Technische Universität München
TUM School of Engineering and Design

Data-driven battery state estimation

Jacob Clay Hamar, M.Sc.

Vollständiger Abdruck der von der TUM School of Engineering and Design der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Ingenieurswissenschaften (Dr.-Ing.)

genehmigten Dissertation.

Vorsitz:		Prof. Dr.-Ing. Markus Lienkamp
Prüfende der Dissertation:	1.	Prof. Dr.-Ing. Andreas Jossen
	2.	Prof. Dr.-Ing. Fridolin Röder

Die Dissertation wurde am 23.05.2024 bei der Technischen Universität München eingereicht und durch die TUM School of Engineering and Design am 01.10.2024 angenommen.

Foreword

This dissertation was made possible by the support of the Techinische Universität München (TUM) and the Bayrische Motoren Werke AG (BMW) under the supervision of Prof. Dr.-Ing. Andreas Jossen and Dr.-Ing Simon Erhard. Financial support was provided by the BMW AG, with academic and laboratory support provided by the TUM.

A special thanks is owed to Prof. Dr.-Ing. Andreas Jossen for the opportunity to pursue this collaborative thesis, for the intensive discussions and guidance and for the great trust to undertake this research effort with the Chair of Electrical Energy Storage Technology.

This thesis afforded the chance to forge many new friendships, with those that helped academically and emotionally, providing a strong foundation for research and in the end making this entire work possible. To Chris Zoerr, Christian Raubitschek, Thomas Kellner, Sebastian Ludwig, Julius Schmitt, Johannes Sturm, Tobias Hofmann, Ludwig Kraft and Marcel Rogge, it has been a pleasure working with each of you and for all of your critical feedback, I am very grateful.

Finally, to my family, who supported me even from across the ocean, and to my wife, the most heartfelt of thanks. Your belief and unrelenting love helped provide the fortitude and strength to follow my ambition, and complete this dissertation.

Jacob Clay Hamar
München, 2023

Abstract

The explosive growth in the electric vehicle market over the last decade has provided both the challenge of encouraging customer acceptance by ensuring safe and robust functionality, as well as, the possibility to develop innovative functions using the massive amounts of data generated by the expanding fleet of electric vehicles. In order to address the issues of safety and reliability, understanding the state of the battery is essential. State estimation of batteries is a broad field of research which will be analyzed primarily through the lens of battery aging. Aiding this research is machine learning– a growing field offering methods for extracting insights into the complex electrochemical system of batteries using large data-sets. In this work battery aging is contextualized using real-world data from battery electric vehicles, and methods for improving state estimation using this data are proposed.

The main body of work is comprised of three papers, each exploring battery state estimation with a focus on battery aging. The first study addresses the consequence of a changepoint. More specifically, it seeks to understand how the alternation from one operation condition to another affects the total aging state, known as path dependent aging. For this, a robust aging study was conducted including six complimentary aging conditions. By defining alternating paths, the aging behavior under various cycling frequencies, temperature and C-rate conditions was explored. The result of this study gives evidence that path dependent aging behavior does exist and should not be dismissed or assumed negligible. Including path dependence in existing models would add to an already complex task, which is where the advantage of machine learning can provide support.

Two papers explore the applicability of various machine learning models to support the complex task of battery state estimation using real-world and electric-vehicle inspired data. The first approach integrates data-driven and conventional battery state estimation methods for state-of-health estimation yielding a semi-empirical model with k-means clustering, an unsupervised learning technique for clustering larger data sets into sub-categories, and a neural network based model that is accurate within 3.4 % under real-world conditions. The final paper showcases the applicability of more modern regression techniques, for predicting non-measurable internal states of the battery, like the anode potential. Also discussed in the paper is the trade-off between the computational resource demand of the data-driven approach and accuracy, with more traditional regression algorithms requiring only a few kilobytes of storage yielding an error of 13 mV, to the previously mentioned random forest requiring over 100 Mb but offering an error of only 2.6 mV.

The broad toolkit within the machine learning space, when coupled with a high quality data set, demonstrates advantages to other methods such as the equivalent circuit model, which can oversimplify complex behavior, or the pseudo two-dimensional model, which is too computationally complex for many applications, even when not considering path-dependent aging. That being said, these methods are best used in tandem and should support each other to provide the most suitable state-estimation approach given the specifics of the task, whether it be online in the battery management system, in the cloud, or somewhere in between.

Contents

List of Publications

Peer-reviewed journal paper contributions (lead author)

a Hamar, J. C.; Stuckenberger, M.; Sturm, J.; Schmitt, J.; Rogge, M.; Erhard, S. V.; Jossen, A.: *Investigating the Path Dependent Aging Behavior of Nickel Cobalt Aluminium Oxide Cathode Batteries During High C-Rate Cycling*, in: *Journal of the Electrochemical Society 2 (171)*, doi.org/10.1149/1945-7111/ad2952, 2024

b Hamar, J. C.; Erhard, S. V.; Zoerr, C.; Jossen, A.: *State-of-health estimation using a neural network trained on vehicle data*, in: *Journal of Power Sources (512)*, doi.org/10.1016/j.jpowsour.2021.230493, 2021

c Hamar, J. C.; Erhard, S. V.; Zoerr, C.; Jossen, A.: *Anode Potential Estimation in Lithium-Ion Batteries Using Data-Driven Models for Online Applications*, in: *Journal of the Electrochemical Society 3 (168)*, doi.org/10.1149/1945-7111/abe721, 2021

Self-written sections of peer-reviewed lead author journal paper contributions are partially contained in this doctoral thesis without further reference in the text.

Selection of conference contributions

a Hamar, J. C.: *Neural Network State-of-Health Estimation from Vehicle Data*, in: *PRiME from the Journal of the Electrochemical Society*, Hawaii, Poster, 2020

Peer-reviewed journal paper contributions (co-author)

a Hofmann, T.; Hamar, J.; Rogge, M.; Zoerr, C.; Erhard, S.; Schmitt, J. P.: *Phyics-Informed Neural Networks for State of Health Estimation in Lithium-Ion Batteries*, in: *Journal of the Electrochemical Society (170)* doi.org/10.1149/1945-7111/acf0ef, 2023

b Hofmann, T.; Li, Jiahao.; Hamar, J.; Erhard, S.; Schmitt, J. P.: *The ΔQ-method: State of health and degradation mode estimation for lithium-ion batteries using a mechanistic model with relaxed voltage points*, in: *Journal of Power Sources (596)*, doi.org/10.1016/j.jpowsour.2024.234107, 2024

Abbreviations

Please note that the list below is based on the main part of this thesis and does not fully cover the abbreviations used in the papers. Each paper itself includes an individual list.

AI . . . . . artificial intelligence

BEV . . . . . battery electric vehicle
BMS . . . . . battery management system
BOL . . . . . begin-of-life

CL . . . . . conductivity loss

DOD . . . . . depth-of-discharge
DVA . . . . . differential voltage analysis

ECM . . . . . equivalent circuit model
EIS . . . . . electrochemical impedance spectroscopy
EU . . . . . European Union
EV . . . . . electric vehicle

IoT . . . . . Internet-of-Things
IQR . . . . . inner quartile range

LAM . . . . . loss of active material
LCO . . . . . lithium cobalt oxide
LFP . . . . . lithium iron phosphate
LLI . . . . . loss of lithium inventory

MAE . . . . . mean absolute error
MAPE . . . . . mean absolute percent error

NCA . . . . . nickel cobalt aluminum
NE . . . . . negative electrode
NMC . . . . . nickel manganese cobalt oxide

OCV . . . . . open-circuit-voltage
OEM . . . . . original equipment manufacturer

P2D . . . . . pseudo two-dimensional
PCA . . . . . principle component analysis

PE . . . . . positive electrode

RMSE . . . . . root mean square error

SEI . . . . . solid electrolyte interphase

Si . . . . . silicon

SiO_x . . . . . silicon oxide

SOC . . . . . state-of-charge

SOH . . . . . state-of-health

SPI . . . . . solid permeable interface

1 Introduction

More than 10 million electric vehicles (EVs) drove on the world's roads in 2020 with the purchase of full battery electric vehicles (BEVs) driving the growth of the electrified fleet [1]. The mobility sector has invested heavily in battery storage technology to facilitate what is known as the 'energy transition' [2], leading to an increased demand in EVs. The transition from combustion drive-trains to electric drive-trains is being propelled by consumer preference, industry production and government support.

Year after year, consumer spending on EVs continues to rise, even as government support stabilizes [3]. One of the main benefits, motivated by consumer demand, is that the increase in EVs has led to a reduction in the global green house gas emission [4]. It is estimated that more than one third of consumers are willing to pay more for sustainability, and 85 % of consumers have become more environmentally conscious [5]. The demand for more sustainable products is sector agnostic, with consumers demanding eco-friendly packaging, energy efficient lighting, and locally-sourced produce, as well as, EVs in order to help combat the looming climate crisis. This new, sustainability-minded, consumer is helping absorb the costs of transitioning to greener transportation as they are willing to pay for the more expensive technology, helping support manufacturers' decision to invest in innovation.

Auto manufacturers are also playing an essential role in the global switch to EVs. Eighteen of the 20 largest original equipment manufacturers (OEMs) have committed to increasing the offer and sale of EVs [1], with some of the most ambitious pledges coming from Ford and Volvo who are committing to 100 % EV sales from 2026 and 2030, respectively. It is projected that the global automotive market share of EVs will reach 43 % by 2030 [6]. EVs alone are responsible for a $50\,\mathrm{MtCO_2}$ reduction in 2020 and their positive contribution is projected to increase up to $410\,\mathrm{MtCO_2}$ in 2030 [4]. Although strengthening consumer demand is helping offset the investment in EV technology, many OEMs have committed to investing in the complete overhaul in their supply-chain and production facilities despite the uncertainty of whether EVs will be the profitable mobility solution in the future.

Governments, for their part, are also helping create policies and incentives to mitigate the risks for the technology transition and help prop-up consumer demand in green solutions. More than 20 countries have electrification targets or internal combustion engine bans for cars, and eight countries plus the European Union (EU) have announced net-zero carbon emission pledges with target dates as early as 2030 [3]. Policy initiatives such as the EU Green New Deal, extension of the New Electric Vehicle subsidy program in China and even the 2020 Covid Stimulus in the US are in large part responsible for the 40 % year over year growth of EV sales in recent years, despite extraordinary economic head-winds [3]. Europe and China continue to push the most progressive policy agendas for EV markets [3]. Europe is leading the way with an EV market share of 34.6 % in Q1/2021, China follows with 11 % and the USA with 7.3 % [7].

The key enabler to this green, sustainable, emission-saving technology is the battery. In early 2021, the European Commission announced a 2.9 billion Euro fund to support battery manufacturing and research [8]. That is on top of an estimated 60 billion Euro that European governments and automakers

invested in EVs and batteries. This significant investment has already been directly attributed to more stringent CO_2 emissions targets [9]. As consumers, producers and governments all align on decarbonizing the mobility sector, enormous pressure is being placed on the electric battery system propelling these vehicles. The battery alone contributes to 45 % of the total vehicle costs [10] and has a significant influence on consumer acceptance by determining the range and lifetime of the vehicle. As such, one of the most effective ways to further support the diffusion of EVs into the market is to maximize the usable battery energy, which can only be achieved in combination with innovative production technologies and highly accurate battery state estimation methods helping to optimize the available energy over the vehicle lifetime.

A by-product of the transition within the automotive sector is the rapid expansion of battery behavior data. The advent of Industry 4.0, or Internet-of-Things (IoT), is not unique to the electric mobility sector, but the generation of petabytes of data-stores has led to a race within the automotive industry to capitalize on data-driven methods to generate a marketable benefit to consumers. According to a Deloitte Research forecast, the fledgling 1 billion USD automotive artificial intelligence (AI) market is projected to grow to 27 billion by 2025 [11] as global OEMs invest in digitalization to stay competitive.

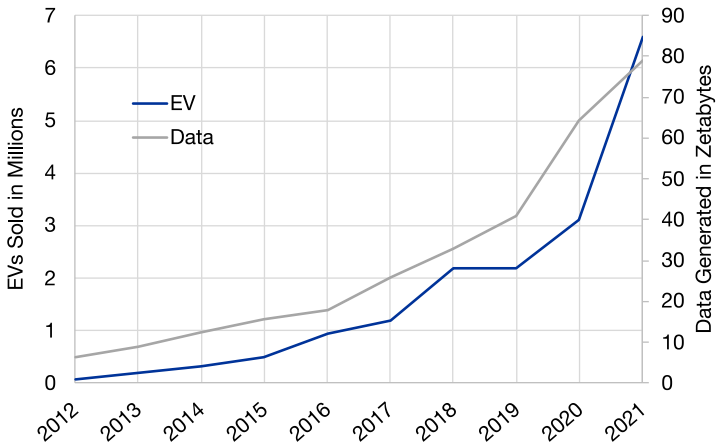

Figure 1.1: Exponential increase of global data generation correlates with the increase in EV sales [11].

The growth of data-driven methods has paralleled the growth of EVs over the last two decades as seen in Fig. 1.1. According to Google's Ngram Viewer the terms 'EV' and 'machine learning' have increased two an astounding 17-fold from 1989 to 2019 [12]. At the same time, the ubiety of 'machine learning' in the title of articles from some of the leading energy research publications has increased from less than half a percent in 2005 to ca. 5 % in 2021 in both the Journal of Power Sources, as well as, the Journal of Nature [13; 14]. The synergies between these two trends are obvious: the large amounts of battery data being produced can be combined with innovative machine learning methods to help improve the understanding of battery behavior.

The explosion of connected systems, sensor networks and simulation techniques has overwhelmed the science community with data-sets [15] and led to breakthroughs across industries and disciplines, from material science to pharmaceuticals [16]. The automotive industry is also aware of the opportunity offered by the large pool of data generated by the over 1 billion vehicles currently on the road, which is estimated to have a monetization value between 450-750 billion USD [10].

Breakthroughs resulting from tapping into the vast data resource have only just started coming to

light. In June 2020, Mark Minevich from Forbes wrote, "The automotive industry is a perfect example of an area that has not yet been perceived as a data source for human good by the mainstream school of thought" [17], however, the digitalization of the automotive industry provides not only the monetary carrot for OEMs to chase, but also the impetus to deliver a direct customer benefit in terms of safety, user experience, and vehicle reliability [18–20].

In order to capitalize on the opportunity digitalization in electric mobility presents, one has to match the benefits of machine-learning to the current problems in battery system development. One area currently growing in interest is the field of battery state estimation. Battery state estimation typically refers to the internal descriptor variables of state-of-health (SOH) and state-of-charge (SOC), though a more broad definition would include all internal parameters required to accurately capture the current and future behavior of a battery system. This thesis focuses mainly on data-driven methods for SOH estimation, as well as, relevant internal parameters, such as the anode potential.

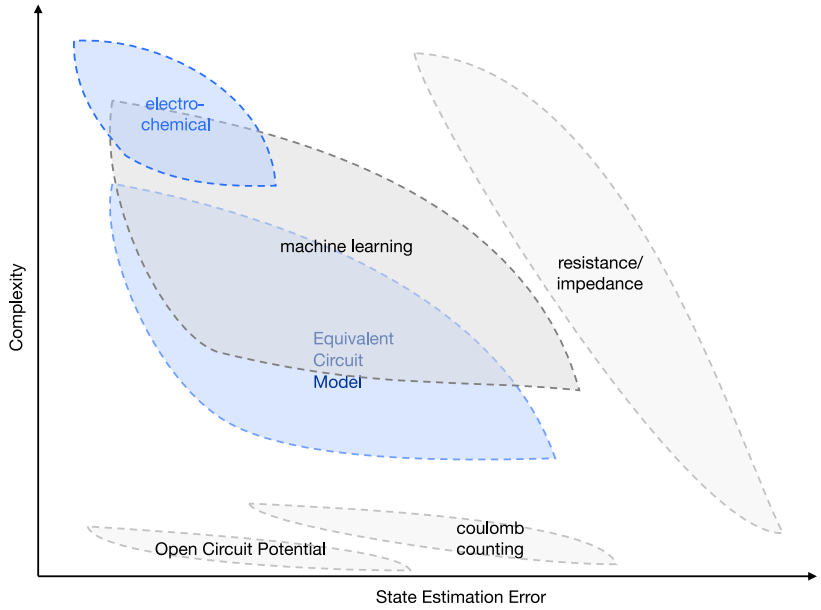

Figure 1.2: The trade-off between complexity and error of common state estimation methods [21].

Battery aging, captured by the SOH metric, is a complex phenomenon involving the coupling of chemical, mechanical and electrical processes. In order to better understand aging within lithium-ion batteries much research has been devoted to developing models and simulation tools capable of describing the multi-physics coupling at varying dimensions and time-scales. The major trade-off between methods is often that of model fidelity and computational efficiency. Figure 1.2 highlights this trade-off of complexity and state estimation method accuracy. Still, even the most advanced methods are not able to fully capture the dependence on more nuanced phenomena in battery aging, such as the occurrence of lithium plating or the coupling between aging mechanisms leading to a path-dependent aging behavior. By incorporating AI and data-driven methods, an improvement in the trade-off can be found where a simplified model can substitute a traditionally highly computationally intensive model, improving model detail and accuracy at the same time.

This thesis focuses on leveraging machine learning for improving battery state estimation. The methods and functions developed in this thesis provide insights into how OEMs can leverage the Big Data generated by the expanding EV fleet to improve battery systems. More accurate battery SOH estima-

tion, and internal state estimation, lead to an improvement in the quality, safety and usability of the EV, which directly affects consumer's willingness to adopt the technology, and improves the financial viability of BEVs for both OEMs and governments.

1.1 State Estimation

The term state estimation broadly refers to methods describing the electrochemical condition of a battery by interpreting and eliminating inaccuracies and errors from measurable data. The states of a battery are many, and can include the current health, charging condition, temperature, and any other measurable, or model-able features. This section provides a foundation for the modeling of a battery and describes the various state equations common in battery science.

1.1.1 Lithium-ion Batteries

Lithium-ion batteries are currently the dominating technology in the automotive power-train transition due to their high energy density, simplicity of mechanical integration and high power capability [22]. All references in this thesis are referring to this general class of battery. The functional principles and components of a lithium-ion battery are listed below and can be seen in the representation provided in Fig. 1.3. Structurally, the lithium-ion battery is composed of two porous electrodes, a separator,

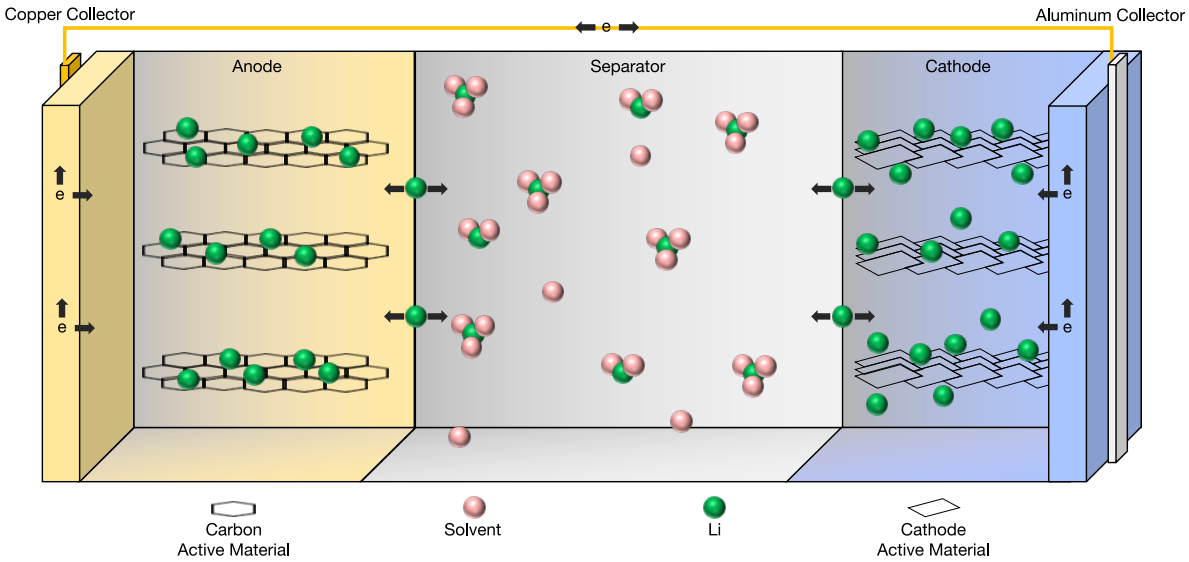

Figure 1.3: The main components of a lithium-ion battery are the anode and cathode electrodes, separator, electrolyte and the current collectors [23; 24].

a liquid medium called the electrolyte and current collectors transferring the electrons out of the electrodes.

The anode, or negative electrode, is typically constructed from a carbon-based graphite material due to carbon's low cost, high thermal stability, and most importantly the low potential against metallic lithium [25; 26]. The graphite structure, LiC_5 [27], offers high flexibility between the graphene layers allowing for easier lithium-ion insertion. Silicon (Si) or silicon oxide (SiO_x) is becoming increasingly common in the anode as it improves energy capacity but at the cost of higher mechanical strain, thus

use as a doping agent is typically limited to compositions of around 1-8 wt. % [28; 29].

The crystalline structure of the electrodes is important for the electron transfer into the non-occupied vacancies and the intercalation of the lithium-ions. The de-/intercalation of the lithium-ions is a result of oxidation or reduction processes depending on the charging or discharging state of the battery [30]. The electron transfer resulting from the reduction and oxidation process occurs between electrodes via the copper current collector at the anode and the (typically) aluminum collector at the cathode, which are electrically connected through a power sink or source.

The aluminum current collector foil of the cathode is adjacent to a substrate of active materials, such as, lithium cobalt oxide (LCO), nickel cobalt aluminum (NCA), nickel manganese cobalt oxide (NMC) and lithium iron phosphate (LFP), to name a few of the most common active materials [31]. The decision for which composition is most suitable results from an evaluation of the energy density, cost and stability of each chemistry [32; 33]. The nickel-rich NCA and NMC offer state-of-the-art energy densities and are becoming a favorite for automotive applications [34]. A strong trend, however, is also seen moving away from cobalt as an active material and favoring ethically traceable, cheaper and more stable technologies, like LFP [35; 36].

Both electrodes are porous allowing for electrolyte– an organic solvent solution containing dissolved lithium salts and other additives– to distribute into the substrate. During the oxidation process (at the positive electrode (PE) when charging or negative electrode (NE) when discharging) lithium-ions are transferred through the electrolyte to the other electrode through the porous separator, which acts as the electronical insulator between the disparately charged electrodes. Closing the electrochemical process at the other end (reduction at the NE when charging or PE when discharging) the associated electrons are released to the sink or source via the current collectors and collected again at the respective opposite electrode. The general redox process follows

$$Li^+ + \Theta_\mathrm{s} + e^- \rightleftharpoons Li\Theta_\mathrm{s}, \tag{1.1}$$

where Θ_s represents the respective electrode host lattice site and e^- is the free electron [30].

1.1.2 State-of-Charge

The amount of cyclable lithium available for discharge during a single cycle is estimated by the SOC. The most common definition represents the SOC of a battery to be between 100 % (full) and 0 % (empty) with

$$SOC = \frac{C_\mathrm{act} - Q_\mathrm{b}}{C_\mathrm{act}}, \tag{1.2}$$

where C_act is the actual total capacity of the cell and Q_b is the charge balance, which is often estimated using coulomb counting as $Q_\mathrm{b} = \int_0 I dt$, where I is the current, and at 100 % SOC $Q_\mathrm{b} = 0$.

Direct integration of the current is the simplest method for implementing SOC estimation, however, it is dependent on an accurate current measurement, as well as, an accurate initialization of C_act [37]. Both of these inputs are subject to measurement error, where the current measurement is dependent on the sensor accuracy– which becomes particularly challenging as larger system architectures reduce measurement granularity and increase system noise– the C_act is estimated based on the voltage limits and SOH of the battery. The difficulties presented in estimating C_act are detailed in the following

sections.

In addition to directly integrating the current, there are also model-based approaches to estimating the SOC. Due to its relative simplicity, the equivalent circuit model (ECM) approach is commonly implemented in the battery management system (BMS). This approach– which is detailed below– requires a mapping from the open circuit voltage and an estimation of the over-potential using multiple impedance elements, leading to an increase in fitting difficulty and modeling error. Other methods, though not often implemented in the BMS, are electrochemcial and machine learning approaches. Both approaches are detailed in further sections, but generally have the disadvantage when compared to direct current integration in that they are more challenging to integrate into software and are computationally expensive.

1.1.3 Overpotential

The measurable voltage of a cell, U_{cell} is a combination of the stoichiometrically dependent open-circuit-voltage (OCV) and the use-dependent overpotential, η [30].

$$U_{\text{cell}} = OCV + \eta \tag{1.3}$$

The OCV is measurable at the cell terminals if no load is, or has been applied for a long period of time. This is also known as a relaxed state. The OCV is a combination of the two half cell potentials of the electrode materials versus Li/Li^+ [30].

$$U_{\text{OCV}} = \phi_{\text{cathode}} - \phi_{\text{anode}} \tag{1.4}$$

The OCV curve is constant within battery formats, and is characteristic for various battery chemistries (LCO, NCA, LFP, etc.). It is common for state estimation techniques like SOC estimation, to map the OCV to the SOC, as this is directly measurable.

The challenge, of course, is that the battery is often operated under loads, giving rise to an overpotential. The overpotential is typically attributed to an ohmic (η_Ω), charge transfer (η_{CT}) and diffusion potential (η_{diff}) [38]. Thus the total measured voltage during load can be re-written as

$$U_{\text{cell}} = U_{\text{OCV}} + \eta_\Omega + \eta_{\text{CT}} + \eta_{\text{diff}}. \tag{1.5}$$

Each overpotential component is dependent on the load and temperature during operation making the fitting and modeling of these polarization elements particularly challenging. Higher temperatures typically reduce overpotentials as conductivity (lower ohmic resistance), charge transfer and diffusion are all aided by improved kinetics [39]. Higher currents lead to higher overpotentials as charge transfer and diffusion processes are increasingly strained by the increased number of lithium-ions.

1.1.4 State-of-Health

The health of a battery over lifetime is a key performance metric for safety and cost considerations, however, the exact definition of battery health, like in many complex physical systems is multi-faceted. The main focus in this thesis is the SOH referring to the capacity retention of the battery. This capacity

is a measure of the available lithium-ions during cycling and is typically measured during a low and constant current charge or discharge process. These capacity check procedures allow for a tracking of the battery capacity fade over time. The ratio of actual remaining capacity and the capacity at the begin-of-life (BOL), $C_{\text{act}}/C_{\text{BOL}}$, is used to define the SOH,

$$SOH = \frac{C_{\text{act}}}{C_{\text{BOL}}}. \tag{1.6}$$

The most common measure of battery health is a measure of the available capacity. There are, however, other considerations for battery-electric systems, such as how much power can be delivered, temperature limits, and likelihood of a failure. For this reason there is a need to expand the definition of battery health to understand underlying aging mechanisms which adversely affect the internal resistance, physical stability and overall longevity of the battery.

Factors leading to a loss of capacity are generally classified into two operation categories: calendric and cyclic [40–43]. Calendric aging refers to non-operation conditions, when the battery is mostly affected by environmental conditions and long-term processes, whereas aging resulting from battery operation is categorized as cyclic aging. The reason for the separation into these two categories is due to the different time-scales and aging mechanisms which are dominant in one or the other operation mode. Aging mechanisms provide a more detailed lens to understand the aging process within the battery as it relates to both the operation mode, as well as, the principle components constituting the battery.

Aging mechanisms are detectable and measurable processes which work to degrade the health and performance of a battery. These mechanisms can often be associated with a particular component and root cause or operating condition. For example, aging mechanisms associated with the anode mainly occur at the electrode/electrolyte interface [44]. One of the most dominant processes is the growth of a passivation layer at this interface, most commonly referred to as the solid electrolyte interphase (SEI) [43; 45]. The formation of the SEI layer is a result of the decomposition of the electrolyte substance and the subsequent reaction of the precipitated molecules with the graphite substrate. Growth of the SEI is a function of time, SOC, temperature and electrolyte/electrode composition, with the fastest growth rate during the first few cycles of a new cell as the presence of an SEI layer typically inhibits further growth, though it never completely stops [44; 45]. A non-linear, precipitous capacity fade in a new battery is typically correlated with the early growth of the SEI layer, but over time, the SEI layer penetrates into the pores or cracks form in the existing SEI layer, causing additional growth and further consuming cyclable lithium [46].

Another aging mechanism, known as lithium plating, is of particular interest for the automotive industry due to its relevance for safety and fast charging ability. The most commonly attributed cause of lithium plating is when the anode potential drops below the standard potential of Li/Li^+ [30; 47; 48]. The result is the deposition of metallic lithium onto the anode surface in a dendrite structure. In addition to the high consumption of cyclable lithium, the dendrite growth poses the risk of bridging through the separator and creating an internal short circuit [49; 50].

The major aging mechanisms affecting the cathode are surface film formation, transition metal dissolution and mechanical degradation [43]. Similar to the growth of the SEI at the anode, a solid permeable interface (SPI) forms at the boundary of the cathode between electrode and electrolyte. The formation of this layer also consumes cyclable lithium, but is often a minor contributor to capacity fade [43].

There are also a number of mechanical stresses occurring inside the battery. The most frequent stress occurs during the de-/intercalation process as lithium-ions strain lattice structures. At the extremes

of the battery SOC, where the electrodes are maximally de-/lithiated the lattice structures experience the highest stresses, leading to cracks in the passivation layers (resulting in new SEI/SPI growth) and structural damage of the electrode lattice leading to a loss of active material [43].

The most dominant aging mechanisms are those listed above, however, a number of other mechanisms are present, including: current collector corrosion, binder decomposition and gas evolution [51]. The effects of each of these major and minor aging mechanisms have been grouped into three degradation modes: conductivity loss (CL), loss of lithium inventory (LLI) and loss of active material (LAM) [51; 52]. LLI aging mechanisms are electrolyte decomposition, lithium plating, SEI formation and formation of lithium-ion grains [51; 53]. CL aging mechanisms are current collector corrosion and binder decomposition [51; 53]. LLI mechanisms are oxidation of the electrolyte, electrode decomposition, intercalation gradient strains in the active particles and crystal structure disorder [51; 53].

1.2 Classic Aging Models

Factors leading to the loss of capacity (and thereby decreases in the SOH) include operation influences, as well as, the environmental conditions. It is the arduous task of aging models to replicate the capacity fade resulting from the interactions between simultaneous and compounding aging mechanisms using measurable inputs from the battery.

Aging models are limited primarily by the lack of available measurable inputs. One of the most widely used and detailed cell behavior models, the pseudo two-dimensional (P2D) electrochemical model, is highly dependent on internal cell parameters which require detailed half-cell measurements over a wide range of operation conditions, as well as, intrusive internal geometry measurements to operate at a single aging state [30; 54–57]. One of the biggest challenges to this method is mapping the influence of various aging mechanisms onto the myriad of fitting parameters.

The simplified cell model, ECM, is most commonly used in applications where computational resources are constrained, such as in the BMS. The general approach is to define a battery voltage response using electric circuit elements such as resistors and capacitors. Aging is integrated into this method by modifying the values of the circuit elements [58].

There exists no single equation, or set of equations, capable of accurately describing the complex reaction of a cell to the interaction of factors leading to capacity loss. For this reason, one of the most common aging models uses linear and non-linear fitting equations, most commonly referred to as semi-empirical models [59]. This method is much more straightforward to implement, requiring only easily measurable parameters, however, it is also this limited measurable parameter set which hinders the fitting of this method, reducing its accuracy on highly dynamic aging profiles. The semi-empirical methods are best suited for laboratory battery aging experiments.

1.2.1 Equivalent Circuit Models

The ECM represents cells with a voltage source, where the internal cell dynamics are modeled by a combination of ohmic, capacitive and inductive electric circuit elements [60; 61]. Figure 1.4 shows the elements of a basic ECM. The goal of these models is to use various combinations of circuit elements to model the cell response to various current perturbations, and to be able to accurately re-create the voltage response at a variety of aging states.

The first element of the circuit is a controlled voltage source. This element provides the OCV of the cell under no loads. The overpotential, or the voltage response resulting from ohmic resistance, electrochemical double layer, charge transfer and ionic diffusion are modeled using a combination of resistors and capacitors [62].

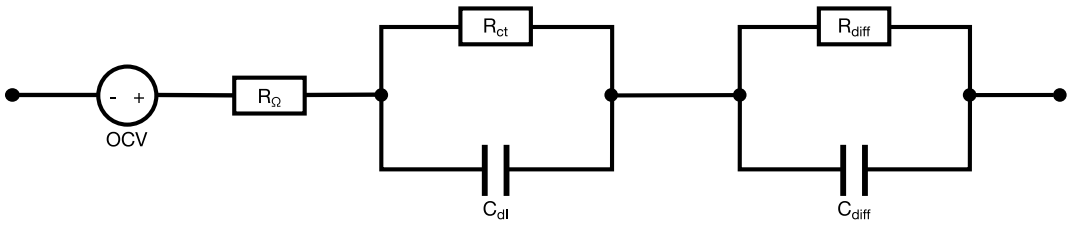

Figure 1.4: Example of an equivalent circuit model.

A series resistor, R_Ω represents the ohmic contribution to the resistance, which is dominant at time scales of $\Delta t \approx 0$. The dynamic elements consisting of a parallel connected capacitor and resistor are used to model processes with longer time constants. The charge transfer resistance, R_{ct}, describes the transition between electronic and ionic conductivity (modeled in the P2D-Model using the Butler-Volmer Equations) [62] and is combined with C_{dl}, which models the double layer effect occurring at phase boundaries between conductors of opposite polarity. As particles accumulate at the boundaries, an electrochemical potential builds up, resulting in a capacitive behavior similar to that seen in dielectric capacitors [63]. These two effects, modeled in parallel, also occur within a range of 10 to 100 milliseconds.

The second pair of circuit elements, R_{diff} and C_{diff}, capture the diffusion resistance in a cell. This is a result of multiple factors, such as the permeability/porosity of the materials, lattice structure and the intercalation process [62]. Diffusion is highly affected by the temperature, leading to a lower diffusion resistance with higher internal energy states at higher temperatures. This process therefore has a less-defined time constant, ranging from seconds to hours.

A common approach to fitting the various elements of the circuit is to use pulse fitting or electrochemical impedance spectroscopy (EIS) [64; 65]. This procedure involves applying an alternating voltage or current perturbation to a cell over a broad spectrum of frequencies, from mHz to kHz. The response results in a characteristic curve in the imaginary and real impedance domain, allowing for an interpretation of the resistances based on their time constants read from the frequencies [66]. Thus the fitting of the ECM is valid only for the conditions these parameters are measured at. EIS is also typically fitted using small excitation currents in order to have cell responses which are linear or pseudo linear [67]. It is also common to combine these methods, as well as, introduce machine learning to fit the RC elements using longer time-series, such as driving profiles. Each of the fitting approaches, however, require large laboratory studies to fit the ECM well enough to be applicable for use in an automotive context.

1.2.2 Electrochemical Models

There are a number of electrochemical models which have been proposed in literature [68], however, the most common are based on the P2D model developed by John Newman and his collaborators [69]. This method is based on the concentrated solution and porous electrode theory which models the reaction kinetics, migration and diffusion inside the lithium-ion cell [70]. The model is, most importantly,

applicable for a variety of lithium-ion cell types.

The term P2D is derived from a macroscopic x-dimension: with two porous electrodes, an insulating separator and the liquid electrolyte and an r-dimension: modeling the diffusion of lithium-ions inside the active particles of the electrodes (see Fig. 1.5 for an illustration of the P2D model). The spatial direction, x, of the electrode stack including the anode, separator and cathode includes equations for the electrochemical reactions and charge and mass transport (see Fig. 1.5). The electrodes are modeled as porous media consisting of inactive and active material. The separator is of electrically insulating solid material and is also porous. The radial direction models the one-dimensional diffusion process of lithium-ions inside the particles, representing active material particles as idealized spheres.

The radial dimension of the model is defined within the active materials of both electrodes, which are modeled as particles. The particles are assumed to be ideal spheres and thus this second dimension can be described along its radius. At each point along the geometrical x-dimension of the electrodes, a geometrical particle dimension is defined but only coupled via a scalar, giving rise to the moniker "pseudo-2D" [71].

A brief overview of the equations is given to provide context to the computational complexity required by a P2D model, as well as, the basis for the simulation tool used to generate data for the anode potential estimation method in Chapter 5. In general, the dependent variables of the P2D model are considered to be the lithium-ion concentration in the liquid and solid phase, c_l and c_s respectively, and the potential in the electrolyte and in the active material composite Φ_l and Φ_s. These are solved by using a number of differential equations. The ionic flux j_n couples the solid and liquid phases [39].

Starting from the separator in the spatial domain (Fig. 1.5), the lithium-ion concentration in the liquid electrolyte domain, c_l is given by

$$\varepsilon_\mathrm{l}\frac{\partial c_\mathrm{l}(x,t)}{\partial t} = \frac{\partial}{\partial x}\left(D_\mathrm{l}^\mathrm{eff}\frac{\partial c_\mathrm{l}(x,t)}{\partial x} + i_\mathrm{l}(x,t)\frac{(1-t_+^0)}{\mathrm{F}}\right), \tag{1.7}$$

where ε_l is the volume fraction of the liquid phase (porosity), $D_\mathrm{l}^\mathrm{eff}$ is the concentration dependent effective electrolyte diffusivity, the transference number of the cations in the solution is t_+^0 and F, the Faraday constant.

In the solid electrode, the concentration of lithium-ions in the active material, c_s is modeled in the particle domain by

$$\frac{\partial c_\mathrm{s}(x,r,t)}{\partial t} = D_\mathrm{s}\left(\frac{\partial^2 c_\mathrm{s}(x,r,t)}{\partial r^2} + \frac{2}{r}\frac{\partial c_\mathrm{s}(x,r,t)}{\partial r}\right), \tag{1.8}$$

with the radius of the active material particles, r, and the diffusivity of the solid phase D_s. The charge transfer at the particle surface boundary from the electrolyte to the active material is defined as

$$\left.\frac{\partial c_\mathrm{s}(x,r,t)}{\partial r}\right|_{r=r_\mathrm{p}} = -\frac{1}{D_\mathrm{s}}j_\mathrm{n}(x,t). \tag{1.9}$$

The ionic flux, is modeled using the Butler-Volmer equation,

$$j_\mathrm{n}(x,t) = \frac{i_0(x,t)}{\mathrm{F}}\left[\exp\left(\frac{\alpha_\mathrm{a}\mathrm{F}}{\mathrm{R}T}\eta(x,t)\right) - \exp\left(\frac{\alpha_\mathrm{c}\mathrm{F}}{\mathrm{R}T}\eta(x,t)\right)\right]. \tag{1.10}$$

The Butler-Volmer kinetics model, relates the reaction rate to the surface overpotential, η, with the ex-

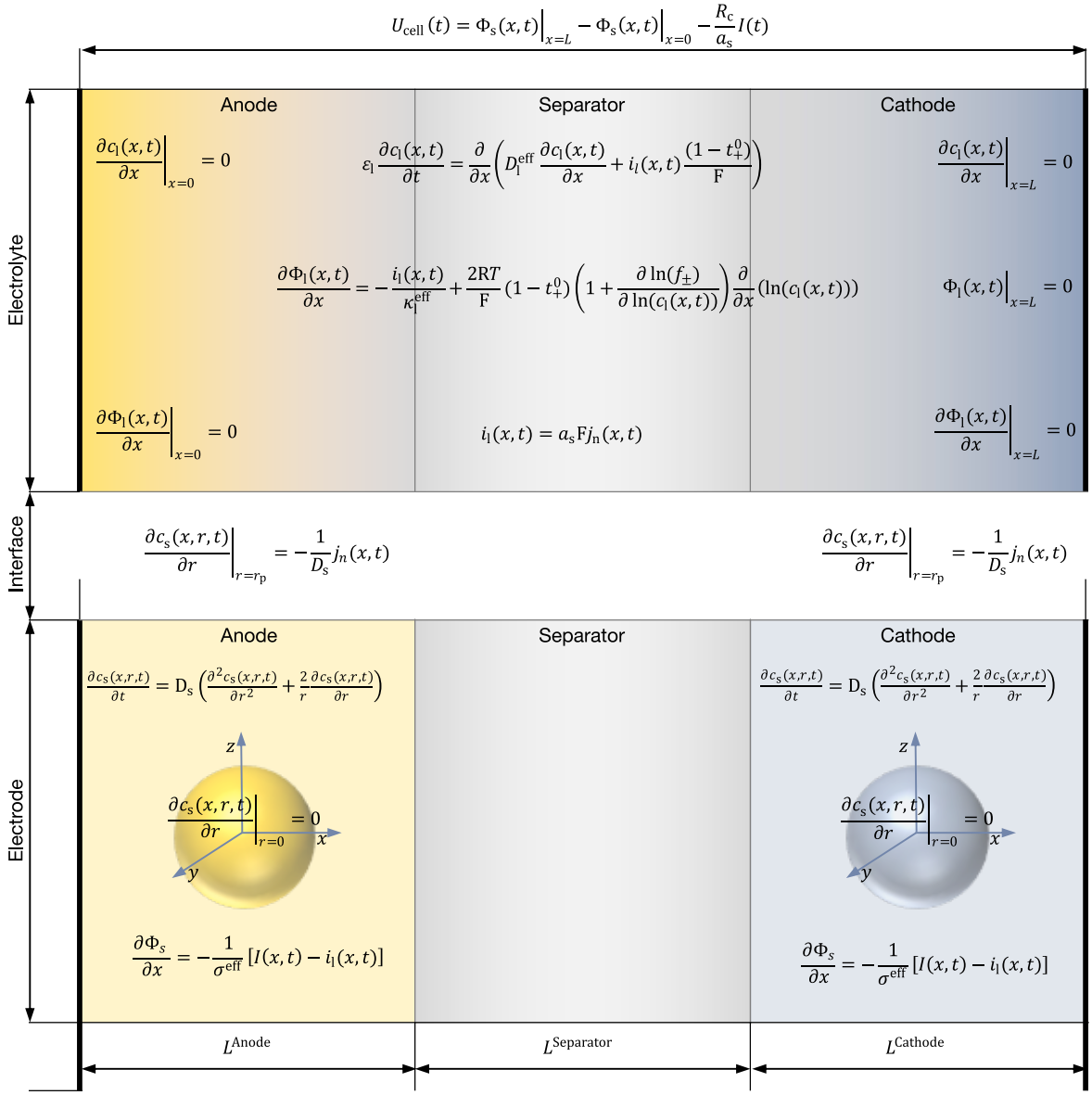

Figure 1.5: Overview of the P2D's differential algebraic equations and the related boundary conditions shown over the relevant main x-dimension and pseudo particle r-dimension reproduced from [72] (according to Ref. [30; 54–57]).

change current density, i_0, which describes the amount of charge that flows in an equilibrium state [71]. The transfer coefficients at the anode and cathode, α_a and α_c, respectively, relate to how an applied potential gradient favors one reaction-direction over another, indicating the reversibility of the reaction. The constants in the equation are the Faraday constant, F, and the Gas Constant, R. The potential difference between the active material composite and the liquid electrolyte, η, is referenced to the electrode potential at the electrode-electrolyte interface at equilibrium state, E_{eq}, by

$$\eta(x,t) = \Phi_s(x,t) - \Phi_l(x,t) - E_{\text{eq}}(\theta(x,t)), \tag{1.11}$$

where $\Phi_s - \Phi_l$ is the electrochemical potential of the electrode, which is also influenced by the electrode stoichiometry, Θ, relating to the degree of lithiation of the active material based on the relationship of the reactants and products.

