the same cells under varying conditions and SOH levels.

3.2 Resistance during acceleration and regenerative braking

The resistance of a vehicle battery pack indicates its power capability and directly impacts the vehicle's acceleration and braking performance. Leveraging instantaneous changes in the pack's voltage and current, and using an RLS estimation scheme, resistance values can be estimated in real-time. By sorting the estimated results based on the sign of the currents at each time step, the resistance can be categorized into two types. Specifically, while positive currents indicate discharging and correspond to the acceleration resistance, negative currents represent the regenerative braking resistance. Using the vehicle with the longest driving distance in the dataset as an example, the estimation results are presented in Fig. 2c and Fig. S1a. Clearly, the data points of resistance exhibit a strong correlation to seasonal temperature variations where the resistance value increases when the ambient temperature decreases. However, due to their severe fluctuation along the accumulative mileage, these resistance values are not informative and almost useless in the existing BMSs, as stated earlier in Section 3.1.

NN estimated battery resistance accuracy. To extract the universal resistance, namely the resistance harmonized on a standardized reference

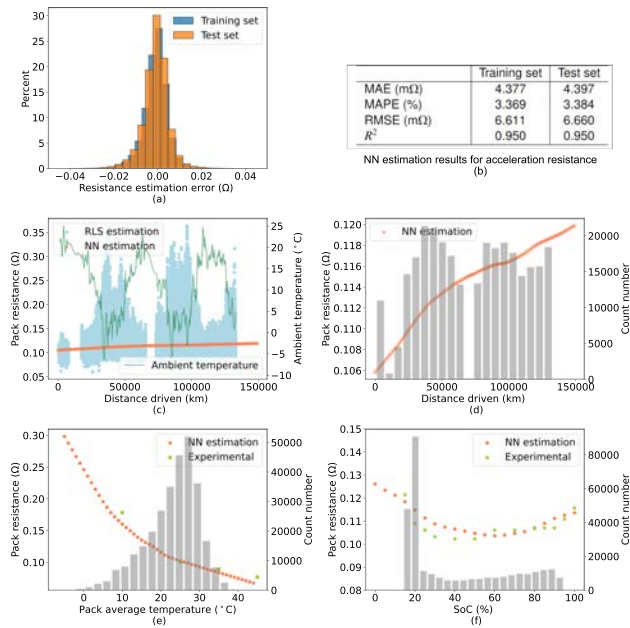


Figure 2: Estimation results of the battery acceleration resistance estimation. (a) Histogram of estimation errors for acceleration resistance using NN on both the training and test sets. (b) Numerical results of NN estimation for acceleration resistance. (c) Comparison of the battery acceleration resistance estimation over the accumulated driving distance of the vehicle using RLS estimation and NN evaluated at the standardized reference point. The vehicle only contains data till around 130,000 km. The orange line exceeds 130,000km, which is the NN resistance prediction value. (d) Detailed view of the NN-assessed resistance development over vehicle distance, alongside a histogram detailing data distribution across distance intervals. (e) NN analysis of temperature dependence on acceleration resistance, accompanied by a histogram quantifying data across temperature intervals. (f) NN evaluation of SoC dependence on acceleration resistance, with an associated histogram showing the distribution of data across current intervals.

space or point, over the battery’s lifetime, we build an NN model using the RLS-based estimates of resistance as targets and the corresponding parametric measurement conditions as inputs. These inputs are extracted from the cur-

rent, SoC, temperature, and accumulated driving distance data. Compared to factors, such as the energy or ampere-hour throughput, that may be used as a feature in the model to capture battery aging, the accumulated mileage is selected due to its direct availability in the field data. The complete dataset was split into training data and test data, corresponding to 80% and 20% of the whole data, respectively. The detailed estimation results on the training and test set are outlined in Fig. 2a and b. The proposed NN model achieves a commendably high overall estimation accuracy, with MAPE consistently below 3.5% over the complete two-and-a-half-year vehicle usage period and the estimation error rarely exceeds ± 20 m Ω on the test dataset. Furthermore, the estimation results on the model unseen test set are only slightly worse than the training data. This low discrepancy suggests that the model has not been overfitted to the training set and maintains high fidelity. However, It is crucial to underscore that the overarching goal of the learned NN is not to replicate the RLS estimation results but rather to learn the relationship between the different impacting factors on the battery resistance. The accurate estimation of the real-time resistance, however, lays the foundation for the later resistance evaluation under the standardized condition. Details of the hyperparameter used in the designed NN are presented in the supplementary materials.

Estimation results of battery resistance over various influencing factors using NN. Fig. 2c compares the resistance over aging using RLS methods with the NN model evaluated at a standardized reference point. Compared to noisy and fuzzy RLS estimation results which can hardly draw any conclusion, the NN model estimation results are much smoother and successfully illustrate the resistance increase over the battery's lifetime. Additionally, due to its exceptional extrapolation capabilities, the NN model can serve as a predictive tool for forecasting future resistance development. Noteworthy is that we have chosen the standardized evaluation point for the resistance to be the mean value of the studied vehicle data, as the region around that point is usually data-rich and representative of general usage. Fig. 2d offers a detailed view of the NN model's estimates, revealing a monotonically increasing trend in battery resistance with accumulated vehicle distance. This observed behavior is consistent with findings from experimental studies on commercial lithium-ion batteries.

Fig. 2e and Fig. 2f illustrate the learned relationship of the battery acceler-

ation resistance results over the temperature and SoC. The results align well with fundamental battery physics principles, indicating the model’s accuracy and reliability.

Firstly, resistance shows a clear, monotonic decrease with rising temperature, consistent with the Arrhenius equation’s exponential relationship. Additionally, compared to the other influencing factors, the temperature is detrimental in determining the resistance values as shown in Fig. S1, where a systematic permutation importance analysis was conducted to study the importance of each impacting factor on the battery resistance. Numerically, the temperature-induced change in the battery pack resistance can result in a variation of up to 0.2Ω as shown in Fig. 2e. Rather reassuringly, we can partially verify the estimation results with the experimental beginning of life (BOL) cell results (green crosses). The NN model demonstrates strong alignment with test results, particularly in regions with abundant training data, underscoring the model’s precision. Minor discrepancies in data-sparse regions are a typical challenge in supervised learning models [38].