The gradient of the electrical state $\partial\Phi_l/\partial x$ is built up from an arbitrary position along the thickness of the electrode stack against a virtual lithium reference electrode in the electrolyte, and represents the liquid potential [73], is defined as

$$\frac{\partial\Phi_l(x,t)}{\partial x} = -\frac{i_l(x,t)}{\kappa^{\text{eff}}} + \frac{2RT}{F}(1-t_+^0)\left(1+\frac{\partial\ln(f_\pm)}{\partial\ln(c_l)}\right)\frac{\partial}{\partial x}(\ln(c_l(x,t))), \tag{1.12}$$

where κ^{eff} is the effective conductivity, and the activity coefficient $\partial\ln(f_\pm)/\partial\ln(c_l)$, which is a function of the lithium-ion concentration in the electrolyte.

The potential in the solid phase, Φ_s, is derived using differential Ohm's law, by equating the solid phase potential (voltage) to the current density in the solid phase, i_s (current), over the electric conductivity, σ^{eff} (resistance), as follows:

$$\frac{\partial\Phi_s(x,t)}{\partial x} = -\frac{1}{\sigma^{\text{eff}}}\left[I(t) - i_l(x,t)\right], \tag{1.13}$$

where i_s is substituted for $I(t) - i_l$ according to Kirchoff's law [71]. The current density in the solid phase in the electrolyte, i_s is derived by Faraday's law and is related to the ionic flux by

$$\frac{\partial i_l(x,t)}{\partial x} = a_s F j_n(x,t), \tag{1.14}$$

where a_s represents the ratio of surface to volume of the spherical active particles.

As can be seen by the representation above, the disadvantage of these methods is that they require computation-costly and time-consuming numerical methods to solve the set of differential algebraic equations [74]. Additionally, in order to achieve the high model fidelity offered by electrochemical modeling methods, age-specific cell parameters are required, such as the material properties of diffusivity in the active material and electrode porosity [75].

1.2.3 Semi-Empirical Models

A common aging estimation approach, semi-empirical modeling, fits a physics-informed function to battery aging data [40; 59; 76; 77]. This is a useful method for estimating the SOH with respect to relevant and measurable variables. Although this approach does not allow for a reproduction of battery dynamics, such as the voltage and temperature response, it does provide a simplified approach for estimating the current or future SOH based on directly measurable parameters.

The general approach for this method requires a large input data-set containing the relevant features. Most often, a holistic aging model is a function of time, SOC, temperature, depth-of-discharge (DOD) and current. It is often common to separate the complete SOH loss into a calendric and cyclic function, where the total aging loss, ΔSOH_{tot}, is a superposition of the calendric loss, ΔSOH_{cal}, and the cyclic

loss, ΔSOH_{cyc}, as seen from the equations below.

$$\Delta SOH_{\text{tot}} = \Delta SOH_{\text{cal}} + \Delta SOH_{\text{cyc}} \tag{1.15}$$

$$\Delta SOH_{\text{cal}} = f(t, T, U) \tag{1.16}$$

$$\Delta SOH_{\text{cyc}} = f(t, I, DOD, T) \tag{1.17}$$

To obtain a final equation out of an aging data-set, the various variables are fit using a regression technique. The term 'semi-empirical' comes from the fact that the form of the fitting equation owes part of its shape to underlying physical equations which describe electrochemical processes occurring within the battery. Two common empirical models use the Arrhenius equation, which describes the dependence between temperature and reaction rates [78] and the Tafel equation, which describes the growth of corrosion layers [79]. It is often a matter of experimental validation providing the degrees fitted to the remaining parameters. Hahn et al. provided a summary of studies related to fitting the time term, t, and concluded $t^{0.75}$ to consistently fit most data [77]. Considering this and the Arrhenius and Tafel equations mentioned above, a holistic aging equation is proposed of the form

$$\Delta SOH_{\text{tot}} = a_1 \overline{U} e^{-\frac{\text{E}_\text{a}}{\text{R}\overline{T}}} t^{0.75} - (a_2 \overline{U}^2 + a_3 \overline{DOD}) Q^{0.5} \tag{1.18}$$

where a_1-a_3 are fitting coefficients, $\overline{U}$ is the average voltage, E_a is the activation energy, R is the gas constant, $\overline{T}$ is the average temperature, $\overline{DOD}$ is the average DOD and $\overline{Q}$ is the total charge throughput. This equation is used as a comparison against a random forest inspired machine learning approach in Chapter 4.

The relatively low number of parameters and simplicity of the resulting equations makes the semi-empirical model an ideal solution for many applications where computation limits are a factor. Fitting the model to a wide variety of aging states presents the largest challenge to this method. In a similar vein, highly dynamic aging, composed of irregular or intermixed calendric and cyclic aging modes significantly challenge the performance of the model.

1.3 The Advent of Data-Driven Modeling

Machine learning is a standard modeling tool best suited for when the underlying behavior is unknown or too complex for traditional approaches and when a sufficient data-set is available. These data-driven methods have been applied to fields such as cyber security, automation and pharmaceutics, and in recent years has been applied to the problem of battery state estimation. The growth of machine learning has expanded to the battery field, with data-driven approaches becoming popular for SOC and SOH estimation [80]. Over the last decade a number of publications have been released where authors have successfully used neural networks, support vector machines and Gaussian regression techniques to estimate SOC and SOH as summarized by Ng et al. [81].

Large collections of data imbue the use of data-driven and machine learning methods. In this section, two main categories of machine learning– supervised and unsupervised– are reviewed in more detail, and a brief discussion of other important terms in the field of data-driven methods are discussed. The necessary terms and approaches which are applied in this work are later detailed.

1.3.1 Fundamental Algorithms

The back-bones of machine learning algorithms are: a loss function, tuneable parameters, an optimization criterion (typically an objective function), and a routine that utilizes the available training data to best satisfy the optimization criteria [82]. Regardless of the machine learning approach the following training loop procedure is followed:

1. Initially (often randomly) parameterize the model tuning parameters (or weights).
2. Compute the loss from estimation to truth.
3. Update weights in the direction that lowers the total loss.

The data-driven techniques used in this work are almost exclusively built around regression (supervised learning). In this section, the fundamental building blocks of linear regression and logistic regression are defined as they are core principles to the neural network approaches used in the subsequent chapters.

Linear Regression

For a regression model, the goal is to predict a value, y, given some set of input features x. A standard representation of a linear model is:

$$f(x) = wx + b, \tag{1.19}$$

where w and b are two parameterization values known as weights and bias, respectively. Provided with a data-set of input and output pairs, $\mathcal{D} = (x_i, y_i)_i^n$, w is a D-dimensional vector the same size as x, and b is a real number. For a regression model, the task is to find the values of w and b corresponding to the predicted value of $\hat{y_i}$ closest to the target value y_i given a set of inputs $\hat{x_i}$, with $\hat{x_i}$ and $\hat{y_i}$ being values used for testing, thus a sub-set of the larger data-set. The solution for w^* and b^* yielding the most accurate result is found by minimizing the objective function,

$$\frac{1}{n}\sum_{i=1}^{N}(f_{w,b}(y_i) - \hat{y_i})^2. \tag{1.20}$$

The objective function is an estimation of the prediction error and can be the absolute, average or root mean square error of the prediction versus the target value, for example. This metric provides feedback to the optimization routine and indicates the level of success any previous change in the model parameters (or weights) had on satisfying the cost function.

Most machine learning algorithms considered in this work (and indeed in practice) use a differentiable optimization function. These differentiable routines employ a gradient decent to identify the local minimum of the objective function. As the linear and non-linear functions used in regression and neural networks can be derived, computing the gradient of loss with respect to the weights, one can tune these parameters in the opposite direction of the calculated gradient. As this is a core attribute of many machine learning algorithms, and is in essence the *learning* part of machine learning, it is worth exploring in more detail.

The objective of regression is to minimize the loss function. Assuming the function is smooth and differentiable, the function's minimum is located where the derivative of the function is equal to zero.

As an example, take the mean squared error as the cost function,

$$C = \frac{1}{N}\sum_{i=1}^{N}(y_i - (wx_i + b))^2, \tag{1.21}$$

where y is the target, x the input and w and b forming the slope and intercept of a linear function. In order to minimize C, the partial derivatives with respect to the tune-able parameters are taken, allowing for the adjustment of those parameters in the direction of decreasing gradient with each iteration. The derivatives are as follows,

$$\frac{\delta C}{\delta w} = \frac{1}{N}\sum_{i=1}^{N} -2x_i(y_i - (wx_i + b)) \text{ and} \tag{1.22}$$

$$\frac{\delta C}{\delta b} = \frac{1}{N}\sum_{i=1}^{N} -2(y_i - (wx_i + b)). \tag{1.23}$$

The function then iterates through all samples, i, and calculates new parameters w and b which approach the minimum of the cost function. A neural network consists of many of these linear functions above which all search for the minimum value based on the input x, and as such the example above provides a key insight into the machinery of many machine learning algorithms.

Logistic Regression

Logistic regression blurs the line between classic regression and classification, as it is a classification algorithm with a binary output $y\epsilon(0, 1)$. The underlying model is still linear of the form seen in Eq. 1.19, where the output could take the form of any value from negative to positive infinity, which is processed by a standard logistic function whose co-domain is between (0,1). The most common standard logistic function is known as the sigmoid function and has the form

$$f_{w,b} = \frac{1}{1 + e^{-wx+b}}. \tag{1.24}$$

In logistic regression, values closer to zero receive a negative label, and values closer to one receive a positive label. In neural networks, mapping the output to binary values is not strictly necessary, however, transforming the unbounded linear function to values between zero and one significantly improves learning performance by scaling all inputs, ensuring that changes between inputs can be equally weighted, and is a key part of the activation function of neurons, as discussed below.

1.3.2 Machine Learning

Machine learning uses data to define a set of rules to describe patterns allowing for predicting the behavior of unknown data. The first commercially successful implementation of a neural network was in 1989 by Yann LeCun from Bell Labs who combined a convolutional neural network with back propagation for the automated reading of ZIP Codes for the United States Postal Service [83]. Machine learning is typically divided into two main categories: supervised and unsupervised. Some of the most common methods per category are shown in Fig. 1.6.

Supervised Learning

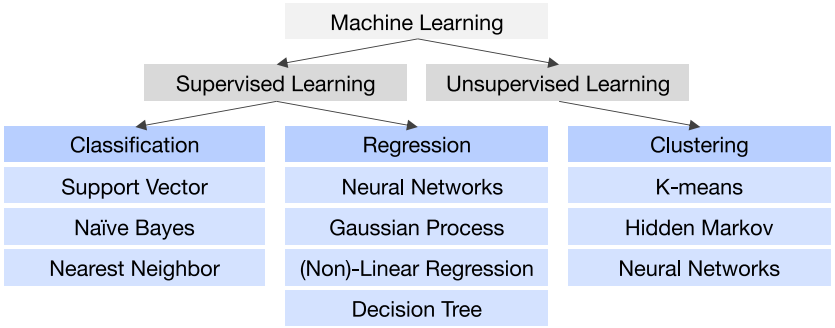

Figure 1.6: Overview of the various machine learning options available and suggestions for algorithm selection based on goals, adapted from Scikitlearn [84].

Sometimes referred to as predictive learning, supervised learning aims to map from a set of inputs, x, to a set of outputs, y. The standard notation describing the basic building block of a machine learning algorithm is a training set of N samples of input and outputs pairs, $\mathcal{D} = (x_i, y_i)_i^N$. The shape of the input, x, depends on the data collected, whether it is vectors and time series, as in this thesis, or pictures, movies or even sentences. The output y also depends on the task at hand, but is typically either a nominal value or a categorical one. When the output is categorical, the machine learning objective is known as classification which assigns labels to previously unlabeled data after learning patterns from training data. When the output is a real-value, the problem becomes one of regression, which builds a model based on patterns from a training set containing a ground-truth of the target variable.

A core element of many machine learning algorithms is the neural network. A neural network learns an output by mapping the dependency of a set of inputs through a neural mesh. Depending on the number of layers in the network, this can also be referred to as 'Deep Learning', where the number of layers is greater than two. The main tasks of the 'machine' in machine learning (the neurons and their connections performing forward propagation) is to: 1. receive an input, 2. perform a transformation to calculate the neuron activation and 3. to adjust the connection strength between neurons to reduce the prediction error and to propagate the value forward.

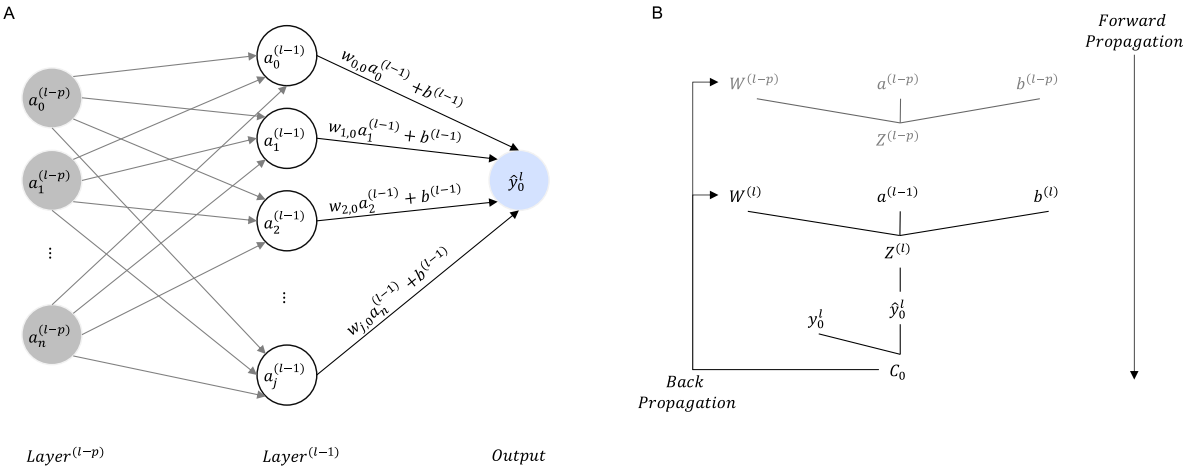

Figure 1.7: A) Neural networks consist of many neurons holding an activation value, connected by mappings, indicating the strength of connections between neurons. B) the 'machine' and 'learning' aspects of a neural network refer to forward propagation and back propagation, respectively. Representations inspired by Burkov et al. [82].

The most basic element of a neural network is the neuron. A neuron is a place-holder for a value, called an activation. The connections between neurons are tuned via linear algebra and are referred to as the activation function. By assigning weights to each of the neuron connections, the strength or relevance of a connection is defined. The weight function is typically a linear function of the same form discussed above in Eg. 1.19, such that for a single neuron (like $\hat{y}$ in Fig. 1.7) the expression becomes

$$\hat{y}_0^l = \sigma(w_n^{(l-1)} a^{(l-1)} + b^{(l-1)}), \tag{1.25}$$

where $\hat{y}$ (or more generally a) represents the activation of a neuron. The weight parameter w, is a matrix with indices n and l corresponding to the index of the current neuron and the index of the previous layer, respectively, and is added to the offset or bias, b. Finally a sigmoid function, σ, is used to scale the value for improved learning performance.

Any layer in a neural network, Z^l, can be determined from the weight matrix, W, the activations or inputs, z, from the previous layer, $l-p$ and the bias vector

$$Z^l = \sigma(W^l z_{l-p} + b^l). \tag{1.26}$$

A fully connected feed-forward neural network is then a nested vector function,

$$y_0^l = f_{(l-p)}(f_{(l-1)}(f_l(x))), \tag{1.27}$$

where each of the vector functions has the form in Eq. 1.26.

In algebraic notation, a neural network takes the form

$$Z^l = \begin{bmatrix} z_0^l \\ z_1^l \\ \vdots \\ z_k^l \end{bmatrix} = \sigma\left(\begin{bmatrix} w_{0,0} & w_{0,1} & \dots & w_{0,j} \\ w_{1,0} & w_{1,1} & \dots & w_{1,j} \\ \vdots & \vdots & \ddots & \vdots \\ w_{k,0} & w_{k,1} & \dots & w_{k,j} \end{bmatrix} \begin{bmatrix} a_0^0 \\ a_1^0 \\ \vdots \\ a_j^0 \end{bmatrix} + \begin{bmatrix} b_0^l \\ b_1^l \\ \vdots \\ b_j^l \end{bmatrix} \right), \tag{1.28}$$

where the subscripts in the weight matrix refer to the neuron index in the previous layer, j, and neuron index in the subsequent layer, k. The representation of a neural network in Fig. 1.7A helps to show the flow of the mapping between neurons and their respective indexing.

With this in mind, it becomes clear that the number of tuning parameters is related to the number of neurons and the number of layers, with each neuron requiring a weight, w, and a bias, b, from each adjoining neuron from the previous layer in a fully connected setup. This matrix operation enables the neural network to parameterize a number of linear functions, and therefore its power to model highly complex functions.

It is common practice to randomly drop connections between neurons, such that the learning does not over-emphasize a single path through the network. The activation value is often scaled to a value between zero and one in order to improve model performance. The scaling of the activation value is aided by one of a number of scaling functions. Common scaling functions are the sigmoid function, hyperbolic tangent function (tanh), step function, and rectified linear activation unit (ReLU) [82].

The sigmoid function, also known as the logistics function, has boundaries between $(0, 1)$,

$$\sigma(x) = \frac{1}{1 + e^{-x}}. \tag{1.29}$$

This function saturates extreme values often leading it to be more susceptible to the vanishing gradient problem. The vanishing and exploding gradient problems are discussed once the back propagation has been introduced as well.

The tanh function is similar to the sigmoid function, including the propensity for the vanishing gradient, except that it has a range between $(-1, 1)$,

$$\sigma(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}. \tag{1.30}$$

The step function is a binary function with two discrete values, typically 0 and 1.

$$\sigma(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases} \tag{1.31}$$

The ReLU function is perhaps the most common sigmoid function used. Similar to the step function, it is a piecewise function where values less than one are saturated to zero, but unlike the other activation functions, there is no upper range for positive inputs. All positive inputs take on a linearly increasing value between zero and infinity, making it less susceptible to the vanishing gradient problem, but more susceptible to the exploding gradient problem.

$$\sigma(x) = \max(0, x) = \begin{cases} x & if\ x > 0 \\ 0 & if\ x \leq 0 \end{cases} \tag{1.32}$$

The learning part of machine learning is then the task of updating the weights and bias to improve performance and is achieved by minimizing a cost function. Figure 1.7B highlights the dependencies between layers and the flow of forward and backward propagation. In the first step, a cost function is defined, for example the squared error

$$C = \sum_{i=0}^{l} \left(Z_i^l - y_i\right)^2, \tag{1.33}$$

where C is the cost, y_i is the true value being trained on, Z_i^l is the estimated value from the neural network (in the regression case $\hat{y}_i$), both being at the index i.

Back propagation is then the minimization of this cost function, allowing for the optimal parameter set by employing gradient descent to change parameters iteratively in the direction of a local minimum and back propagating on each layer. To achieve this, the partial derivative of the cost function is taken with respect to each parameter utilizing the chain rule, a, w, and b, and successively for each layer, following the flow diagram in Fig. 1.7B.

$$-\nabla C = -\nabla \Sigma_{i=0}^{l} \left(Z_i^l - y_i\right)^2 = \frac{\delta C}{\delta W^l} \cdot \frac{\delta C}{\delta a^{(l-p)}} \cdot \frac{\delta C}{\delta b^l} \tag{1.34a}$$

$$\frac{\delta C}{\delta W^l} = \frac{\delta Z^l}{\delta W^l}\frac{\delta a^l}{\delta Z^l}\frac{\delta C}{\delta a^l} \tag{1.34b}$$

$$\frac{\delta C}{\delta a^{(l-p)}} = \frac{\delta Z^l}{\delta a^{(l-p)}}\frac{\delta a^l}{\delta Z^l}\frac{\delta C}{\delta a^l} \tag{1.34c}$$

$$\frac{\delta C}{\delta b^l} = \frac{\delta Z^l}{\delta b^l}\frac{\delta a^l}{\delta Z^l}\frac{\delta C}{\delta a^l} \tag{1.34d}$$

As mentioned above about the activation functions, machine learning models can often suffer from a vanishing gradient problem or conversely, an exploding gradient. This arises during the back propagation step in the learning. As the cost function calculates the error gradient through each of the layers, the gradients of each subsequent layer have less impact on the overall model error. As the sigmoid, tanh and step function are saturating the inputs between a more narrow range, the model does not have as much variance to learn from. The opposite occurs if the activation function provides too large of a range, where the gradients during back propagation keep expanding indefinitely. This is known as the exploding gradient problem [82].

The neural network is a powerfully simple use of linear algebra that functions as a keystone to machine learning. The size and complexity of the network is in large part dependent on the topology of the network, allowing the user to define the optimal numbers of neurons and layers. In comparison to linear or non-linear regression– where the form of the function is explicitly defined based on an assumed dependency between the inputs and the output, often limiting the number of tune-able parameters to a small handful– the neural network offers a high level of flexibility, requiring no knowledge of the physical dependencies being modeled.

Two other types of machine learning methods worth of note are semi-supervised learning and reinforcement learning. Semi-supervised learning contains both labeled and unlabeled data. In some cases the labels are generated by heuristic algorithms using the input data. Autoencoders are often used in this type of learning which are one way of generating temporally supervised learning, for example, learning the context of a video frame using the following frames for more context, or by learning the last word of sentence by using the other words in a sentence. Reinforcement learning takes a state vector as an input and outputs an action which optimizes a perceived reward. This special class of machine learning is well adept at solving sequential problems with long-term goals such as resource management, supply chain logistics or strategy games [85–87].

Unsupervised Learning

Unsupervised learning does not have a defined output y and instead labors to identify patterns or other useful information contained within the training data $\mathcal{D} = (x_i)_i^N$. As there is no target value or ground-truth to compare predictions to, evaluating model accuracy is a complex challenge in this application, however, it does offer the opportunity to uncover previously unknown patterns. The most common applications of unsupervised learning are clustering, dimensionality reduction and structure detection [88]. Supervised methods are also adequate at performing these tasks, however, they require labeled training data, which often requires manually filtering the data before hand, or acquiring specific data-sets.

One of the most common uses of unsupervised learning is clustering. The working principles of the K-means algorithm are detailed here as they are later central to the work in Chapter 4. K-means clustering identifies groups based on the euclidean distance from a centroid. The objective function,

J, defining the K-means algorithm is

$$J = \underset{k=1,\ldots,K}{\operatorname{argmin}} ||x_i^k - c_k||^2, \tag{1.35}$$

where i is the index of a point within the total sample set, k is the cluster under evaluation within the defined cluster set, K, x_i is a single sample and c_k is the centroid for the cluster under evaluation. The minimization iterates through various cluster toplogies until the minimum total distance between each cluster, given the specified amount of clusters is found.

The number of centroids is an input to the algorithm, and can be any real number. Typically, the clustering accuracy decreases with increasing number of clusters. There are a few methods for choosing the optimal amount of clusters: average silhouette method and elbow method, however, these are often imprecise science [82].

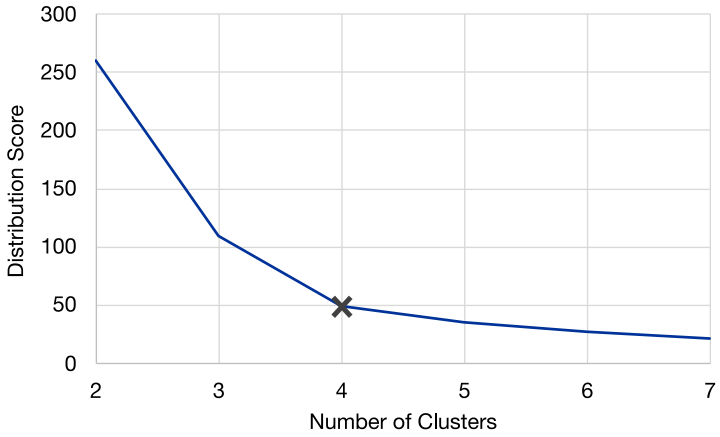

Figure 1.8: The elbow method is a graphical approach to determining the optimal number of K-means clusters, by finding the ”elbow” in the distribution score plot [82]. In this hypothetical example the optimal number of clusters is four.

The average silhouette method is a measure of how similar objects within a cluster are compared to another cluster. The value between (-1,1), with negative values being falsely identified, or outliers, 0, indicating points on the boarder between two clusters and 1 indicating a perfect cluster fit. The optimal number of clusters found via this method is considered to be the number of clusters which return the highest average silhouette score for all points in the data-set. The elbow method plots the average distance within the clusters (see Fig. 1.8). As the number of clusters increase, the average distance within the cluster to the cluster's centroid decreases. The elbow refers to the point in this plot where the average distance within the clusters remains relatively stable.

Other important unsupervised methods utilized to uncover undefined structure in the data are often rule-based. These methods, such as the Apriori algorithm, are also similar to the random forest method (which is described in more detail in Chapter 4). In this method, data are split into smaller groups following binary rules until sufficiently small groups are formed. One last unsupervised method which is highly relevant for feature engineering is dimensionality reduction. Principle component analysis (PCA) is one of the most popular approaches for this, where the correlation of each feature to one another is used to define component directions, helping combine multiple correlated features into a new component.

1.3.3 The Machine Learning Process

The term machine learning is a broad term comprising many very different models. Although each of these models can be as different as a P2D and an ECM model, the machine learning process follows a fairly standard template comprising the following key steps:

Data Processing
The accuracy of a machine learning model depends mostly on the data preparation and collection [89]. In Chapter 2 a detailed exploration of an automotive data-set provides an indication of the quality of the available data-set for use in data-driven applications. In this step, the data is first evaluated for quality, removing any outliers or otherwise corrupted data.

Feature Engineering
The term feature is synonymous with input variable. Feature engineering is used here to generally refer to the selection and pre-processing of the raw data for use in the machine learning model. Depending on the type of data, various methods such as one-hot encoding, binning or normalization are effective means of reducing the data to only the most numerically efficient component [82].

Model Training
Perhaps the most important step in model training is first selecting which model is to be trained. Depending on the task at hand, a number of methods are applicable and multiple may even be suitable (see Fig. 1.6 for an overview of potential methods). Considerations can also include desired training speed, categorical vs. numerical features, prediction/training speed and explainability. Another very important consideration is the potential for over or underfitting– which often results from either too complex of a model considering the available data or one that is not complex enough, respectively.

Hyperparameter Tuning
In addition to the weights learned during model training, there are a number of hyperparameters which specify the model training strategy. Common training parameters include the test/train split (see below), learning rate, activation function, layer type, number of epochs and batch size. A robust hyperparameter tuning will assess various combinations of these parameters to find the set which yields the most accurate model. Effective strategies for finding the ideal parameter-set include grid search, random search and Baysien hyperparameter optimization [83].

Validation
For model development there are typically three data-sets, the training, validation and test data-set. It is best practice to randomly shuffle the total data-set from the feature engineering step into the three sub-data-sets. The largest of these three should be the training data-set, with the validation and test data-set being roughly the same size. The model only sees the training data-set during development with the test data-set being used for tuning of the hyperparameters and the validation data-set being used for final model assessment.

1.3.4 Data-Driven Approaches in Battery Applications

A growing area of research focuses on battery state-estimation using data-driven methods. These methods have grown in popularity because of their ability to map a given input to a desired output using linear algebra, instead of a pre-determined function. This method is particularly useful when deployed in areas where the underlying function which is being modeled is unknown or has too many interdependencies to capture analytically.

By far, the neural network, and its variations are the dominant data-driven method in estimating

the internal states of batteries in the literature [90; 91]. One of the first publicly available large data-sets was issued from NASA's Ames Prognostics Center of Excellence (PCoE), which spurred a number of research papers applying data-driven techniques onto the data-set [92–97]. Other notable data repositories include ones from Sandia National Lab, Everlasting and ones from the universities of Oxford, Berkeley and Karlsruhe (KIT) [98–102].

A small summary of recent work utilizing data-sets for state estimation is provided in Table 1.1. Indeed these approaches are showing great promise in the fields of state estimation of SOH [93; 103–108] and SOC estimation [109–112]to fault detection [113; 114] and thermal management [115]. Common challenges addressed in each of these works, like data quality, model selection and optimization are also encountered in the further chapters of this thesis. What this current work additionally addresses is the added complexity when dealing with highly variable and dynamic use conditions, such as those from non-laboratory studies and path dependent aging.

Table 1.1: Summary of applied data-driven methods for battery state estimation with the reported errors using root mean square error.

Method	**Target**	**Error / %**	**Input Values**
Support Vector Machine [109]	SOC	0.7	U, I, T
Gaussian Process Regression [110]	SOC	0.2	U, I, T
Feed-Forward Neural Network [111]	SOC	0.8	U, I, T
Recurrent Neural Network [116]	SOC	0.7	U, I, T
Recurrent Neural Network [112]	SOC	0.9[1]	U, I, T
Recurrent Neural Network [117]	SOC	1.8[1]	U, I, T
Support Vector Machine [104]	SOH	2.5	ΔT
Support Vector Machine [105]	SOH	0.5	E, Ah, Q
Support Vector Machine [106]	SOH	1.4[2]	SOC, Q, U
Long Short-Term Memory [107]	SOH	0.8[1]	U
Long Short-Term Memory [118]	SOC/SOH	3.2	U
Long Short-Term Memory [119]	SOH	2.5	U, I, T
Convolutional Neural Network [108]	SOH	1.1	U, I, T

[1] These errors are listed in as mean absolute error (MAE).
[2] These errors are reported in mean absolute percent error (MAPE), all others are root mean square error (RMSE).

1.4 Structure of the Dissertation

The composition of this work can be broken into three main parts. In order to provide a quick overview of the contents of the thesis, a graphical outline is provided in Figure 1.9. To help highlight the degree of focus from the more classical battery modeling and those with a more data-driven approach, the graphical outline is color-coded, with blue corresponding to battery modeling and green representing the inclusion of data-driven methods. Chapter 1 should have provided the context for the scope of this work, detailing fundamentals of battery state estimation and machine learning. Chapter 2 focuses on the automotive context and provides a detailed description of the real-world data-set from BEVs which is used to help structure the path dependent measurements found in Chapter 3 and which was available for the data-driven approaches of Chapter 4. This chapter is a statistical evaluation of data and is therefore strictly focused on the data, its distribution and generation. Although the data is analyzed through the lens of battery operation, there is no direct application of classic battery

modeling methods in this section (and is therefore classified as green according to the legend used in the graphical abstract). The main contribution in this work comes in the form of three peer-reviewed papers. The first paper, published via the Journal of the Electrochemical Society [120], investigates path dependent aging from a measurement study and is included as Chapter 3. This paper relies solely on the state-of-the-art differential voltage analysis methods, and is therefore the only section completely classified as blue using the legend in the graphical abstract. In order to improve battery aging model accuracy– with the potential to compensate for path dependent aging– an investigation into the application of machine learning on automotive field data is carried out in Chapter 4. This paper has been published by the Journal of Power Sources [121], and highlights the power of machine learning for battery state estimation, at least with respect to aging. A second application of machine learning is investigated in Chapter 5, namely the potential to improve fast charging, and was published in the Journal of the Electrochemical Society [72]. Chapter 5 presents three methods for online estimation of the anode potential during charging operation. As these two papers contain both classic and data-driven modeling, they are shown as both green and blue in Fig. 1.9. Finally, Chapter 6 details the results and discusses their relevance in a larger context, as well as, elaborates on the connection between the publications.

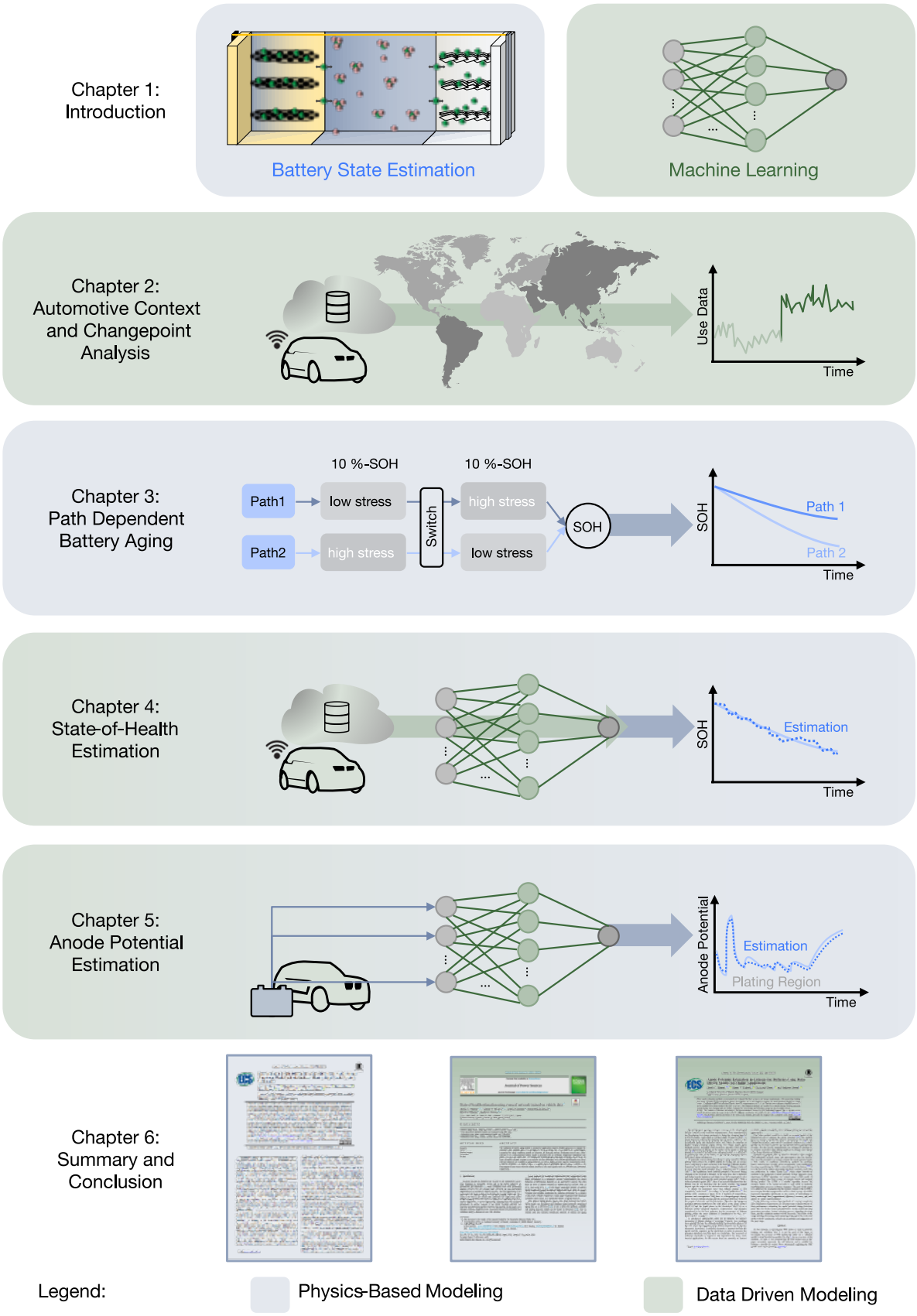

Figure 1.9: The graphical outline of the thesis highlighting the contribution of battery state domain knowledge (blue) and machine learning domain knowledge (green).

2 Aging Behavior in the Automotive Context

Figure 2.1: Crop from the thesis structure highlighting the focus of this chapter.

The automotive context referred to in this thesis describes vehicles driven in real-world conditions with no artificial influence on the driving behavior or environment. A large data-set from the BMW i3 was available for analysis to inform data-driven battery model development [122; 123]. In this section, the statistical lens through which the data is interpreted, as well as, a presentation of this automotive context is provided. To the best of the author's knowledge, the analysis in this section is conducted on the largest and most complete collection of BEV measurements. The BMW i3 fleet is one of the first BEVs from the modern era, offering over eight years of collection history. The purpose of this section is to first explore the data and provide an overview of effective data analysis techniques saving a more confirmatory analysis for future sections.

2.1 Data Acquisition

In this section, the data collection context is established. Data generation is a key step to machine learning, and all model-based approaches for battery state estimation, therefore, an understanding of the origin and acquisition process of the data is tantamount to a proper handling and interpretation.

2.1.1 The BMW i3

The BMW i3 is one of the oldest modern BEVs. Production for the BMW i3 started on Sept. 18th, 2013, and ended in 2022. In October 2020, the 200,000th BMW i3 was produced. There were three distinct generations of BMW i3 models, the 60 Ah, 94 Ah and 120 Ah battery systems. For this section only the 60 Ah version is presented. This first BMW i3, produced from 2013-2017 is estimated to have sold over 98,000 units in over 74 countries [124]. The wide distribution offers a varied data-set reflecting the diverse conditions experienced by BEVs. What distinguishes the available data-set from the state-of-the-art is access to hundreds of thousands of measurements from thousands of vehicles in non-laboratory conditions. In contrast, the maximum number of cells measured from a publicly available battery data-set is 233 cells [91].

2.1.2 Online Measurements and Embedded Systems

All modern battery systems are accompanied by a BMS. The BMS provides battery control and monitoring functions for the safe and optimal operation of the energy storage system and vehicle. Common BMS functions include: cell balancing within a pack, over and under voltage protection, fault detection, SOC estimation, SOH estimation, communication with other onboard vehicle controllers and external charging controllers, as well as others [125; 126]. The BMS connects with the numerous sensors installed in the energy storage system responsible for measuring the current, voltage and temperature. In addition to raw data collection from the multiple sensors, basic control functions are programmed into the control unit to provide some of the functionality above based on these measured inputs.

As the central controller for safe and robust battery operation, the BMS is required to meet numerous industrial standards such as IEEE 1679.1, UL 1973 and IEC 62619, all relating to the safety specifications for lithium-ion battery operation [127]. Logged data must be permenantly available on the BMS and stored in a flash-resistant manner, such that it is not lost during routine software updates. The demand for system robustness requires premium materials resulting in a high cost-to-performance ratio. For auto manufacturers, ensuring the safest and most reliable battery and BMS involves defining efficient data collection and data handling functions to help minimize the cost of the BMS. This means that every byte of data collected is evaluated for its contribution to the overall safety and cost of the system.

This trade-off between data collection and system cost often results in a more narrow sub-set of data which is optimized for achieving or monitoring the functions listed above. In addition to the strict focus on relevant parameters, customer privacy also restricts potentially relevant data. Data collection should be limited to information regarding the performance and safety of the system, and as such, superfluous data reflecting personal or private information from the user of the system is excluded from collection. The effect on the data-set presented in this chapter is that a few of the relevant battery information signals: SOC, temperature, C-rate and power, are collected in the form of time-integrated binned histograms. These histograms obfuscate any information which could be gained by collecting the data as a time-series, showing the behavior of a user over a linear time-segment, and instead shows only how many seconds a vehicle was operated in a specific region.

2.1.3 Cloud-Data Collection

Cloud computing and connected vehicles offer the opportunity to understand collected data in a larger context. As mentioned above, the BMS is limited due to costs, restricting the potential functionality of an onboard BMS. Additionally, these onboard systems can only view the behavior of the specific energy system they are connected with. Cloud-based platforms are able to store limitless amounts of data in a flexible system. Cloud platforms help connect these single systems to a larger network offering the potential to learn macro-trends and develop functions based on a more broad set of data.

As seen in Fig. 2.2 data is first collected from the BMS. The BMS collates multiple measurement signals describing the current state of the vehicle to be used for safety and reliability analysis of the fleet. These data packages from the field are collected at a central storage server in the Cloud. Combining this battery data from all available vehicles enables Big Data analytics and the application of machine learning methods, which will be further discussed later in this work.

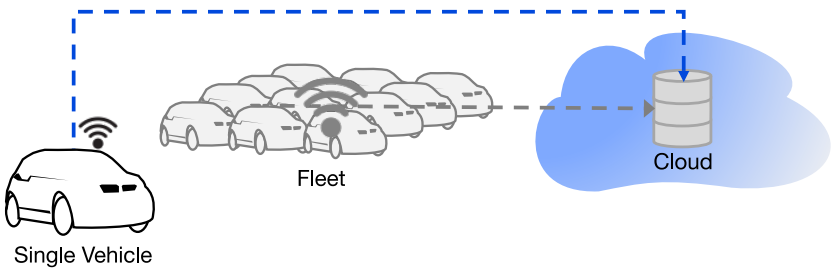

Figure 2.2: A standard process diagram depicts how raw data is measured and processed by an onboard battery management system and then transferred to the cloud for fleet analytics.

2.2 Statistical Data Review

Before the data is used for model development, a non-bias evaluation of the data is first considered. In this section the major statistical properties of the data is explored. This not only introduces the data to the reader, but it also highlights strengths and weaknesses contained within the data-set.

2.2.1 Statistical Evaluation of a Data-set

Terminology and notation used in data science varies between, and sometimes even within, disciplines. In this thesis a few terms will be discussed for standardization and familiarization for the readers.

Raw data is collated into a relational database assuming a rectangular form. The raw data is a mixture of continuous, discrete and categorical types. Continuous data is bound within a specified interval and can be any value in this interval, such as temperature or vehicle age. A discrete data-set can only be specific integer values and is sometimes referred to as integer or count data types. One example of discrete data from this data-set is the number of vehicles in a specific country. Finally, categorical data types ascribe a label from a specified list within a category, as the name would imply. Categorical data types are also known as enumerated or factors. A common categorical data in this work is the origin country for the vehicle.

The rectangular database is referred to as a data frame in this work and can be thought of as a spreadsheet with each column, or feature, coming from a raw signal input from the vehicle. Other terms used to describe these features, which will be used interchangeably in this work, are input and variable. A measurement or readout makes up the rows of the data frame. Other common terms for this are record, observation and sample.

The process of data exploration was first defined by John Tukay in 1977, and was a first departure from classic probability theory to approach data-set with no confirmatory requirements [128]. Data exploration can be broken down in three categories to get a feel for what the data contains: location, variability and distribution [129].

The location provides an estimate of the central tendency of the data and includes measures of mean, median and outliers. For data analyzed in this thesis, two mean estimations are used: the classic mean and a weighted mean which is used for calculating the mean of time-based histogram information. The mean is given by

$$\bar{x} = \frac{\Sigma_i^n x_i}{n}, \tag{2.1}$$

and the weighted mean,

$$\bar{x}_w = \frac{\Sigma_i^n w_i x_i}{\Sigma_i^n w_i}, \tag{2.2}$$

where x represents the feature, n is the number of observations, i indicates the observation index and w the weight (which is time for much of the subsequent analysis) [129].

Handling variability in a data-set is a core tenet to data analysis. The key metrics used to discuss the variably in a feature are the variance, standard deviation, range, percentile and inner quartile range. The variance, σ^2, is calculated as the squared standard deviation, σ, by

$$\sigma^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1}, \tag{2.3}$$

and standard deviation as

$$\sigma = \sqrt{\sigma^2}. \tag{2.4}$$

The range is the spread of the values from minimum to the maximum and the percentile, represents the value at which at least P-percent of the values are equal to or greater than. The inner quartile range (IQR) is a common term denoting the span of values between the 25^{th} and 75^{th} percentile [129].

The final data exploration technique used in this work is to investigate the distribution of the data using box-plots, histograms and density plots. These graphical techniques help characterize the location and variability of the data-set as well as provide information on the skewness and kurtosis [128]. Figure 2.3 shows three common graphical techniques used to investigate a data-set. Figure 2.3A shows randomly generated normally distributed data to highlight a standard distribution and the correlation to the location metric, mean, and the variance, where 95.4 % of the data lies within 2σ. If more data is lying outside the 2σ threshold then the data set is considered to have a high kurtosis, and as well, if the mean is not centered, the data-set is said to be skewed (see Fig. 2.3C).

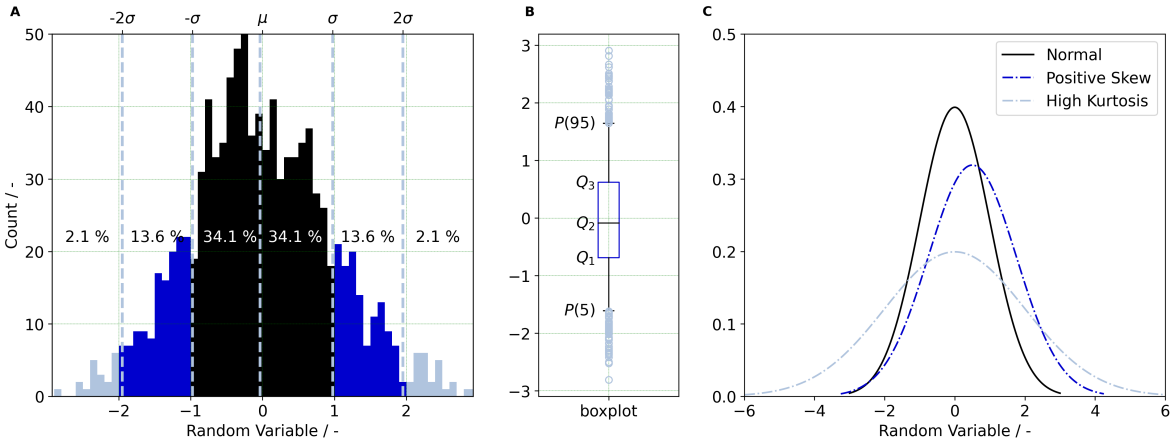

Figure 2.3: A randomly generated dataset shows three classic graphical techniques for data exploration: A) variance plot, B) boxplot and C) distribution plot.

Skewness is a measure of asymmetry. A normal distribution will have zero skew, where the left and right side distribution of the mean are equal [128]. The skewness coefficient λ_G is negative for left skewness and positive for right skewness. One method often used for computing skewness is the Galton skewness

which is a function of the quartiles, where:

$$\lambda_G = \frac{Q_1 + Q_3 - 2Q_2}{Q_3 - Q1}, \tag{2.5}$$

with Q_2 representing the median and Q_1 and Q_3 representing the upper and lower quartile or 25^{th} and 75^{th} percentile, respectively. The boxplot is also a common method for showing this distribution, also providing a good visualization of outliers in the data, as demonstrated in Fig. 2.3B.

Kurtosis measures the tail distribution. A normal distribution has a kurtosis value of 3, which is why it is common to measure the excess kurtosis as

$$\lambda_P = \frac{\Sigma_{i=1}^{n}(x_i - \bar{x})^4/n}{\sigma^4} - 3. \tag{2.6}$$

Using this definition, a 'heavy-tailed' distribution will have a positive value, and a 'light-tailed' distribution will have a negative value of kurtosis.

2.2.2 Fleet Data Availability

One of the enabling factors for this work was the access to a large store of BEV data from the automotive sector. The data collected and used throughout this section was recorded from BMW i3 vehicles operating in the field. Access to this data is required for monitoring fleet safety and alert to any issues which may affect the fleet as a whole. One requirement for the data collection process is that the customer's privacy must be protected at all times. This restriction causes some challenges when using the data for algorithm development, however, some insights into general driving trends can be gained. A quick overview of the total number of measurements from the available BMW i3 is provided in Fig. 2.4.

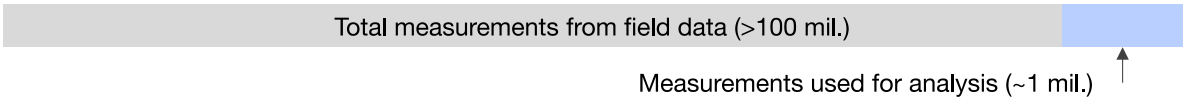

Figure 2.4: One million readouts from over 4000 BMW i3 60 Ah vehicles were used to analyze driving behavior in the automotive context.

The era of big-data means that over 100 million measurements were recorded from the total BMW i3 60 Ah fleet. Available fleet in this context refers to the accessible data used for monitoring purposes. In total, there are ca. 4000 unique vehicles available for analysis which have sufficient data quality and are not restricted due to privacy or other concerns. From these ca. 4000 vehicles, ca. 1,000,000 measurements are available from the total set of 100 million. A measurement is considered as a single readout from a vehicle which includes a package of relevant operation information which will be discussed in more detail shortly. On average there are ca. 260 measurements per vehicle as seen in Fig. 2.5A, with an average time between readouts of ca. 10 days as seen in Fig. 2.5B. In Fig. 2.6 the availability is shown also with respect to the global distribution of data. In the global map, it can be seen that a significant portion of vehicles are reporting from the US, with Europe as a whole providing over 1000 vehicles. The rest of the vehicles are distributed around the world, with grey shaded countries being excluded from this analysis due to low data availability.

A breakdown of the readout frequency, or the time between two measurements from a vehicle is shown in Figure 2.5B. Many of the measurements are recorded daily and 95 % of them occur within 30 days

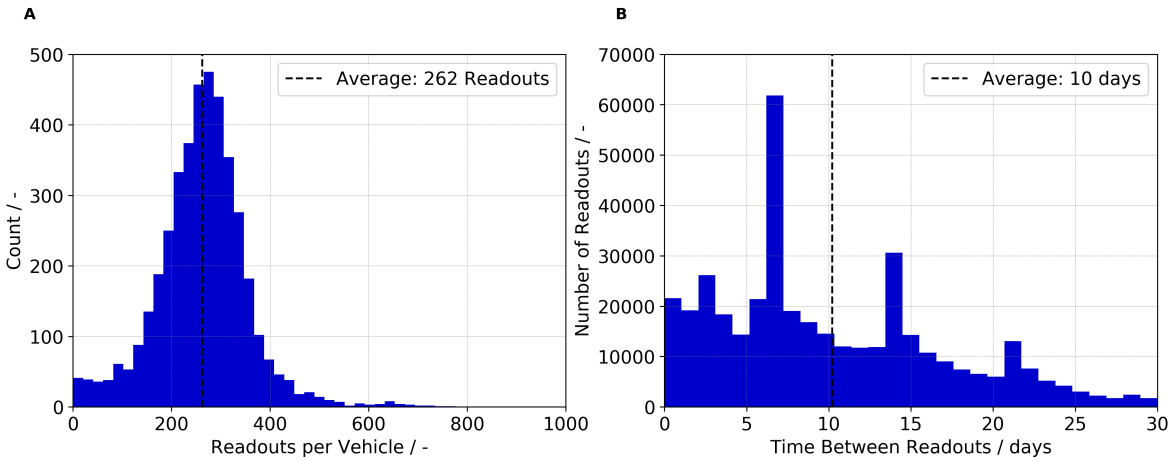

Figure 2.5: A) A distribution of the number of readouts per vehicle from the i3 data-set. B) Measurements are transmitted from the vehicles typically every ten days. A small portion of readings however are separated by over 30 days.

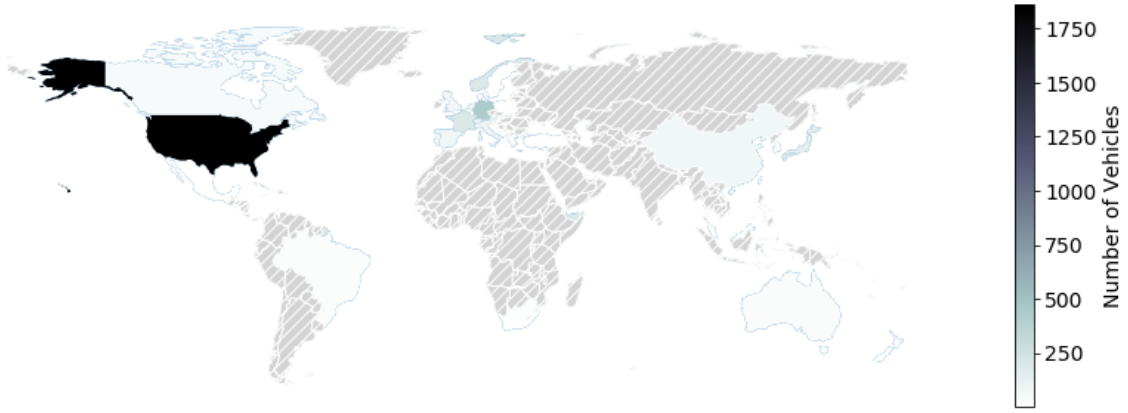

Figure 2.6: A global map shows the number of vehicles per country, and where the data is distributed.

of the previous measurement.

In order to put the data analysis in the next section into the proper context, it is important to understand what is recorded in a measurement. Table 2.1 is a reproduction of a small snippet from the collected data-set where the vehicle and measurement ID have been anonymized. From this table it can be seen that each measurement and each vehicle have a unique identifier which allows for the grouping of information for a specific vehicle, as shown in Fig. 2.7. The Measurement ID never repeats, whereas the Vehicle ID does. The Vehicle ID *JPA488C7*, found in Tab. 2.1 is repeated multiple times. Using the Vehicle ID, the history for a single vehicle can be analyzed and extracted from the larger data-set (for example see Fig. 2.7).

Data from the fleet is continually collected and updated. The data analyzed in this section was last updated in 2023 and as such contains measurements from over 8 years of BMW i3 data. As can be seen from the single vehicle in Fig. 2.7, measurements are collected regularly, however, there are also extended periods where no data is available for some vehicles.

A few of the variables recorded in the readout are included in Tab. 2.1 to explore a readout in more detail. One can see from Tab. 2.1 how this data can be used to understand the history of a vehicle.

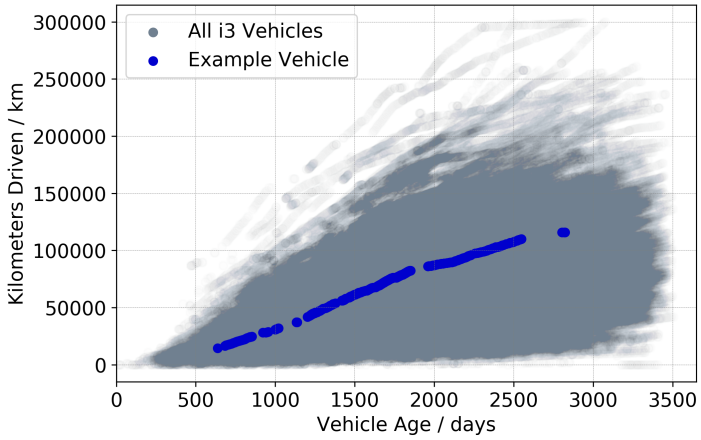

Figure 2.7: A scatter plot of the driven kilometers over the vehicle age from all measurements from every vehicle provides a cloud of information. A single vehicle is highlighted showing the history which can be rebuilt using the collected data. The single vehicle, and larger trend from the data cloud show a steady increase in the kilometers driven over vehicle age.

Tracking the *JPA488C7* vehicle, an increase in kilometers and decrease in SOH can be observed between the two measurements. As mentioned above, there are over 1 million measurements for the BMW i3 used for this analysis and 4000 unique vehicles. To help represent this massive cloud of data, Figure 2.7 plots each available measurement of kilometers driven over the vehicle age at the time of the readout. This representation is useful to see the range of values expected in the field data and highlight macro-trends in the data. Isolating a single vehicle, using the Vehicle ID, shows the collected historical data from a single vehicle (Fig. 2.7) showing how the history of a single vehicle can also be tracked.

Table 2.1: Snippet from the data-set showing key details from the measurements. Vehicle and measurement IDs have been anonymized.

Measurement ID	**Vehicle ID**	**Kilometers**	**Temperature / $°C$**	...	**State-of-Health / %**
BTGJSMAP	G2BB6LDS	42876	16.7	...	97.7
BMSP52WP	3FGRBLWT	23430	17.4	...	98.8
LLA4SDHF	JPA488C7	17403	16.3	...	99.3
...	...	...	...	...	...
57H4U7YQ	AVDVW2JN	92010	20.5	...	96.2
62ZVKJDK	JPA488C7	18309	16.8	...	99.3

As the data is collected from each vehicle, it is aggregated for the entire vehicle life, which is why the kilometers driven are continually increasing for the single vehicle in Fig. 2.7. This aggregation also occurs for other variables which are collected in time-based histograms, including temperature, charge and discharge C-rate and SOC. As is potentially best illustrated in the temperature plot in the subsequent section, Fig. 2.10, a cumulative average over all readouts flattens significantly the values as compared with taking the average between two readouts. For this reason, attention is given to the measurement presentation in this section, labeling values reflecting a *Cumulative* or *Readout Average* as such. Figure 2.8 is provided to show the readout difference over age when considering either the cumulative or readout average values.

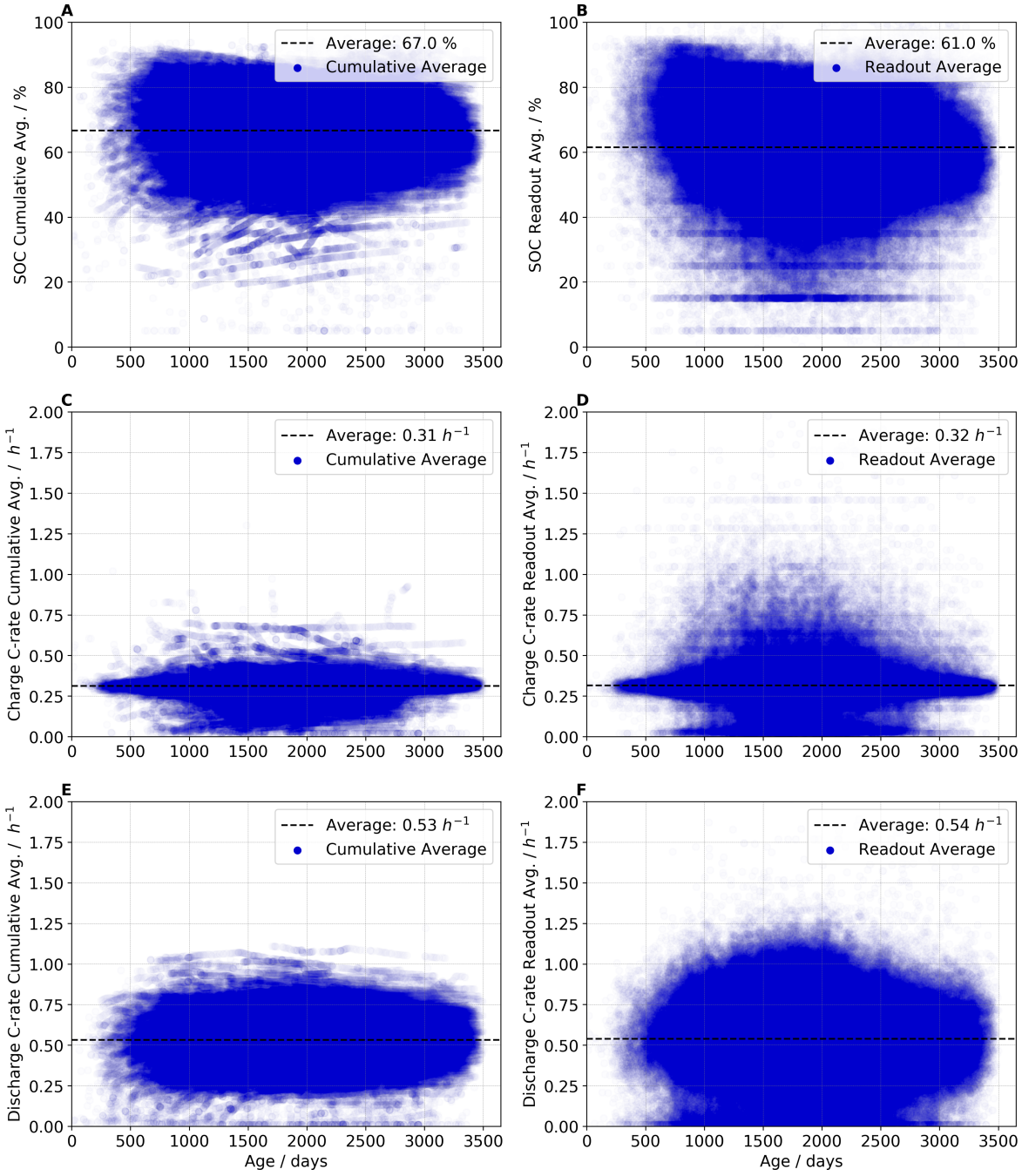

Figure 2.8: Time-aggregated histograms are used to calculate the average of a variable in two ways: the cumulative average representing the total average up until the last readout (A, C and E) and the average during only one readout period (B, D and F).

2.2.3 Operation Data

Batteries used in the automotive context experience a wide range of operation conditions. In addition to the environmental factors detailed above, factors such as driver aggressiveness, charging habits and driving frequency significantly influence the battery. In this section data reflecting the operation conditions of the vehicle is presented. Table 2.2 describes the measured variables in more detail. An accompanying histogram plot, Fig. 2.9, helps visualize the spread of some of the meta-data parameters

taken from the last readout per vehicle.

Table 2.2: An overview of the measured variables available for model development.

Variable Name	Units	Variable Description
Age	years	Each vehicle saves the date of production, marking the first time the high voltage system (HVS) is activated. This date is used as the beginning of life for the HVS. Using this date and the date of each readout, the current age of the HVS can be calculated.
Readout	date	Each measurement is readout with a corresponding time stamp including the day, month and year.
Kilometers	km	The kilometers are saved just as they are displayed on the vehicle odometer and represent the total accumulated number of driven kilometers.
Current Throughput	Ah	Two values are saved for current throughput, one corresponding to the flow in the positive direction, or charging of the battery, and the other in the negative direction, or discharging. The integral counters are registered in ampere-hours and when taken together represent the total current flow through the battery.
State-of-Health	%	The SOH is a ratio of the remaining battery capacity over the nominal capacity as estimated by the onboard BMS. The recorded SOH value represents the cumulative degradation the battery experienced up to the readout.
State-of-Charge	%	The SOC is a ratio of the remaining charge over the total available charge as a function of the SOH as estimated by the onboard BMS. The recorded SOC value represents the cumulative average SOC the vehicle has experienced up to the readout measured when the vehicle is on and off. The value is calculated by recording the amount of time the battery has experienced each SOC value in a histogram, and then calculating a time weighted average.
C-rate	h^{-1}	A normalized measure of charge and discharge current, C-rate, is calculated by dividing the cumulative average current by the nominal battery capacity. Both charging and discharging measurements are recorded only when the vehicle is in operation. This feature is originally collected as a histogram, where the values analyzed in this thesis represent the time weighted average of the histogram.
Temperature	$°C$	A temperature sensor records the cumulative time average over the life of the vehicle from multiple sensors located within the battery system. This temperature is very closely related to the environmental temperature the vehicle experiences during aging. This feature is originally collected as a histogram, where the values analyzed in this thesis represent the time weighted average of the histogram.

The histograms from Fig. 2.9 each contain 100 bins. The x-axis of the histogram shows the value of the corresponding variable and the y-axis the count of how often it was recorded in the last readout per vehicle data-set. In addition to the occurrence frequency of each value from a variable, the histogram also provides an indication on the distribution of the collected data. Two distributions stand apart from the more normally distributed variables of Fig. 2.9, the vehicle age (Fig. 2.9C) and the throughput per kilometer (Fig. 2.9D), with the age histogram having a significant right-ward skew, and the throughput per kilometer showing a left tail. As these readouts come from the the last readout per vehicle and reflect cumulative values, it makes sense that the age skews to higher values. The left tail seen in the throughput per kilometer plot, however, may indicate vehicles which either do not have sufficient data

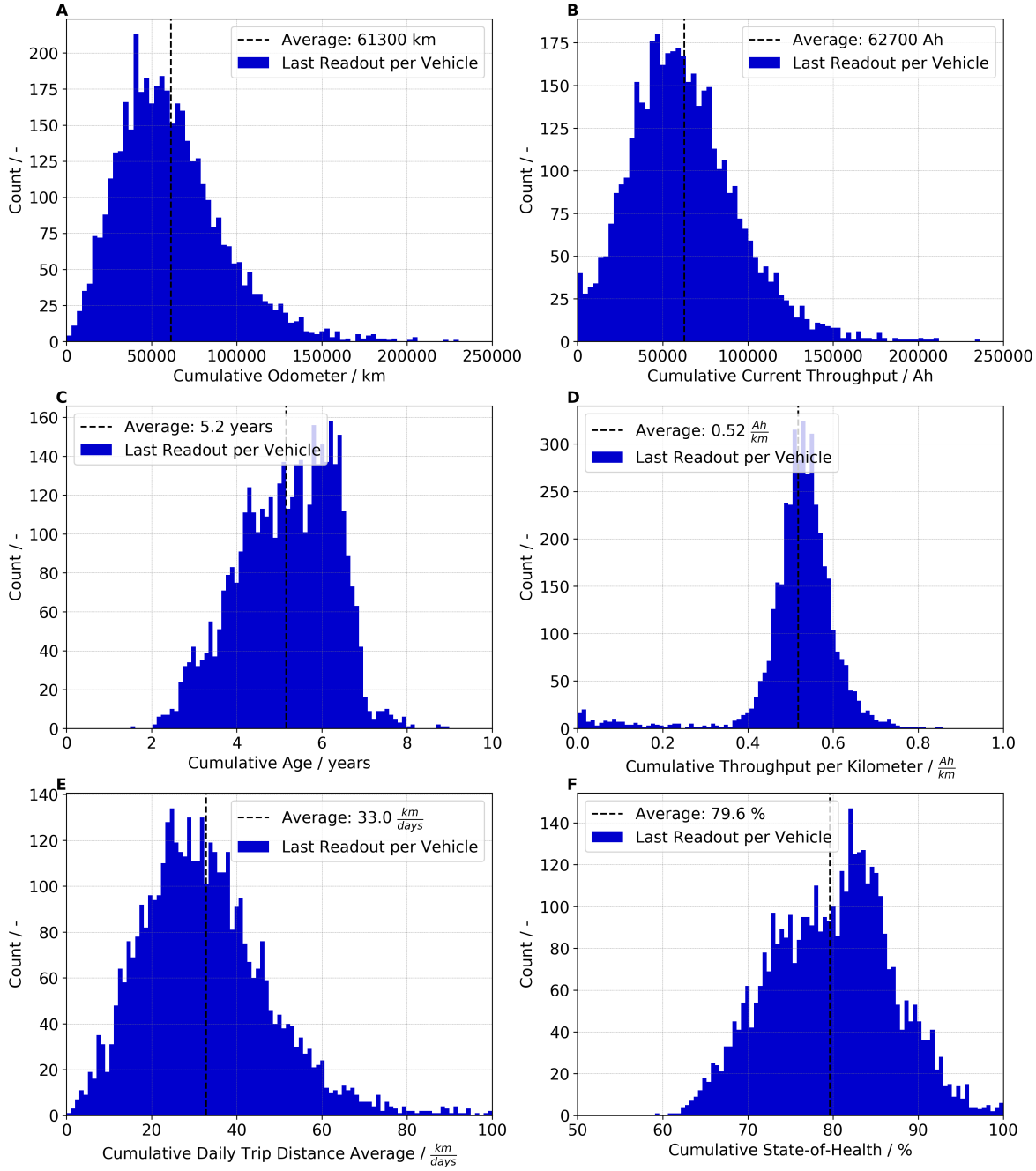

Figure 2.9: A histogram showing the distribution of various cumulative variables as read from the most recent readout per vehicles: A) kilometers driven, B) current throughput, C) vehicle age, D) throughput per kilometer, E) daily trip distance and F) state-of-health.

or are otherwise reporting outlier behavior.

One interesting trend shows that the total kilometers driven and the total current throughput do not have as significant of a skew towards the higher end of the range as seen in the vehicle age (Fig. 2.9A-C). There is, however, a tail in the distribution pulling the averages to the right from vehicles experiencing a significant amount of driving. There are few vehicles driving over 150,000 km and the average total kilometers driven is 61,000 km. According to two studies investigating driver behavior in the US and EU, the average yearly distance for drivers is between 18,000-21,000 km which would indicate an

expected average between 108,000-126,000 km at six years [130; 131]. In this data-set, however the average driven kilometers per year is only 12,000 km, which can be read from Fig. 2.9E by scaling the daily distance to a whole year. This lower than expected trip distance maybe explained by the early generation of EV having smaller batteries and early customers having range-anxiety, however, the i3 is also by its DNA a city-vehicle and therefore would not likely be used for longer trips [132].

The low kurtosis seen in the cumulative throughput per kilometer plot in Fig. 2.9D indicates the narrow range of efficiency from various users in the field. Looking to the tails, a user demanding 0.8 $\frac{\text{Ah}}{\text{km}}$ is likely driving with much higher C-rates on average (possibly a commuter with a high portion of highway) than one driving with 0.4 $\frac{\text{Ah}}{\text{km}}$, but that the count of occurrences is very low indicates that the i3 fleet maintains the 0.52 $\frac{\text{Ah}}{\text{km}}$ efficiency over a very broad range of use conditions.

The SOH is also an interesting plot (Fig. 2.9F) to compare with the distribution from age, as it could be expected that the SOH would have a mirrored left-ward skew to the skew seen in the age plot. The correlation of SOH and the other measured variables is presented in Chapter 4.

More user-specific variables, C-rate and SOC are presented in Fig. 2.8 with both their cumulative averages over lifetime and the averages as recorded between single readouts. This comparison helps to highlight how the time-aggregated histogram data could obfuscate more dynamic behavior, as the variability in the readout average plots are significantly higher compared with the cumulative averages.

The cumulative average SOC reflects the effective aging SOC. Similar to the cumulative average temperature, the cumulative average SOC settles to a more constant value over the life of the vehicle due to the nature of cumulative time averages. It is also possible to take the difference between two cumulative readouts and analyze the behavior between readouts. As shown in Fig. 2.10, the cumulative average temperature for a vehicle settles to an equilibrium, whereas the average per readout shows significant seasonal temperature variation.

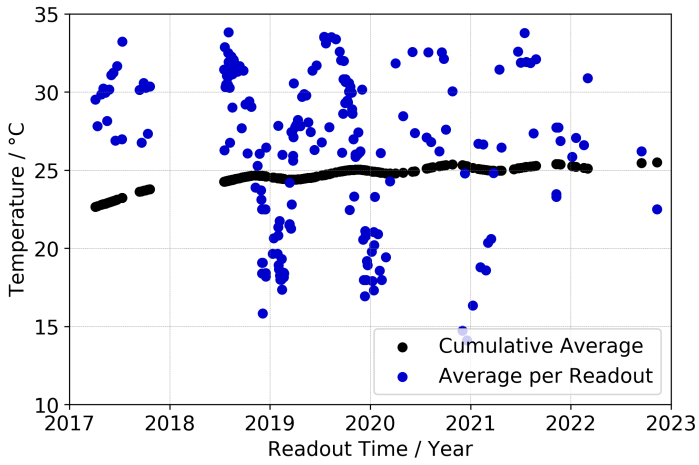

Figure 2.10: The cumulative average temperature and the average temperature per readout show the stabilization to a lifetime average temperature and seasonal variation, respectively.

The charge and discharge C-rate histograms (Fig. 2.8C and Fig. 2.8E) reflect the total lifetime average C-rate seen in the vehicle fleet. The charge C-rate includes regenerative charging during braking and is significantly more narrow with values centering around the mean 0.3 h^{-1}. This is likely due to the fact that the vehicles are most often charged with a level-2 charger as these are the most popular chargers for private use [133] and with an average power output around 7.2 kW, the 400 V, 60 Ah battery would

experience a C-rate of ca. $0.3\,h^{-1}$. The discharge C-rate is much more varied and shows the extreme differences in driver aggressiveness with a spread between $0\,h^{-1}$ and $1.25\,h^{-1}$.

Due to the fact that the average time between readouts is 10 days this operation data does not reflect the behavior over a single trip. At best, the average data provides a snapshot into the lifetime operation of a vehicle from the cumulative averages, and a slightly more detailed snapshot of the average behavior over a week when using only the averages per readout. The data-set provides significant insights into macro trends in the operation conditions for the i3. Where this data-set does not provide adequate resolution, however, is identifying or assessing specific events in the driving history, as these are lost due to the sampling frequency and collection method using the time histograms. There is, for example, no possibility of identifying how often the vehicle was above a specified C-rate and at a specified temperature as the histogram data does not combine these features.

2.3 Case Studies in Data Analysis

This section uses a few case studies to explore the information contained in the available data-set. As mentioned in the previous section, there are limitations to the data collected, however comparing different members from the data-set can offer some insights into larger trends. The methods developed in this section to analyze the data will help inform further insights into aging behavior which will be discussed in a later section.

2.3.1 Influence of Climate on Collected Data

One of the most influential factors affecting the health of a battery is the environment it is operated in. This data-set represents a global amalgamation including data from each continent which helps capture the extreme variety of possible environments experienced by BEVs.

From the list of possible environmental factors including: humidity, temperature and elevation– temperature is often the only condition considered in battery aging experiments [43; 134–136]. From the global map in Fig. 2.11 the difference in average temperature measured from the vehicles is 16 °C. This average temperature represents the total cumulative average temperature of the vehicle over the vehicle life. In Fig. 2.11, all vehicles in each country were grouped together and the average temperature from the most recent readout was calculated. The cumulative time average temperature represents a stable value after many years as the variation due to daily, weekly or seasonal swings are smoothed and as such, this value is a good indicator for the average aging temperature experienced by the battery, which is further analyzed in Section 2.3.1.

This case study selects two countries representing the two extremes of the data collected. As temperature plays an important role in the health and safety of the battery, it is critical to understand the environmental conditions the vehicles find themselves in. The average temperature per country is shown in Fig. 2.11 and Tab. 2.3 with the two extremes being Malaysia (including Singapore) and Norway. Spain was also included in the case study to represent a moderate country with a mediterranean climate which is different than the tropical climate of Malaysia or the nordic climate of Norway. Additionally, three groups were created out of vehicles with a cumulative average temperature of 14 °C, 21 °C and 27 °C. From this group the affect of temperature on the setting time of the cumulative average temperature can be investigated.

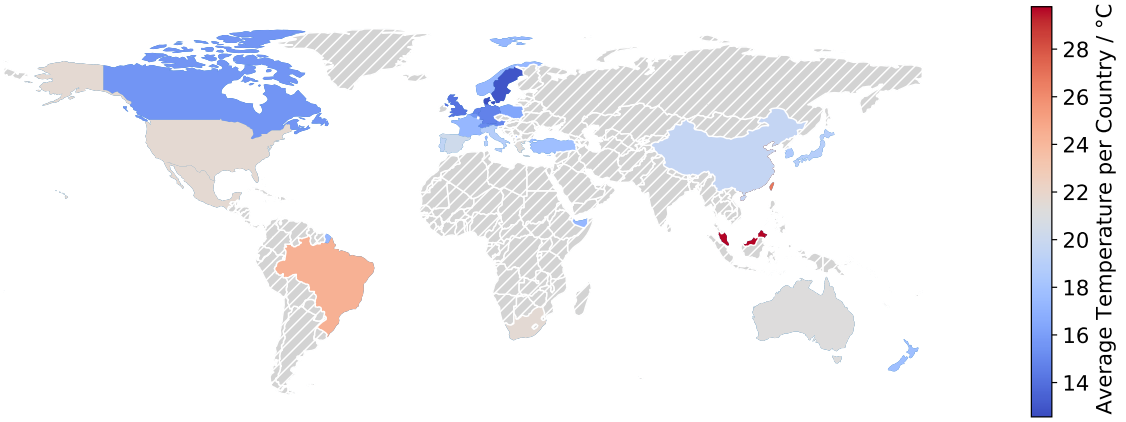

Figure 2.11: A temperature heat map shows the recorded average temperature from the vehicles grouped by country. This view also provides insight into the global distribution included in the data-set.

Table 2.3: Recorded temperature data from three countries used for the case study on temperature.

Group	**Nr. Readouts**	**Nr. Vehicles**	T_{avg} / **°C**	T_{min} / **°C**	T_{max} / **°C**
Malaysia	569	4	27.6	14.3	32.5
Spain	610	5	19.9	6.1	31.4
Norway	567	3	11.1	-7.3	26.8
$T_{avg=27°C}$	411	3	26.7	14.3	32.5
$T_{avg=21°C}$	1380	8	21.7	6.1	35.4
$T_{avg=14°C}$	1440	9	14.3	-2.8	30.4

From each of the three countries, 3-5 vehicles were randomly selected which had regular measurement frequency for multiple years to better show the seasonal dependency in those regions. In total there are 1746 measurements with an average of 145 measurements per vehicle. The readout period shows three full years of data from 2017-2020, capturing the periodicity of the seasons in the data.

The four vehicles from Malaysia have an average of 27.6 °C with a spread from 14.3-32.5 °C. A maximum temperature of 40.2 °C was recorded from the global data-set. Vehicles experiencing a more moderate mediterranean climate, represented by Spain, show an average temperature of 19.9 °C and fall within a range of 6.1-31.4 °C. The coldest average temperature was seen in Norway, with an average of 11.1 °C and low/high split from -7.3-26.8 °C.

In Fig. 2.12A the seasonal variation in temperature for the three countries is displayed. Using the average temperature per readout, the affect of these seasons can still be captured. Although the maximum temperature the vehicles experienced is lost due to the time histogram averaging, significant differences between these three groups can still be observed. The vehicles aging in a warm country experience a relatively constant average temperature with low fluctuations during the season. This average temperature is also located at the upper extremes of the collected data. In contrast, the vehicles from the warm country span the entire range of data. The cooler country of Norway also has more pronounced temperature extremes with a temperature difference of ±17 °C from the average. This swing is the largest, compared to ±12.5 °C for the warm country and only ±9.1 °C from the hot country.

The temperature swings from the seasons also impact the cumulative average temperature. To in-

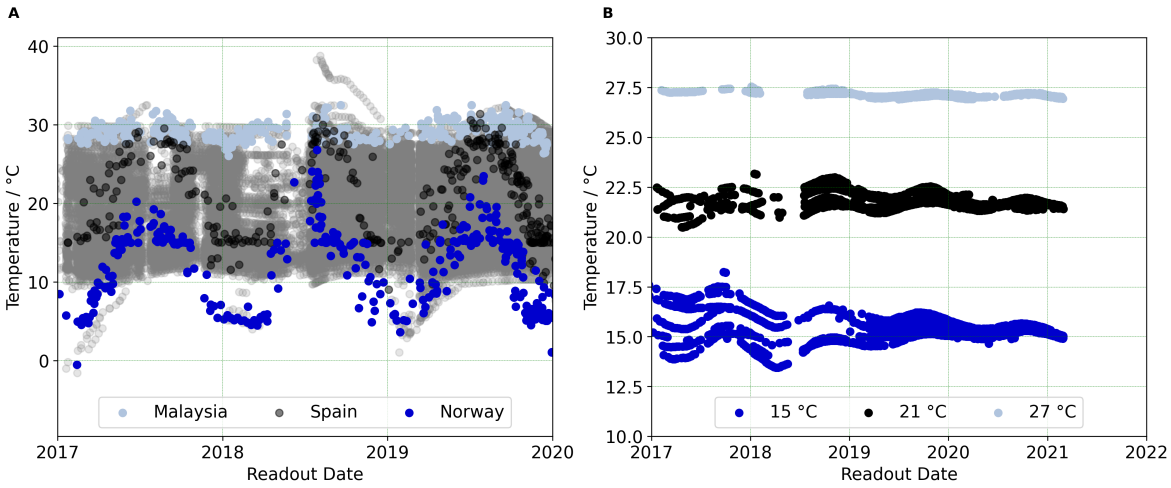

Figure 2.12: A) The average temperature per readout from vehicles in three example countries shows the effect of seasonal variability. B) Seasonal variability is smoothed-out using the cumulative average temperature value. The settling time for when this value remains within 0.5 °C depends on the average temperature experienced.

vestigate this effect, three groups at various cumulative average temperatures at the time of the last readout were collected. For example, in the group labeled as $T_{avg=27°C}$, only vehicles with a cumulative average temperature at the time of the last readout within 0.5 °C of 27 °C were collected. This grouping criterion resulted in around 1000 readouts per category from 15 vehicles. The highest temperature grouping has the smallest data-set with 411 readouts from three vehicles. As the behavior from this group does not deviate from what is seen using the average temperature per readout from the hot vehicle group discussed above, the relative small size of the data-set still helps to confirm the trend observed.

The cumulative average temperature smooths out the variation between readouts as this is a time-based average. It can be seen form Figure 2.12B that the higher temperature variability seen in the warmer and cooler climates leads to a longer settling time in the data. This settling time can be described as a damped sinusoidal with

$$T_s = \bar{T_c} e^{-\lambda t} cos(\omega t + \phi), \tag{2.7}$$

where the settling time, T_s is a function of the initial seasonal variation, T_c, the period ω and a phase shift ϕ. The seasonal variation, T_c can be taken from the country within which the vehicle is operating as this provides a good approximation for most vehicles. The ω and ϕ represent the seasonal swing and time of year the function is applied. The period for the settling time is one year, $\omega = 2\pi/365$, and assuming the vehicle data for fitting starts in the summer at the peak of the amplitude the phase shift is set to zero. Once the data is fit with the damped sinusoidal, a criterion for when the temperature is settled can be defined. After this period, the strong seasonal swings no longer pull the cumulative average outside of this criterion.

In general, the settling time increases with decreasing temperature. At one end of the spectrum, vehicles experiencing an average of 14 °C require up to four years before the cumulative average stays within 0.2 °C of the final settling temperature. For warmer climates, as apparent from the relative stability of the temperature, the average value can be taken after 2.5 years. Knowing when the value

reliably represents the lifetime average is useful for aging estimation, as this input is a main factor affecting the aging rate.

The temperature data collected from the time histograms is able to capture both seasonal effects and the effective lifetime aging temperature of the vehicle. There are significant differences between countries, mostly a result of the climate, reflecting the need to further break down the data from larger countries into specific climate zones.

2.3.2 Car Sharing in Germany

In this case study a group of 54 former DriveNow [137] i3 vehicles is available for further analysis. The available vehicles are located in Germany and as such a comparison between the i3 data collected from German vehicles and the DriveNow vehicles is performed. The DriveNow vehicles offer an interesting comparison to the other data collected in the field because the usage profile for on-demand mobility is assumed to be different than that from private ownership. The typical car-sharing customer is limited to driving within major urban centers as most mobility service providers do not operate in more rural areas. Another important distinction is that the vehicles are not privately owned and as such drivers may behave differently. DriveNow customers have totaled over 200 million miles driven [138], and in this study three million of those kilometers are analyzed.

One of the first steps to understanding the differences between these two groups is to look at the current conditions of the vehicles. Starting with the first two subplots from Fig. 2.13A-B it can be seen that the control group of German cars was selected to have similar age and temperature characteristics by filtering for all vehicles between two and six years of age within an average temperature range between 13 °C and 17 °C. Indeed the selected German cars experience very similar climates as the average temperature is only 0.2 °C apart. Where they differ slightly is in the average age, with the German subset being five months older on average. This is most likely due to the fact that the i3 fleet is older on average and there are significantly more vehicles in the German car data-set, containing 170 vehicles. The average total kilometers within each group is also similar with 60,700 km for the DriveNow vehicles and 52,300 for the German i3s.