Secondly, the internal resistance’s dependency on SoC is accurately captured, exhibiting a characteristic parabolic curve. This phenomenon is ascribed to the anode’s charge transfer predominance at high SoC levels and the cathode’s charge transfer dominance at low SoC levels [23], [39]. The NN model effectively mirrors the experimental trends, although minor deviations are noted. These may be attributable to SoC’s relatively lower impact on resistance—approximately one-tenth that of temperature—resulting in a variation of about 0.02Ω making it difficult to extract an exact relationship.

Finally, as shown in Fig. S2, the battery resistance decreased as the current level increased, mainly due to the activation overpotential increasing slowly at higher current density, as elucidated by the Butler-Volmer equation [40]. This behavior also harmonizes with most of the experimental results probing the relationship between battery resistance and applied current level [41]. Notably, the impact of current on the battery resistance is around 0.04Ω , considerably lower than the other impacting factors.

For the regenerative braking resistance, the overall estimation results are presented in Fig. S3. The conclusions that we draw on the acceleration resistance are also valid here, the regenerative braking resistance exhibits similar to the one we found during acceleration on different impact factors.

Estimation robustness demonstration. In data-driven methodologies,

the volume of training data significantly influences the performance of ML algorithms. This importance is especially evident in our study, where the primary goal is to understand how resistance relates to various impact factors, rather than merely achieving high accuracy. Therefore, it is crucial to examine how sensitive the learned resistance behavior is to the amount of training data. To illustrate this, we explored the aging aspect of battery resistance changes. Figure 3a displays the NN’s learned relationship between battery resistance and the vehicle’s accumulated driving distance, using varying amounts of training data. We analyzed ten scenarios, each with a 10% incremental increase in training data, reflecting real-world conditions where vehicles accumulate more data over time. The results indicate that with limited data, particularly at the BOL, it is challenging to accurately learn the aging behavior of battery resistance. Specifically, data from the initial 30,000 km is insufficient to capture the resistance changes accurately, as early resistance variations are primarily influenced by temperature, SoC, and current levels. However, as more data is collected over a longer period, the NN model demonstrates a more consistent learning of resistance aging behaviors. This is evidenced by the robust results obtained when using more than 30% of the data, corresponding to an accumulated driving distance of over 40,000 km.

Due to the inherent stochasticity in the training process and the intrinsic characteristics of the NN algorithm, the performance of the trained model exhibits variability, especially when the model’s objective extends to learning functional relationships. Consequently, the robustness and consistency of the model should be assessed. Fig. 3b illustrates the results of NN training repeated for 100 different random seeds. The depicted analysis reveals that the learned resistance behavior over vehicle accumulated driving distance is rather robust, predominantly falling within a $\pm 3 \text{ m}\Omega$ boundary across multiple runs, affirming the trustworthiness of the established relationship. Additionally, looking at the model accuracy evaluated on the test set shown in Fig. 3c, the MEAP of the 100 NN varies only from 3.3%-3.55% with a standard deviation of 0.06%. This consistency further underscores the robustness exhibited by the model.

Unifying real-world field data. Current in-vehicle data logging and storage remain at an early stage of development. Different car manufacturers, and even different products within the same company, employ varied strategies regarding which data to save and the sampling rates to use. Although

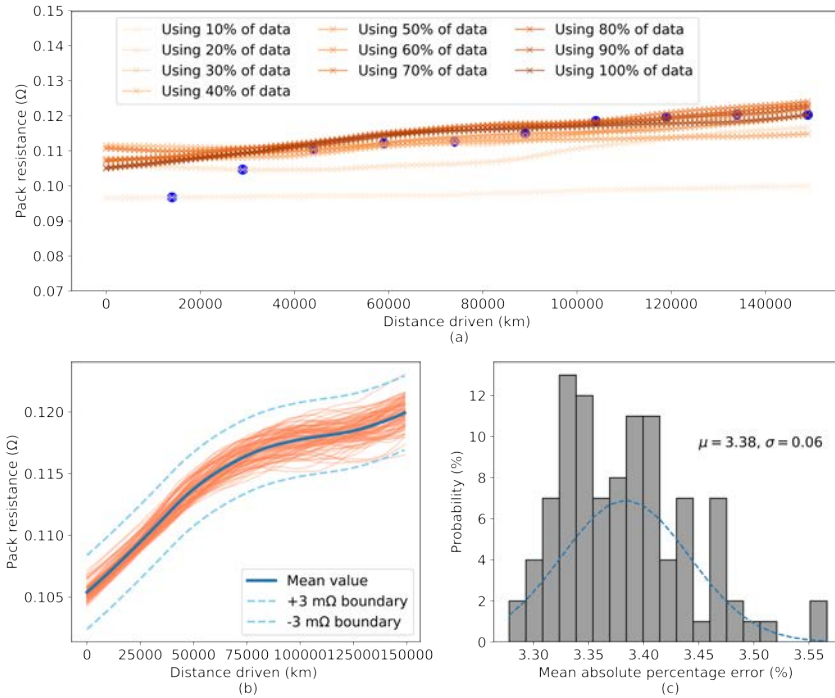


Figure 3: Analysis of fleet-level pack resistance performance over the battery’s lifetime. (a) Sensitivity analysis showing the impact of dataset size on NN resistance estimation accuracy, with the blue dot indicating the last data point used for training (data preceding this point is included in the training datasets) under various dataset size scenarios. (b) Variability in resistance aging curves generated from 100 NN models, each trained with different random seeds. (c) Aggregated estimation accuracy across the 100 NN models, demonstrating predictive reliability.

regulatory authorities have mandated battery information transparency and traceability—such as the EU Battery Passport initiative—detailed technical specifications are still under development. Here we present a systematic comparison study of extracting battery performance indicators using raw data with different sampling rates. The numerical estimation results are presented in Table 1. As sampling rates increase, the NN estimation accuracy decreases, likely due to the combined factors of reduced training data and the failure to

capture fast dynamics during vehicle driving. Such results mandate the requirement of striking a balance between how fast the data needs to be logged and the richness of the saved data information.