The next comparison looks into the battery state parameters of SOC and discharge C-rate and can shed insight into the driving behavior. The DriveNow vehicles show one of the most significant differences in the average SOC with values 17 % lower compared to the German subset (Fig. 2.13C). It can be inferred from this difference that the DriveNow vehicles both experience higher DOD and are often parked at lower SOC leading to a lower average. The Charge C-rate showed no significant difference with a mean charging C-rate around $0.32\,\mathrm{h}^{-1}$ and for the German car and DriveNow data-sets, respectively, and was therefore not included in the figure. For the discharging C-rate, Fig. 2.13D, a larger difference was observed with the German car average of $0.59\,\mathrm{h}^{-1}$, and the DriveNow average of $0.47\,\mathrm{h}^{-1}$. It is supposed that the higher C-rates seen in the German car data-set comes from more frequent highway driving.

The final comparison which can be pulled from the data reflects the usage of the vehicles. As mentioned, the DriveNow subset is slightly younger on average, however, the kilometers driven are more. This is reflected in the average kilometers per day in Figure 2.13E. The consumption, shown in Fig. 2.13F and measured in Ampere-hours per kilometer (in the discharge only) is also higher in the DriveNow subset. These two plots indicate that the vehicles are driven more frequently and most likely experience higher inefficiencies by starting and stopping.

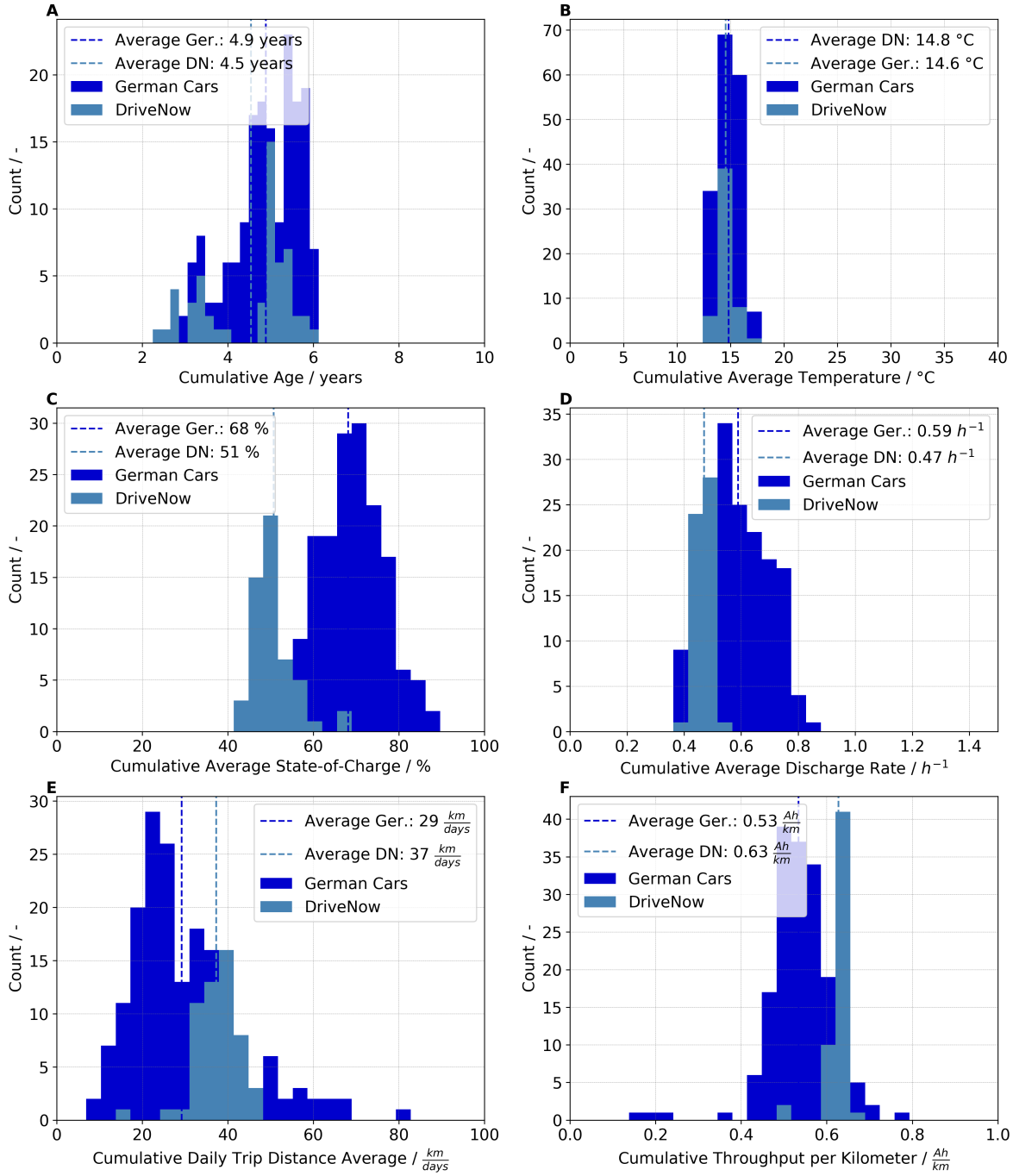

Figure 2.13: A comparison between shared vehicles and privately driven i3 vehicles shows differences in the collected data. A) cumulative age, B) cumulative average temperature, C) cumulative average state-of-charge, D) cumulative average discharge rate, E) cumulative average trip distance and F) cumulative throughput per kilometer.

From the case study it can be concluded that there is a noticeable difference in operation between the i3 DriveNow vehicles in Germany and the i3 vehicles not operating as part of a car-sharing service. The trends in kilometers per day and throughput per kilometer for the DriveNow vehicles would suggest a higher degradation rate and wear on the vehicles, however with lower average SOC, discharging C-rate and age, no significant difference in SOH is seen, with the average SOH of the two data-sets being 84 %

and 85 % for the German Cars and DriveNow, respectively (see Fig. 2.14).

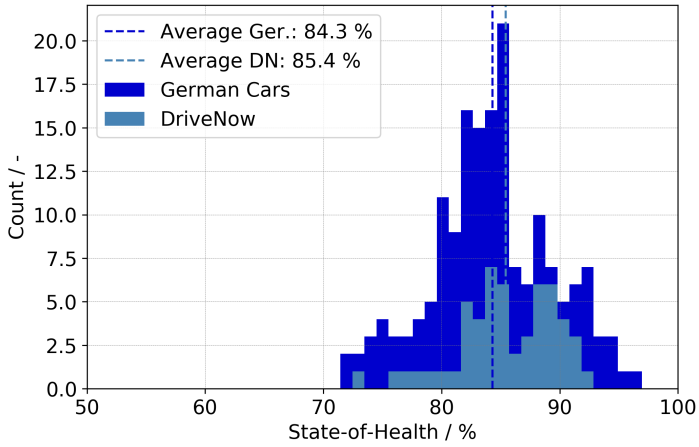

Figure 2.14: An SOH histogram for two distinct use-classes for the i3 show similar aging characteristics.

What this analysis does show, however, is that aging is a complex phenomenon that is a combination of multiple influences, from environment, driving behavior and use-frequency. By piecing the vehicle history data together, one can start to build an understanding for which types of clusters or aging conditions are most relevant in the automotive industry, and which factors have contributed most to the degradation behavior.

2.4 Path Dependence in the Automotive Context

An overview of the aging conditions seen in the automotive context is provided in Fig. 2.9 from Section 2.2. Table 2.4 reviews the statistical properties of the collected driving data and highlights the upper and lower bounds. From this analysis the average and the extreme driving behavior can be quantified. From the data available, the average driver of an i3 has been driving for five years, driven 61.300 km with an average SOC around 66 %, discharged with about 0.53 h^{-1} C-rate and charged with 0.31 h^{-1} C-rate.

Table 2.4: Statistical properties of collected vehicle data.

Variable	Units	Average	σ	Min	Max
Vehicle Age	years	5.2	1.1	1.5	8.9
Total Kilometers	km·10^{-3}	61	30	0.5	229
Trip Distance	km/day	32.7	15.2	0.3	139
Average SOC	%	66.8	15.2	19.5	90.4
Average Discharge C-Rate	h^{-1}	0.31	0.12	0.02	1.1
Average Charge C-Rate	h^{-1}	0.53	0.04	0.01	0.68

Due to the variability in real-world aging conditions, it is difficult to isolate specific aging paths or data-sets which differ only in one variable, i.e., vehicles driving the same distance, with the same C-rate and SOC but with different average temperatures. Due to this variability a controlled experiment using real-world conditions is extremely challenging. In lieu of this, a laboratory study was used to collect path dependent aging data, which is discussed in Chapter 3. Using the available data, however,

it was at least possible to identify significant changes in driving behavior using a changepoint detection algorithm.

In this section, a 'changepoint' detection algorithm from the R (programming language) package [139] is combined with electric vehicle driving data, to illustrate that dynamic aging conditions can be grouped into clusters of similar aging behaviors. This grouping could then be used to further inform and improve battery models or identify changes in behavior patterns which may indicate a deviation from prescribed operation conditions.

The data available for the changepoint analysis is described in detail in Section 2.2. For this analysis multiple years of vehicle data from the i3 was available and only vehicles which were at least 1 year old, had over 175,000 kilometers and over 100 readings were selected. The final sample size for this application was 300 vehicles. Three variables were analyzed: kilometers per day, discharge rate and temperature. The variable kilometers per day provides information on driving frequency and has a significant impact on the calendric and cyclic contribution to battery aging. The discharge rate is selected to showcase a variable relevant to battery safety systems, as well as, aging. Finally temperature was also analyzed as this is known to have a cyclic changing pattern dependent on the seasons.

This is best visualized in Fig. 2.15A where a significant jump in the average discharge rate can be seen in the measurement signal from one vehicle. A measurement index, seen as the x-axis in Figure 2.15 and Fig. 2.16 refers to one measurement period (typically 6-7 days).

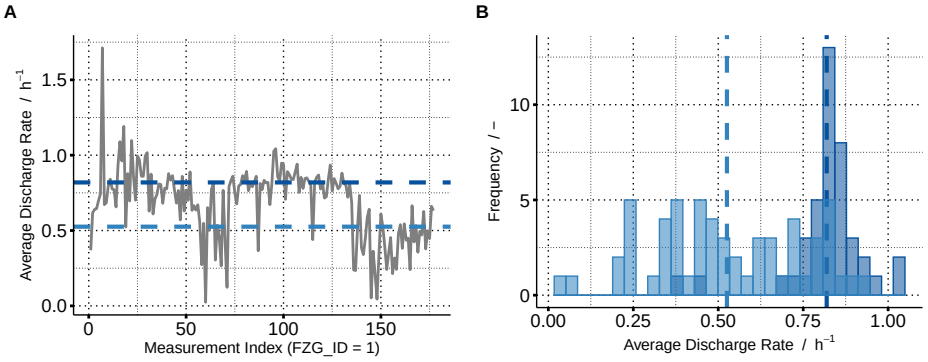

Figure 2.15: A) An example measurement for discharge C-rate from a representative vehicle. B) Two distinct histograms seen within the signal– an indicator for a changepoint.

As was shown in Sec. 2.2, the measurement data can be broken down into a distribution function, where $p(x = x|\mu, \Theta)$ is the probability density function consisting of two parameters: mean, μ, and standard deviation, σ. An example of a measured C-rate signal (see Fig. 2.15A) is shown to highlight a statistical variation in the data. In this figure a binomial distribution is clearly evident (also see Fig. 2.15B) representing two distinct probability functions with their own respective mean and standard deviation.

The changepoint algorithm was developed to identify these sections within a measurement signal. Given a time-series measurement $_{1:n} = (y_1, ..., y_n)$ a changepoint is said to occur if there exists a time period $\tau \in 1, ..., n-1$, such that the properties of two subsets $_1, ..., _\tau$ and $_{\tau+1}, ..., _n$ are significantly different.

It first achieves this by defining the test statistic using the likelihood ratio method [140]. This method requires the calculation of the maximum log-likelihood of a distribution, $\log p(y_1 : n)|\hat{\Theta}$, where $p(.)$ is

the probability density function associated with the distribution of the data and $\hat{\Theta}$ is the maximum likelihood estimate of the mean and standard deviation parameters. The maximum likelihood for a given time τ_1 with $\tau_1 \in 1, .., n-1$ is given by

$$ML(\tau_1) = logp(y_{1:\tau_1}\hat{\Theta}_1) + logp(y_{(\tau_1+1):n}\hat{\Theta}_2), \tag{2.8}$$

where the maximum over all possible changepoint locations is used to determine if a changepoint is greater than some threshold, c, using the test statistic

$$\lambda = 2[\max_{\tau_1} ML(\tau_1) - logp(y_{1:n}|\hat{\theta})], \tag{2.9}$$

where a changepoint is confirmed if $\lambda > c$. The selection of the threshold is used to tune the sensitivity of the algorithm to deviations in the data. Smaller thresholds result in more changepoints.

For detecting multiple changepoints a cost function, C, is introduced,

$$\Sigma_{i=1}^{m+1}[C(y_{(\tau_{i-1}+1):\tau_1})]\beta f(m), \tag{2.10}$$

where $\beta f(m)$ is a penalty to guard against overfitting. The changepoint algorithm implements multiple methods for solving this cost function which allows further user input such as the number of expected changepoints or the expected segment length [141; 142].

The application of the changepoint detection algorithm on the provided data found 20 vehicles with significant deviations in driving behavior. Within these 20 vehicles, an average of one distinct driving behavior change occurred every 3.2 years (excluding temperature). The results of the analysis can be found in Table 2.5. The average change in the means between two significantly different driving behaviors was 2 % for kilometers/day, 13 % for discharge rate and 12 % for temperature.

Table 2.5: Average number of change points and average change for relevant aging related variables during vehicle lifetime.

	Kilometers per Day	**Discharge Rate**	**Temperature**
Years per Changepoint	3.2	3.1	0.64
Average Changepoint	2 %	13 %	12 %

Figure 2.16 shows how the algorithm performed for one of the 20 vehicles. In this plot there are several changes in the discharge rate, one change in the kilometers per day and seasonal changes in temperature. Temperature changepoints were identified with a much higher frequency, as these correlate with seasonal differences more so than driving behavior. The average time between changepoints of 0.64 years for the temperature case reflects the diurnal seasonal cycle.

This method was found to be successful in detecting changes in discharge rate (Fig. 2.16C), average trip distance (Fig. 2.16D) and temperature (Fig. 2.16B) from real-world driving measurements. In some cases it was found that changes in behavior could deviate by up to 50 % of their averages (see Fig. 2.16D). The ability to detect these variations can help improve aging modeling performance, as well as, allow for potential fault detection use cases.

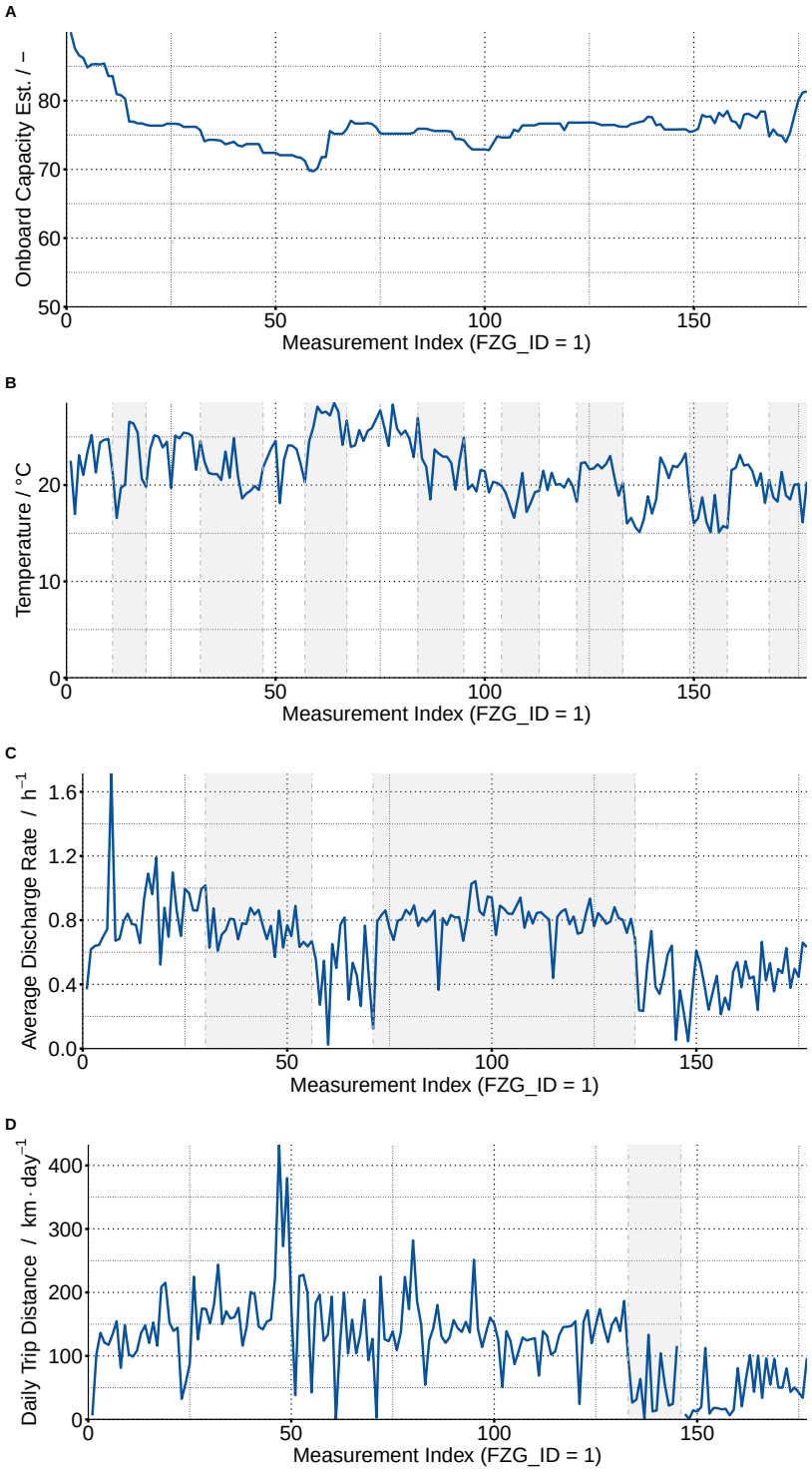

Figure 2.16: Example of a significant changes in an aging-related variable during driving operation. A) An estimation of the SOH, B) periodic changes in temperature likely related to seasonal fluctuations, C) two distinct discharge rates, with 50 % deviation from average and D) a change in trip distance.

3 Path Dependent Battery Aging During Race-Inspired Cycling Conditions

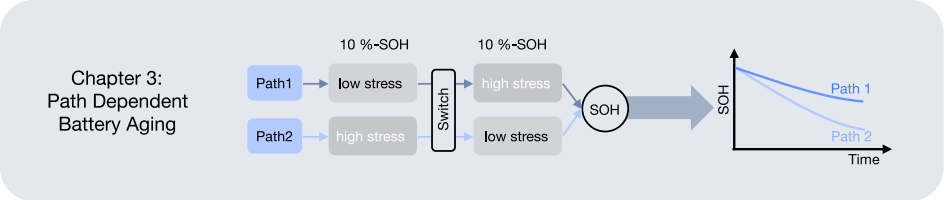

Figure 3.1: Crop from the thesis structure highlighting the focus of this chapter.

It is well evidenced that battery degradation is influenced by usage conditions resulting in various aging mechanisms which often present themselves in hierarchies and can even occur simultaneously [41; 43; 51; 143–145]. To which extent these hierarchical and simultaneous mechanisms are coupled with one another, however, is not well understood [41; 136; 146; 147]. In this work, an investigation into the degradation modes present during aging, and their potential influence on path dependent aging is investigated.

Path dependence refers to the time-dependent and sequential influence of aging conditions, and more specifically for this research, refers to the switching of a stress condition after an equal amount of cycles, or PD II as defined by Roeder et al. [148; 149]. The premise of this investigation is that one can not simply superimpose various degradation events because the order in which these events occur has a direct influence on the total degradation. This experiment is set-up to test this hypothesis, posing two complimentary aging paths against each other. By allowing the cells to experience the same aging events, only in different sequences, coupling between event sequences can be identified.

This research is of particular interest for holistic aging models which often assume that the end aging state of a battery is an addition of time spent or stress accumulated due to calendric or cyclic aging [40; 59; 150]. The validity of this assumption has a direct influence on the accuracy of the aging model. Further relevance of this study is that it breaks down the aging modes into main aging mechanisms, offering more insights into the reason for coupled aging. The finer resolution, compared to just calendric and cyclic aging modes, means that the impact the operation conditions have on the battery can be better distinguished.

In order to identify the dominant degradation modes, the relative change in the differential voltage curves can be analyzed using the differential voltage analysis (DVA) method [43; 151–157]. DVA requires a low constant current charge or discharge cycle (C-rate $<$ C/5), and plots the derivative of the voltage over the derivative of capacity on the y-axis and the capacity on the x-axis [52; 158]. By analyzing the shift of characteristic peaks and valleys resulting from this plot, the aging modes can be characterized. The study design is set-up such that the dominant degradation modes occurring within an interval can be quantified and compared to the aging mechanisms from the complimentary path. If there was no path dependence, for example, than the end DVA curve from both complimentary paths

would also look the same.

The profiles chosen for analysis were ones relevant to BEVs aging. Boundary use-cases were chosen where the identification of the dominant aging mechanism should be easier. These conditions included various combinations of high C-rates and high temperatures. The result was that some paths led to dominant LLI, such as the fast charging profile at lower temperatures, and some lead to more accelerated LAM, such as those with higher discharging C-rates. In particular, categorizing aging mechanisms into aging modes for the cycled NCA cells helped to highlight a potential path dependence between LLI and LAM [46; 51; 159–168].

Acknowledgement I would like to thank all contributors to this paper, without whom this would not have been possible. A breakdown of each author's individual contribution is listed as per the Contributor Role Taxonomy (CRediT) guideline [169].

- Author: Jacob Hamar, Credit: Conceptualization, Methodology, Software, Validation, Formal Analysis, Data Curation and Writing.
- Co-Author 1: Magdalena Stuckenberger,
  Credit: Data Curation, Formal Analysis, Software, Writing- Original Draft.
- Co-Author 2: Johannes Sturm,
  Credit: Data Curation, Software.
- Co-Author 3: Julius Schmitt,
  Credit: Formal Analysis, Software, Writing- Original Draft.
- Co-Author 4: Marcel Rogge,
  Credit: Data Curation, Formal Analysis, Software, Writing- Original Draft.
- Co-Author 5: Simon V. Erhard,
  Credit: Conceptualization, Funding acquisition, Writing- Review and Editing, Supervision.
- Co-Author 6: Andreas Jossen,
  Credit: Conceptualization, Project Administration, Resources, Supervision.

Path Dependence Battery Aging During Race-Inspired Cycling Conditions

Jacob Hamar, Magdalena Stuckenberger, Johannes Sturm, Julius Schmitt, Marcel Rogge, Simon V. Erhard and Andreas Jossen

Accepted by the Journal of The Electrochemical Society, February 14, 2024.

Permanent Weblink:
https://iopscience.iop.org/article/10.1149/1945-7111/ad2952

Journal of The Electrochemical Society, 2024 **171** 020555

Investigating the Path Dependent Aging Behavior of Nickel Cobalt Aluminum Oxide Cathode Batteries During High C-rate Cycling Conditions

Jacob C. Hamar,[1,2,z] Magdalena Stuckenberger,[2] Johannes Sturm,[2] Julius Schmitt,[2] Marcel Rogge,[1] Simon V. Erhard,[2] and Andreas Jossen[1]

[1]*Technical University of Munich, School of Engineering and Design, Department of Energy and Process Engineering, Munich, Bayern 80333, Germany*
[2]*BMW Group, Munich, Bayern 80788, Germany*

Nickel-rich cathode material batteries such as the nickel cobalt aluminum oxide with graphite anode are common in the automotive sector due to their high energy density and power performance capabilities, however, they are limited in lifetime performance. Accelerated degradation in specific operation conditions, such as high C-rates are known to significantly contribute to this issue. To better understand a chronological dependence of these high-stress conditions occurring during battery lifetime, a path dependent aging study is performed using profiles inspired by high-performance race operation. The focus on C-rates >1C during charging and discharging using real-world inspired profiles- known to induce particle cracking or lithium plating- addresses the need to understand aging behavior in less studied, but highly relevant high-performance automotive context. The aging conditions were tailored to induce specific degradation modes to better identify potential degradation mode coupling. Using differential voltage and capacity analysis techniques, degradation modes were identified which likely contributed to path dependent aging behavior, including loss of lithium inventory and loss of active material in the positive electrode. Two critical paths were identified in this study where later high C-rate conditions in the form of fast-charging at temperatures of 25 °C resulted in accelerated aging and where loss of lithium inventory induced stresses in the positive electrode, accelerating loss of active material.
 [DOI: 10.1149/1945-7111/ad2952]

Manuscript submitted October 7, 2023; revised manuscript received January 26, 2024. Published February 28, 2024.

Understanding battery degradation is essential for the commercial success of battery electric vehicles (BEV) as accurate degradation state estimation improves the reliable operation, range and power availability in the vehicle. In the automotive context, there is a significant challenge posed by the myriad of dynamic aging conditions resulting from the unique demands required by the BEV operators.[1] By investigating these degradation pathways and understanding the coupling of various degradation modes, a more complete and accurate picture of the vehicle capacity based state-of-health (SOH) and internal resistance (IR) is expected.

A degradation path is defined as the usage schedule or specific sequence of operation conditions leading to an observed degradation mode.[2] Much work has been done to define the degradation modes into two main categories: loss of lithium inventory (LLI) and loss of active material (LAM) using incremental capacity analysis (ICA) and differential voltage analysis (DVA).[3–5] The categorization helps to identify the underlying aging mechanisms present during the aging process,[6,7] and has been used effectively to identify degradation modes under dynamic conditions.[8,9] Less work, however, has been done investigating the chronological dependence of the degradation paths.[10]

In this work, the path dependence of aging is investigated. Using the definition from Roeder et al. path dependence refers to a state change depending on the sequence of conditions within a use schedule.[10] An example of this comes from Bauer et al. who systematically switched the temperature path experienced by a battery during aging resulting in a significantly different power and capacity fade at the end of testing, but they stopped short of concluding anything about path dependence.[11] One of the first papers found mentioning path dependence comes from Gering et al. who investigated aging using profiles with the same average power, start state-of-charge (SOC) and average temperature, to see if there would be a dependence on the exact profile used during cycling.[12] A follow-up study was conducted by Dubarry et al. and a similar study investigating the temperature influence on cyclic aging from Werner et al. found that despite the same total average temperature the cumulative aging varied significantly between temperature paths.[2,13]

For calendric aging, many studies have found little to no contribution of a path dependent aging effect. Two similar studies modeling calendric aging under dynamic conditions with alternating temperature and SOC conditions concluded, that since a model fit on static conditions could be superimposed and provide an accurate estimate for the dynamic conditions, path dependent affects could be neglected.[14,15] However, in a follow-up study by Sarasketa-Zabala et al. under cycling conditions it was found that alternating use schedules led to significant deviations in overall degradation.[16] Su et al.[17] provided one exception, where an internal resistance difference was identified with varying SOC conditions during calendric aging positing that a change in the diffusion process could have resulted in path dependent aging behavior.

Under cyclic aging conditions, more evidence has been found for a path dependence. Dubarry et al. observed aging behavior under various duty cycles inspired by grid operation and concluded that high C-rates resulted in path dependent aging.[2] They used ICA to conclude that high loss of active material of the negative electrode (LAMne) degraded the kinetics of the negative electrode (NE) and could have lead the an aggravated LLI. More recently, Raj et al. and Karger et al. have suggested a path dependence exists under higher C-rate cycling[18,19] with the former also using DVA and ICA to confirm that LAMne was dominant in the path dependent aging condition, possibly as a result of particle cracking and loss of electrical contact.

The nickel cobalt aluminum oxide (NCA) active material is highly relevant for high-performance BEV, recently achieving 300 $Wh \cdot kg^{-1}$.[20] As this study uses 21 700 NCA cells with a silicon-graphite anode, it is important to briefly review the main degradation mechanisms present during aging. It is well documented that one of the dominating degradation mechanisms of NCA/graphite cells is LLI due to side reactions at the anode.[20,21] Particle cracking is also a significant cause for capacity and power

[z]E-mail: jacob.hamar@tum.de

Journal of The Electrochemical Society, 2024 **171** 020555

fade.[22,23] This phenomenon occurs in both the NE and positive electrode (PE) and is strongly influenced by the depth-of-discharge (DOD) and SOC ranges during cycling.[24] Aging effects on NCA/graphite cells when cycled at high DOD are related to surface phase transitions of NCA particles and micro-crack formation in their surface which increase the charge transfer resistance and reduces the overall cell performance.[25] The main calendric aging contributions are assumed to be side reactions in the graphite anode.[26,27] The addition of silicon in the graphite helps to deliver a higher energy density cell, however, it is known to also induce high volumetric changes during cycling resulting in a loss of electrical contact with the active material and increased SEI formation.[28] This affect can be mitigated by narrowing the potential range during cycling and using blended silicon and graphite electrodes.[28,29]

In this work, an investigation of the path dependence induced by various degradation modes of 21 700 NCA cells cycled under race-inspired high-performance conditions, including fast charging, is performed. Using ICA and DVA techniques degradation paths are identified and analyzed for their potential effect on path dependent aging behavior. This work adds to the existing literature by targeting specific degradation modes during cycling: LAMne and loss of active material of the positive electrode (LAMpe) at high SOC and C-rates and LLI at high temperatures and fast charge (FC) at lower temperatures- then analyzing the degradation modes for potential path dependent influences. If critical paths are identified, that information is vital for improving SOH and IR estimation, by providing more insights into the interaction of degradation mechanisms, and also optimizing vehicle operation, by avoiding conditions which may accelerate aging due to a negative feedback of degradation mechanism.

Experimental Procedure

To further investigate path dependent aging behavior, a laboratory aging experiment using cylindrical 21 700 type silicon-graphite/NCA was performed. The aging conditions were inspired by high-performance race operation scenarios of electric vehicles. A confirmation of path dependence for this study refers to a difference in the end aging state between two paths within an aging condition. A systematic alternating of aging allows for an investigation of the sequence of aging, and whether the order of the stress conditions affects the overall aging of the cell.

The cell used in this study has a rated capacity of 4.8 Ah, a rated internal resistance of 27.5 mΩ (at 50% SOC and 25 °C) and a nominal voltage of 3.6 V. The maximum allowed voltage and temperature are 4.15 V and 60 °C. The minimum allowed voltage and temperature are 2.5 V and −20 °C. A plot of the half cell voltages is provided in Fig. 1.

A total of 24 cells were cycled with four cells per aging condition and two cells per path, providing a modicum of redundancy. From an initial batch of 60 cells, groups of cells were selected with the most similar capacity and internal resistance values. Within the 60-cell batch, the average capacity was 4.72 Ah ± 0.01 Ah and an average internal resistance of 22.5 mΩ ± 0.36 mΩ. Six aging conditions were considered. An example of the path set-up and cell distribution within one aging condition is shown in Fig. 2. One aging condition contains two complimentary paths, which in turn experience the same two stress conditions. Initially, one path starts at a low stress factor, while the other path starts at a high stress factor. The cells are aged until one cell reaches about 90% SOH. Once it is ensured that the cells are cycled the same amount of equivalent full cycles (EFC) at their starting stress factor by cycling any remaining cycled capacity difference from the lagging path, the stress factor is switched. Thus the cells originally experiencing the low aging cycle are now cycled under the high stress factor and vice-versa. Cycling was continued until each cell achieved the same amount of EFC as was undergone during the first half of aging.

The cycling was carried out using BaSyTec CTS cyclers and data acquisition software. Cycling proceeded in 50 cycle increments, where after each increment a check-up was performed. Since various stress factors led to various amounts of EFC, some cells received more check-ups than others. The check-up included a capacity check with two C/5 full discharge and charge cycles with a C/20 cutoff CV phase between the voltage limits 2.5 V and 4.15 V. A cells' SOH was calculated from the average of the total charge throughput (Ah) from the two complete discharge cycles divided by the initial capacity of each cell. The IR was calculated by discharging the cell to 50% SOC and applying a 10 s pulse with C/2 current in the discharge direction, and using the change in voltage over the change in current during the pulse as the IR value. This is the resistance value referenced in the subsequent discussion and figures. Using the BaSyTeC data acquisition software, it was possible to record the voltage, current, temperature and other relevant parameters for further analysis. It is through the collection of this data that the total capacity throughput was monitored and kept constant between the two groups of cells after a switch in the aging path.

The DVA measurements were performed using a C/5 discharge and for Paths 7–12 (defined in the next section) an additional C/20 discharge was performed to allow for a more robust DVA analysis. Due to a restriction in the available resources, baseline aging studies could only be performed for three of the stress factors under non-path depending aging conditions. These three baselines are also discussed below.

Profile Definition

The profiles were selected based on the goal of emulating real automotive operation conditions inducing a variety of degradation modes. This was achieved by the definition of the stress factor. Within this study, Path 1/2 investigated high C-rates, Paths 3/4 investigated temperature variation and Paths 5/6 investigated frequency of high C-rates. Paths 7–12 were added after the initial study started to further explore the effect of high discharge C-rates.

Consequently, the six aging conditions with their pair of stress factors are (see Fig. 3):

- **Path 1/2** [C-rate]: the maximum C-rate applied to the cell was alternated between a 1C/1C discharge/charge profile and the highest C-rate profile, the high-performance profile II (HPII) profile.
- **Paths 3/4** [Temperature]: using the high-performance profile I (HPI) profile, the two groups of cells were switched between cycling at 25 °C and 40 °C.
- **Paths 5/6** [Frequency]: cells were alternated between the low frequency charger (LFC) and high frequency charger (HFC) mimicking a change in racing frequency.
- **Paths 7/8, Paths 9/10 and Paths 11/12** [Discharge C-rate]: These paths isolate the effect of discharging C-rate on path dependence by considering three switching cases combining the 1C/1C profile with a 1C/FC (Paths 7/8), 2C/FC (Paths 9/10), and 4C/FC (Paths 11/12), respectively.

The two racing profiles, HPI and HPII, were defined to provide a realistic and extreme discharging condition derived from a measurement trace from a driven Nuerburgring (NBR) lap (Fig. 4a). Breaking down the NBR trace into a binned histogram (see Fig. 4b), where the normalized frequency shows the amount of time spent in each bin with the total adding to one, it is evident that a racing profile is similar to a duty cycle where the majority of the operation time is spent at maximum discharge and maximum recuperation. This duty cycle operation inspired two profiles, HPI and HPII to be used as aggressive C-rate aging profiles where the maximum allowed charge and discharge currents were pulsed for discharge and recuperation.

The HPII profile maximizes the time spent at the max allowed discharge/charge C-rate by adhering to the pulse limits and temperature limits of the cell. This results in a duty cycle discharge with a maximum discharge pulse for 30 s, followed by a 10 s rest period and another 10 s recuperation at max charge current, as can be seen in Fig. 3c. The peak

Journal of The Electrochemical Society, 2024 **171** 020555

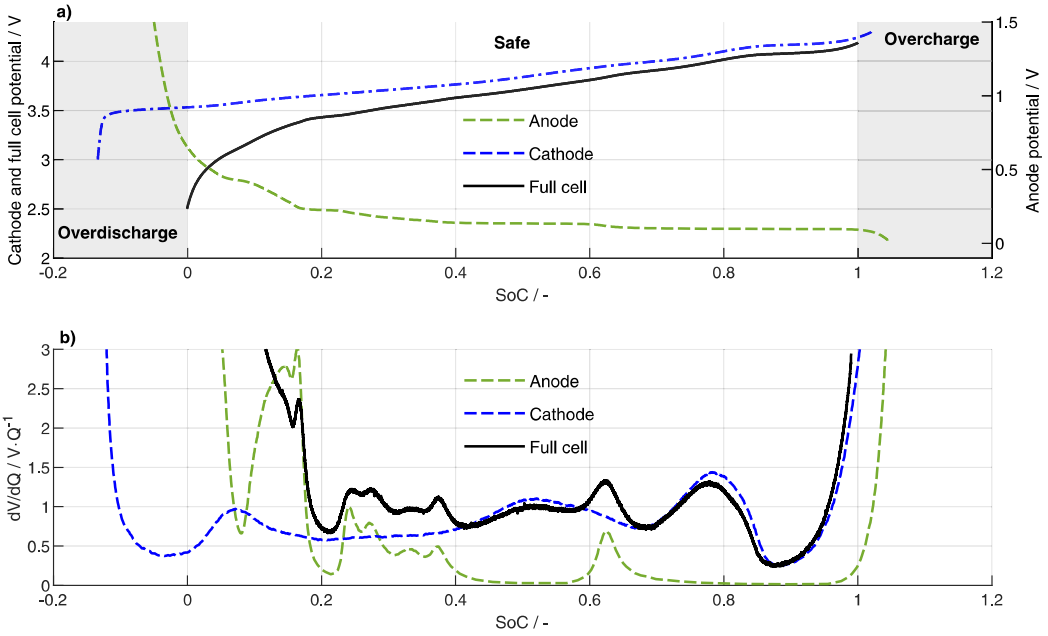

Figure 1. (a) The cathode half cell potential and full cell potential are scaled on the left *y*-axis with the anode half cell potential on the right *y*-axis. (b) A dV/dQ plot shows the differential voltage of each half-cell, as well as the combined full cell dV/dQ.

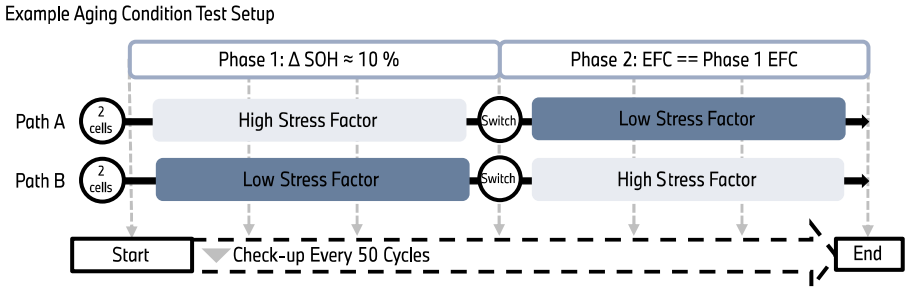

Figure 2. For each of the three path dependent conditions, two levels of severity were defined: low stress factor and high stress factor. Two independent groups of two cells were started at one of the two stress factors and after ca. 10% SOH the two cells were switched to the other stress factor.

discharge and charge C-rate were 4C and 1C respectively. During cycling the maximum temperature reached during discharging was 45 °C, or a temperature increase of ca. 15 °C. A 1C was used for the recuperation current as this was the highest allowed current, also at higher SOC. The HPI profile follows the same principle but with a reduced maximum discharge C-rate of 3C, resulting in a somewhat less aggressive profile compared to HPII (Fig. 3b). The maximum temperature reached during HPI discharging was 48 °C, or a temperature increase of ca. 18 °C. These duty cycles were repeated until the cell discharged 53%-DOD based on capacity throughput. The 53%-DOD was used to ensure that the cell did not reach the operation voltage limits during the cycle which would require a constant voltage phase to maintain same capacity throughput between the profiles. Due to the lower C-rate, the HPI profile has more pulses during the discharge phase to achieve the same total capacity throughput. Also due to the shorter discharge times, the average cycling temperature was very similar between the HPI (46.2 °C) and HPII (46.9 °C) profiles with a difference of less than 1 °C, which was ca. 7 °C higher than the average cycling temperature for the baseline 1C/1C profile.

Each high-performance profile was accompanied by a FC profile shown in Fig. 3a. The FC procedure was provided by the cell manufacturer and follows a multi-step constant current profile with a maximum charging C-rate of 4C and was charged to the same starting voltage using a constant voltage phase. The FC profile resulted in a temperature increase of 13 °C.

In order to investigate the effect of high stress cycling frequency on aging, two additional profiles were defined: HFC and LFC, shown in Figs. 3e and 3f, respectively. For the LFC, a ratio of one HPII profile per nine 1C/1C cycles was used. The 1C/1C is also shown in Fig. 3d. For the HFC, a more aggressive ratio of three HPII profiles per seven 1C/1C cycles profiles was used.

Baseline Test Results

Initial insight into capacity loss from the three stress factors used throughout Paths 1–6 are provided by the SOH over EFC plot in Fig. 5. The 1C/1C profile ages the most significantly after 500 EFC. Despite the lower C-rates, it seems that the 100% DOD may have lead to accelerated aging as seen by Watanabe et al.[22] who also concluded that cell performance over lifetime was maintained when the DOD was restricted to less than 60% due to reduced formation of micro-cracks in the active material.

The HPI profile maintains fairly consistent aging regardless of the set climate chamber temperature. It should be noted, however,

Journal of The Electrochemical Society, 2024 **171** 020555

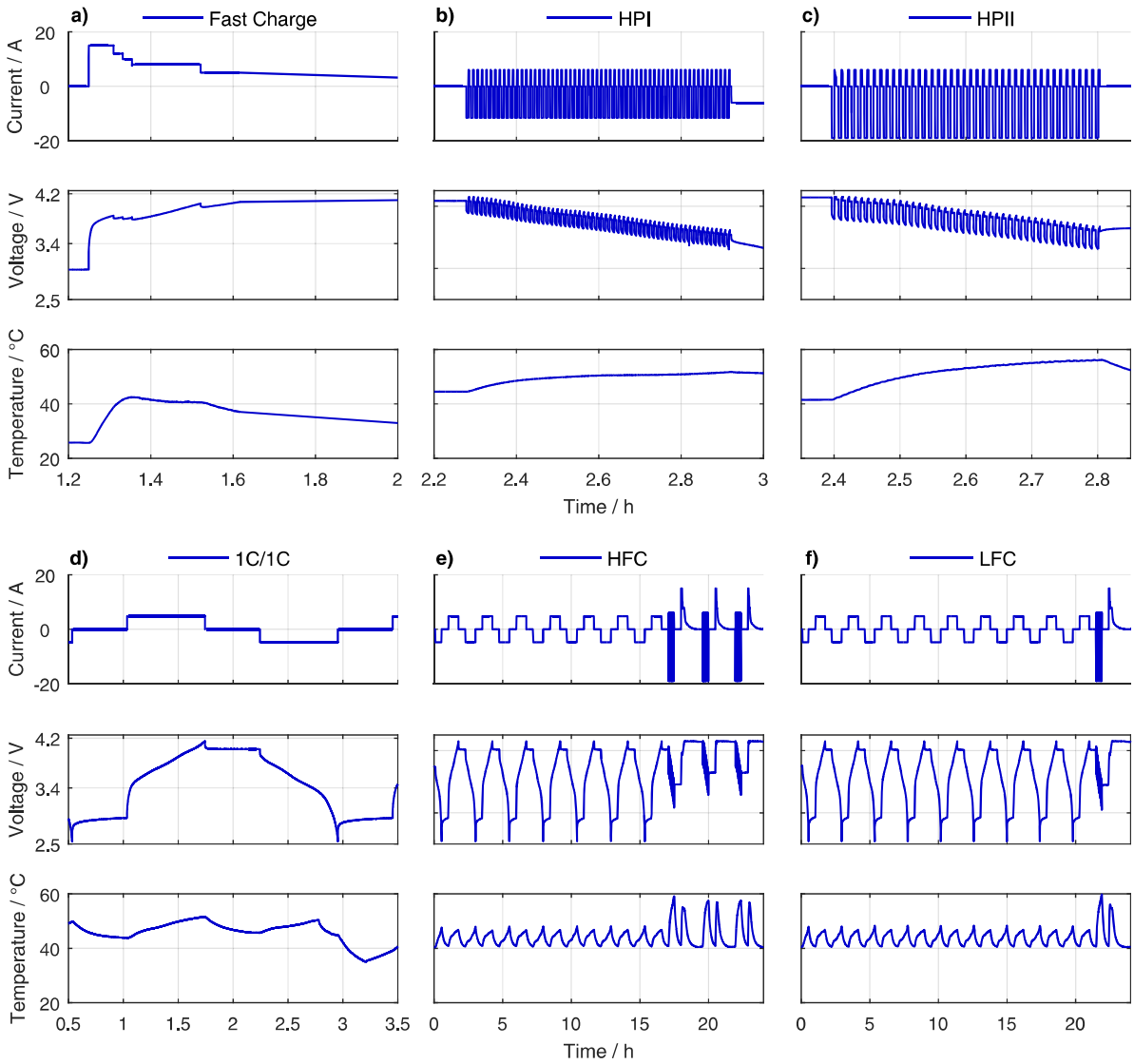

Figure 3. Examples of the voltage, temperature and current during each of the cycling profiles including: (a) fast charge (FC), (b) high-performance I (HPI), (c) high-performance profile II (HPII), (d) 1C/1C charging and discharging, (e) high frequency charger (HFC) and (f) low frequency charger (LFC).

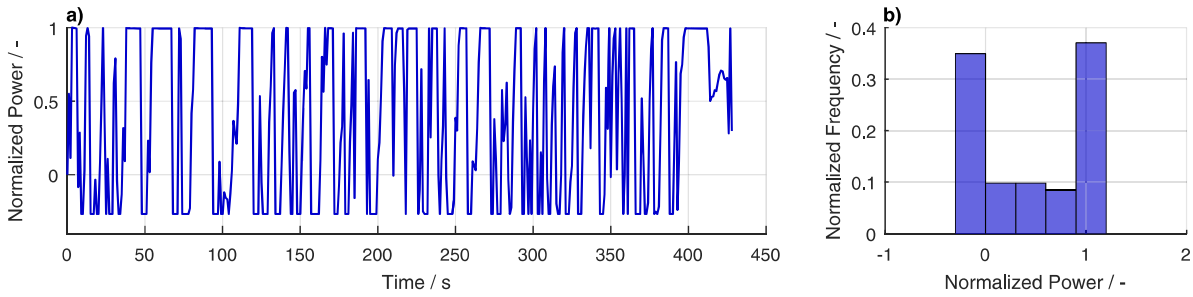

Figure 4. (a) A normalized power profile for a driven Nuerburgring (NBR) racing cycle which inspired the HPI and HPII profiles. (b) A histogram of the normalized power from the NBR profile.

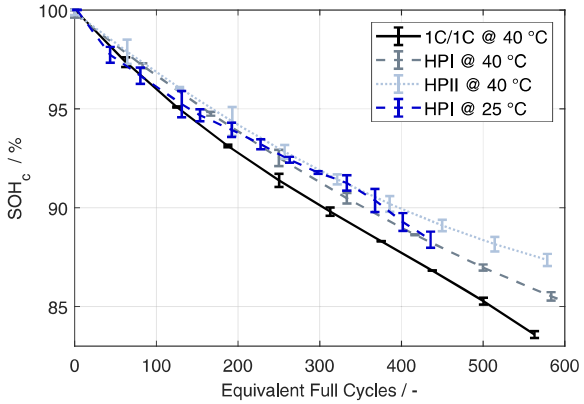

Figure 5. Three baseline aging profiles (1C/1C, HPI and HPII) at 40 °C and the HPI at 25 °C are shown.

that one from three cells used rapidly lost capacity after only 300 EFC at 25 °C and was therefore excluded. There was, in general, more variability in the cells running FC at 25 °C and it is assumed that lithium plating affected performance in these conditions.

It was expected that the higher discharge C-rates used in HPII compared to HPI would have lead to accelerated aging, however, as evidenced by Fig. 5, this is not the case. One explanation could be that the cycle depth is effectively lowered by the higher amount of regeneration pulses.[30] The difference is marginal between the various profiles, with a about 2% SOH after 500 EFC. The baseline profiles of 1C/1C, HPI and HPII at 40 °C have an average absolute error between the three cycling cells of: 0.25, 0.31 and 0.73%, and the HPI at 25 °C has an average absolute error or 0.34%.

Path Dependence Capacity Loss

The SOH and IR plots for all six of the aging conditions are analyzed in this section. Out of the six testing conditions, three seem to exhibit some form of path dependence. From the degradation plots for the first three conditions shown in Fig. 6, the C-rate condition, in Fig. 6a), shows a difference in the total accumulated degradation and the temperature condition, Fig. 6b, shows a significant departure between the Path 3 and Path 4 in both the SOH and IR. The test condition switching the relative frequency of high-performance cycling, Fig. 6c, did not seem to lead to a path dependence. The lower capacity fade, but higher resistance increase in Path 1, Fig. 6a, may be an early indicator that LAM is more prevalent as this is a strong indicator for resistance increase.[6]

The ΔSOH and ΔIR between check-ups are shown as box plots before and after the switch condition for each of the paths in Fig. 6. The racing frequency condition demonstrates the expected behavior in these plots if no path dependence is evident (Figs. 6c.i–6c.ii). In this case, aging rate of the more aggressive stress factor before the switch flips positions to have a lower rate after the switch leading to Path 5 and Path 6 to come together in the end. It can be seen that this flip does not happen in the cases where a path dependence is evidenced. In Path 1 and 2, this flip does not occur, and instead Path 1 After and Path 2 After maintain the same average rate of capacity fade. A similar trend is seen after the switch in the temperature condition in Path 3 and Path 4 (Figs. 6b.i–6b.ii) where Path 3 always has the highest rate. The evidence of cell failure also becomes clearer in this plot, where 0.06% SOH is lost per cycle.

For the varying discharge C-rate conditions of Paths 7–12 only one seems to exhibit a path dependence: Paths 11/12 cycling with the most aggressive charge and discharge profile (see Fig. 7). Although a significant difference seems to be present in the ΔIR/cycle plot for Paths 7/8, no conclusion regarding path dependence in this case is drawn. Instead Path 11, after the switch condition is not able to support the 3C/FC profile and loses an average of 0.04% SOH/cycle.

Incremental Capacity and Differential Voltage Analysis

Two techniques are used to investigate the degradation modes occurring during the path dependent measurement series: ICA and DVA. The use of ICA to identify degradation modes was been pioneered by Matthieu Dubarry, Bor Yann Liaw and their associates at the University of Hawaii,[5] with Ira Bloom and company publishing some of the first applications of the use of DVA for degradation mode identification.[3,4] There are typically five degradation modes which can be interpreted using ICA and DVA including LLI, LAM, with LAM divided into losses from the PE and NE in both lithiated and delitiated states.[31] In this work, since all check-ups

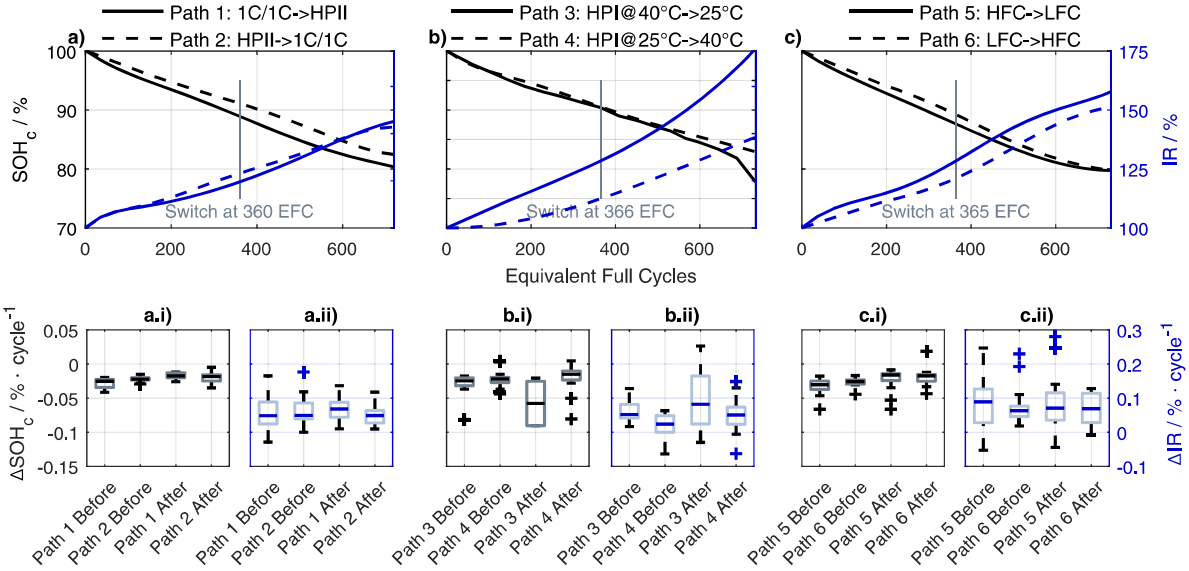

Figure 6. (a) Paths 1/2 aging at 1C/1C and HPII cycling before and after switch. (b) Path 3/4 aging under HPI cycling at alternating 25 °C and 40 °C. (c) Paths 5/6 aging compares LFC and HFC aging rates. The lower subplots, (a.i)–(c.ii), show the aging rate in terms of %-SOH/cycle and %-IR/cycle.

Journal of The Electrochemical Society, 2024 **171** 020555

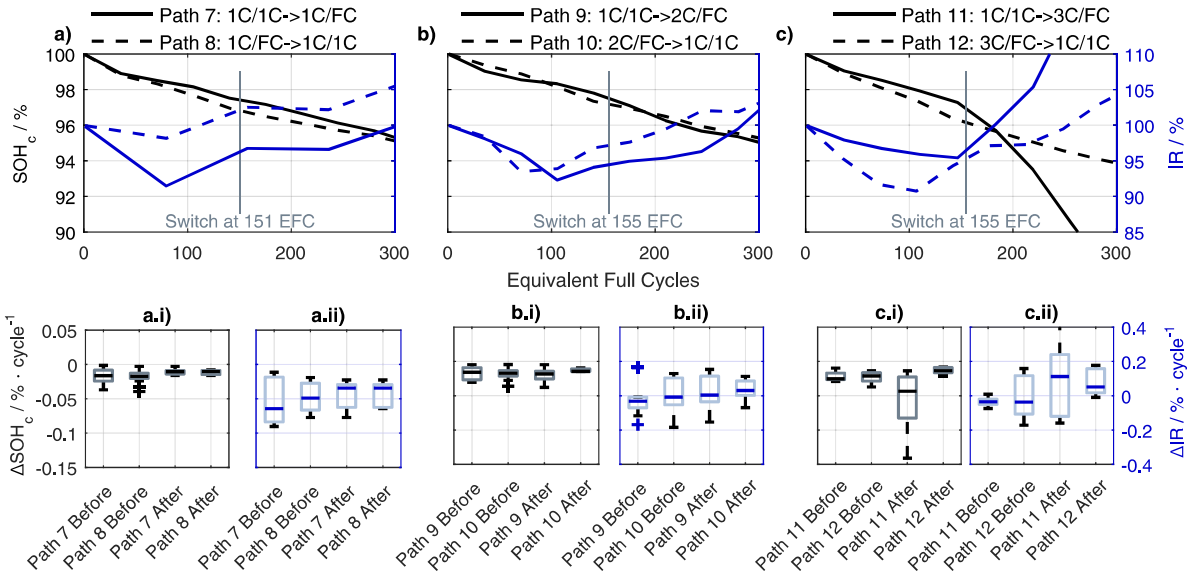

Figure 7. (a)–(c) show the degradation behavior of the three path dependent profiles switching between 1C and fast charging and varying discharging rates 1C, 2C, and 3C, respectively (Paths 7–12). The lower subplots, (a.i)–(c.ii), show the aging rate in terms of %-SOH/cycle and %-IR/cycle.

were performed in the discharge direction, only lithiated LAMpe and delithiated LAMne are considered. In general, LLI refers to any process consuming lithium ions removing them from the intercalation and de-intercalation processes, for example, parasitic reactions or electric isolation in the active material if containing lithium particles, and LAM refers to the loss of insertion sites due to, for example, particle cracking or structural disordering.[6,7] These groupings, however, also are often complicated by the coupling and feedback of mechanisms, where for example, transition metal ions from dissolution at the PE are deposited at the NE accelerating SEI growth. This example also highlights how LLI can result from lost active material being lithiated.[32]

Mathematically, ICA uses the gradient of the capacity (Q) with respect to voltage (V), dQ/dV, plotted against the charge or discharge voltage (see Fig. 8a) with DVA using the inverse dV/dQ plotted against the capacity (Fig. 8b).

A number of publications have applied these techniques to aged NCA cells offering insights to the relevant features exhibiting in the characteristic ICA/DVA curves.[33,34] The peaks of the dQ/dV curve represent the charge transfer during the reaction, where displacement and restructuring increase the cell capacity but not the potential, showing as a voltage plateau on the cell voltage curve.[5] The peaks of the dV/dQ are related the phase transformation (staging) within the electrode, and are seen as voltage changes in the full cell voltage curve.[31] These curves being inverse of each other means that the valleys of the dV/dQ curve are the peaks of the dQ/dV curve, and vise-versa. The ICA/DVA use the change of these features to discern the degradation modes affecting the specific electrode (or general lithium inventory) since each feature reflects various half-cell reactions.

Features G, F and E in the dQ/dV curve (Fig. 8) are related to the well known staging of the silicon-graphite anode.[35] Around 3.5 V, 3.9 V, and 4.05 V the cathode first transitions from hexagonal to monoclinic (H1-M), monoclinic to hexagonal (M-H2), to finally hexagonal to hexagonal (H2-H3).[20,24] From these phase transformations, two features can be unambiguously attributed to the PE, namely the largest peak, or Feature A (H2-H3 staging), and Features B and C which are both related to the M-H2 phase transition, separated by the graphite staging II.[5] This leaves Feature D, which takes place under reactions at both the PE and NE.

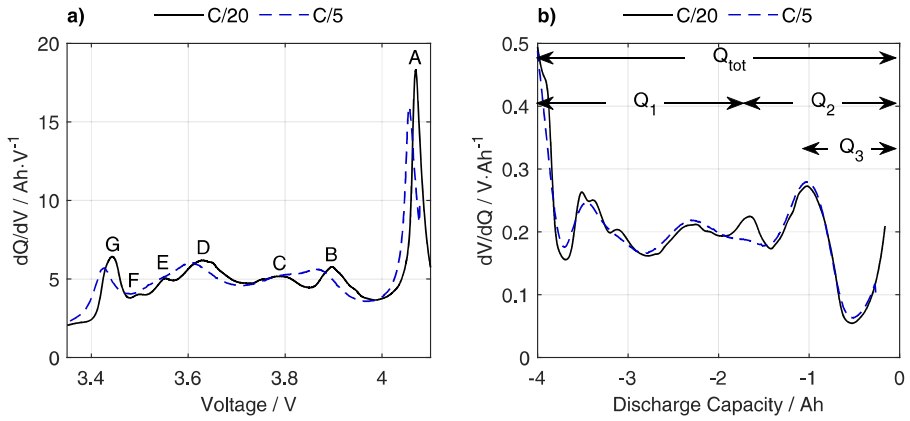

Figure 8. (a) Features of interest are shown on the dQ/dV plot against voltage at two C-rates, C/20 and C/5 from a new cell at 100% SOH. (b) Characteristic capacities are identified with their position on the dV/dQ plot against the discharge capacity, also at C/20 and C/5 C-rates.

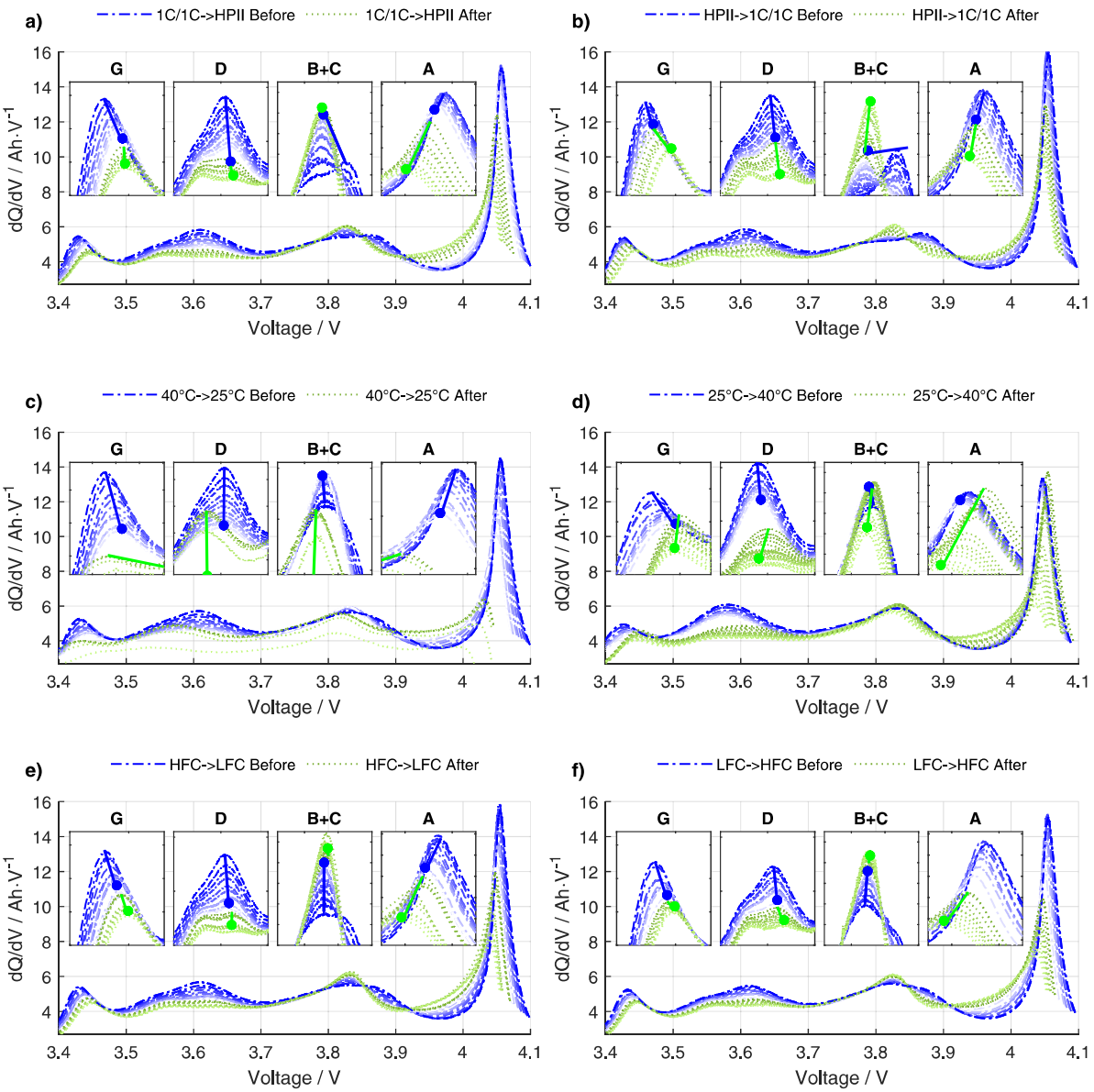

Figure 9. A dQ/dV showing the key degradation Features G-A, where features G reflect changes in the NE, B + C and A changes in the PE and D a combination from both. (a) Path 1: 1*C*/1*C* → *HPII*, (b) Path 2: *HPII* → 1*C*/1*C*, (c) Path 3: 25 °C → 40 °C, (d) 40 °C → 25 °C, (e) Path 5: *HFC* → *LFC* and (f) Path 6: *LFC* → *HFC*.

The dQ/dV features are not re-named in the valleys of the dV/dQ in Fig. 8b as this curve will be interpreted using the distance between the peaks as an indicator for the degradation modes as presented by Keil et al.[34] In this method Q_1-Q_3 measures the capacity lost between electrode reactions. Q_{tot} measures the total cell capacity during discharge. Q_1 relates to the LAMne as both the lower SOC limits and the graphite staging II, around 60% SOC are dominated by the NE. Conversely, Q_3 uses only markers dominated by reactions occurring in the PE, and can therefore be used to identify LAMpe. Q_2 uses both PE and NE markers and represents the balance between the electrodes, which is affected by the degradation mode LLI.

Figure 8 shows both the dQ/dV and dV/dQ at C/20 and C/5 C-rates. For Paths 1–6, check-ups were only done with C/5 discharge currents. This eliminates the potential use of the Q_1-Q_3 with the dV/dQ curve as the central graphite lithiation peak is not discernible. This is a result of resistive and diffusion polarization effects increasing at higher currents.[36] Other markers are also lost from the dQ/dV curve, with Features B and C combining due to that peak's absence, and the smaller anode staging reactions also combining to the more significant Feature D, leaving only Feature G as a marker for the NE and Features A-C for the PE. Despite the loss of some minor features, the ICA can still be useful in the qualitative evaluation of degradation modes.[37,38]

A qualitative assessment of the shifting of the peaks with respect to one another can provide insight into the dominant degradation modes. The ICA signatures of the first six paths are displayed in Fig. 9. In general, a decrease in peak height combined with a shift to

Journal of The Electrochemical Society, 2024 **171** 020555

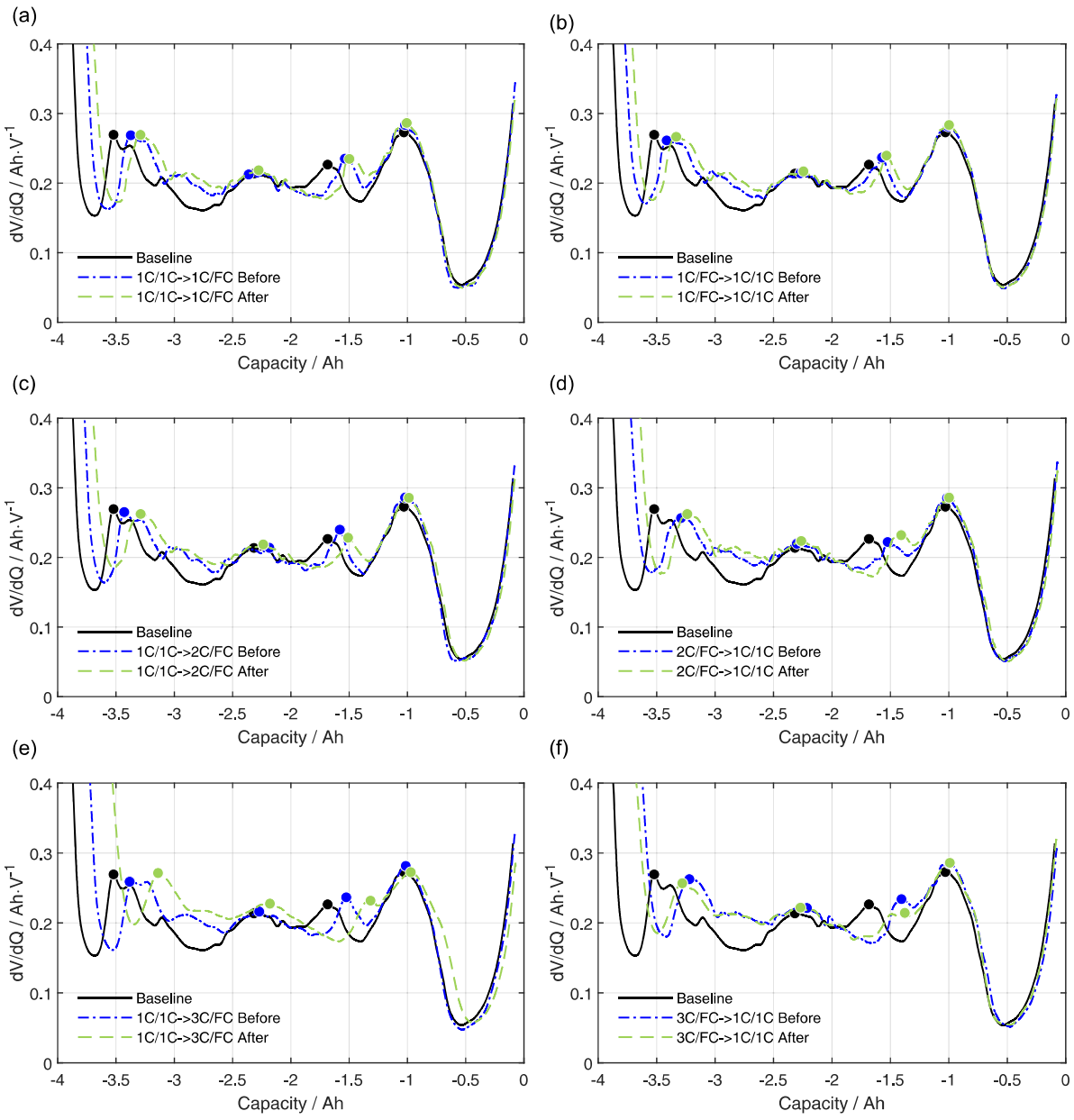

Figure 10. This Figure shows the initial dV/dQ, the dV/dQ before the switch and the dV/dQ after the switch, with the peaks highlighted which were used for the Q_1-Q_3 method. (a) Path 7: 1*C*/1*C* → 1*C*/*FC* (b) Path 8: 1*C*/*FC* → 1*C*/1*C*, (c) Path 9: 1*C*/1*C* → 2*C*/*FC*, (d) Path 10: 2*C*/*FC* → 1*C*/1*C*, (e) Path 11: 1*C*/1*C* → 3*C*/*FC*, (f) Path 12: 3*C*/*FC* → 1*C*/1*C*.

higher or lower voltages is an indicator for LLI, and a decrease in peak height at approximately constant voltage is an indicator for LAM.[7] According to Devie et al. who analyzed NCA cells using ICA, a disappearance of A and a shift of all other peaks toward lower potential would be an indicator for LAMne, a decrease in the intensity for all peaks, but G and D are less affected, indicates LAMpe and a shift of G toward higher potentials, whereas D is disappearing, is a marker for LLI.[33]

Figure 9 shows each of the six paths with the change in color and linetype indicating the switch to the alternate aging conditions. A zoomed-in view of the key ICA features A, B + C, D and G are provided for convenience.

The dominant aging signature seen in almost all the paths is a reduction and shift to higher potentials of G, a reduction and shift to higher potentials in D, an increase of B + C with a shift to lower potentials and a reduction and shift to lower potentials of A. This can be seen in Path 1 Before, Path 2 After, Path 3 Before, Path 4 Before and both Before and After from Path 5 and Path 6. This signature is indicative of LLI as all peaks are reduced and shifted, except for B + C.[33] The exceptional increase from Feature B + C likely comes from the electrode balance shift causing the Stage II peak of the anode to shift past the cathode M+H1 staging.

The fact that Path 1 Before, Path 2 After and all Path 5 and Path 6 show similar degradation mode signatures is unsurprising

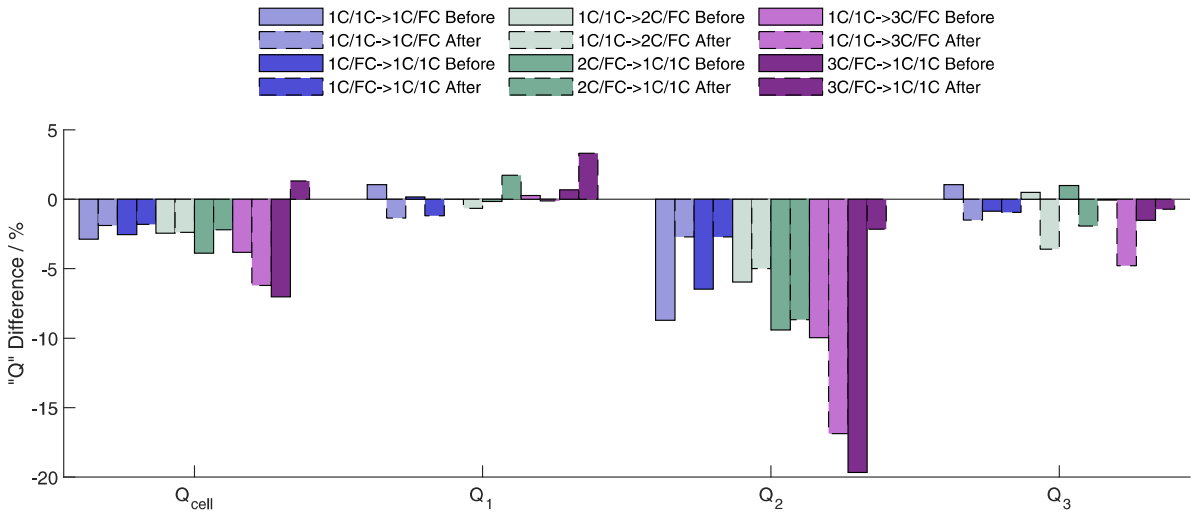

Figure 11. The change in the characteristic capacities "Q" are analyzed to determine the contributing degradation modes resulting in aging during Paths 7–12. Here Q_1 relates to losses in the LAMne, Q_2 to the balancing between the electrodes and Q_3 to the losses in the LAMpe.

considering the dominance of the 1C/1C cycling in the LFC and HFC profiles. A departure from 1C/1C degradation mode signature is seen in Path 2 Before, or when cycling the high C-rate profile HPII. Here, instead of increasing, B + C is decreasing, along with A. As these peaks are attributed to the M+H1 staging and H2+H3 staging, it is a strong indication that LAMpe is occuring.[24] The complementary path, Path 1 After, does not show as significant a reduction in the B + C feature, however, this could be due to the overall higher aging state, where changes between check-ups are in general smaller.

The next extreme exception comes from Path 3 After, where the switch from 40 °C to 25 °C led ultimately to a cell failure. Also during the baseline tests, one of three cells failed while cycling the HPI profile at 25 °C after 400 EFC. The rapid disappearance of A has been tied to lithium plating,[33] as the key NE feature clearly can no longer be lithiated to the same extent. This is possibly resulting from the IR drop at the NE during charging making the potential <0 V with respect to Li^+, or if the LAMne is high enough to result in overcharging.[31] Dubarry et al. note that initial LAMne aging could trigger plating, however, LAMne could not be explicitly identified from the Path 3 Before signature.[2] Interestingly, Path 4 Before cycling first at 25 °C does not show a particularly aggravated degradation, but as the degradation mode indicates LLI, lithium plating cannot be ruled out. After the switch, Path 4 seems to be dominated by LAMpe as indicated by the steep decrease of A and B + C with the concurrence of LAMne, marked by the shift to lower potentials of all peaks, indicating that LAMpe is a degradation mode in both the higher C-rate profiles.

The DVA signatures from Paths 7–12 investigating the discharge C-rate can be seen in Fig. 10. The dV/dQ has the major advantage over the dQ/dV in that the signature can be modeled as a linear combination of individual half cells, allowing for a more clear picture of the electrode balancing.[21] With the possibility of using the characteristic capacity method from Keil et al.[34] due to the visibility of the stage II graphite peak, an alternative analysis method can potentially shed a different perspective to the observed degradation modes. A look at the results from the changes in the characteristic capacities in Fig. 11 show that LLI is the leading cause of capacity fade for these aging conditions. From every subplot, the anode markers at high anode delithiation and the stage II graphite peak shift significantly to the right indicating a shift in the electrode balance caused by the NE or LLI.[39] In Fig. 11 the solid outlined bars represent the differences calculated between the baseline or reference dV/dQ taken from five cells in pristine condition (black line in Fig. 10)–and the dV/dQ measured just before switching to the second stress factor. The dashed outlined bars are then the difference in the characteristic capacity as measured from the switching point. The first set of characteristic capacities (solid outline in Fig. 11) are scaled based on the initial characteristic capacities, and the second set (dashed outline in Fig. 11) is scaled with respect to characteristic capacities at the switching point.

A slight trend in increasing Q_3 after the switch condition for all profiles with initial starting conditions of 1C/1C seems to indicate an accelerated presence of LAMpe. It is also observed that there is a slight increase in Q_1 of the 2C/FC and 3C/FC after the switch condition, seemingly indicating an increase in LAMne when switching to the 1C/1C cycling regime.

Conclusions

In this study, a path dependence analysis of silicon-graphite/NCA cells under high-performance cycling is performed. From a capacity perspective, some evidence of path dependence was seen for Paths 1/2, Paths 3/4 and Paths 11/12, with indications of diffusion limitations from a significant difference in internal resistance development while switching temperatures between Paths 3/4. Both ICA and DVA confirmed LLI to be the dominant degradation mode for cycling at higher C-rates, which aligns with many studies finding side reactions at the graphite surface, such as solid electrolyte interface (SEI) growth/repair to be the leading aging mechanism for nickel-rich-cathode/graphite cells.[3,4] These finding add to the growing amount of literature suggesting that the NCA cell exhibits path dependent behavior depending on C-rate,[2,18,19] however, due to similar aging rates, with only ca. 2% SOH-difference between various paths, a definitive case for path dependence in all conditions could not be made.

In the case of Paths 3/4 which experience lower temperatures, lithium plating is assumed. The path dependence from this aging condition lead to Path 3 (40 °C → 25 °C) failing. The trigger of plating is attributed to a potentially kinetically limited SEI growth removing cyclable lithium, decreasing the anode potential with respect to Li^+ causing an increased risk of plating.[37,40,41] In the compliment path, Path 4 (25 °C → 40 °C), there is likely less plating as the anode overpotential is less during initial cycling, and after the

switch to higher temperatures the improved ionic mobility and diffusion rate reduces plating potential.[42,43]

For the two path dependent cases where higher C-rates were the main stress factors (Paths 1/2 and Paths 11/12), an accelerated LAMpe in the second half of aging when switched to the higher C-rates from the lower stress conditions was evident. The coupling of degradation modes could be a result of the anode shifting to higher potentials with respect to the cathode, inducing overcharge stresses in the PE which lead to accelerated LAMpe mechanisms as micro-cracking and side reactions.[21,24]

From the ICA and DVA signatures in Fig. 8, low SOC regions are not shown as no features were detectable, making the analysis of the contribution from the silicon in the anode difficult. It is assumed that at least some of the reduction of G in Fig. 9, and rightward shift in Fig. 10 are contributions from the silicon degradation. It is also expected that due cycling in higher SOC ranges during the high-performance profiles of HPI and HPII, the aging due to degradation of the silicon is not a significant factor. Peaks around 20% SOC are attributed to the graphite stages, while a silicon contribution would be expected around 20%.[44] For the 1C/1C profile, cycling in lower SOC ranges, it is likely the silicon leads to increased SEI growth particle cracking which would correspond to higher LLI and LAMne respectively.[45]

The qualitative and quantitative approaches presented in this work offer an insight into degradation mode identification. A more quantitative approach is planned for a deeper investigation of the affect of specific aging modes on the OCV signature behavior using an OCV model fitting approach like the one proposed by Dubarry et al. using the ALAWA Toolbox.[31,46]

Acknowledgments

This work was funded by the BMW Group and was performed in cooperation with the Technical University of Munich.

ORCID

Jacob C. Hamar https://orcid.org/0000-0002-0845-3667
Johannes Sturm https://orcid.org/0000-0001-8876-9989
Julius Schmitt https://orcid.org/0000-0002-0815-7788
Marcel Rogge https://orcid.org/0000-0002-5805-4840
Simon V. Erhard https://orcid.org/0000-0002-6523-2502
Andreas Jossen https://orcid.org/0000-0003-0964-1405

References

1. J. C. Hamar, S. V. Erhard, A. Canesso, J. Kohlschmidt, N. Olivain, and A. Jossen, "State-of-health estimation using a neural network trained on vehicle data." *Journal of Power Sources*, **512**, 1–8 (2021).
2. M. Dubarry, G. Baure, and A. Devie, "Durability and reliability of ev batteries under electric utility grid operations: Path dependence of battery degradation." *J. Electrochem. Soc.*, **165**, 1 (2018).
3. I. Bloom, A. N. Jansen, D. P. Abraham, J. Knuth, S. A. Jones, V. S. Battaglia, and G. L. Henriksen, "Differential voltage analyses of high-power, lithium-ion cells: 1. technique and application." *Journal of Power Sources*, **139**, 295 (2005).
4. I. Bloom, J. Christophersen, and K. Gering, "Differential voltage analyses of high-power lithium-ion cells: 2. applications." *Journal of Power Sources*, **139**, 304 (2005).
5. M. Dubarry, V. Svoboda, R. Hwu, and B. Y. Liaw, "Incremental capacity analysis and close-to-equilibrium ocv measurements to quantify capacity fade in commercial rechargeable lithium batteries." *Electrochem. Solid-State Lett.*, **9**, A454 (2006).
6. C. R. Birkl, M. R. Roberts, E. McTurk, P. G. Bruce, and D. A. Howey, "Degradation diganostics for lithium ion cells." *Journal of Power Sources*, **341**, 373 (2017).
7. 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." *Journal of Power Sources*, **360**, 301 (2017).
8. G. Baure, A. Devie, and M. Dubarry, "Battery durability and reliability under electric utility grid operations: path dependence of battery degradation." *J. Electrochem. Soc.*, **166**, A1991 (2019).
9. Z. Ma, J. Jiang, W. Shi, W. Z. Chunting, and C. Mi, "Investigaiton of path dependence in commercial lithium-ion cells for pure electric bus applications; aging mechanism identification." *Journal of Power Sources*, **2015**, 29 (2015).
10. F. Roeder and S. Ramasubramanian, "A review and perspective on path dependency in batteries." *Energy Technology*, **10**, 1 (2022).
11. M. Bauer, C. Guenther, M. Kasper, M. Petzl, and M. A. Danzer, "Discrimination of degradation process in lithium-ion cells based on the sensitivity of aging indicators towards capactiy loss." *Journal of Power Sources*, **283**, 494 (2015).
12. K. L. Gering, S. V. Sazhin, D. K. Jamison, C. J. Michelbacher, B. Y. Liaw, M. Dubarry, and M. Cugnet, "Investigation of path dependence in commercial lithium-ion cells chosen for plug-in hybrid vehicle duty cycle protocols." *Journal of Power Sources*, **2011**, 3395 (2011).
13. D. Werner, S. Paarmann, A. Wiebelt, and T. Wetzel, "Inhomogeneous temperature distribution affecting the cyclic aging of li-ion cells. part ii: analysis and correlation." *Batteries*, **6**, 12 (2020).
14. E. Sarasketa-Zabala, I. Gandiaga, L. M. Rodriguez-Martinez, and I. Villarreal, "Calendar ageing analysis of a lifepo4/graphite cell with dynamic model validations: towards realistic lifetime predictions." *Journal of Power Sources*, **272**, 45 (2014).
15. M. Naumann, M. Schimpe, P. Keil, H. C. Hesse, and A. Jossen, "Analysis and modeling of calendar aging of a commercial lifepo4/graphite cell." *Journal of Energy Sources*, **2018**, 153 (2018).
16. E. Sarasketa-Zabala, I. Gandiaga, E. Martinez-Lasema, L. M. Rodriguez-Martinez, and I. Villarreal, "Cycle ageing analysis of a lifepo$_4$/graphite cell with dynamic model validations: Towards realistic lifetime predictions." *Journal of Power Sources*, **275**, 573 (2015).
17. L. Su, J. Zhang, J. Huang, H. Ge, Z. Li, F. Xie, and B. Y. Liaw, "Path dependence of lithium ion cells aging under storage conditions." *Journal of Power Sources*, **2016**, 35 (2016).
18. A. Raj, M. F. Rodrigues, and D. P. Abraham, "Rate-dependent aging resulting from fast charging li-ion cells." *J. Electrochem. Soc.*, **2020**, 1 (2020).
19. A. Karger, L. Wildfeuer, D. Ayguel, A. Maheshwari, J. P. Singer, and A. Jossen, "Modeling capacity fade of lithium-ion batteries during cycling concideringpath dependence." *Journal of Energy Sources*, **52**, 1 (2022).
20. A. Zuelke, Y. Li, P. Keil, R. Burrel, S. Belaisch, M. Nagarathinam, M. P. Merce, and H. E. Hoster, "High-energy nickel-cobalt-aluminum oxide (nca) cells on idle: Anode- versus cathode-driven side reactions." *Batteries & Supercaps*, **4**, 934 (2021).
21. W. M. Dose, C. Xu, C. P. Grey, and M. F. L. De Volder, "Effect of anode slippage on cathode cutoff potential and degradation mechanisms in ni-rich li-ion batteries." *Cell Reports Physical Science*, **1**, 1 (2020).
22. S. Watanabe, M. Kinoshita, T. Hosokawa, K. Morigaki, and K. Nakura, "Capactiy fade of lial$_y$ni$_{1-x-y}$co$_x$o$_2$ cathode for lithium-ion batteries during accelerated calendar and cycle life test (surface analysis of lial$_y$ni$_{1-x-y}$co$_x$o$_2$ cathode after cycle tests in restricted depth of discharge ranges)." *Journal of Power Sources*, **258**, 210 (2014).
23. S. Watanabe, M. Kinoshita, T. Hosokawa, K. Morigaki, and K. Nakura, "Capacity fading of lial$_y$ni$_{1-x-y}$co$_x$o$_2$ cathode for lithium-ion batteries during accelerated calendar and cycle life tests (effect of depth of discharge in charge-discharge cycling on the suppression of the micro-crack generation of lial$_y$ni$_{1-x-y}$co$_x$o$_2$ particle)." *Journal of Power Sources*, **260**, 50 (2014).
24. K.-J. Park, J.-Y. Hwang, H.-H. Ryu, F. Maglia, S.-J. Kim, P. Lamp, C. S. Yoon, and Y.-K. Sun, "Degradation mechanism of ni-enriched nca cathode for lithium batteries: are microcracks really critical?" *ASC Energy Letters*, **4**, 1394 (2019).
25. F. Benavente-Araoz, M. Varini, A. Lundblad, S. Cabrera, and G. Lindbergh, "Effect of partial cycling of nca/graphite cylindrical cells in different soc intervals." *J. Electrochem. Soc.*, **167**, 1 (2019).
26. S. Watanabe, M. Kinoshita, and K. Nakura, "Comparison of the surface changes on cathode during long term storage testing of high energy density cylindrical lithium-ion cells." *Journal of Power Sources*, **196**, 6906 (2011).
27. S. Watanabe, M. Kinoshita, and K. Nakura, "Capactiy fade of lial$_y$ni$_{1-x-y}$co$_x$o$_2$ cathode for lithium-ion batteries during accelerated calendar and cycle life test. i. comparison analysis between lial$_y$ni$_{1-x-y}$co$_x$o$_2$ and licoo$_2$ cathodes in cylindrical lithium-ion cells during long term storage tests." *Journal of Power Sources*, **247**, 620 (2014).
28. K. P. C. Yao, J. S. Okasinski, K. Kalaga, J. D. Almer, and D. P. Abraham, "Operando quantification of (de)lithiation behavior of silicon-graphite blended electrodes for lithium-ion batteries." *Adv. Energy Mater.*, **9**, 1 (2019).
29. T. Yoon, C. C. Nguyen, D. M. Seo, and B. L. Lucht, "Capacity fading mechanisms of silicon nanoparticle negative electrodes for lithium ion batteries." *J. Electrochem. Soc.*, **162**, A2324 (2015).
30. P. Keil and A. Jossen, "Imact of dynamic driving loads and regenerative braking on the aging of lithium-ion batteries in electric vehicles." *J. Electrochem. Soc.*, **164**, A3081 (2017).
31. M. Dubarry, C. Truchot, and B. Y. Liaw, "Synthesize battery degradation modes via a diagnostic and prognostic model." *Journal of Power Sources*, **2012**, 204 (2012).
32. J. S. Edge et al., "Lithium ion battery degradation: what you need to now." *Royal Society of Chemistry*, **23**, 8200 (2021).
33. A. Devie and M. Dubarry, "Durability and reliability of electric vehicle batteries under electric utility grid operations. part 1: Cell-to-cell variations and preliminary testing." *Batteries*, **2**, 1 (2016).
34. P. Keil and A. Jossen, "Calendar aging of nca lithium-ion batteries investigated by differential voltage analysis and coulomb tracking." *J. Electrochem. Soc.*, **164**, A6066 (2017).
35. J. R. Dahn, "Phase diagram of li$_x$c$_6$." *Physical Review B*, **44**, 1 (1991).
36. A. Fly and R. Chen, "Rate dependency of incremental capacity analysis (dq/dv) as a diagnostic tool for lithium-ion batteries." *Journal of Energy Sources*, **29**, 1 (2020).
37. J. Li, J. Harlow, N. Stakheiko, N. Zhang, J. Paulsen, and J. Dahn, "Dependence of cell failure on cut-off voltage ranges and observation of kinetic hindrance in lini$_{0.8}$co$_{0.15}$al$_{0.05}$o$_2$." *J. Electrochem. Soc.*, **165**, 1 (2018).