However, with the proposed battery performance indicator extraction framework, we can potentially avoid such a dilemma by only saving the trained NN model or simply the relationship between the interested performance indicators and impacting factors. Fig. 4 shows the relationship between battery resistance over the accumulated driving distance, pack average temperature, and SoC. Overall, we can conclude that the learned relationship is plausible and convincing across different raw data sampling rates. Specifically, resistance increases with aging, decreases as temperature rises, and exhibits a parabolic trend within the SoC window. It is noteworthy that with an increasing sampling rate, the learned resistance value also rises due to charge transfer and diffusion reactions within the battery, resulting in a higher overall resistance. Therefore, once the sampling rates are fixed, as is often the case in commercial BMS, these NN-learned functional relationships can provide unique and necessary information to indicate battery performance, allowing for comparison with other vehicles' batteries. Furthermore, it is recommended to use a sampling rate of less than 1 second for current, voltage, and temperature measurements to ensure a more reliable and accurate relationship.

Compared to transmitting and storing vast amounts of raw measurement data from individual batteries, saving the learned NN models or the functional relationships of performance indicators is far more energy and cost-efficient. Additionally, this information can be directly integrated into the BMS or other vehicle controllers for immediate use without further processing, thereby unifying onboard battery field data. We foresee this framework becoming the standard for logging battery information in the future.

Sampling rates (s)	0.5	1	2	4	8
MAE	4.390	4.523	5.522	6.894	7.994
RMSE	6.610	7.164	8.527	10.301	11.329
MAPE	3.396	3.396	3.953	4.813	5.634
R^2	0.942	0.942	0.919	0.880	0.838

Table 1: Comparison of NN acceleration resistance estimation results using a dataset with different sampling rates

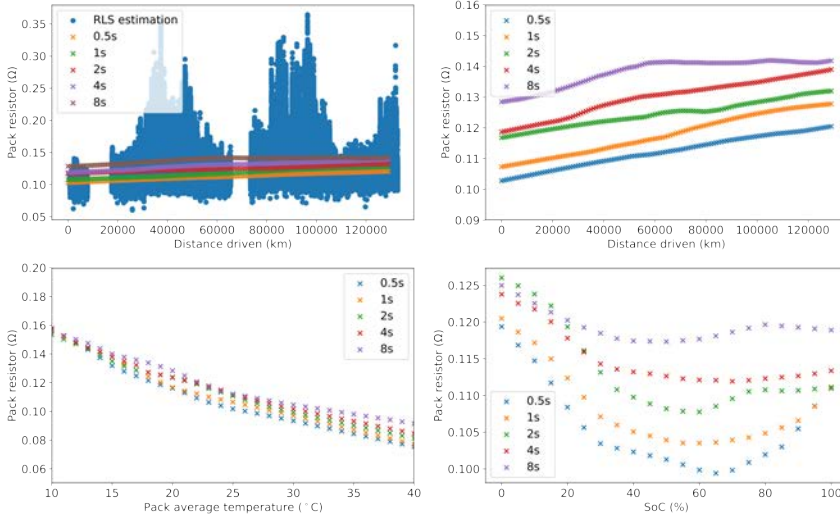


Figure 4: Comparison of NN-learned performance indicators using different raw data sampling rates. (a) Battery acceleration resistance estimation over the vehicle’s accumulated driving distance, comparing RLS estimation and NN evaluation at the standardized reference point. (b) Detailed comparison of NN-assessed resistance development over vehicle distance using raw data with different sampling rates. (c) NN analysis of temperature dependence on acceleration resistance using raw data with different sampling rates. (d) NN evaluation of SoC dependence on acceleration resistance using raw data with different sampling rates.

3.3 Capacity during vehicle charging

In contrast to the highly stochastic and dynamic discharge part of battery usage in EVs, the charging process is generally standardized as the engineers usually predefine the charging currents and target SoC levels. Therefore, numerous studies focus on leveraging the charging segment of vehicle usage to assess the battery health status, especially the battery pack’s capacity [32]. However, similar to battery resistance, the battery’s available capacity is significantly influenced by the applied current and temperature. Evaluating the capacity under standardized operating conditions, such as those defined in an RPT, is therefore imperative. Here, the available capacity of the battery is estimated using a commonly used Ampere hour counting method coupled with

an open circuit voltage (OCV)-adjusted SoC method. Comprehensive details regarding the computation of the available battery capacity are outlined in Section 5.4.

Fig. 5a compares the results acquired using the traditional Ampere hour counting method and those derived from the trained NN, evaluated at the standardized operating conditions. The NN model results effectively capture the capacity change over the vehicle’s accumulated driving distance, demonstrating a reasonable monotonic decrease over the course of the vehicle’s lifetime. However, we also notice that the capacity retention rate over the vehicle mileage varies. Rather reassuringly, the fast decrease of the available capacity in the initial phase aligns well with most laboratory experimental results on comparable commercial automotive batteries [42]–[44].

Additionally, the capacity retention rate is slower between 30,000 km and 70,000 km compared to the remaining range. While experimental verification for in-service vehicles is practically challenging, this observed phenomenon is consistent with battery physics, as lower temperatures tend to result in a reduction of available capacity. By examining the usage history during this period, we hypothesize that the lower capacity retention during such a period is due to considerably milder ambient temperatures and battery temperature compared to the rest of the period, as shown in Fig.5c–d. According to a large automotive-grade lithium-ion battery (with the same cathode and anode material and similar rated capacity as in this study) aging test campaign conducted by Stadler, etc. [42], the most healthy optimal temperature for such cells is between 20°C and 25°C with increasing degradation if the battery experiences lower or higher temperature. In our case, the considerably longer and warmer temperatures experienced in the remaining periods, other than 30,000km to 70,000km, may contribute more to the battery’s aging and, consequently, faster capacity retention. Nevertheless, one of the major challenges working with field data is the lack of validation [11]. Thus, verifying such a hypothetical assumption is practically challenging as no periodic battery performance calibration was planned for the studied vehicle.

Furthermore, Fig. 5b illustrates the relationship between the battery available capacity and the temperature when the vehicle has covered a mileage of 50,000 km. In general, with an increasing pack average temperature the available pack capacity is also increasing which aligns well with the existing theory of how battery available capacity changes with the temperature.

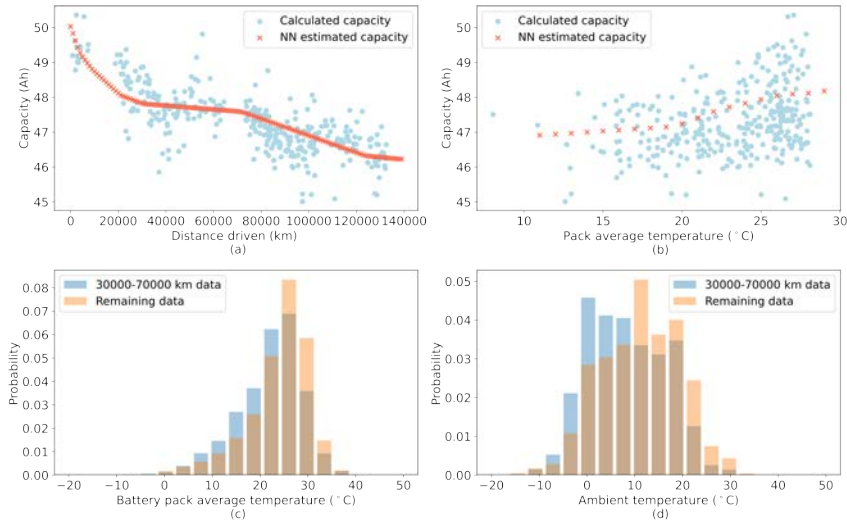


Figure 5: Comprehensive analysis of battery capacity estimation. (a) Comparison of estimated available battery capacities over vehicle driving distance, employing both Ampere-hour counting and NN models, evaluated at standardized reference points. (b) Depiction of NN-estimated available pack capacities under varying average pack temperatures. (c) Histogram of average battery pack temperatures for vehicles with accumulated driving distances between 30,000 km and 70,000 km, along with remaining data. (d) Histogram of ambient temperatures for the same driving distance range and remaining data.

3.4 Fleet level evaluation

In addition to a standardized and systematic assessment of individual EV battery performance indicators, another crucial assessment involves comparing vehicles of the same type within the entire fleet. Such comparative analysis provides valuable insights into the impact of usage behavior on battery performance and can effectively identify premature failures in specific battery packs. Moreover, by fusing the raw data of all the vehicles within the fleet, an average fleet model can be established to serve as a reference benchmark to compare the performance of each individual vehicle. Fig. 6a presents the comparative results of how the battery resistance changes over the vehicle mileage across a total of six vehicles in the studied fleet, alongside the average performance

model results using all vehicle data. It is noteworthy that Vehicle 01 and Vehicle 03, possessing more extensive datasets than the rest of the fleet, influence the average model to align closely with their battery resistance aging behaviors. This alignment is particularly evident for mileage values exceeding 40,000 km, where other vehicles have hardly accumulated any driving data.

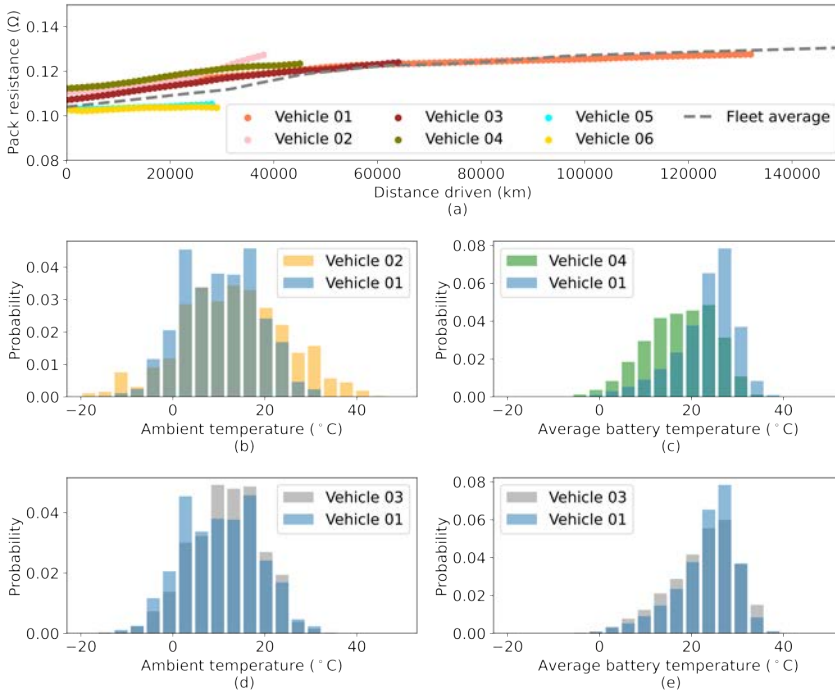


Figure 6: Comparison and analysis of the battery acceleration resistance degradation behavior. (a) Overview of the resistance degradation of individual vehicles within the fleet alongside the fleet average aging behavior, evaluated at the standardized reference point. (b) Comparison of the ambient temperature histogram between Vehicle 01 and Vehicle 02. (c) Comparison of the battery pack average temperature histogram between Vehicle 01 and Vehicle 04. (d) Comparison of the ambient temperature histogram between Vehicle 01 and Vehicle 03. (e) Comparison of the battery pack average temperature histogram between Vehicle 01 and Vehicle 03.

In general, all vehicles within the studied fleet exhibit a discernible increase

in resistance over the vehicle-driven distance, with some vehicles displaying a bit different incremental rates. This may be attributed to the varied cycling and operating conditions experienced by the batteries. Vehicle 01, having accrued the most driving experience and closely mirroring the fleet’s average resistance degradation behavior over the entire studied distance, serves as the benchmark for comparison.