38. M. Dubarry, C. Truchot, B. Y. Liaw, K. Gering, S. Sazhin, D. Jamison, and C. Michelbacher, "Evaluation of commercial lithium-ion cells based on composite positive electrode for plug-in hybrid electric vehicle applications. part ii. degradation mechanism under 2 c cycle aging." *Journal of Power Sources*, **196**, 10336 (2011).
39. P. Keil, S. F. Schuster, J. Wilhelm, J. Travi, A. Hauser, R. C. Karl, and A. Jossen, "Calendar aging of lithium-ion batteries." *The Journal of the Electrochemical Society*, **2016**, A1872 (2016).
40. 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.*, **166**, 1 (2019).
41. S. F. Schuster, T. Bach, E. Fleder, J. Müller, M. Brand, G. Sextl, and A. Jossen, "Nonlinear aging characteristics of lithium-ion cells under different operational conditions." *Journal of Energy Sources*, **1**, 44 (2015).
42. T. Waldmann, M. Wilka, M. Kasper, M. Fleischhammer, and M. Wohlfahrt-Mehrens, "Temperature dependent ageing mechanisms in lithium-ion batteries—a post-mortem study." *Journal of Power Sources*, **2014**, 129 (2014).
43. A. Tomaszewska et al., "Lithium-ion battery fast charging; a review." *eTransportation*, **1**, 1 (2019).
44. J. Schmitt, M. Schindler, A. Oberbauer, and A. Jossen, "Determination of degradation modes of lithium-ion batteries considering aging-induced changes in the half-cell open-circuit potential curve of silicon-graphite." *Journal of Power Sources*, **532**, 1 (2022).
45. D. Anseán, G. Bauer, M. González, I. Cameán, A. B. García, and M. Dubarry, "Mechanistic investigation of silicon-graphite/lini$_{0.8}$mn$_{0.1}$co$_{0.1}$o$_2$ commercial cells for non-intrusive diagnosis and prognosis." *Journal of Power Sources*, **459**, 1 (2020).
46. HNEI. Alawa central, 2023.

4 State-of-Health Estimation Using a Neural Network Trained on Vehicle Data

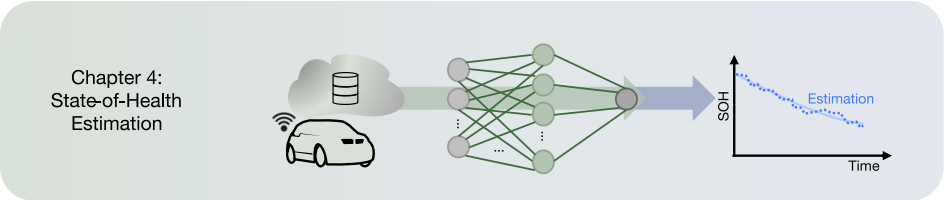

Figure 4.1: Crop from the thesis structure highlighting the focus of this chapter.

The degradation of a battery is an extremely complex process, even under controlled laboratory conditions, let alone in the automotive context. The advent of BEVs has not only generated new opportunities to help the transition from fossil fuels, but also new areas of battery research [170–172]. Understanding the aging of batteries, especially those used in the mobility sector, is essential to the feasibility and financial success of the electric drive train. Much work has been focused on building holistic aging models describing battery capacity fade over lifetime [40; 59; 77; 80; 150; 173–175]. These models often use a semi-empirical approach which superimposes a calendric and a cyclic degradation function to form the total loss equation.

Data-driven models do not rely on underlying physical equations, but instead rely on the availability of adequate data which describes the complete range of potential aging behaviors. This is where the explosive growth of electric vehicles provides the most potential for advancement in battery modeling, because data is now becoming available which describe highly diverse and dynamic aging conditions, which in turn can be used to train data-driven models.

The paper in this chapter compares some of the state-of-the-art modeling techniques, like the semi-empirical model, with that of a machine learning approach trained on automotive data under real-world operation conditions. In a first approach, a holistic aging model was coupled with K-means clustering to better fit the semi-empirical model [176; 177]. K-means is an unsupervised learning method which mathematically groups like measurements together. In this case, the same features which were used to train the semi-empirical model and the neural network are used for grouping the K-means clusters. As a result of the clustering, it is possible to select and fit a holistic aging function more accurately. This approach is most similar to fitting a holistic aging function using various input data-sets from specific temperature, C-rate or load profiles, as is often done in literature [40; 59; 77; 150], but the benefit of K-means is that a combination of multiple features can be used to form these clusters. This reduces the influence of manual cluster selection while still allowing for a better fitting of data sub-sets.

A second data-driven technique used a neural network to fit the aging data. Although there have been recent advancements in using data-driven modeling for improving the performance of battery health estimation, few have been validated in the dynamic and complex conditions experienced in the automotive context, and instead are restricted to constant laboratory conditions [80; 93; 175; 178].

The machine learning was implemented using the Keras framework in Python [82; 83; 179; 180]. Neural networks, like the one used in this study, have been shown to be highly effective at solving regression problems, like battery degradation [104–107; 118; 119]. What helps this approach succeed is the availability of a training label. One of the biggest challenges to data-driven approaches is the lack of access to labeled reference points to train against. While this issue is particularly relevant for some internal parameters of battery cells, the possibility to directly measure the SOHs of a battery means that generating a suitable training data-set is only a matter of time and cost, and not feasibility.

Acknowledgment I would like to thank all contributors to this paper, without whom this would not have been possible. A breakdown of each author's individual contribution is listed as per the Contributor Role Taxonomy (CRediT) guideline [169].

- Author: Jacob Hamar, Credit: Conceptualization, Methodology, Software, Validation, Formal Analysis, Data Curation and Writing.
- Co-Author 1: Simon V. Erhard,
  Credit: Conceptualization, Data Curation, Writing- Review and Editing, Supervision.
- Co-Author 2: Angelo Canesso,
  Credit: Investigation, Writing- Original Draft.
- Co-Author 3: Jonas Kohlschmidt,
  Credit: Investigation, Writing- Original Draft.
- Co-Author 4: Nicolas Olivain,
  Credit: Investigation, Writing- Original Draft.
- Co-Author 5: Andreas Jossen,
  Credit: Conceptualization, Supervision.

State-of-health estimation using a neural network trained on vehicle data

Jacob Hamar, Simon V. Erhard, Angelo Canesso, Jonas Kohlschmidt, Nicolas Olivain, and Andreas Jossen

Journal of Power Sources 512, 230493, 2021.

Permanent Weblink:
https://doi.org/10.1016/j.jpowsour.2021.230493.

Journal of Power Sources 512 (2021) 230493

Contents lists available at ScienceDirect

Journal of Power Sources

journal homepage: www.elsevier.com/locate/jpowsour

State-of-health estimation using a neural network trained on vehicle data ☆

Jacob C. Hamar [a,b,*], Simon V. Erhard [a,**], Angelo Canesso [a], Jonas Kohlschmidt [a], Nicolas Olivain [a], Andreas Jossen [b,**]

[a] BMW Group Research and Innovation Center, Knorrstrasse 147, 80788 Munich, Germany
[b] Technical University of Munich, Arcistrasse 21, 80333 Munich, Germany

HIGHLIGHTS

- Battery aging studied using eight years of battery electric vehicle data.
- Two data-driven methods trained and evaluated using BEV data.
- Estimation error using a semi-empirical model was 3.4%-SOH.
- Estimation error using a neural network model was 3.0%-SOH.

ARTICLE INFO

Keywords:
Aging
Machine learning
Automotive
Big data

ABSTRACT

The validation of battery aging models in automotive applications requires reliable aging data to compare the accuracy of each proposed model. Using a sample of 704 vehicles aged up to eight years under diverse nominal conditions two aging estimation models are proposed. By analyzing relevant automobile battery data a more relevant fit of a semi-empirical holistic model is provided with an Arrhenius temperature dependence and pseudo-Tafel voltage dependence. As a comparison, a neural network capturing the aging behavior using the most correlated variables available in the data-set was also developed. Over 110,000 measurements from seven relevant indicators are available as aging predictors, as well as, highly-accurate capacity measurements which is used as the ground truth capacity targets to train and validate the proposed models. Against these points the Semi-Empirical and Neural Network models achieved a root mean squared error of 3.4%-SOH and 3.0%-SOH, respectively.

1. Introduction

Modern lithium-ion batteries are critical to the automotive industries transition to renewable energy and to the global reduction of greenhouse gas emissions [1]. Many companies have been producing electric vehicles (EV) for several years, allowing for two interesting opportunities in automotive battery research. The first opportunity is to investigate the aging behavior under nominal driving conditions. This work is the first to use vehicle data collected over eight years under nominal driving conditions to validate novel state-of-health (SOH) estimation algorithms. A second opportunity resulting from the increased production of electric vehicles is the access to large data-sets for training and deploying machine learning data models. In this work, two machine learning supported SOH estimation models are presented and validated using a data-set of over 700 vehicles.

Aging behavior in automotive applications vary significantly from those reproduced in a laboratory setting. Understanding the aging behavior of lithium-ion batteries in an automotive context has long been an area of intense research as summarized by reviews from Si et al. and Waag et al. [2,3]. The major challenges present in battery aging estimation within automotive applications arise from the highly dynamic and variable conditions the vehicles experience. As is shown in this work, vehicles experience a wide range of physical and electrical demands which result in a significant spread of aging behavior.

Most holistic aging models, which treat both calendric and cyclic aging modes, are developed and validated using artificially generated aging data on a cell level [4–6] or use a select few publicly available cell testing data-sets which can be found in literature [7,8]. Due to intense time and resource limitations inherent in battery cell aging

☆ This document is the results of the research funded by the Bayerische Motoren Werke AG.
* Corresponding author at: Technical University of Munich, Arcistrasse 21, 80333 Munich, Germany.
** Corresponding authors.
E-mail addresses: jacob.hamar@tum.de (J.C. Hamar), simon.erhard@bwm.de (S.V. Erhard), andreas.jossen@tum.de (A. Jossen).
URL: http://linkedin.com/in/jacob-hamar-41a10068 (J.C. Hamar).

https://doi.org/10.1016/j.jpowsour.2021.230493
Received 13 July 2021; Received in revised form 23 August 2021; Accepted 3 September 2021
Available online 16 September 2021

data generation, this often limits studies to a patch-work of testing conditions requiring significant interpolation and extrapolation. The availability of aging information from the highly dynamic and various conditions that exist in the current data-set provide a much more colored picture to the range of possible aging scenarios and represent the most accurate validation metric for a proposed holistic aging model.

The methods developed for this work were necessarily analytic in nature. As a first approach, a well-established semi-empirical model method was fit to the data [6,9–11]. This method attempts to analytically describe the aging behavior based on known stress factors such as temperature and current, and as such, is more closely related to the physics resulting in aging. The second model proposed uses a neural network framework to learn aging behavior of the data-set with no *a priori* information from the underlying physics. This method is chosen as it has proven capable of accurate capacity estimation using cell tests [8,12] and has long been a tool for regression analysis using larger data-sets. Many novel methods have been presented with respect to battery SOH estimation as summarized from the reviews of Li et al. and Vidal et al. [5,6] where models have ranged from feed-forward, recurrent and radial bias neural networks, among others. In this work, a single feed-foreword neural network method is presented.

This research addresses the gap in SOH estimation methods validated on automotive batteries aged under the highly dynamic nominal driving conditions. The methods proposed do not attempt to capture potential path dependence or sequence effects in aging as the modeling of these effects and the consensus of the relevance of any path dependence is still an area of open research with few studies supporting path dependent aging [13,14], while few others offer contrary results [15–17]. With only a few notable exceptions, most aging models are validated on data generated under artificial conditions [5,6,9,18]. Two methods were chosen to highlight the advantages of incorporating machine learning techniques for SOH estimation when analyzing larger data-sets.

In the following sections a detailed review of the methodology (Section 2)– including a review of the acquired data-set and the two models chosen for this study– will provide the context in which the estimation results were generated. The later sections will highlight these estimation results (Section 3) and discuss the implications, short-comings and next steps in the conclusion (Section 4).

2. Method

In this section, the data-set will be described, including the aging data collected (Section 2.1) and the method of obtaining the Off-board Capacity Measurements (OCM) (Section 2.2) which were used as the ground truth for validating and training the analytic models. Next, data handling methods including a correlation analysis (Section 2.3) and a clustering of similar vehicles, are detailed. Finally, both estimation algorithms model are presented: the Semi-Empirical (Section 2.5) and the Neural Network (Section 2.6).

2.1. Vehicle data

For this investigation, a large data-set representing 704 vehicles was analyzed. The bulk of the data available for this work includes over 110,000 measurements of seven variables over the life of the vehicle. These variables include: kilometers, current throughput, state-of-charge (SOC), discharge/charge C-rate and temperature (see Table 1). Each vehicle was indicated with a unique identification number allowing for the tracing of each measurement variable over the life of the vehicle, as well as, connect the field measurements to the corresponding OCM. The data collected is from the high-voltage battery pack, and not from individual cells. The temperature is collected from the outside out of storage system. The decision to use the temperature data from this sensor location is motivated in large part by the fact that this feature is found in the calendric portion of the semi-empirical model see Section 2.5, and the chosen temperature sensor reflects the average aging temperature over the life of the vehicle.

Table 1
An overview of the measured variables available for model development.

Variable Name	Variable Description
Kilometers	Accumulated number of driven kilometers.
charge throughput	Integral of the charge and discharge current.
State-of-Charge (SOC)	Cumulative average of the ratio of the remaining capacity over the nominal capacity.
C-rate	Both charging and discharging measurements are cumulative averages of the ratio of current over battery capacity, first registered as a binned histogram.
Temperature	Cumulative average of the temperature as measured on the outside of the battery pack.
Age	Total vehicle age.

Weekly uploads of essential battery information is provided by the selected vehicles as part of maintenance and safety observation. Much of the collected data comes in the form of binned histograms, where the cumulative amount of time in each bin is registered. This measurement aggregation is a strong departure from traditional laboratory aging data, which is recorded as time-series values.

From these histograms, an average value was extracted and used for model development. Variables which are registered as histograms and then averaged are: SOC, C-rate and temperature. The reminding variables, kilometers, charge throughput and time, are registered as a cumulative amount up to the moment of the reading. Since these measurements occur under nominal vehicle operation, the data includes some errors resulting in missing values ('NaN') or faulty measurements ('Inf'). Reasons for faulty signals can include: flash update of the on-board computer, transcription error of data, faulty sensor, as well as others.

An overview of the data is provided in form of a cumulative distribution function (CDF) plot Fig. 1. The CDF is an alternative representation to the scatter distribution and the histogram (lower triangle and diagonal of Fig. 3, respectively). The y-axis of the CDF shows the range of the values. The x-axis is the percent of readings. This is useful for determining, for example, what percent of readings are under the mean, or as indicated on the plot, where the 5th or 95th percentile are found. For data with a normal distribution, the mean would fall at 50%. Similarly, for normally distributed data, the tails characteristic of the Gaussian distribution would also be seen in the CDF.

The discharge C-rate shown in Fig. 1D is most representative of a normal distribution, which reflects the wide range and highly dynamic operation of the vehicles during driving. As a comparison, the charging C-rate has a flat CDF function, reflecting the relatively standard charging procedures. From this plot, a comparison of the samples with and without OCM can be performed quickly to identify significant variations in the data used for learning the model and the expected global data-set. The distributions of SOC, charge and discharge C-rate are consistent between the two data-sets. The OCM data has a slightly higher average kilometers driven and a higher average charge throughput, though the distribution (left-skew Weibull) is consistent. Similarly, the average temperature from the OCM measurement data-set is slightly lower, however, the distribution pattern remains are very similar.

2.2. Off-board capacity measurements

Among these data are a relatively small sub-set of samples which include results from an Offboard Capacity Test (OCM). From the larger data-set of over 110,000 measurements, 704 measurements contain the additional OCM measurement. When combined with the vehicle data (Section 2.1), a powerful data-set is created which contains the aging history of a vehicle (same as recorded in the larger data-set) with an accurate estimation of the capacity fade at some time during the life

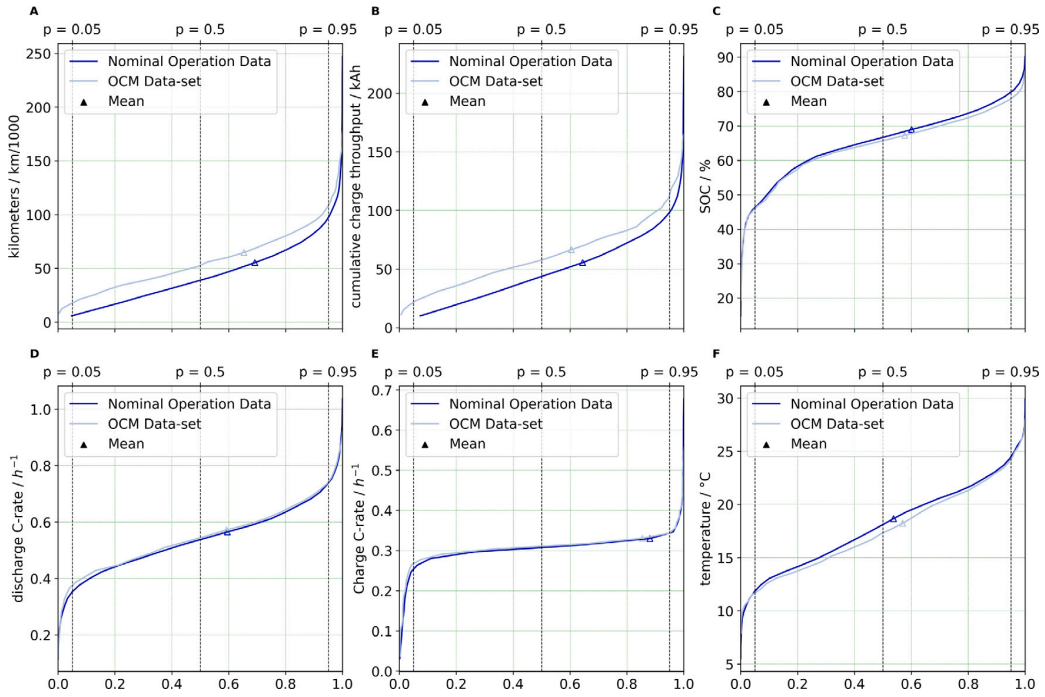

Fig. 1. Six variables (seven including time) which were recorded over the life of the vehicles, are available for use in capacity estimation (dark blue). These measurements are combined with Off-board Capacity Measurement readings (light gray) to provide a detailed picture of vehicle aging under nominal driving conditions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of the vehicle. The highly accurate OCM occurs only when the vehicle is taken to a service center, and as such there is a limited number of these measurements available.

The OCM procedure is standardized to be the same across all global service centers, and should therefore be a equivalent for all vehicles. Each test was performed with a constant current constant voltage (CCCV) discharge procedure to the nominal operation voltage. With a low discharge rate at a standard temperature, a highly accurate and consistent capacity measurement data point is generated. The exact pre-conditioning, tempering and current protocols cannot be described in more detail as they are covered by intellectual property, but this procedure can be compared to similar methods used in laboratory settings to estimate cell capacity [10]. This capacity measurement is combined with the aforementioned aging history up to the point of the measurement and provides the most detailed snapshot of vehicle use. These 704 read-outs are used as the ground truth for the state-of-health (SOH) estimation.

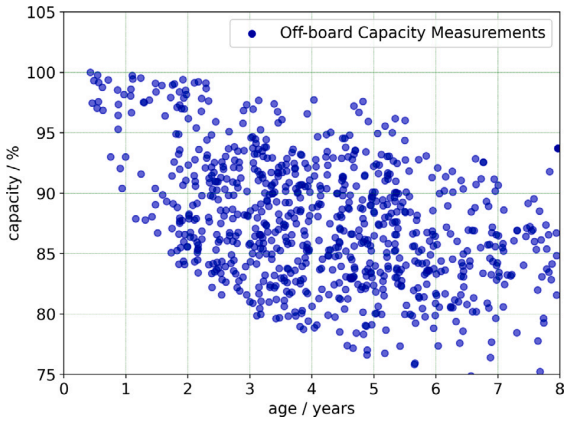

Fig. 2. Cumulative distribution functions are used to compare the 704 samples with the OCM measurement and the larger data-set of over 110,000 measurements to highlight any significant difference in the aging history (kilometers driven, charge throughput, average state-of-charge, average discharge/charge C-rate during operation and average temperature) collected in both data-sets.

2.3. Correlation analysis

An evaluation of each (independent) variable as it correlates to the capacity (predictor variable) is performed using the Pearson correlation coefficient. For this analysis, only the OCM data-set was used as this data-set contains the dependent variable which is the target of the estimation algorithms. The Pearson correlation coefficient gives a value between −1 and 1 indicating the degree and direction of the correlation [19]. Fig. 3 shows, in the first column, the scatter plot of each variable against the capacity, and also in the first row, a list of the corresponding correlation coefficients (with higher absolute values corresponding to higher correlations). This metric is highly relevant for selecting independent variables for a data-driven model ensuring that the variance of the dependent variable is modeled by the independent variables.

Also relevant for the data-driven method is understanding the co-linearity of the independent variables, which is a measure of the redundancy of information contained within a pair of variables. The use of colinear variables should be avoided in order for a regression fitting to have a well-defined solution [19]. From Fig. 3, colinearity was determined from the Pearson's correlation coefficient of the independent variables (all rows except row one) as any r-value above 0.7.

An initial analysis of the correlation of the independent variables reveals that the highest correlation with capacity exists between the kilometers ($r = -0.69$) and the charge throughput ($r = -0.67$). With a consideration of the colinearity, only one variable of either charge throughput or kilometers is selected for use with the Neural Network. Two other independent variables are considered as they demonstrate

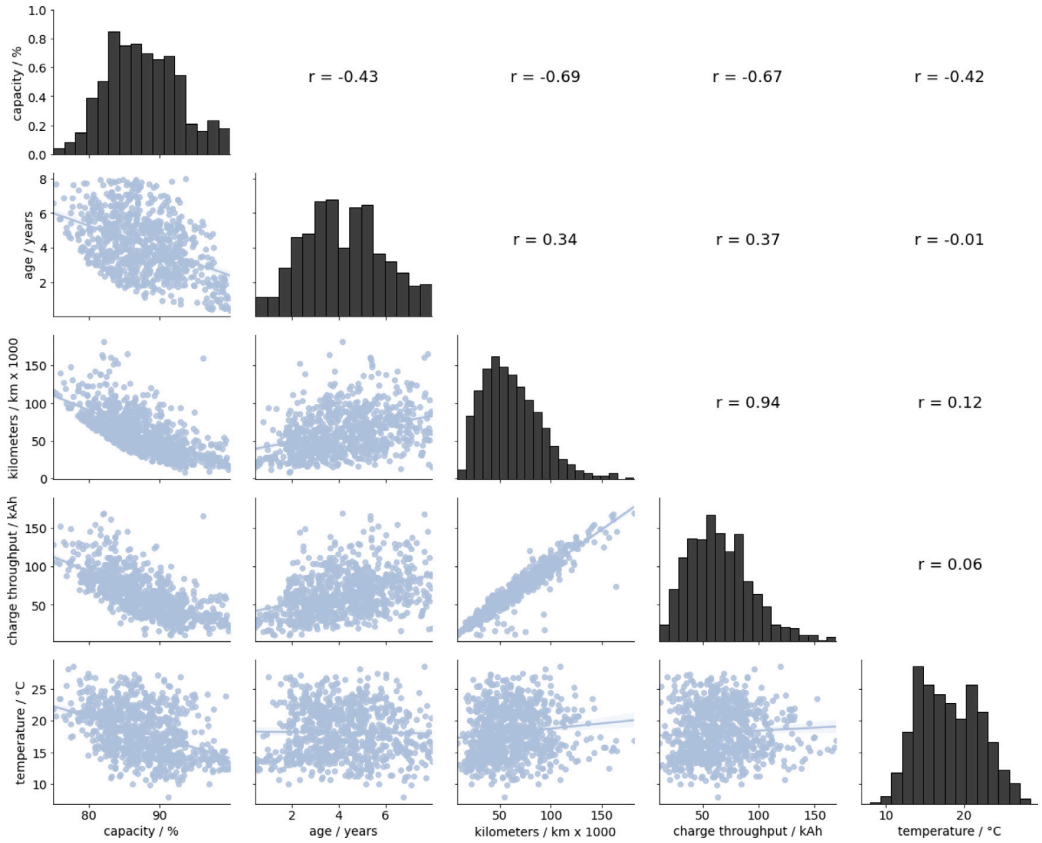

Fig. 3. A correlation analysis of each of the available input feature to the target variable (SOH) and each other helps select the leanest set of predictor variables. The lower triangle of the figure shows each variable plotted against the others. The central diagonal is the histogram distribution (with a relative frequency between $0-1$ as seen on the y-axis of the first plot in row one and column one, and the bins shown on the figure's x-axis). The upper corner of the figure lists the Pearson correlation coefficient, r. Charge/discharge C-rate and state-of-charge were not included in the plot as their correlation was not significant ($r < 0.1$).

reasonably high correlation: age ($r = -0.43$) and temperature ($r = -0.42$) and do not show significant correlation between the other variables. Three variables (not shown in Fig. 3) were not considered for use in the Neural Network model as they did not demonstrate a high correlation with the dependent variable: discharge C-rate ($r = -0.16$), charge C-rate ($r = -0.06$) and SOC ($r = -0.05$). This should not be interpreted, however, as the C-rate and SOC do not play a significant role in the SOH values over lifetime, but rather that the averaged values based on the specific binning of the histograms collected from the vehicles in this work did not adequately capture this correlation in the data sample analyzed. This result is seen by the authors as an indication that the average values collected from the discharge C-rate histogram and possible the SOC histogram did not reflect the contribution either feature has on aging as seen from more controlled laboratory experiments [8,20]. It should also be noted that the Pearson correlation coefficient of $r = -0.16$ for the discharge C-rate does indicate some degree of dependency with the target variable, but was excluded due to the fact that the discharge C-rate exhibited significant co-linearity with other predictor variables.

2.4. Clustering

As can be seen in Fig. 2 the variance, or spread, of the SOH distribution increases with the age of the vehicle. This increasing variance would lead to decreasing regression accuracy over time with a single regression function. In order to reduce the effect of this spread, a K-means clustering algorithm was applied [21,22].

K-means allows for splitting the data into a set $S = S_1, ..., S_k$ of k clusters of similar behavior. Once clustered, a regression can be performed on each cluster of the data improving the overall regression accuracy (see the implementation of the cluster fitting using the Semi-Empirical regression model from below (Section 2.5).

The K-means algorithm then seeks to minimize the average squared distance between points $X = (x_1, \ldots, x_n)$ with $x_i \in \mathcal{R}^d$ within the same cluster k with the following objective function:

$$\underset{S}{\operatorname{argmin}} \sum_{i=1}^{k} \sum_{x_j \in S_i} \|x_j - \mu_k\|^2 \tag{1}$$

where μ_k is the centroid of cluster k. Six variables were selected for clustering as they are the same inputs selected for the Semi-Empirical function (see Section 2.5). The variables used for clustering are: time, charge/discharge rate, SOC, charge throughput and temperature. For this clustering task, four clusters were specified. The number of clusters is a variable parameter which can be changed by the developer and is commonly identified by using the elbow and/or silhouette methods [23]. It was determined by both the elbow and silhouette method that the optimal number of clusters is four according to the respective distance scores. Additionally, when iterating the model for accuracy optimization, there were no improvements when choosing a different number of clusters. Since many programming languages offer specific K-means packages, the implementation of the clustering algorithm is quite accessible, however there are a few disadvantages. K-means assumes that clusters are spherical and that each cluster has roughly an equal numbers of observations. Although the nominal dataset is skewed over time due to the fact that there are fewer older

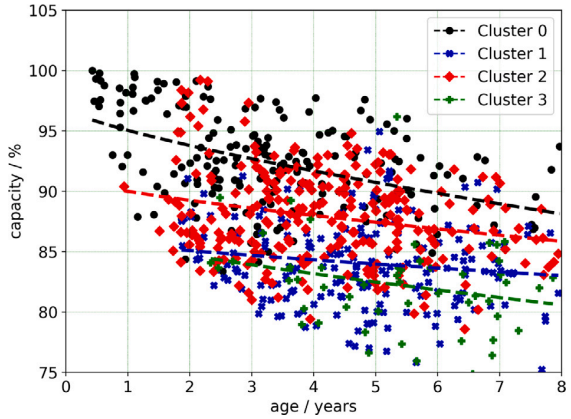

Fig. 4. The K-means algorithm clusters data points to the nearest centroid with respect to multiple input criteria. The four clusters are also fit with the Semi-Empirical regression, with the color corresponding to the cluster. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

vehicles to sample from, the OCM data is normally distributed (this assumption was satisfied after considering the results of the Shapiro–Wilks test [24], which tests the assumption that a sample comes from a normally distributed population). By training the K-means clustering algorithm on the OCM data, it is possible to fit a clustering function on a normally distributed data-set which is representative of the nominal data-set the algorithm is applied to (see Fig. 1). The results of the clustering can be see in Fig. 4.

2.5. Semi-empirical model

When estimating battery behavior (including aging), pseudo-2D models are often considered a highly accurate method with an acceptable computation demand, however, these models require an extensive characterization set and a series of complex differential equations [25]. For this reason, semi-empirical models have been suggested as a compromise between capturing physical dependencies and allowing for simple parameterization [10]. The Semi-Empirical analytic model used in this work was inspired by the works of Schimpe et al. Hahn et al. and Schmalstieg et al. [10,11,26] and uses both the Arrhenius temperature dependency during calendric aging, as well as, a charge dependency during cyclic aging.

The proposed model is listed below. A more interested reader in the identification of these dependencies from cell tests is encouraged to find more information there; this work, however, focuses on the suitability of such an analytic method for use with automotive data and therefore does not consider the derivation of the model.

$$SOH_c = 100 - SOH_{cal,c} - SOH_{cyc,c} \tag{2}$$

$$SOH_{cal,c} = a_1 \overline{U}_c e^{-\frac{6975}{\overline{T}_c}} t_i^{0.75} \tag{3}$$

$$SOH_{cyc,c} = (a_2 \overline{U}_c^2 + a_3 \overline{DOD}_c) \overline{Q}_c^{0.5} \tag{4}$$

The model (Eq. (2)) consists of two terms, calendric aging (Eq. (3)) and cyclic aging (Eq. (4)), contributing independently to the SOH calculation. The Semi-Empirical model captures the dependency on the aging conditions by considering: average voltage ($\overline{U}$), average temperature ($\overline{T}$), time (t), average depth-of-discharge ($\overline{DOD}$) and average total charge throughput ($\overline{Q}$). The average terms are calculated per cluster and are then fixed coefficients in Eq. (2) per cluster, as denoted by the subscript "c". These computed averages are also listed in the parameterization set found in Table 2. An estimated activation energy from literature, divided by the gas constant yields the exponent constant $-6975K$ [10]. One term from the set of inputs is not a constant, and that is the time.

In the calendric portion of the aging equation (Eq. (3)), the Arrhenius temperature dependency is modeled with the factor $\overline{T}$, which represents the average temperature the battery experienced up to the moment of the readout. A voltage dependency using the variable $\overline{V}$, is also included, as suggested by Schimpe et al. [10]. Although voltage was not directly measured, it was calculated from the measured $\overline{SOC}$ using a proprietary mapping based on the cell voltage range giving $\overline{V}$. This same $\overline{V}$ factor is used in the cyclic aging portion of the aging equation (Eq. (4)). The cyclic aging term also includes a depth-of-discharge dependency $\overline{DOD}$, which is calculated as $1 - \overline{SOC}$. Although the SOC value is used for generating the clusters and the complement $1 - \overline{SOC}$ is used in the Semi-Empirical calculation, since these values contain the same information, it is reasonable to assume that using the SOC for clustering did not negatively impact the results. The final term in the cyclic aging term is $\overline{Q}$, which represents the total charge throughput from charging and discharging of the vehicle.

For the fitting of each cluster, the cluster specific averages are used– in combination with the Levenberg–Marquardt method of solving the non-linear least square problem– to derive the fitting parameters: a_1, a_2, and a_3. The final regression line for each cluster (seen in Fig. 4) is computed using these fitting parameters and the cluster specific averages. In order to calculate the SOH of a new vehicle measurement at an instant in time, t_i, using the method above, the associated cluster must first be determined. This is accomplished by inputting the measured state-of-charge (SOC_i), temperature (T_i) and charge throughput (Q_i) at the time-step (t_i) at the measurement instance (i) into the fitted K-means clustering algorithm. Once a cluster has been assigned for the time-step, the SOH value is determined using Eq. (2).

The fitting of Eq. (2) is carried out on each cluster resulting from the K-means clustering (see Section 2.4). The clustering algorithm yields a grouping of 172 vehicles in Cluster 0, 156 in Cluster 1, 151 in Cluster 2 and 225 vehicles in Cluster 3. A visual of the clustered OCM is shown in Fig. 4. From this figure, the resulting fitted Semi-Empirical function is also plotted as a colored dashed-line corresponding to each of the four clusters. The fitting parameters and average cluster values for each cluster are listed in Table 2.

The three fitting parameters, a_1-a_3, combined with the average input values per cluster, fit the Semi-Empirical function to varying aging trends. Cluster 3 results in nearly the highest average charge throughput and has the highest average temperature which leads to the empirical model fitting for low SOH values. The cluster with the least aggressive fit is Cluster 0, which contains more moderate values for temperature and average charge throughput, however, the average SOH for this cluster is 94, which is the highest from all cluster, showing the relevance of each parameter used for clustering. The most variation between the clusters was seen in the average charge throughput and the average temperature. The relatively constant voltage and DOD may reflect a superfluous inclusion in the k-means clustering algorithm, however, as these were inputs to the Semi-Empirical model, these values may still be relevant, especially when considering a different data set where these averages are likely to deviate more than was seen in the data collected for this work.

The fitting parameters, a_1-a_3 indicate that each term in the fitting function of Eq. (2) lead to a lowering of the SOH, which is to be expected. The parameter a_1 is not negative due to the negative sign in the exponent of the term. The a_1 term is the highest in Cluster 0, which indicates the calendric aging portion is more relevant. Together with the moderate average temperature and low average charge throughput values from the cluster, this line represents vehicles experiencing lower calendric aging. Cluster 3, conversely, has the lowest value for a_1, and with a significant charge throughput and temperature average, indicates vehicles with a larger overall aging contribution. The fitting parameters of a_2 and a_3 impact the calendric portion of the aging equation, and the contribution of the average voltage and DOD, receptively.

Table 2
A complete list of the parameterization values fitted to the Semi-Empirical model for each cluster.

Fitting Parameter	Unit	Cluster 0	Cluster 1	Cluster 2	Cluster 3
a_1	$VKs^{(-\frac{1}{4})}$	$1.31 \cdot 10^5$	$3.71 \cdot 10^4$	$3.10 \cdot 10^4$	$2.66 \cdot 10^4$
a_2	V^-2	$-1.41 \cdot 10^2$	$-9.00 \cdot 10^1$	$-5.55 \cdot 10^2$	$-1.18 \cdot 10^1$
a_3	$Ah^-0.5$	$-1.30 \cdot 10^2$	$-8.80 \cdot 10^1$	$-5.34 \cdot 10^2$	$-1.12 \cdot 10^1$
$\overline{Q}$	Ah	0.41	0.44	0.50	0.48
$\overline{U}$	V	3.62	3.65	3.64	3.63
$\overline{DOD}$	%	64	66	65	65
$\overline{T}$	$°C$	17.3	17.9	16.7	18.1

In this regards, Cluster 3 again stands out as having the lowest values, again indicating the dominant role the calendric term a_1 has in the fitting equation. Finally, as seen with the average values per cluster, there is little difference between the average voltage and DOD fitting coefficients of a_2 and a_3 which reflects the low variation in the average DOD and voltage values from each cluster.

2.6. Neural network model

Neural networks have proven highly effective at modeling the capacity fade from batteries aged in a variety of conditions [5,12]. These methods are adept at filtering irrelevant information from noisy data-sets and building strong connections based on patterns displayed in the training data-set. For the 704 samples available for training, a deep learning approach is not ideal due to over-fitting, however, a shallow feed-forward neural network is able to provide valuable insights into the viability of purely data-driven methods for predicting SOH using measured automotive data.

A neural network with a single hidden layer was employed to carry out a regression of the remaining capacity. The sequential model was developed in Keras using Python programming language [27,28]. Using the sequential implementation a hyper-parameter tuning using a search matrix was performed. The tuning parameters investigated are summarized in Table 3.

The Neural Network model consists of three inputs: age, kilometers and temperature which were chosen due to their high correlation value and low colinearity (see Section 2.3). During the development of this model, other combinations of input features were investigated but did not yield significantly improved results. The results of the search matrix also indicated that the combination of a densely connected input layer of 27 neuron layer and a single hidden layer with 300 neurons was the most accurate. The size of the hidden layer had the largest impact on overall model accuracy, as well as model size. The final combination resulted in a model with 11,545 trainable parameters (assuming 32-bit integers, the model would occupy 369 kB of memory). A *relu* activation function was used at each layer in the model as these help reduce the vanishing gradient problem (where the derivative of the loss function goes to zero) by extending the activation range linearly and not squishing it between 0 and 1, and it also encourages sparseness by assigning a binary 0 at the low end. Both features then contribute to improved learning performance compared with other sigmoid activation functions [23,29]. A single drop-out layer was also introduced with a 30% dropout rate to reduce over-fitting [29]. A more detailed review of how neural networks function is not within the scope of this paper, however, a well-written book on the subject can be read from Francois Chollet (the founder of Keras) [29].

The final neural network model was trained using a train-test random split of 75−25 % with 528 samples for training and 176 samples for testing from the OCM data-set. For the model training and estimation all values were scaled to between 0 and 1 to improve training performance. The Keras Regressor model was optimized using the *adam* optimizer as this is well suited to dealing with sparse gradients (see above) and noisy data [28]. The final optimization parameters are the number of epochs, 500, and the batch size, one, which also showed the best results from the search matrix.

3. Results

This section considers the validation method for the proposed models. The evaluation of the accuracy relies heavily on the OCM data-set as these are considered ground truth SOH values with respect to the vehicles aging under nominal automotive conditions. A comparison of the error between the models estimation is discussed. Finally a representative selection of four vehicles is used to show the estimation behavior over vehicle life.

One of the most reliable methods for evaluating the accuracy of an estimation model is to consider the root mean squared error (RMSE) between the estimation and the ground truth [23]. Fig. 5 shows various model estimations against the OCM, as well as, the RMSE of the error between the OCM and the Semi-Empirical model (Fig. 5A) and the Neural Network model (Fig. 5B). The end-of-service life of traction batteries is typically considered to be 70 %-SOH. An improvement of 0.4%-SOH RSME error from the Semi-Empirical model to the Neural Network model reflects a 1.3% improvement in lifetime usage of the battery. Additionally, the Pearson correlation coefficients provide an estimate for how well the estimated SOH correlates with the OCM and are 0.75 and 0.8 for the Semi-Empirical and Neural Network models respectively. The linear $R = 1$ line was plotted to help visually demonstrate the skew present in the models. It can be seen from the Pearson correlation coefficient that the Semi-Empirical model has significant variance which is unaccounted for in the model. The r^2 value for each model is 0.52 and 0.65 for the Semi-Empirical and Neural Network model, respectively. The r^2 is a metric showing the variance of the dependent variable explained by the model, therefore, the Semi-Empirical model explains 54% of the variance in the SOH, whereas the Neural Network covers 65% [30].

Another metric used to evaluate the error was to plot the residuals (or error) against the available input features as seen in Fig. 7. These residual plots help highlight any correlation of an existing input feature with the residuals, highlighting any variance which has not yet been captured in the structure of the model. A balanced model will have residuals evenly distributed above and below the zero-line, indicating a noisy or random distribution of the prediction error. In addition to the distribution about the zero-line, the darkness of the point indicates the corresponding state-of-health of the data-point. In these plots, only the residuals of the Semi-Empirical model plotted against temperature exhibit a noticeable correlation $r = 0.47$ using the Pearson correlation coefficient which suggests the Semi-Empirical model using the Arrhenius temperature dependence does not sufficiently capture the influence of temperature on aging using the data provided. This correlation shows that the Semi-Empirical model is under-predicting at lower temperatures and over predicting at higher temperatures. Temperature plays an important role in battery aging in the automotive industry due to the extended periods of calendric aging. The color-mapping to SOH did not reveal any error dependency with respect to new or old aged vehicles — the error remains evenly distributed throughout the life of the vehicle.

For data-driven models, there is most often a trade-off between accuracy and storage requirements. A comparison of the storage requirement and accuracy for the two selected models is shown in Table 4. The Semi-Empirical model would require less than 1 kB in order to parameterize the prescribed formula with a few 32-bit integers, however, the final model described in this work includes a clustering algorithm requiring an additional 79 kB, for a total model size of 80 kB. In comparison, the Neural Network requires 369 kB to completely describe the mapping of the input to the output layer.

In addition to the estimation accuracy of the OCM, it is important to evaluate the model estimation performance over the useful life of the battery. For this reason, an arbitrary selection of four vehicles is used to understand the behavior of each model in more detail. From Fig. 6, one of the biggest differences between the modes becomes evident: the step behavior of the Semi-Empirical model. As described in Section 2.4,

Table 3
A search matrix of hyper-parameters was used to identify the optimal set for the Neural Network model.

Neurons (Input layer)	Neurons (Hidden layer)	Dropout rate	Epochs	Batch size
3, 9, 27, 63	100, 300, 900, 1500	0.2, 0.3, 0.4, 0.5	50, 250, 500, 1000	1, 3, 10, 25

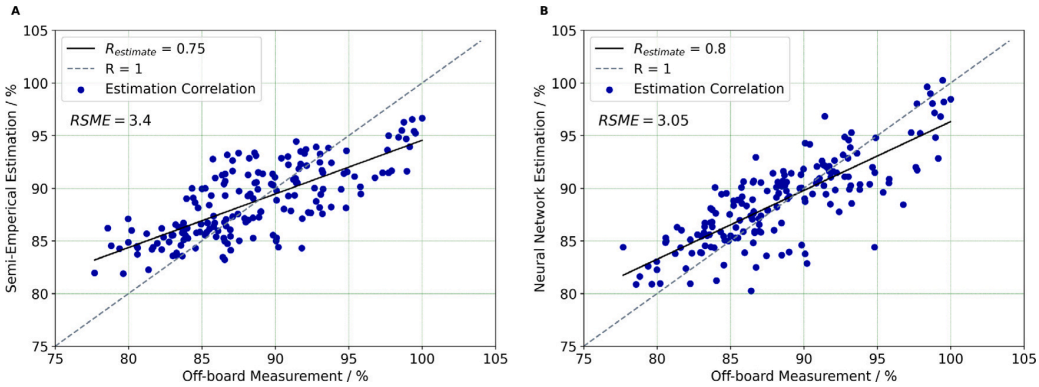

Fig. 5. A validation method plotting the correlation between the (A) Semi-Empirical Model and (B) Neural Network Model against the Off-board Capacity Measurements allows for a quick assessment of the respective prediction accuracy. The root mean squared error is also calculated for each method yielding 3.4% and 3.0% error, respectively.

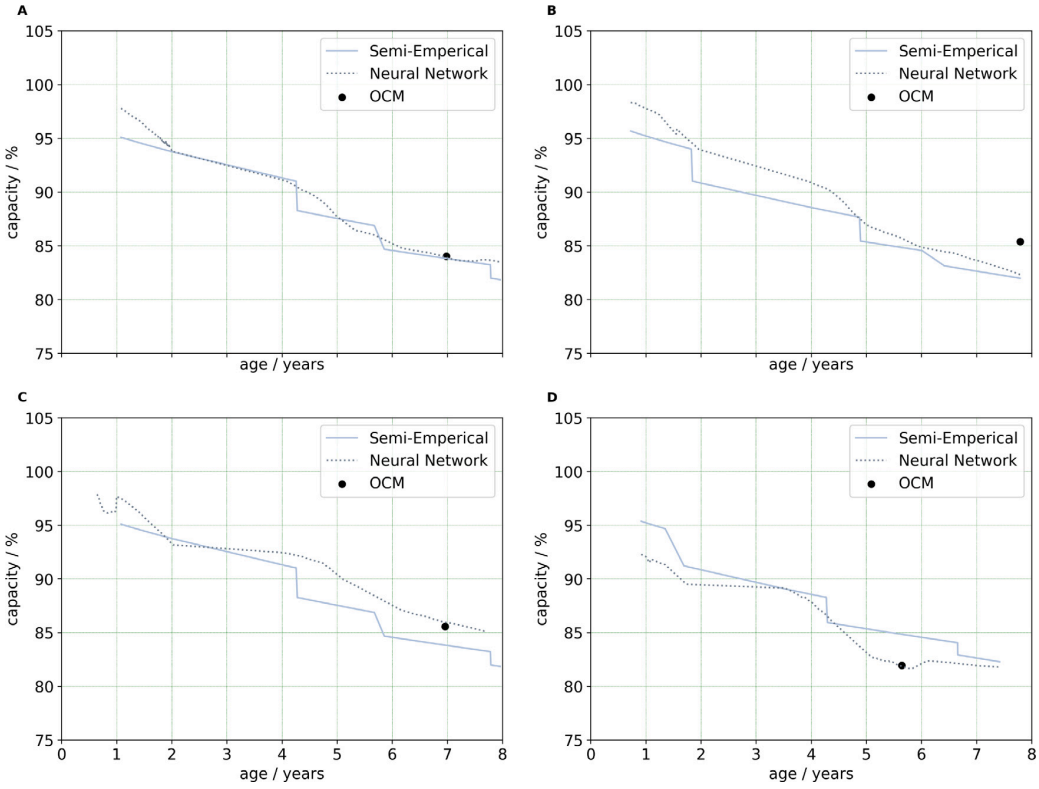

Fig. 6. The capacity estimation over the lifetime of four selected vehicles from the data-set (A–D) using the Semi-Empirical model (solid line) and the Neural Network model (dashed line) is shown. The black singular dot denotes an Off-board Capacity Measurement, and is considered the ground truth for validation; all other points are estimations.

each point is assigned a cluster and follows the corresponding fit of the Semi-Empirical model. The visible steps result when there was a change in the cluster.

The Neural Network model also exhibits a noticeable susceptibility to minor fluctuations in the input data. Whereas the Semi-Empirical model follows a pre-determined trajectory from the fitted function, using only time and the assigned cluster as the input for new measurements estimation, the Neural Network model uses the raw measured input features. As this is data recorded direct from vehicles in nominal conditions the input data can be highly dynamic, which directly impacts the SOH estimation as seen in the early estimation period in Fig. 6B and C. The accuracy difference between the two models is highlighted in Fig. 6. The average difference in SOH estimation from

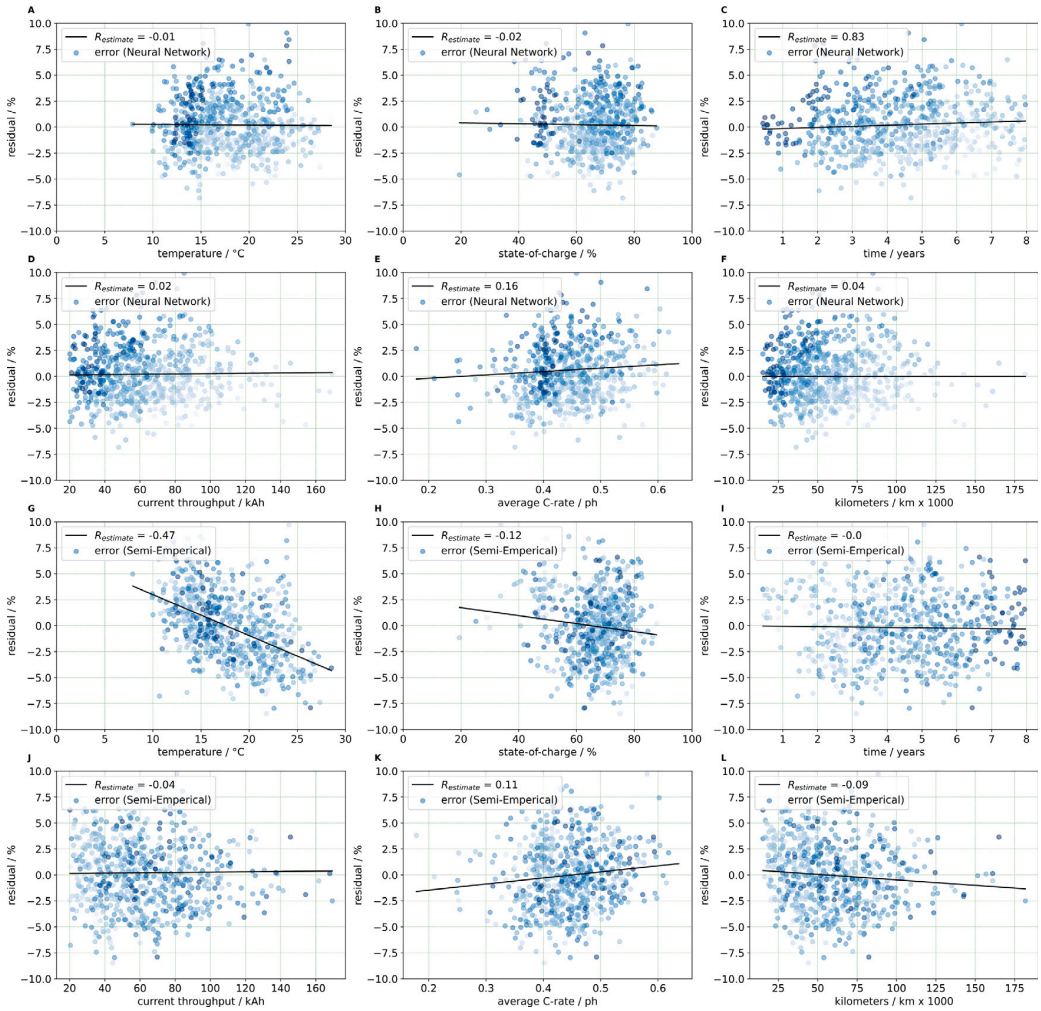

Fig. 7. The residual of the Neural Network model (A–F) and the Semi-Empirical model (G–L) is plotted against each of the available input parameters.

Table 4
Each model is evaluated in terms of its accuracy and hardware resource requirements.

Model	Accuracy (RMSE)/%	Storage Requirements/kB
Semi-Empirical	3.4	80
Neural Network	3.0	369

the Neural Network model to the Off-board Capacity Measurements for these four vehicles is 1.5%-SOH whereas the estimation difference for the Semi-Empirical model is 2.5%-SOH which reflects the larger trend from Fig. 5.

4. Conclusions

The main focus of this work was to provide insight into the modeling of automotive battery aging under nominal aging conditions. A more detailed interpretation of the model results and a critique of potential short-comings of the methods, as well as, a look into future work related to automotive battery SOH estimation using novel estimation algorithms on vehicle data follows.

When considering the accuracy of the presented models, it was assumed that the OCM were a reliable ground truth and that the input feature data was accurate. This study did not focus on an evaluation of the data in terms of quality or to interpret the aging data to analyze driving behavior. The datasets were evaluated to show the location and distribution of the values contained in each feature vector, however, since the normality of the predictor variables is not a pre-requisite for regression problems, the location and skew of the data was not used for further model development [31]. As such the effects of changing driver behavior, or sampling frequency is not evaluated. Although this is certainly relevant for battery aging models, especially when considering path dependent aging behavior, and as such, more detailed investigation into the influence of driving behavior is saved for future work.

With respect to the model performance, the goal of this work is to provide competing models achieving a similar level of accuracy, and was not to promote a single best-method. Each model has its respective advantages and disadvantages. The Semi-Empirical model with a clustering algorithm achieved an RSME value of 3.4% which is comparable to similar proposed aging models [5,6]. As seen in Section 3, a significant change in the SOH estimation occurs at the boundary between two clusters. This step could be reduced with a higher resolution of clusters, however, as the available training data-set contained 704 samples, increased clustering resulted in each cluster being less relevant to an actual aging mode. Additionally the Semi-Empirical model tends to over-estimate SOH values below 85 %-SOH and underestimate values above 95 %-SOH (see Fig. 5).

This skew represents poor training performance of the model. This Semi-Empirical model includes a cyclic and calendric model, however, due to the highly dynamic aging behavior inherent in automotive conditions, the traditional semi-empirical methods do not have enough degrees of freedom to capture all potential aging modes. Semi-empirical methods are better suited to fitting laboratory data where the aging behavior is more rigorously controlled. One potential error source was identified in the residual plot of Fig. 7G, which suggests that the temperature dependence is not adequately captured by the Semi-Empirical formula. As these models are reliant on the underlying physics being described by the formula, they are highly influenced by the complexity of that physical process. As mentioned, the Arrhenius exponential dependence on temperature was used in the Semi-Empirical model following the advice of numerous papers on the subject [6,9–11]. It was not within the scope of this paper to identify a more suitable relation to the temperature which may better fit the data, however the asymmetric distribution of the model prediction error plotted over temperature highlights that the Semi-Empirical formulation would benefit from further investigation. One possible source of error is that the activation energy term used in the exponent is estimated from literature, and may not reflect the true activation energy of the cell chemistry used in the vehicles investigated.

One motivation for introducing a clustering feature for the Semi-Empirical model was to help reduce the overall variance in the training data, which would improve fitting performance by collecting similarly aged vehicles together similar to how laboratory measurements focus on distinct aging profiles. Another factor further motivating the use of K-means was to help deal with the high level of variation in the input data. By using the average values ($\overline{U}_c$, $\overline{T}_c$, $\overline{DOD}_c$ and $\overline{Q}_c$) instead of the measurement values at the current time-step (U_i, T_i, DOD_i and Q_i), the model produced much more reliable and stable results.

One significant advantage, however, is that the fitted Semi-Empirical model can be efficiently stored with 1 kB, and coupled with the existing K-means algorithm, a modest 80 kB are required, compared to the 369 kB required for the Neural Network model. In addition to the storage requirements, the output of the Semi-Empirical model is predictable and stable, as the output– depending only on the cluster and time– will follow a given trajectory. One could also assume that for a given amount of time in the future, the cluster would remain the same, and have a simple prediction for the SOH. On the data-set provided, a Semi-Empirical model with out a clustering algorithm would not have a reasonable accuracy as there is considerable variation in the dependent variable and the input features. A semi-empirical model is best suited to more static conditions, such as in laboratory experiments, where factors such as temperature, average SOC and charging conditions are kept within specified ranges. The improvement of the Semi-Empirical model proposed in this work by adding the k-means clustering (from an RSME value greater than 10 %-SOH, to the 3.4%-SOH presented here) helps highlight the role data-driven methods can have in battery state estimation.

The Neural Network model is not without its own advantages. This model allows for a much finer estimation as it is not restricted to a set amount of clusters, and is able to capture more complicated inter-dependencies of the input variables than the holistic aging function provided by the Semi-Empirical model, as demonstrated by the lower RSME value of 3.0% and the symmetric distribution of residuals over temperature as seen in Fig. 7A. This improved accuracy is likely to increase further given a larger data-set as the neural network is able to extract more meaningful connections between neurons, however, it also comes with a cost of higher computational resources and storage requirements. In addition, as seen in Section 3, the model can behave unexpectedly given a more dynamic input, which, under nominal operating conditions could be expected. There are methods to reduce this effect such as: training the model with noise, filtering input signals and filtering the estimation output.

It is also presumed that the accuracy using neural networks should improve as the amount of input data increases. A small investigation into this was performed during model development by first training models on only 100, 300, 500 and finally all available 704 samples, to understand how accuracy is changed with increasing data. The results of this study showed that both methods were equally poor in estimation performance when 300 or fewer training samples were available, with an RMSE of less than 6 %-SOH. At 500 samples, the Neural Network model started to out-perform the Semi-Empirical model at 3.5%-SOH compared to 4.2%-SOH, finally, still showing the marginal gains seen above when all samples are made available.

The collection method of the data would also play a critical role in the models prediction ability, as some information could be over-represented or missing from the training data. For example, assessing the distributions in the CDF of Fig. 1, or the histograms in Fig. 3, it can be seen that the SOH values collected skew to higher SOH values, with few points available for aging less than 80 %-SOH. Similarly, the observable left-skew in kilometers and temperature may suggest the sample data over-represents lower-aged vehicles with fewer driven kilometers at lower temperatures, and may contribute to the residual error in temperature seen in Fig. 7G.

One of the major advantages of neural networks is their ability to capture dependencies with out explicit knowledge of these relationships. As demonstrated in this work, data-driven methods are well adept to estimating battery state parameters and are increasingly relevant the more data is available and the more complex analytic models become. When considering methods for automotive applications, the data-driven models can learn from a large collection of behavior from the fleet, however, the computational resources of these methods are not well suited for implementation on modern automotive controllers. For online and onboard applications, analytic models carry the advantage as these models can be stored efficiently in few bytes.

This study has focused on two novel holistic aging estimation algorithms for use on automotive battery data collected under nominal conditions which provide upwards of 3% estimation error. As the automotive sector transitions to a more sustainable and electric future, many companies, and customers can benefit from the insights into aging data to enable more accurate aging estimations and predictions.

CRediT authorship contribution statement

Jacob C. Hamar: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing. **Simon V. Erhard:** Conceptualization, Data curation, Writing – review & editing, Supervision. **Angelo Canesso:** Investigation, Writing – original draft. **Jonas Kohlschmidt:** Investigation, Writing – original draft. **Nicolas Olivain:** Investigation, Writing – original draft. **Andreas Jossen:** Conceptualization, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Energy Policy Review, International Energy Agency (Ed.), Germany, 2020, https://www.iea.org/t&c/, 2020.
[2] Xiao-Sheng Si, Wenbin Wang, Chang-Hua Hu, Dong-Hua Zhou, Remaining useful life estimation- a review on the statistical data driven approaches, European J. Oper. Res. (2011) 1–14, http://dx.doi.org/10.1016/j.ejor.2010.11.018, 2011.
[3] Wladislaw Waag, Christian Fleischer, Dirk Uwe Sauer, Critical review of the methods for monitoring of lithium-ion batteries in electric and hybrid vehicles, J. Power Sources (2014) 321–339, http://dx.doi.org/10.1016/j.jpowsour.2014.02.064, 2014.

[4] M. Berecibar, I. Gandiaga, I. Villarreal, N. Omar, J. Van Mierla, P. Van den Bossche, Critical review of state of health estimation methods of li-ion batteries for real applications, Renew. Sustain. Energy Rev. (2016) 572–587, 2016.

[5] Carlos Vidal, Pawel Malysz, Phillip Kollmeyer, Ali Emadi, Machine learning applied to electrified vehicle battery state of charge and state of health estimation: State-of-the-art, IEEE Access (2020) 52796–52814, 2020.

[6] Yi Li, Kailong Liu, Aoife M. Foley, Alana Zülke, Maitane Berecibar, Elise Nanini-Maury, Joeri Van Mierlo, Harry E. Hoster, Data-driven health estimation and lifetime predictino of lithium-ion batteries: A review, Renew. Sustain. Energy Rev. (2019) 1–18, 2019.

[7] Phillip Kollmeyer, Panasonic 18650pf li-ion battery data, 2018, http://dx.doi.org/10.17632/wykht8y7tg.1.

[8] Kristen A. Severson, Peter M. Attia, Norman Jin, Nicholas Perkins, Benben Jiang, Zi Yang, Michael H. Chen, Muratahan Aykol, Patrick K. Herring, Dimitrios Fraggedakis, Martin Z. Bazant, Setphen J. Harris, William C. Chueh, Richard D. Braatz, Data-driven predictino of battery cycle life before capactiy degradation, Nature Energy (2019) 383–391, http://dx.doi.org/10.1038/s41560-019-0356-8, 2019.

[9] E. Sarasketa-Zabala, E. Martinez-Lasema, M. Berecibar, I. Gandiaga, L.M. Rodriguez-Martinez, I. Villarreal, Realistic lifetime prediction approach for li-ion batteries, Appl. Energy (2016) 839–852, http://dx.doi.org/10.1016/j.apenergy.2015.10.115, 2016.

[10] M. Schimpe, M.E. von Kuepach, M. Naumann, H.C. Hesse, K. Smith, A. Jossen, Comprehensive modeling of temperature-deendent degradation mechanisms in lithium iron phosphate batteries, J. Electrochem. Soc. (2018) A181–A193, http://dx.doi.org/10.1149/2.1181714jes, 2018.

[11] Johannes Schmalstieg, Stefan Käbitz, Madeleine Ecker, Dirk Uwe Sauer, A holistic aging medel for li(nimnco)o2 based 18650 lithium-ion batteries, J. Power Sources (1865) 325–334, 2014.

[12] Robert R. Richardson, Michael A. Osborne, David A. Howey, Gaussian process regression for forcasting battery state of health, J. Power Sources (2017) 209–219, 2017.

[13] Kevin L. Gering, Sergly V. Sazhin, David K. Jamison, Christopher J. Michelbacher, Bor Yann Liaw, Mathhiew Dubarry, Mikael Cugnet, Investigation of path dependence in commercial lithium-ion cells chosen for plug-in hybrid vehicle duty cycle protocols, J. Power Sources (2011) 3395–3403, http://dx.doi.org/10.1016/j.jpowsour.2010.05.058, 2011.

[14] Laisuo Su, Jianbo Zhang, Jun Huang, Hao Ge, Zhe Li, Fengchao Xie, Bor Yann Liaw, Path dependence of lithium ion cells aging under storage conditions, J. Power Sources (2016) 35–46, http://dx.doi.org/10.1016/j.jpowsour.2016.03.043, 2016.

[15] E. Sarasketa-Zabala, I. Gandiaga, L.M. Rodriguez-Martinez, I. Villarreal, Calendar ageing analysis of a lifepo4/graphite cell with dynamic model validations: Towards realistic lifetime predictions, J. Power Sources (2014) 45–57, http://dx.doi.org/10.1016/j.jpowsour.2014.08.051, 2014.

[16] Maik Naumann, Michael Schimpe, Peter Keil, Holger C. Hesse, Andreas Jossen, Analysis and modeling of calendar aging of a commercial lifepo4/graphite cell, J. Energy Sour. (2018) 153–169, http://dx.doi.org/10.1016/j.est.2018.01.019, 2018.

[17] Heinz Wenzl, André Haubrock, Hans-Peter Beck, Degradation of lithium ion batteries under complex conditions of use, Z. Phys. Chem. (2013) 57–71, http://dx.doi.org/10.1524/zpch.2012.0170, 2013.

[18] Marius Bauer, Clemens Guenther, Michael Kasper, Mathias Petzl, Michael A. Danzer, Discrimination of degradation process in lithium-ion cells based on the sensitivity of aging indicators towards capactiy loss, J. Power Sources (2015) 494–504, http://dx.doi.org/10.1016/j.jpowsour.2015.02.130, 2015.

[19] Peter Bruce, Andrew Bruce, Practical Statistics for Data Scientists: 50 Essential Concepts, Vol. 2017, O'Reilly, 2017.

[20] Robert R. Richardson, Michael A. Osborne, David A. Howey, Gaussian process regression for forcasting battery state of health, J. Power Sources (2017) 209–219, 2017.

[21] J. MacQueen, Some methods for classification and analysis of multivariate observations, in: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, MSP ed., University of California Press, Berkeley, Calif, 1967, pp. 281–297, URL https://projecteuclid.org/euclid.bsmsp/1200512992.

[22] D. Arthur, S. Vassilvitskii, K-means++: The advantages of careful seeding, in: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, ACM-SIAM ed., SODA '07, Society for Industrial and Applied Mathematics, USA, 2007, pp. 1027–1035.

[23] Andriy Burkov, The Hundred-Page Machine Learning Book, 2019, URL http://themlbook.com/.

[24] S.S. Shapiro, M.B. Wilk, An analysis of variance test for normality, Biometrika (1965) 591–611, http://dx.doi.org/10.1093/biomet/52.

[25] S.V. Erhard, P.J. Osswald, J. Wilhelm, A. Rheinfeld, S. Kosch, A. Jossen, Simulation and measurement of local potentials of modified commerical cylindrical cells: Ii: Multi-dimensional modeling and validation, J. Electrochem. Soc. (2015) A2707–A2719, http://dx.doi.org/10.1149/2.0431514jes.

[26] Severin Lukas Hahn, Mathias Storch, Ramanathan Swaminathan, Björn Obry, Jochen Bandlow, Kai Peter Birke, Quantitative validation of calendar aging models for lithium-ion batteries, J. Power Sources (2018) 402–414, http://dx.doi.org/10.1016/j.jpowsour.2018.08.019.

[27] Python Software Foundation, Python, 2020, URL http://www.python.org.

[28] Francois Chollet, Keras, 2020, URL https://keras.io/.

[29] Francois Chollet, Deep Learning with Python, Manning Publication, 2018.

[30] Christopher H. Achen, Interpreting and using Regression: Series: Quantitative Applications in the Social Sciences, Sage Publications, Inc., 1982.

[31] John O. Rawlings, Sastry G. Pantula, David A. Dickey, Applied Regression Analysis: A Research Tool, Springer, New York, 1989.

5 Anode Potential Estimation in Lithium-Ion Batteries Using Data-Driven Models for Online Applications

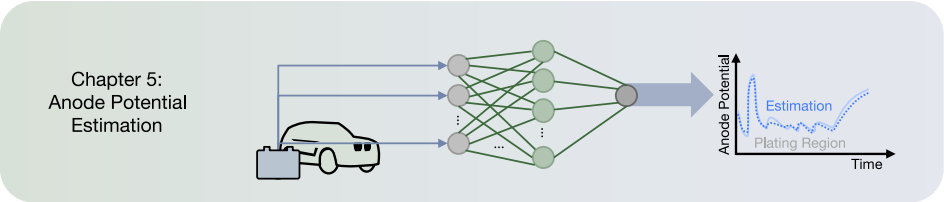

Figure 5.1: Crop from the thesis structure highlighting the focus of this chapter.

One of the biggest challenges facing many battery management systems is the inability to measure internal states of the battery. The anode potential, or potential difference at the graphite/electrolyte interface, is a highly relevant internal state parameter for improving fast charge performance. The anode potential is directly linked to lithium plating, which not only leads to accelerated aging of the battery but could also lead to a short circuit if the plated lithium dendrite structure grows long enough to puncture through the separator [181; 182]. The anode potential is of particular importance during fast charging, as fast charging conditions exacerbate the likelihood of plating onset. In fact, most fast charging algorithms are based on a regulation of the anode potential [183].

Lithium plating is caused by a break-down of the intercalation process. During intercalation, a charge transfer between the lithium-ion and graphite structure occurs allowing for a lithium-ion to settle between the layered electrode structure. Under conditions where this charge transfer is limited, such as low temperatures and high current flux, the inability of ions to settle into intercalation sites leads to a build up of ions at the interface. Finally, as the potential difference at the interface drops, lithium-ions become increasingly likely to attach directly to the graphite surface and each other, forming a dendrite structure, as shown in Figure 5.2.

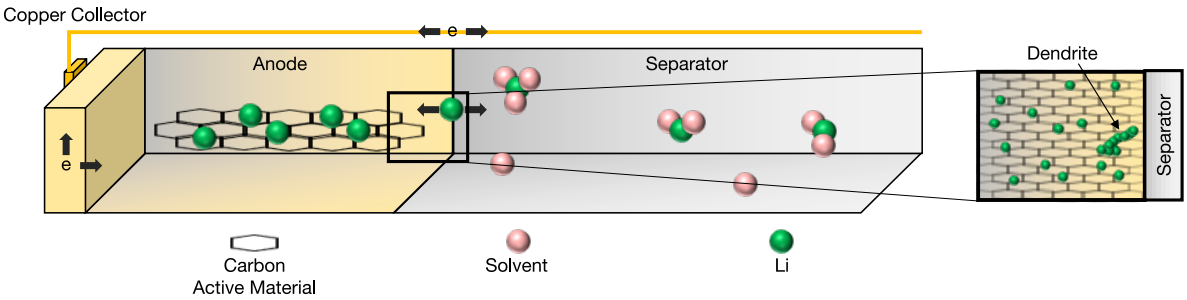

Figure 5.2: The lithium plating process occurs at the anode electrode interface and can result in dendrite formations puncturing the separator causing a short circuit.

Chapter 5 highlights the power of machine learning for internal battery state estimation, at least with respect to the anode potential. In this chapter a second application of machine learning is investigated, namely, the potential to improve fast charging. In order to reduce range anxiety in potential

customers, many auto manufacturers are working to reduce charging time to under 30 minutes [184]. One of the largest challenges to achieving shorter charging times is to avoid the potential lithium plating during fast charging [44; 143; 185–188]. Disadvantages of lithium plating are that it leads to a significant capacity fade and can also lead to a short circuit if the dendrite lithium deposit punctures the separator [49; 50].

Estimating the anode potential is often handled with electrochemical models or measured directly with invasive and sensitive reference electrodes [54; 189–191]. When the anode potential is known, batteries can be charged with higher currents, only reducing the current when diffusion becomes critical, thus decreasing the overall charging time [192–194]. The recent rise in data-driven state-estimation methods in battery development offers an alternative to these time and resource intensive methods [82; 88; 129; 178; 195–202].

As was recently shown by Lin et al., anode potential can be estimated using Long Short Term Memory neural networks [200]. This paper takes an empirical approach to fit a linear and non-linear equation and compare the results against a simple machine learning method, random forest, to better understand the accuracy versus complexity trade-off between various data-driven models. The training data in this paper was generated using an electrochemical simulation model [54; 189; 190; 203]. As a first step, a correlation analysis coupled with a variable inflation factor was performed to extract the relevant features for fitting the models [204; 205].

Acknowledgement I would like to thank all contributors to this paper, without whom this would not have been possible. A breakdown of each author's individual contribution is listed as per the Contributor Role Taxonomy (CRediT) guideline [169].

- Author: Jacob Hamar, Credit: Conceptualization, Methodology, Software, Validation, Formal Analysis, Data Curation and Writing.
- Co-Author 1: Simon V. Erhard,
  Credit: Conceptualization, Funding Acquisition, Writing- Review and Editing, Supervision.
- Co-Author 2: Christoph Zoerr,
  Credit: Data Curation, Formal Analysis, Software, Writing- Original. Draft.
- Co-Author 3: Andreas Jossen,
  Credit: Conceptualization, Supervision.

Anode Potential Estimation in Lithium-Ion Batteries Using Data-Driven Models for Online Applications

Jacob Hamar, Simon V. Erhard, Christopher Zoerr, and Andreas Jossen

Journal of The Electrochemical Society 168 (3), 030535, 2021.

Permanent Weblink:
https://doi.org/10.1149/1945-7111/abe721.

Journal of The Electrochemical Society, 2021 **168** 030535

Anode Potential Estimation in Lithium-Ion Batteries Using Data-Driven Models for Online Applications

Jacob C. Hamar,[1,2,z] Simon V. Erhard,[2] Christoph Zoerr,[2] and Andreas Jossen[1]

[1]*Technical University of Munich, Munchen, Baveria 80333, Germany*
[2]*BMW Group, Munchen, Bavaria 80788, Germany*

Three anode estimation methods are presented and evaluated for their accuracy and storage requirements. After generating training data using a Pseudo-2D Physiochemical model, these models are fit and trained to estimate the anode potential during fast charge events. A simplified linear and non-linear model show an estimationerror of ca. 13 mV and the lowest memory demand, however, a novel random forest model reduces the error to 2.6 mV. The empirical methods are suitable for a lithium plating warning detection system during fast charging and are further evaluated for over-fitting and robustness using an out-of-sample dataset.
 [DOI: 10.1149/1945-7111/abe721]

Manuscript submitted December 6, 2020; revised manuscript received January 16, 2021. Published March 22, 2021.

One of the most pressing customer concerns to the adoption of electric vehicles is the expected charging time. Auto manufacturers are developing fast charge protocols to bring the charging time of a 60 kWh battery (equivalent to 320 km) under 30 mins by 2025.[1] A major barrier to reducing the charging time of electric vehicles is the accelerated aging and increased safety risk which are inherent in higher current charging. Aging during fast charge results from multiple factors, including solid electrolyte interface (SEI) growth and the deposition of lithium on the surface of the anode, or lithium plating.[2,3] In order to develop faster charging profiles it is necessary to monitor the state of the battery to prevent the damaging effects from these aging mechanisms.

Of particular concern during fast charge is aging caused by lithium plating which reduces the available lithium for cycling and, in extreme cases, could lead to a short circuit due to the build up of lithium dendrite formations on the anode puncturing the separator.[3–5] Plating is believed to occur when the anode potential, $\Phi_{an,s}$ is reduced to near 0 V against Li/Li^+.[6] The equilibrium anode half-cell potential is decreased during charging as the electrode is lithiated. At the same time, due to diffusion and charge transfer limitations and ohmic loses, the over-potential is also increased, further decreasing the anode potential against Li/Li^+. With a reduced potential against Li/Li^+, there is an increased likelihood that the lithium ions do not intercalate into the anode structure, but instead deposit onto the surface at the electrode-electrolyte interface.[7–9]

It should be mentioned here, that lithium plating is not completely understood.[6] According to recent publications[10] the plating onset condition is likely to be a function of temperature, pressure and concentration. Still there is a thermodynamic dependency between lithium plating and the potential difference between the negative electrode and the electrolyte. Therefore the minimum potential difference—between the solid phase at the anode surface ($\Phi_s(x,\ t)$) and the liquid phase in the electrolyte ($\Phi_l(x,\ t)$) as a function of the electrode thickness (x-dimension) and time—is considered to be the best indicator for the occurrence of lithium plating, where the onset condition is considered to be $\Phi_s(x,\ t) - \Phi_l(x,\ t) \leqslant 0$ V vs Li/Li^+.

In automotive applications, there are no methods for directly measuring if lithium plating is occurring. Capacity loss resulting from plating can only be confirmed during a post-mortem analysis of the cell, where the formation of the lithium deposits can be directly measured. In practice, a reference electrode is inserted between the anode and the separator in the electrolyte in order to measure the potential difference between these two electrodes. The insertion of reference electrodes is expensive and impractical for many commercial applications. for this reason there are currently no sensors available which can reliably detect lithium plating in automobile applications.

One other method would be to build an accurate model of the lithium-ion cell to simulate the anode potential.[11–13] This method however requires considerable effort to parameterize the model and validate the accuracy, and with few exceptions (Sturm et al.[14]), these models are too large to operate on conventional battery management systems. For this reason, manufacturers are looking for alternatives to estimate or predict when plating might occur during a fast charge event using data-driven methods.

Data-driven methods offer an attractive alternative when coupled with sufficiently accurate, large, and representative data sets because they have the potential to accurately estimate the anode potential under a wide variety of conditions. Data-driven methods for modeling battery behavior have recently grown in popularity with contributions focusing on predicting the SOH or state-of-charge of the battery[15–21] It is to the best of the authors knowledge that there currently exists only one publication, by author Xianke Lin[20] which adapts data-driven methods to the prediction of anode potential. In their approach a long short-term (LSTM) algorithm was trained using a P2D model to generate training data from a range of constant current and standard driving profiles. The LSTM is a suitable algorithm because the memory feature in the neural networks is well suited to modeling time-dependent processes in battery dynamics. The goal of this paper is to addresses the assumption that the memory feature of the LSTM is necessary in data-driven approaches and to evaluate a range of suitable regression algorithms specifically in the context of fast-charging to better understand their computational efficiency, accuracy and suitability for online implementation.

In the following sections three methods of varying complexity and computation requirements are presented and evaluated based on their performance estimating the anode potential during fast-charging. The core of this work is detailed below. In this section the data generation procedure, feature selection process, algorithm development, and the validation method will be discussed. The results of the work and their discussion can be found in at the end of the work and, along with the conclusion, will provide a summary and suggestion of the next steps.

Method

In this method, a conventional P2D model is used to generate training and validation data. It is not the goal of this paper to investigate the accuracy of P2D models but rather use a validated model to train data-driven models (See Fig. 1 for an overview of the method). As such, it was assumed that the P2D model used in this paper accurately represents the cell behavior and is suitable for training a data-driven model. More information regarding the P2D model used can be found in Appendix.

[z]E-mail: jacob.hamar@tum.de

Journal of The Electrochemical Society, 2021 **168** 030535

P2D model parameterization.—The use of a P2D model offers a reliable estimation of cell behavior including: current, voltage and temperature, as well as other variables which can not be easily measured, such as anode potential. In addition, the P2D model offers a quick and flexible alternative to data generation when compared to laboratory measurements. The P2D model used in this paper is based on the Newman model[11–13] The parameterization set (see Table IV), is derived from a lithium-ion cell with an nickle-manganese-cobalt (NMC) cathode. For a deeper discussion into the equations used in the P2D model see the Appendix or Ennifar et al.[22]

Training data simulation.—The anode potential estimation models were trained and validated on P2D simulation data. This has the significant advantage that data can be collected quickly and cost effectively compared to laboratory measurements. One of the largest drawbacks, however, is if the simulation conditions or model do not adequately capture the cell performance under real-world conditions. To help mitigate any error arising between simulation and real-world conditions it would be advisable to include measurement data along with the simulation data for model training. One difficulty of generating measurement data for model training though is that, in addition to significant time and cost considerations, the anode potential is not easily measured during operation. Due to time and cost restrictions, it was not possible to include validation on measurement data.

The training data generated represents a range of expected operating conditions for fast-charge events which could occur during the normal lifetime of an electric vehicle. The training data matrix is composed of various starting SOC and temperatures which are chosen at intervals broad enough to avoid over-fitting, but also narrow enough to provide accurate and robust predictions. The profiles simulated were based on the multiple-step constant current charging procedure (MSCC). The MSCC is a charging profile designed specifically to avoid conditions inducing lithium plating by reducing heat generation and reducing mechanical stress when lithium diffusion is constrained.[6,23] In order to limit the required amount of training data, only charging (MSCC) conditions were used for training and testing. The simulation-space focused on profiles with high C-rates, as well as temperatures above 15 °C as these are likely profiles to be encountered in automobile applications. By not considering other charging profiles, temperature and currents, the algorithms performance can only be evaluated within the variable ranges provided in Table I.

An overview of the initial conditions for the simulations can be seen in Table I. Each training data point (marked with an 'x' in Table I) consists of five fast charge profiles with varying initial currents, and changes in the step duration of the MSCC profile. These profiles were defined similar to those implemented by charge controllers in the automobile industry. Between the training intervals, out-of-sample validation data was generated in order to test the model performance on data not seen during training. These points are marked with an open circle "o" in in Table I. Interruption data, (marked with an "x" in Table I), were also simulated to further test the robustness of the data-driven models under unlikely operation conditions. These profiles include: linearly increasing current, current disruption, sawtooth current and a rapid current reduction. More detail on these profiles can be found below.

The limits of the training matrix exclude conditions not expected during normal operation, and which would be prohibited by most charging controllers. Initial starting temperatures below 15 °C were excluded due to the severe risk of lithium plating at cooler temperatures. For the specific Li-NMC cell simulated, currents above 2.0 C were considered to be outside of the normal operation range, and were therefore also excluded. Initial currents below 0.9 C and initial SOC above 65% do not qualify as fast charge events for the training of this model. The matrix resulted in 80 fast charge profiles, at four initial temperatures and four initial SOC, with five profiles simulated at each point.

Each simulation has a sampling rate of 1 Hz, resulting in 58033 samples of 7 independent variables and the simulated anode potential (see Table II for a full list of simulation variables).

Feature Selection.—The P2D model outputs several variables which could be used as inputs for a data-driven model (Table II). In this work, the same set of variables were used as independent variables for each of three the data-driven models. This section provides an overview on the variable selection process which includes a correlation analysis and a measure of collinearity.