Fig. 6b-e presents detailed histogram comparisons of specific usage conditions across different vehicles in the fleet. The higher resistance degradation rates observed in Vehicles 02 and 04 can potentially be attributed to the significantly elevated ambient temperatures and lower average battery temperatures encountered throughout their usage history. Nevertheless, similar aging patterns were discovered between Vehicles 01 and 03. Encouragingly, the ambient and average battery temperatures experienced by these vehicles, in terms of occurrence probability, are similar, with minor variations that partially validate the hypothesis concerning the different aging behaviors observed in Vehicles 02 and 04. Another notable finding is that Vehicles 05 and 06 exhibit lower beginning-of-life (BOL) resistance values compared to the rest of the fleet. This can be attributed to the fact that the batteries in these two vehicles come from a different production batch,

4 Conclusion

This study introduces a novel NN-based framework that systematically assesses lithium-ion battery performance indicators using real-world field data. Our framework harmonizes performance indicators such as internal resistance and capacity, correlating them with environmental and operational variables including temperature, current rates, SoC, and vehicle mileage. This method provides a deeper understanding of battery performance across a wide range of operating conditions and throughout the entire lifespan of the battery. By projecting performance indicators onto arbitrary reference conditions and harmonizing noisy estimations across various operating conditions and different battery packs, our approach significantly enhances the reliability of these predictions. This aids in the development of more effective battery management and maintenance strategies. Furthermore, this study addresses a critical gap in the use of field data for battery health assessment—namely, the absence of regular reference performance tests that can provide a ground truth for

training other predictive models. By acquiring standardized metrics, we reduce data volume for storage and transmission which unifies data storage formats and sampling rates. Moreover, the obtained indicators can systematically accommodate parameters in model-based methods, maintaining high model fidelity under varying conditions and aging levels. In conclusion, this work contributes to advancing battery technology by offering a robust tool for the optimization of the lifespan and efficiency of battery systems in various applications. It positions the framework as a potential standard for lifelong battery management, predictive maintenance, and extended battery service life, thereby supporting the sustainable use of resources and minimizing environmental impact.

5 Method

5.1 Resource availability

Lead contact

Further information and requests for resources and materials should be directed to and will be fulfilled by the lead contact, Changfu Zou (changfu.zou@chalmers.se)

Materials availability

This study did not generate new materials.

Data and code availability

The detailed fleet data containing six PHEVs battery pack real-time measured current, voltage, temperature, and SoC data during driving and charging that are used and analyzed in this paper will be made publicly available upon completion of the review of this manuscript.

5.2 Data processing

In this study, several data processing procedures were implemented to mitigate the volume of data requiring processing and enhance the subsequent speed of ML training. Initially, log data pertinent to driving or charging

modes were selectively retained based on the vehicle’s usage mode, encompassing driving, charging, idling, and servicing, among others. Addressing the signal synchronization challenge highlighted in Section 2, all discharging data were interpolated to every half-second and charging data to every second using linear interpolation. The reason for this dissimilarity in interpolation intervals is that only minor changes in current are made during the charging phase. The selection of a half-second sampling rate for the driving represents a compromise between the reliability of resistance estimation and processing speed. A comprehensive comparative analysis of the real-time resistance estimation results using different sampling rates is available in the supplementary materials. Subsequent to these processing steps, a data sanity check was executed to eliminate evident errors within the dataset, such as instances where default values were employed instead of authentic measured values during vehicle initialization.

The cumulative raw data retrieved from the fleet database amounted to approximately 1TB. Following the execution of the aforementioned data processing pipeline, the data underwent compression, resulting in a reduced size of 5.6GB. The entire procedure was carried out utilizing the company’s cluster infrastructure at Zeekr Technology Europe AB and took approximately 150 hours.

In this work, we utilize only the pack-level voltage and current data, treating all 96 series-connected cells in each battery pack as a big, lumped cell. Consequently, the derived performance indicators and parameters represent the entire battery pack. Notably, the proposed framework is generic and can be equally applied to handle cell-level performance indicators and parameters once cell voltage and current information become available.

5.3 Real-time battery resistance estimation

For the real-time estimation of battery resistance, the focus is placed on the driving sector data, leveraging the abrupt changes in current characteristics of dynamic driving events to identify on-site resistance values. The acceleration and regenerative braking events are differentiated by examining the sign of the current, with positive current indicating regenerative braking and negative representing acceleration events. The RLS estimation method, similar to the one introduced in [45], is employed for online battery resistance estimation.

Given a sufficiently short time interval, the battery ohmic resistance R_0 can

be approximated as

$$R_0 = \frac{U(k) - U(k-1)}{I(k) - I(k-1)}, \quad (\text{E.1})$$

where U is the terminal voltage, I is the applied current and k is the time instant. Note that the denominator needs to be non-zero, which will be handled during the conditioning check. The resistance, then, can be estimated using the RLS method, i.e.,

$$\hat{R}_0(k) = \hat{R}_0(k-1) + K(k)(\Delta U(K) - \hat{R}_0(k-1)\Delta I(k)), \quad (\text{E.2})$$

where $\Delta U(K) = U(k) - U(k-1)$ and $\Delta I(K) = I(k) - I(k-1)$. The gain $K(k)$ and the covariance $P(k)$ undergo updates according to the following equations:

$$K(k) = \begin{cases} \frac{P(k-1)\Delta I(k)}{\lambda + \Delta I(k)^2 P(k-1)}, & \text{if } |\Delta I(k)| > \sigma \\ 0 & \text{otherwise} \end{cases} \quad (\text{E.3})$$

$$P(k) = \frac{(1 - K(k)\Delta I(k))P(k-1)}{\lambda}. \quad (\text{E.4})$$

In these equations, σ is selected to be 25% of the maximum current value in the studied driving log file. Additionally, if the maximum difference in current ΔI is less than 50A throughout an entire driving log file, such a log file is excluded from the calculation. The estimated resistance value R_0 , along with the corresponding operating conditions (including current level, battery and ambient temperature, SoC, pack voltage, and vehicle accumulated driving distance) at the specific estimation point, are preserved. These data points are retained to function as input features for subsequent NN training.

5.4 Real-time battery available capacity estimation

Despite substantial efforts invested in the development of precise and resilient battery capacity estimation methods by both industry and academia, their application to real-world usage conditions remains a formidable challenge. This challenge is particularly pronounced in our scenario, where access to laboratory aging cycling data is unavailable, and the only available information consists of time-series measurements obtained from onboard vehicle systems.

Consequently, the methods available for probing the battery’s available capacity are inherently constrained.

Compared to the highly dynamic discharge data, charging is more controlled, usually with engineer-predefined charging current and SoC levels. For this reason, charging log files are used to estimate the battery’s available capacity. Here, we adopt an Ampere hour counting method coupled with an OCV-adjusted SoC method. Mathematically, the capacity C can be calculated as

$$C = \frac{\int_0^t I(t)dt}{SoC(t) - SoC(0)}, \quad (\text{E.5})$$

where $SoC(t)$ and $SoC(0)$ denote the SoC at the end and beginning of charging, respectively. These values are estimated based on the SoC-OCV look-up table. This map, illustrated in the Supplementary materials, was derived from laboratory testing of the same battery cell under conditions of $25^\circ C$, discharging at a rate of $1/20C$ from $4.2V$ to $2.6V$. It is acknowledged that both the operating temperature and battery aging can alter the OCV curve. However, we assume that the OCV remains unchanged due to resource limitations. Fortunately, there is some consolation in the fact that the OCV-SoC curve undergoes only marginal changes over the course of battery aging, as suggested by experimental results for NMC-graphite automotive-grade cells [46]. To further mitigate potential effects, only battery average temperatures within the range of $10^\circ C$ to $40^\circ C$ during charging are considered in this study, thereby limiting the impact of temperature on the reshaping changes of the OCV curve. Furthermore, to prevent a small number in the denominator of (E.5) from disproportionately inflating the overall capacity estimation error, we restrict the inclusion of charging data in the study to instances where $SoC(0)$ is less than 0.25 and $SoC(t)$ is greater than 0.9 .

5.5 Estimating battery internal parameter aging dependency using NN

Individual vehicle-specific estimation

To systematically evaluate the degradation of battery resistance and capacity for a standardized reference point, we propose constructing a model capable of capturing real-time changes in these parameters. This involves formulating a

supervised learning problem, with the influencing factors as the model inputs and real-time estimated resistance or available capacity as the output targets. For this purpose, we opt for a fully connected layer NN. Such a choice is motivated by its exceptional ability to handle non-linear and complex relationships between the input and output variables. Despite maintaining reasonable computational requirements compared to more intricate deep NN architectures, it proves effective in capturing the nuanced dependencies present in the data. Moreover, the NN's notable generalizability is a crucial asset for our task. Our objective extends beyond achieving high accuracy; we aim to extrapolate well on unseen data. In essence, the NN should learn the functional relationship between the target variables (resistance and capacity) and the influencing factors, showcasing adaptability to varying conditions and scenarios. This ensures the model's reliability in predicting degradation trends and performance even when confronted with previously unobserved data.

For battery resistance, mathematically, we build a model with the following format

$$R_0(k) = f_R(I(k), U(k), SoC(k), T_{\text{avg}}(k), T_{\text{amb}}(k), d(k)), \quad (\text{E.6})$$

where $T_{\text{avg}}(k)$ is the pack average temperature, $T_{\text{amb}}(k)$ is the ambient temperature at time instant k , $d(k)$ represents the vehicle accumulated driving distance at time k . The output $R_0(k)$ is calculated using the method described in Section.5.3. In the same way, the capacity of the battery is also trained using a NN with the following format

$$C(n) = f_C(I_{\text{avg}}(n), T_{\text{avgc}}(n), T_{\text{ambc}}(n), d(n)), \quad (\text{E.7})$$

where $I_{\text{avg}}(n)$ is the average current during the n th charging cycle, $T_{\text{avgc}}(n)$ and $T_{\text{ambc}}(n)$ are the average pack temperature and average ambient temperature during the n th charging cycle. The output $C(n)$ is calculated using the method introduced in Section 5.4. Since only one capacity $C(n)$ value can be estimated for each charging profile we use the input average of the complete charging profile instead of point measurements as in the resistance case.

Following the training of the NN model, our initial assessment involves utilizing previously unseen test data to evaluate the model's accuracy. Our premise is based on the notion that if the model's estimation accuracy surpasses a predefined threshold, the inferred functional relationship between

aging indicators and influencing factors can be deemed reliable. Nevertheless, it is essential to acknowledge the challenge of conclusively determining the minimum accuracy level that ensures sufficient accuracy and reliability of the learned relationships. To address this uncertainty, we employ sensitivity and robustness analysis through Monte Carlo simulation, providing a partial validation of the acquired insights.

In fact, the learned NN can potentially be evaluated for arbitrary operating conditions simply by inputting specific values and allowing the NN to compute the corresponding outputs. However, it is essential to acknowledge that, akin to other data-driven methods, the model's accuracy is most pronounced in regions where the training data are abundant. Furthermore, such operating points are also of interest as they better reflect real-world usage scenarios. Hence, we evaluate each individual vehicle by considering the mean values across its entire dataset. The calculations for the reference resistance ($R_0(m)$) and capacity ($C(m)$) are then expressed as follows:

$$R_0(m) = f_R(\bar{I}, \bar{U}, \bar{SoC}, \bar{T}_{avg}, \bar{T}_{amb}, d(m)), \quad (\text{E.8})$$

$$C(m) = f_C(\bar{I}_{avg}, \bar{T}_{avgc}, \bar{T}_{ambc}, d(m)), \quad (\text{E.9})$$

where \bar{I} , \bar{U} , \bar{SoC} , \bar{T}_{avg} , \bar{T}_{amb} , \bar{I}_{avg} , \bar{T}_{avgc} and \bar{T}_{ambc} are the mean value of all feature inputs, m denotes the number of evaluation points, and $d(m)$ is the represented accumulated vehicle driving distance, e.g., 1000, 2000, 3000, \dots , 150000 km.

Fleet level estimation

Continuing with the methodology outlined in the preceding section, each individual vehicle will have its own NN model and is assessed at distinctive operating points specific to that vehicle. In contrast to training individual models, a fleet-level average model can also be acquired by combining all relevant data into a centralized database. Subsequently, an NN is trained using this comprehensive dataset, allowing for evaluation at average operating points representative of the entire fleet. To gauge the aging behavior of the entire fleet, the model is then assessed at population-level average operating points, utilizing the mean values derived from the centralized database as feature inputs. This fleet-average aging behavior serves as a benchmark, facilitating a comparative analysis of the degradation performance across vehicles within

the fleet. It is imperative to note that for a fair and consistent comparison, individual models must be evaluated using population-level average operating points rather than their respective individual averages. This ensures an equitable assessment of the vehicles' degradation performance against the fleet benchmark.

Acknowledgements

This work was supported by the Swedish Energy Agency within the Vehicle Strategic Research and Innovation Program (Grant No. P2023-00611). The computations were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) at Chalmers, partially funded by the Swedish Research Council through grant agreement no. 2022-06725.

References

- [1] M. Meinshausen, J. Lewis, C. McGlade, *et al.*, “Realization of paris agreement pledges may limit warming just below 2 C,” *Nature*, vol. 604, no. 7905, pp. 304–309, 2022.
- [2] X. Chen, H. Zhang, Z. Xu, C. P. Nielsen, M. B. McElroy, and J. Lv, “Impacts of fleet types and charging modes for electric vehicles on emissions under different penetrations of wind power,” *Nat. Energy*, vol. 3, no. 5, pp. 413–421, 2018.
- [3] B. Nykvist and M. Nilsson, “Rapidly falling costs of battery packs for electric vehicles,” *Nat. Clim. Change*, vol. 5, no. 4, pp. 329–332, 2015.
- [4] M. Chen, X. Ma, B. Chen, *et al.*, “Recycling end-of-life electric vehicle lithium-ion batteries,” *Joule*, vol. 3, no. 11, pp. 2622–2646, 2019.
- [5] E. A. Olivetti, G. Ceder, G. G. Gaustad, and X. Fu, “Lithium-ion battery supply chain considerations: Analysis of potential bottlenecks in critical metals,” *Joule*, vol. 1, no. 2, pp. 229–243, 2017.
- [6] D. L. Wood, J. Li, and S. J. An, “Formation challenges of lithium-ion battery manufacturing,” *Joule*, vol. 3, no. 12, pp. 2884–2888, 2019.

- [7] M. Woody, M. Arbabzadeh, G. M. Lewis, G. A. Keoleian, and A. Stefanopoulou, “Strategies to limit degradation and maximize Li-ion battery service lifetime-critical review and guidance for stakeholders,” *J. Energy Storage*, vol. 28, p. 101 231, 2020.
- [8] J. S. Edge, S. O’Kane, R. Prosser, *et al.*, “Lithium ion battery degradation: What you need to know,” *Phys. Chem. Chem. Phys.*, vol. 23, no. 14, pp. 8200–8221, 2021.
- [9] J. Vetter, P. Novák, M. R. Wagner, *et al.*, “Ageing mechanisms in lithium-ion batteries,” *J. Power Sources*, vol. 147, no. 1-2, pp. 269–281, 2005.
- [10] M. Broussely, P. Biensan, F. Bonhomme, *et al.*, “Main aging mechanisms in Li ion batteries,” *J. Power Sources*, vol. 146, no. 1-2, pp. 90–96, 2005.
- [11] V. Sulzer, P. Mohtat, A. Aitio, *et al.*, “The challenge and opportunity of battery lifetime prediction from field data,” *Joule*, vol. 5, no. 8, pp. 1934–1955, 2021.
- [12] G. Pozzato, A. Allam, L. Pulvirenti, G. A. Negoita, W. A. Paxton, and S. Onori, “Analysis and key findings from real-world electric vehicle field data,” *Joule*, vol. 7, no. 9, pp. 2035–2053, 2023.
- [13] F. von Bülow and T. Meisen, “A review on methods for state of health forecasting of lithium-ion batteries applicable in real-world operational conditions,” *J. Energy Storage*, vol. 57, p. 105 978, 2023.
- [14] X. Hu, Y. Che, X. Lin, and S. Onori, “Battery health prediction using fusion-based feature selection and machine learning,” *IEEE Trans. Transp. Electrification.*, vol. 7, no. 2, pp. 382–398, 2020.
- [15] 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, no. 2, pp. 338–371, 2023.
- [16] J. Wang, P. Liu, J. Hicks-Garner, *et al.*, “Cycle-life model for graphite-LiFePO₄ cells,” *J. Power Sources*, vol. 196, no. 8, pp. 3942–3948, 2011.
- [17] L. Lam and P. Bauer, “Practical capacity fading model for Li-ion battery cells in electric vehicles,” *IEEE Trans. Power Electron.*, vol. 28, no. 12, pp. 5910–5918, 2012.

-
- [18] W. He, N. Williard, M. Osterman, and M. Pecht, “Prognostics of lithium-ion batteries based on Dempster-Shafer theory and the Bayesian Monte Carlo method,” *J. Power Sources*, vol. 196, no. 23, pp. 10 314–10 321, 2011.
- [19] Y. Zou, X. Hu, H. Ma, and S. E. Li, “Combined state of charge and state of health estimation over lithium-ion battery cell cycle lifespan for electric vehicles,” *J. Power Sources*, vol. 273, pp. 793–803, 2015.
- [20] P. Shen, M. Ouyang, L. Lu, J. Li, and X. Feng, “The co-estimation of state of charge, state of health, and state of function for lithium-ion batteries in electric vehicles,” *IEEE Trans. Veh. Technol.*, vol. 67, no. 1, pp. 92–103, 2017.
- [21] X. Hu, H. Yuan, C. Zou, Z. Li, and L. Zhang, “Co-estimation of state of charge and state of health for lithium-ion batteries based on fractional-order calculus,” *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 10 319–10 329, 2018.
- [22] J. Christensen and J. Newman, “A mathematical model for the lithium-ion negative electrode solid electrolyte interphase,” *J. Electrochem. Soc.*, vol. 151, no. 11, A1977, 2004.
- [23] L. Wildfeuer, N. Wassiliadis, C. Reiter, M. Baumann, and M. Lienkamp, “Experimental characterization of Li-ion battery resistance at the cell, module and pack level,” in *2019 Fourteenth International Conference on Ecological Vehicles and Renewable Energies (EVER)*, 2019, pp. 1–12.
- [24] X.-G. Yang, Y. Leng, G. Zhang, S. Ge, and C.-Y. Wang, “Modeling of lithium plating induced aging of lithium-ion batteries: Transition from linear to nonlinear aging,” *J. Power Sources*, vol. 360, pp. 28–40, 2017.
- [25] M. Dubarry, D. Howey, and B. Wu, “Enabling battery digital twins at the industrial scale,” *Joule*, vol. 7, no. 6, pp. 1134–1144, 2023.
- [26] M.-F. Ng, J. Zhao, Q. Yan, G. J. Conduit, and Z. W. Seh, “Predicting the state of charge and health of batteries using data-driven machine learning,” *Nat. Mach. Intell.*, vol. 2, no. 3, pp. 161–170, 2020.

- [27] H. Rauf, M. Khalid, and N. Arshad, "Machine learning in state of health and remaining useful life estimation: Theoretical and technological development in battery degradation modelling," *Renew. Sustain. Energy Rev.*, vol. 156, p. 111 903, 2022.
- [28] Y. Li, C. Zou, M. Bercibar, *et al.*, "Random forest regression for online capacity estimation of lithium-ion batteries," *Appl. Energy*, vol. 232, pp. 197–210, 2018.
- [29] R. R. Richardson, M. A. Osborne, and D. A. Howey, "Gaussian process regression for forecasting battery state of health," *J. Power Sources*, vol. 357, pp. 209–219, 2017.
- [30] 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," *J. power sources*, vol. 482, p. 228 863, 2021.
- [31] A. Nuhic, T. Terzimehic, T. Soczka-Guth, M. Buchholz, and K. Dietmayer, "Health diagnosis and remaining useful life prognostics of lithium-ion batteries using data-driven methods," *J. Power Sources*, vol. 239, pp. 680–688, 2013.
- [32] Y. Zhang, T. Wik, J. Bergström, M. Pecht, and C. Zou, "A machine learning-based framework for online prediction of battery ageing trajectory and lifetime using histogram data," *J. Power Sources*, vol. 526, p. 231 110, 2022.
- [33] Z. Deng, L. Xu, H. Liu, X. Hu, Z. Duan, and Y. Xu, "Prognostics of battery capacity based on charging data and data-driven methods for on-road vehicles," *Appl. Energy*, vol. 339, p. 120 954, 2023.
- [34] Q. Huo, Z. Ma, X. Zhao, T. Zhang, and Y. Zhang, "Bayesian network based state-of-health estimation for battery on electric vehicle application and its validation through real-world data," *IEEE Access*, vol. 9, pp. 11 328–11 341, 2021.
- [35] C. She, Z. Wang, F. Sun, P. Liu, and L. Zhang, "Battery aging assessment for real-world electric buses based on incremental capacity analysis and radial basis function neural network," *IEEE Trans. Ind. Inform.*, vol. 16, no. 5, pp. 3345–3354, 2019.

-
- [36] L. Zhou, Y. Zhao, D. Li, and Z. Wang, “State-of-health estimation for lifepo 4 battery system on real-world electric vehicles considering aging stage,” *IEEE Trans Transp Electrifi.*, vol. 8, no. 2, pp. 1724–1733, 2021.
- [37] A. Aitio and D. A. Howey, “Predicting battery end of life from solar off-grid system field data using machine learning,” *Joule*, vol. 5, no. 12, pp. 3204–3220, 2021.
- [38] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [39] P. S. Sabet, G. Stahl, and D. U. Sauer, “Non-invasive investigation of predominant processes in the impedance spectra of high energy lithium-ion batteries with nickel-cobalt-aluminum cathodes,” *J. Power Sources*, vol. 406, pp. 185–193, 2018.
- [40] W. Waag, C. Fleischer, and D. U. Sauer, “On-line estimation of lithium-ion battery impedance parameters using a novel varied-parameters approach,” *J. Power Sources*, vol. 237, pp. 260–269, 2013.
- [41] W. Waag, S. Käbitz, and D. U. Sauer, “Experimental investigation of the lithium-ion battery impedance characteristic at various conditions and aging states and its influence on the application,” *Appl. Energy*, vol. 102, pp. 885–897, 2013.
- [42] J. Stadler, C. Krupp, M. Ecker, J. Bandlow, B. Spier, and A. Latz, “Investigation and modeling of cyclic aging using a design of experiment with automotive grade lithium-ion cells,” *J. Power Sources*, vol. 521, p. 230952, 2022.
- [43] A. Mikheenkova, A. J. Smith, K. B. Frenander, *et al.*, “Ageing of high energy density automotive Li-ion batteries: The effect of temperature and state-of-charge,” 2023.
- [44] L. Wildfeuer, A. Karger, D. Aygül, N. Wassiliadis, A. Jossen, and M. Lienkamp, “Experimental degradation study of a commercial lithium-ion battery,” *J. Power Sources*, vol. 560, p. 232498, 2023.
- [45] B. Fridholm, T. Wik, and M. Nilsson, “Robust recursive impedance estimation for automotive lithium-ion batteries,” *J. Power Sources*, vol. 304, pp. 33–41, 2016.

- [46] A. Klintberg, E. Klintberg, B. Fridholm, H. Kuusisto, and T. Wik, “Statistical modeling of OCV-curves for aged battery cells,” *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 2164–2168, 2017.