When building a data-driven regression model, the independent variables must be related to the dependent variable. One of the most widely used indicators of this dependence, or correlation, between two variables, y and x, is the Pearson Correlation Coefficient, r. The Pearson Coefficient measures the degree of which the predictor variable increases (or decreases) as a response to an increase (or decrease) of an independent variable with values ranging from -1 to 1 indicating the strength and direction of the correlation.[24]

$$r = \frac{\Sigma_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{(N-1)\sigma_x\sigma_y} \quad [1]$$

where N is the number of samples, i is the index within the sample set, and $\bar{x}$ and $\bar{y}$ are the average of x and y. The variable σ is the standard deviation,

$$\sigma = \sqrt{\frac{\Sigma_{i=1}^{n}(x - \bar{x})^2}{N-1}}. \quad [2]$$

By plotting the available independent variables from the simulation against the dependent variable, see Fig. 2, and calculating the Pearson Coefficient, a few trends can already be seen. First, there are a number of variables which are highly correlated with the anode potential.

The variables with the highest correlation are: voltage ($r = -0.83$), capacity ($r = -0.64$), OCV ($r = -0.78$), SOC ($r = -0.64$), current ($r = -0.39$), temperature ($r = 0.30$) and time

Table I. Training matrix indicating the initial conditions for the P2D simulations. The P2D simulation generated the data necessary to train and validate the data-driven methods.

		Temperature/°C							
		15	20	25	29	33	38	43	
SOC/%	75	x		x		x		x	
	65		i				o		x—Training
	50	x		x		x		x	o—Validation
	30	i			o			i	i—Interruption
	25	x		x		x		x	
	15		o				i		
	0	x		x		x		x	

Journal of The Electrochemical Society, 2021 **168** 030535

Table II. An overview of the output variables from the P2D model which are selected from to use as independent variables for each of the data-driven models.

Variable Name	Variable Description
Cell Voltage	OCV plus over-potentials: ohmic, diffusion resistance and electrode kinetics.
Capacity	Integral of the charging current
OCV	Average equilibrium potential of cathode minus average equilibrium potential of anode.
SOC	Concentration of lithium ions in the cathode related to the maximum concentration.
Temperature	Temperature considered uniform between anode and cathode in the x dimension.
Current	Current at the current collector of the elementary cell.
Time	Simulation time.

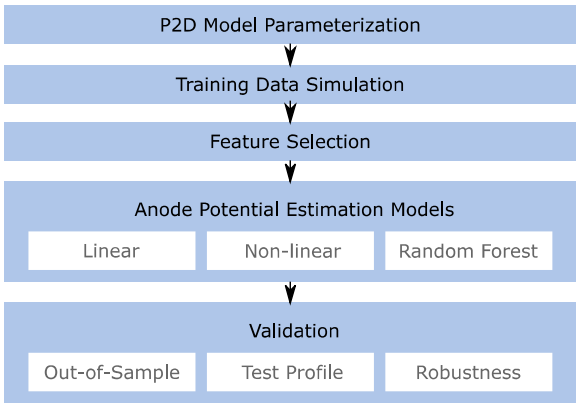

Figure 1. The proposed method includes five steps: (2.1) P2D Model Parameterization- adapting a validated P2D model to a known cell, (2.2) Training Data Simulation- generating data required for training and validating data-driven models, (2.3) Feature Selection- analysis of potential independent variables for regression models, (2.4) Anode Potential Estimation Models- development of multiple data-driven models, and (2.5) Validation- compare models against one another and their suitability for automobile applications

($r = 0.50$). The Pearson Correlation assumes a linear correlation, however, the dependence of anode potential on some of the factors may be better described by a non-linear relationship. The effects of assuming linear dependency is discussed at the end. The statistical hypothesis test, p, is also evaluated with a null hypothesis that the true correlation between the two variables is zero, with p-values less than 0.01 indicating a high probability that the correlation predicted is likely.

Figure 2 also highlights potential issues regarding collinearity between variables. Collinearity means that two or more independent variables are nearly linearly dependent on each other, which leads to the regression coefficients being unstable and sensitive to small random errors.[25] The output variables from the P2D simulation are tested for their correlation with the anode potential, and each other to check for collinearity.

When selecting the independent variables for the regression model the minimum number of factors required to explain the variance in the dependent variable should be selected. When independent variables are highly correlated with themselves there tends to be a large standard errors for the partial regression coefficients of the independent variables, and also a reduction in the statistical significance of both.[25] Removing collinear variables not only improves the stability of the model, it also reduces RAM and CPU required.

Aside from looking at the Pearson Coefficent between the dependent variables, estimating the Variable Inflation Factor (VIF) is the standard method for identifying dependent variables which are collinear within a model.[25] The VIF is calculated from the regression coefficient, r as follows, with larger values indicating a higher instability:

$$VIF = \frac{1}{1 - r_i^2} \quad [3]$$

It was found that when considering all variables in a multiple regression model: voltage ($VIF = 3.5 \cdot 10^2$), capacity ($VIF = 3.5 \cdot 10^6$), SOC ($VIF = -4.4 \cdot 10^2$), and current ($VIF = 3.5 \cdot 10^2$) are collinear, and as such only one should be selected as a factor in the linear regression model. Additionally, the low VIF for temperature ($VIF = 3.2 \cdot 10^0$) and current ($VIF = 3.1 \cdot 10^1$) indicate they describe unique variance in the residual of the dependent variable, and should both be considered.

Taking into account the Pearson Coefficient and the Variable Inflation Factor, three variables were chosen for the use in this study: voltage (U), current (I) and temperature (T). Voltage was chosen over similarly highly correlated factors, such as SOC and capacity because it is readily measurable during vehicle operation.

One omission to the potentially relevant set of predictor variables is the battery state-of-health (SOH). As the battery ages, the development and growth of various aging mechanisms will influence the charging behavior of the battery, for example, more heat will be generated due to higher internal resistance or a sudden roll-over in capacity loss.[26,27] In order to address the influence the change in the electro-chemical behavior has on the anode potential during fast charging, training data over a wider range of SOH could be incorporated, or, as many battery management units also provide an SOH estimation, having the SOH as an input parameter would also help to capture any variance caused by an aging cell.

Anode potential estimation models.—Three data-driven regression models were developed to predict the anode potential from multiple variables. Each method was implemented using R statistical programming software[28] and are described in detail in this section.

Linear model.—As a first approach, a multiple linear regression model was used to predict anode potential. Using three dependent variables: current, voltage and temperature, a simple model was fit on the training data and tested with an out-of-sample data set. Multiple linear regression is a widely applied method to generate fitting functions to data, however, they are most appropriate under a few conditions: the regression model is linear in its parameters, no perfect multicollinearity exists, residuals demonstrate homoscedastitcity and no autocorrelation.[29] Each of these points will be investigated in more detail in the later sections.

The proposed multiple linear regression equation for anode potential estimation is as follows,

$$\hat{\phi}_{an} = \hat{b}_0 + \hat{\beta}_U U + \hat{\beta}_T T + \hat{\beta}_I I \quad [4]$$

x where $\hat{\phi}_{an}$ is the predicted anode potential, U, T and I are independent variables, $\hat{b}_0$ is the y-intercept and $\hat{\beta}_U$, $\hat{\beta}_T$, $\hat{\beta}_I$ are the fitted coefficients for the dependent variables. The linear model in

Journal of The Electrochemical Society, 2021 **168** 030535

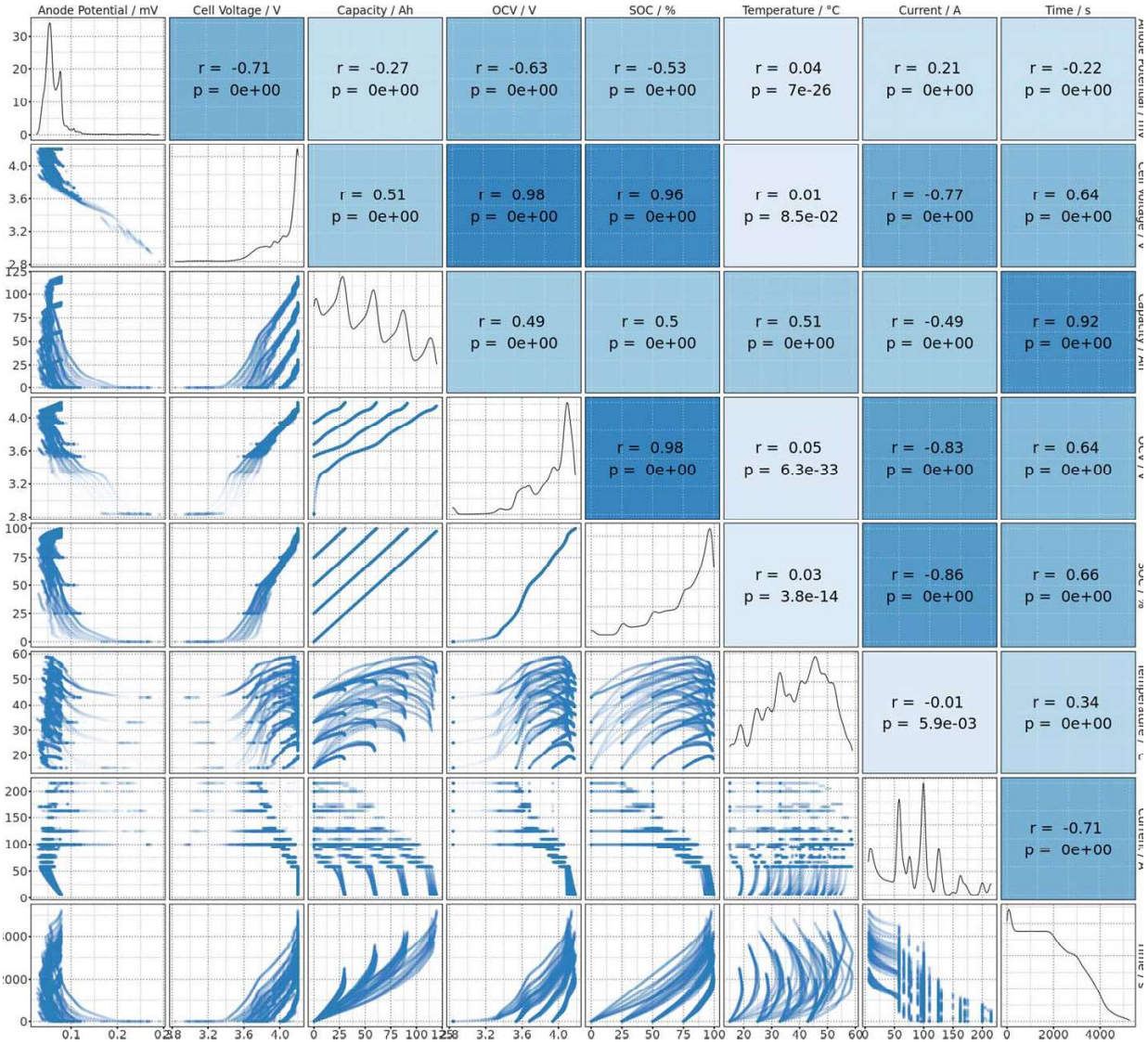

Figure 2. Using all of the result from the P2D model, each variable is plotting against the other to evaluate the correlation and collinearity of the independent variables. The darker the (blue) box, and the higher the r-value, indicate a higher correlation. The p-value is also measured to check the assumption that the variables are somehow dependent on each other.

this work was fit using a least square regression function. Least sum of square models minimize the residual sum of squares function,

$$RSS = \Sigma_{i=1}^{n}(Y_i - \hat{\phi}_{an})^2. \quad [5]$$

The resulting equation has the following regression coefficient fits: voltage ($\beta_U = -170$), temperature ($\beta_T = 0.11$) and current ($\beta_I = -0.44$), with a y-intercept of 796.

Although two parameters, voltage and temperature, show distinct non-linear dependence, the relatively high Pearson Coefficient suggests that there exists sufficient linear dependence to justify this assumption, however, this assumption will be further discussed with the non-linear model.

Non-linear model.—Non-linear models should be used when the independent variables show a non-linearity with respect to the dependent variable or a known coupling between two variables exists. As discussed in the previous section, voltage and temperature do not exhibit purely linear behaviors (see Fig. 2). The following section will introduce a non-linear model for predicting the anode potential and will serve to investigate in more detail the dependence of anode potential on voltage and temperature.

The same independent variables used for the development of the linear model were again used for the non-linear model. This method requires an assumption of the form of the model and is sensitive to the initial parameters. The form of the function was manually determined based on a visual analysis of the independent variable correlation with the dependent variable from Fig. 2.

The suggested model is the superposition of the three non-linear models from the three independent variables:

$$\phi(U) = b_1 + M_1 U + Asin(UB), \quad [6]$$

Journal of The Electrochemical Society, 2021 **168** 030535

$$\phi(T) = b_2 + \frac{C}{T}, \quad [7]$$

$$\phi(I) = b_3 + M_2 I. \quad [8]$$

yielding a combined model,

$$\phi(U, T, I) = b_0 + M_1 U + Asin(UB) + \frac{C}{T} + M_2 I \quad [9]$$

where b_0 ($b_0 = 797$) is the superposition of all three y-intercepts (b_1, b_2 and b_3), M_1 ($M_1 = 170$) and M_2 ($M_2 = -0.42$) are the slopes of linear equations and A ($A = 2.88$), B ($B = -623$) and C ($C = -115$) are constants, each fitting an independent variable (voltage (U), current (I) and temperature (T)) to the dependent variable ϕ.

The single variable non-linear models in Eq. 9 were chosen to model the dependency between the selected independent variable and the anode potential. The voltage shows an almost linear dependency with a slight sinusoidal perturbation. Temperature shows a reciprocal relationship with the anode potential. Finally, a linear model was chosen to represent the dependence on current as no discernible alternative was seen from the data (see Fig. 2).

While this method attempts to capture the non-linear influence of voltage and temperature, it is still limited by the ability of the chosen functions to mimic the physiochemical behaviors dictating the anode potential. Equations 6–8 are one possible set of candidate functions but there are potentially infinite functions which could yield a more accurate model.

In this approach, the form of the equation must be specified which would lead to a time-consuming process to manually identify the most accurate combination of variables and their exponents. One alternative approach would be to use Principal Component Analysis,[24] which automatically condenses all available variables into a new set, drastically reducing the parameter space for model identification. The focus in this work however, is to introduce the multiple non-linear approach and validate its applicability for online anode potential estimation, and as such a more in depth optimization is left for a future work.

The accuracy of this non-linear model and the comparison against the linear model is continued in below.

Random forest model.—Random forest models are often the go-to algorithm for many machine learning regression problems because of their intuitive implementation, robust accuracy and applicability to a diverse problem-set.[30] The random forest method is chosen in this paper as an example algorithm to highlight the core trends of increasing model accuracy and complexity when switching from linear/non-linear models to machine learning methods. Many common programming languages including R, Python and Matlab have built-in packages to make using a random forest algorithm easily accessible. Other methods, such as neural networks, would also be applicable, however, these methods are often better suited to computer vision tasks and over-complicate simple regression problems. In this section the general theory of random forest models will be explained, as well as, the specific implementation with the P2D simulation data.

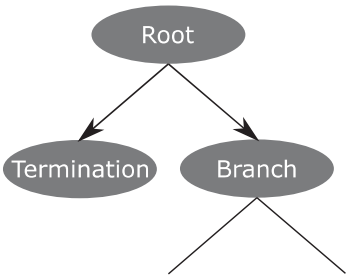

Figure 3. A random fores is comprised of multiple decisions trees which divide random partitions of the training data into branches based on binary splitting conditions which then result in a final value a the termination node.

A random forest is a collection of decision trees which are assigned a random partition of the overall data-set. In this section, the basic principles of a random forest model are discussed; for a more comprehensive review, refer to Murphy et al.[31] A decision tree, shown in Figure 3, stems from the root, splitting into multiple branches and ending at a termination node, or leaf. The entire data-set (or the random partition of the data-set, in the random forest case) starts at the root. The decision tree algorithm then evaluates all possible binary splits in the node which result in the lowest variance of the dependent variable according to

$$\min e(x_i) = \frac{1}{n}\Sigma_i^n (x_i - \bar{x}_i)^2 \quad [10]$$

where e is the residual or variance of the dependent variable x, n is the number of samples in the split and $\bar{x}$ is the average of the variable. The decision tree continues to split into branches with fewer samples in each node until an end criterion is reached-typically a specified number of branches. If there is no split at a node (or leaf), then the there is a value assigned based on the function

$$m(x) = \Sigma_i^l (k_i I(x \epsilon (D_i))) \quad [11]$$

where m is the average, k is a constant, $I(.)$ is a binary operator returning 1 or 0, l is a leaf and D is a randomly selected partition of the original data, which gives a fitted average of all the dependent variable values grouped in the termination node. A forest is then the

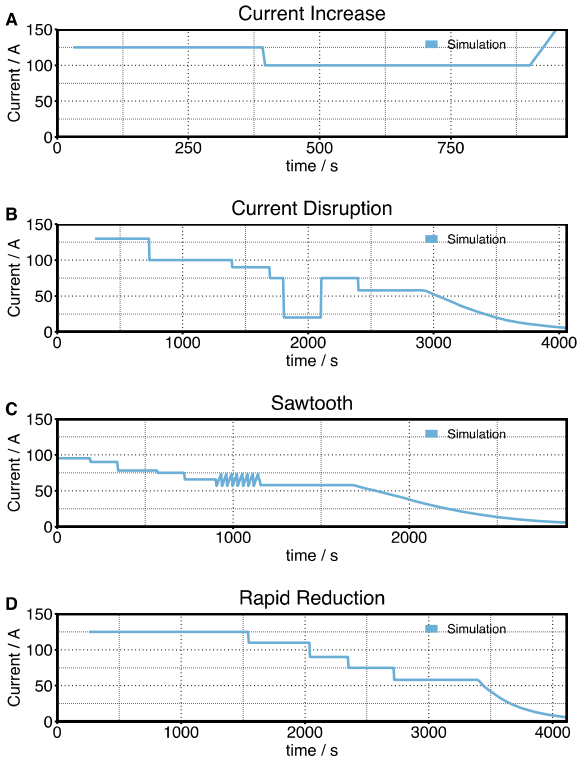

Figure 4. Four current profiles were used to test the robustness of the data-driven models to uncharacteristic charging behaviors. (A) current increase near end of charging, (B) current disruption for 6 minutes during charging, (C) sawtooth wave form and (D) a rapid reduction in steps.

Journal of The Electrochemical Society, 2021 **168** 030535

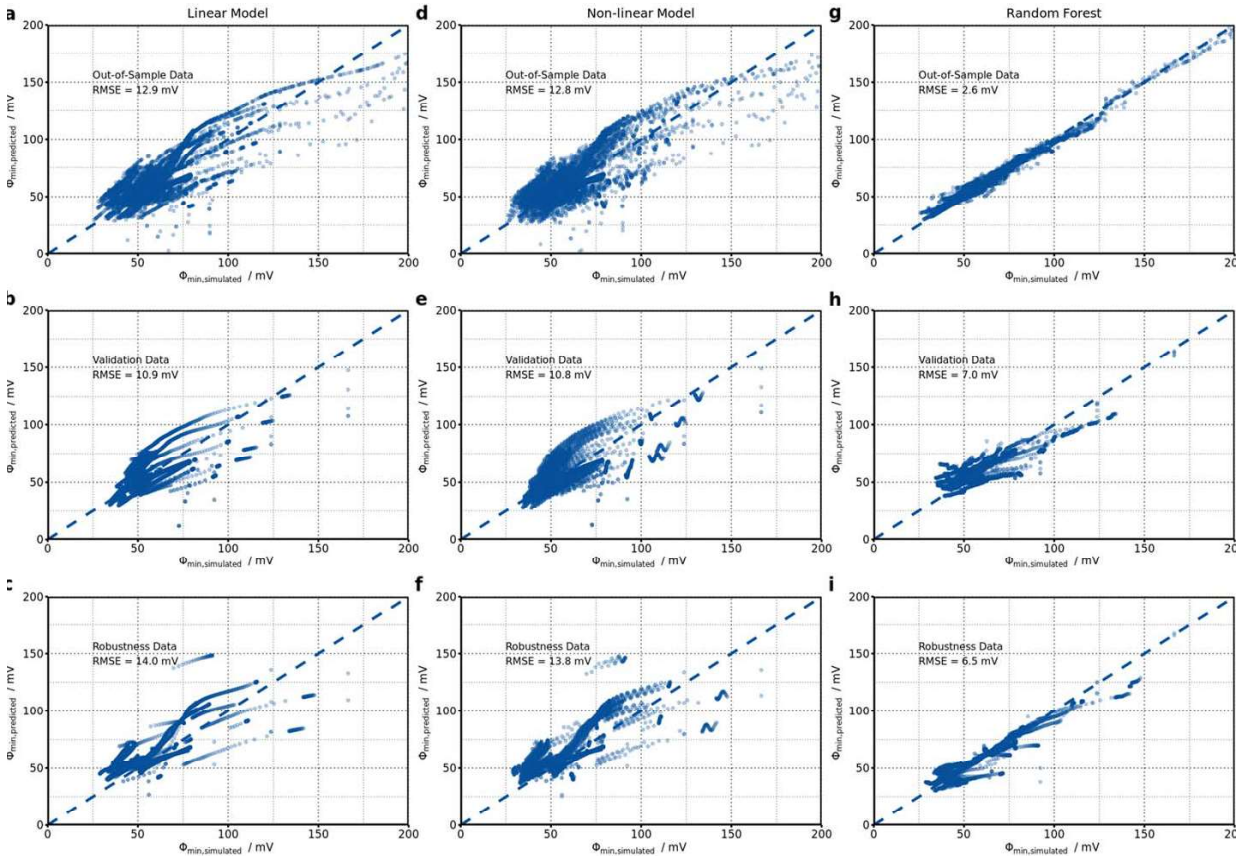

Figure 5. Validation matrix showing the predicted vs simulated anode potential over the three validation data-sets for the three models. (a-c) show the results for the linear model, (d-f) show the results for the non-linear model and (g-i) show the results for the random forest model.

combination of the results from many decision trees, each being initialized with a random partition of the overall dataset. This distributed method helps the algorithm achieve a minimum cost with reduced over fitting to the dataset as well as a reduced computation cost per tree. The number of trees, branches and the cost function are typical hyperparameters which can be optimized for a given problem. For the estimation of the anode potential, an intuitive explanation of how the algorithm works is as follows. First, the training dataset is broken up into multiple random partitions each including measurement samples of voltage, temperature, current and the simulated anode potential. From this first branch, a binary split occurs on one of the input features, for example, all measurements with a temperature below 20 °C would follow the left branch, and all above would follow the right branch. This binary branching will continue following the criterion from Eq. 10 until the designated number of branches is reached. At the end of this tree, there are groups of readings all with similar voltage, temperature and current, and it is from these samples that an average is taken from their simulated anode potential. When a new reading is fed to the random forest model, it will follow the binary split until it reaches the termination node and will pull the estimated anode potential from this leaf. In this respect, a random forest is a highly efficient algorithm at generating a large look-up table.

Validation.—Since no anode potential measurement was available for validation, the ground truth used to evaluate the model accuracy was the anode potential simulation from the P2D model. In order to evaluate the performance of the various models, three groups of out-of-sample data sets were generated along side the training data using the P2D model. In addition to splitting the training data into two groups for training and validating the random forest model, a second validation data set was generated at temperatures and starting SOC which were not included in the training data set for an additional investigation into model performance on unknown data. A third validation set was also generated, again at temperatures and starting SOC outside of the training data set, with charging profiles behaving significantly differently from those used in the training dataset. The initial conditions of these two groups of validation sets can be found in Table I. The profiles are described in more detail below and the results can be seen in the following section.

Out-of-sample.—In this method, all data used for the training of the model was split into two groups. A random selection of 75% (43, 524 samples) was reserved for a new training data-set and the remaining 25% (14, 509 samples) for validation using the base *sample* function in R. Each model was then re-trained with the smaller training data set. This approach helps to identify any autocorrelation in the prediction because the validation data no longer contains any time dependence.

Profiles.—A group of simulations were also run at starting SOC and temperatures which were not included in the training data set (marked as "o" in Table I). These profiles help to test the sparseness of the training matrix to better understand if a finer mesh is needed for accurate prediction.

Robustness.—The last group of profiles are intended to test the response of the prediction algorithms to severe interruptions and

Journal of The Electrochemical Society, 2021 **168** 030535

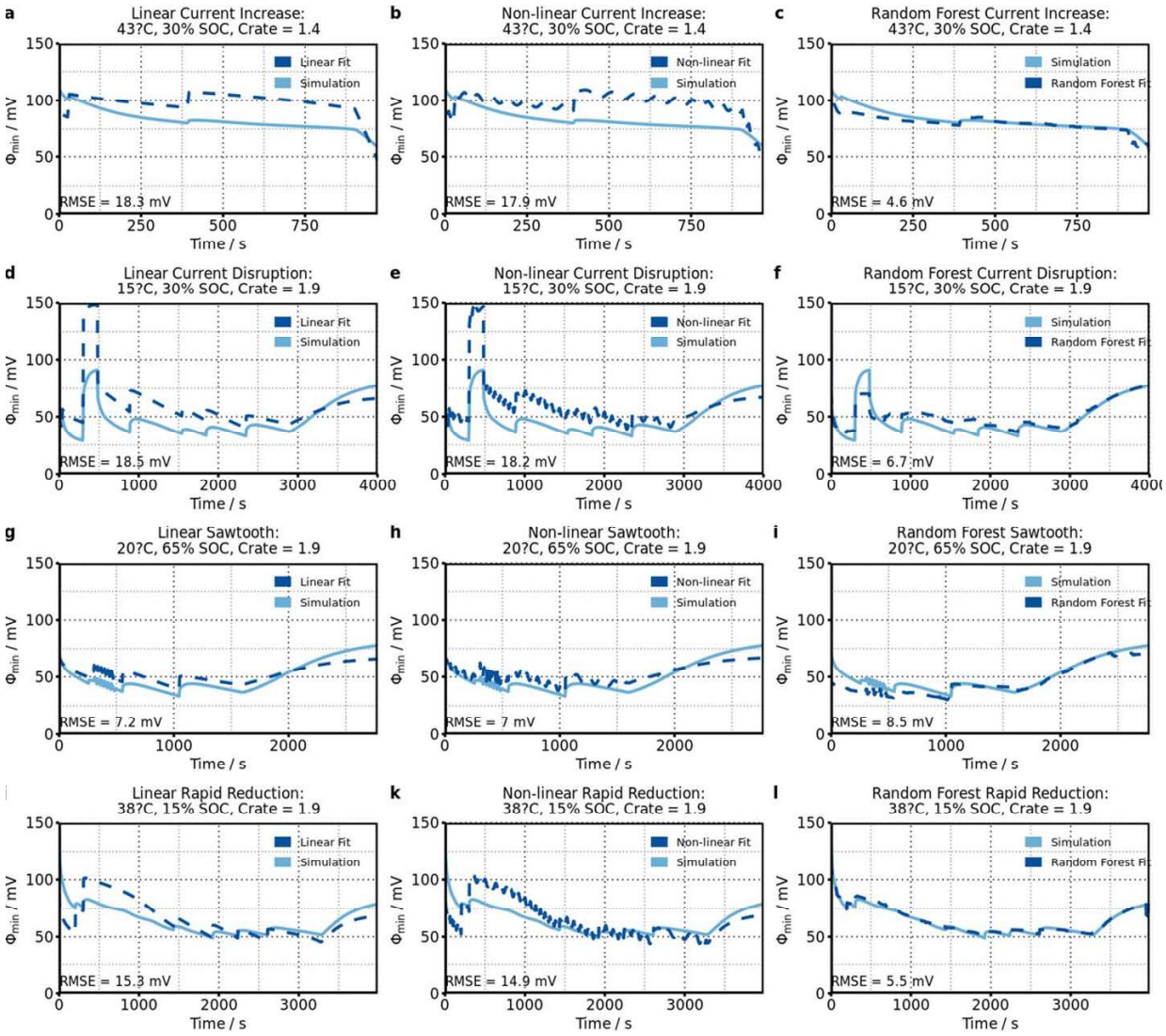

Figure 6. The results of the robustness tests: (a-c) Current Increase, (d-f) Current Disruption (g-i) Sawtooth, (j-l) Rapid Reduction—for each of the tree models: (a,d,g,j) linear model, (b,e,h,k) non-linear model, (c,f,i,l) random forest.

unexpected cell behavior not captured in the training data. These are meant to be worst case scenarios which should be detected and avoided by the charge controller. Four interruption charge profiles were used: current disruption, sawtooth, increasing current, and a rapid reduction. Each of the profiles can be seen in Figure 4. For the current disruption case, current drops to zero. The sawtooth profile has an amplitude of 10 A and a frequency of 1 min^{-1}. The increasing current profile, increases with a slope of 10 $A \cdot min^{-1}$. Finally, the rapid reduction profile switches from between multiple profiles in quick succession. The robustness profiles have various starting SOC and temperatures according to the training matrix in Table I (robustness profiles are marked as "i"). Each disruption lasts for 5 minutes, except for the last case, where the profile simply changes. A total of 16 robustness profiles are simulated.

Results and Discussion

Three models were developed to estimate the anode potential during a fast charge event. Each model was fit using the same data-set generated from the P2D model describe in the Appendix. In order to compare model accuracy, the RSME was calculated for the predicted values against simulation values using three sets of validation data detailed in the previous section. The RSME is calculated by

$$RMSE = \frac{\sqrt{\Sigma_{i=1}^{2}(y_i - \hat{y}_i)^2}}{n}, \qquad [12]$$

where y_i is the dependent variable, $\hat{y}_i$ is the prediction, n is the number of samples and i is the measurement index.

The results for each model over each testing method are shown in Fig. 5. The multiple linear regression model has the highest average RMSE over the three validation sets with the highest RMSE (14.0 mV) shown when predicting the robustness profiles. Marginal improvement is seen using the non-linear model with an RMSE value of ca. 0.2 mV improvement compared to the linear model. The lowest RMSE value results when using the random forest model with 500 trees, showing

Journal of The Electrochemical Society, 2021 **168** 030535

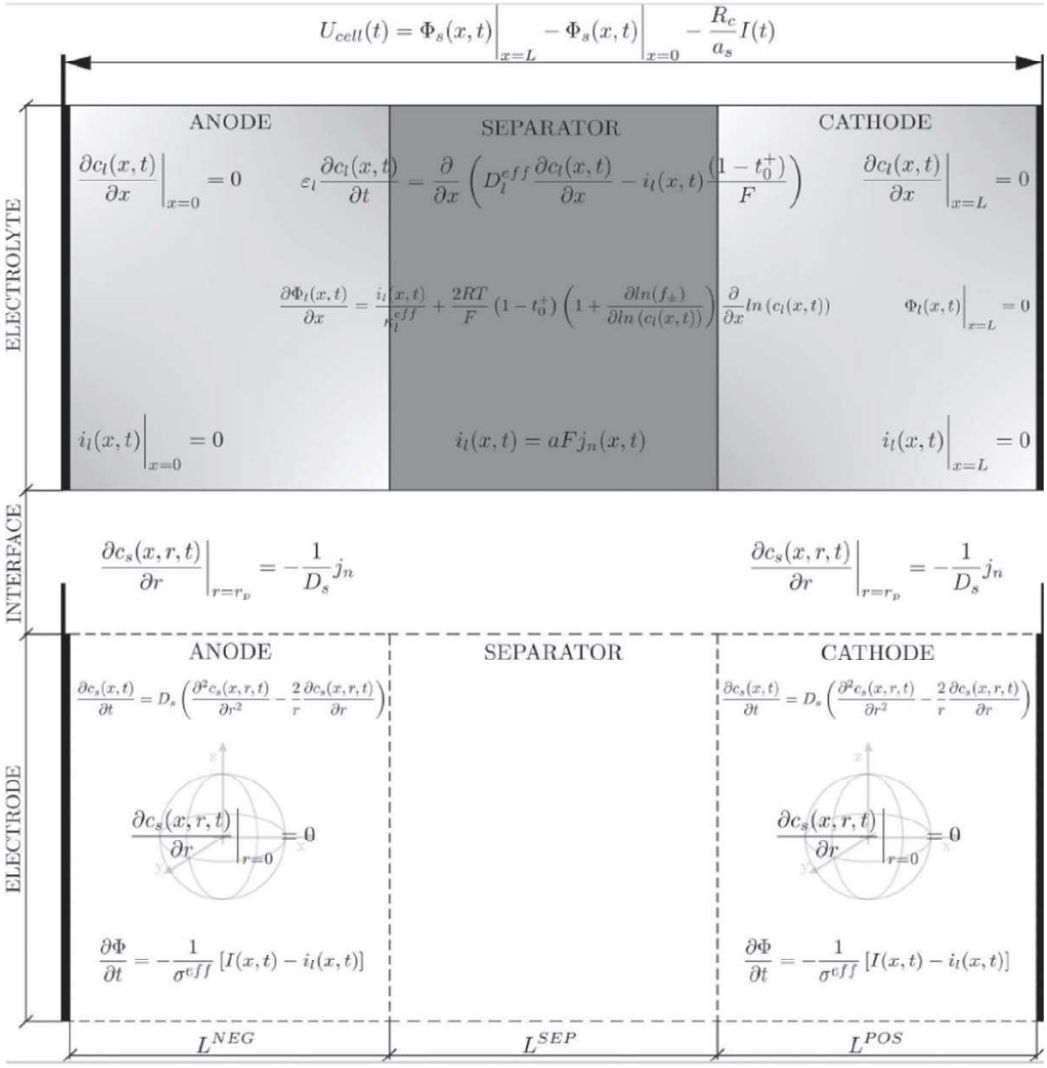

Figure 7. Dimension related equations of the so-called Newman-type P2D model (According to from Ref. 32).

less than 2.6 mV accuracy when tested using the out-of-sample method. The results testing the validation and robustness profiles are similar at 7.0 mV and 6.5 mV, respectively.

The linear and non-linear models show poor prediction accuracy especially at higher anode potentials, which occur at the beginning of a charging event (see Fig. 5). The prediction error was also particularly poor for these two models when considering higher temperatures and lower charge rates as seen in Fig. 6 for the charging profile at 43 ° C and 30% SOC with an initial charge C-rate of 1.4 $h^{-}1$. This is likely due to the fact that these profiles exhibit a relatively high anode potential, where these two models are known to perform poorly, as seen in the RMSE error plots. This is likely due to the fact that the bulk of the training data is in regions with lower anode potentials, where the average anode potential of the training data is 63 mV potentials leading to an under-fitting at higher anode potentials. This discrepancy also results in higher errors at the onset of fast-charging.

For the non-linear and linear fit models, there is higher inaccuracy at the on-set of the charging event (within the first few minutes), with improved accuracy as the voltage stabilizes. The accuracy decreases again during the CV phase. The inaccuracy at anode potentials above ca. 50 mV is most certainly a result of the high cell dynamics within the cell as it responds to a sudden change in condition. This behavior is dominated by short-term ohmic resistance and the ability of the electrode to accept more ions (concentration gradients). As the cell warms and the diffusion reaches equilibrium, the dynamics stabilize.

Since there is little improvement by assuming a non-linear model it could be concluded that the assumption of sufficient linear correlation between the independent variables and the dependent variable is valid. It should also be mentioned that the non-linear model could be potentially improved by choosing a different fitting function, however, capturing the highly irregular behavior of the anode potential with one function would be a daunting task, which is one reason why machine learning methods, such as the proposed random forest model are so powerful, because they are able to model this non-linear behavior with no prior information to this behavior required.

The linear model, and the non-linear model were significantly out-performed by the random forest model. The average error from Fig. 5 does not capture large prediction errors which occur during a few of the simulated fast charge profiles. The maximum error of the random forest model (using 250 trees) on the simulated robustness profiles was

Journal of The Electrochemical Society, 2021 **168** 030535

Table III. Training matrix initial conditions for P2D simulation.

Model	RMSE	Training time	Prediction time	RAM
Linear Model	12.9 mV	0.05 s	0.001 s	3 kb
Non-Linear Model	12.9 mV	2.4 s	0.001 s	8 kb
Decision Tree, 1 Tree	22 mV	16 s	0.01 s	1068 kb
Random Forest, 30 Trees	15 mV	41 s	0.08 s	7288 kb
Random Forest, 100 Trees	11 mV	130 s	0.10 s	21416 kb
Random Forest, 250 Trees	5.5 mV	348 s	0.15 s	52034 kb
Random Forest, 500 Trees	2.6 mV	660 s	0.15 s	103382 kb

15 mV, also during the current disruption. As can be seen from the results in Fig. 5, this method yielded highly accurate results, even when considering profiles well outside of the expected use conditions.

The random forest model significantly out-performs either the linear or non-linear models. This increase in accuracy, however, comes with added computation cost. For both the linear and non-linear models less than 10 kb of memory is required compared with 103 Mb for the random forest model with 500 trees. The offline fitting time of the models, as well as the estimated online prediction time are less than 10 ms for the linear and non-linear models. Depending on the desired accuracy and storage capability, the random forest model accuracy improves exponentially whereas the RAM, training time and prediction time increase linearly. A summary of these results can be seen in Table III. As a comparison, an on board P2D model presented by Sturm et al.[14] would require 10 Mb RAM and one minute to compute. The on board model, however, consists of a system of highly non-linear differential equations, each requiring a time consuming order-reduction to be able to run on a micro-processor. An alternative machine learning approach was taken by Lin et al., which results in ca. 3.5 mV error using long short-term memory regression algorithms.[20] Once trained, their algorithm could predict the anode potential, using the same input variables used in this work, with an on-board algorithm requiring 345 Mb.[20] The accuracy of the linear and non-linear models show, however, that a sufficient early warning detection for lithium plating could be implemented with minimal computation costs and modeling effort as it could detect anode potentials nearing a threshold. The low accuracy might render this method ineffective for producing optimized fast-charge protocols which are highly sensitive to the anode potential accuracy.

Conclusions

With a prediction accuracy between 2.6 mV and 12.9 mV the models would be able to alert the battery control module as the anode potential nears 0 mV against Li^+. As plating does not occur instantaneously as the anode potential dips below 0 mV, a warning system with any of the above the proposed anode potential prediction models would be capable of detecting sudden drops or anode potentials within a threshold, and as such it would be sufficient to alert the controller to a risk of plating. As plating is a long-term process, by estimating the amount of time the battery has a low anode potential, the controller can track and estimate the severity of plating over time. The more accurate the estimation is, for example with the random forest estimation, the better the plating accumulation estimation will be. The low computation and storage requirement and estimation time, even from the more complex random forest models, indicate good online applicability. For the non-linear and linear models, the storage requirement of only a few kilobyte means that there should be no difficulty embedding the function of a battery management unite. Although battery management units are used in a wide variety of applications with different capabilities and limitations, the 100 Mb required by the 500 tree random forest would be more challenging to implement in automotive applications as there is a strict limit and competition for available controller space. In order to further improve the linear and non-linear models, the heteroscedasicity of the model should be considered.[29] Heteroscedasicity refers to relationship of the residuals of the independent variables to the dependent variable. An appropriate model will show the same random variation, or noise, across all variables. Typical factors influencing the heteroscedasicity of a model are autocorrelation, missing independent variables and skew in the independent variable distribution.[29]

Further research will include an expanded investigation into the training data used for the machine learning model development, the incorporation of a lag factor to offset minor autocorrelation dependencies and hardware validation of the machine learning algorithm in a real system. Additionally, a study on the applicability of this method for use over the life-time of a battery will also be investigated in including SOH as an independent variable.

Acknowledgments

This work was funded by the BMW Group and was performed in cooperation with the Technical University of Munich.

Appendix: P2D Model

The literature reveals plenty of models to describe the behavior of a Li-ion battery.[33] In this work the generally accepted and extensively discussed Newman-type physiochemical model is used. The so-called pseudo-two-dimensional (P2D) model describes the Li-ion cell on a macroscopic scale (x-dimension) with two porous insertion electrodes, an insulating separator and the liquid electrolyte. The diffusion of Li-ions within the electrode's active material particles is modeled by an additional pseudo-dimension (r-dimension).

The DUALFOIL model, which is a Fortran based version of the P2D model, and is derived from the work of Doyle[11] is taken as a reference for this study. The isothermal model is coupled to a zero-dimensional thermal model to treat the temperature dependency of the reaction kinetics, and the transport parameters accurately. Since the Newman-type model approach is well-known in literature, it will only be given a short introduction to the key equations of the model here. A list of all P2D parameters and thermal model parameters used in this model and their values are given in Table IV and Table V. Thermalparameters assumed for the NMC-graphite cell respectivly. A more precise description is given in the publications of Doyle, Fuller and Newman.[11–13]

In general, the dependent variables of the P2D model are considered to be the Li-ion concentration in the liquid and in the solid phase c_l and c_s, the potential in the electrolyte and in the active material particles Φ_l and Φ_s, the current density in the liquid and in the solid phase i_l and i_s, as well as the ionic flux (pore-wall flux) j_n.

The Li-ion concentration in the liquid electrolyte domain c_l is given by

$$\varepsilon_l \frac{\partial c_l(x,t)}{\partial t} = \frac{\partial}{\partial x}\left(D_l^{eff} \frac{\partial c_l(x,t)}{\partial x} - i_l(x,t)\frac{(1-t_0^+)}{F} \right) \quad [13]$$

with a volume fraction of the active material ε_l, a concentration dependent effective electrolyte diffusivity D_l^{eff}, and a transference number of the cations in the solution t_0^+.

Journal of The Electrochemical Society, 2021 **168** 030535

Table IV. P2D model parameterization set for the investigated NMC-graphite cell.

Electrochemical Model *Geometry*		*Anode*	*Separator*	*Cathode*
Thickness L	m	7.35e-05	1.40e-05	5.55e-05
Particle radius r_P	m	7.00e-06		7.00e-06
Active material volume fraction ε_s	···	0.7164		0.6465
Inert filler volume fraction $\varepsilon_{s,na}$	···	0.0900		0.0840
Porosity ε_l	···	0.2350	0.3900	0.2560
Bruggeman coefficient α_B	···	2.60	2.50	2.10
Thermodynamics				
Stoichiometry at 100 % SOC $\theta_{100\%}$	···	0.8330		0.1890
Stoichiometry at 0 % SOC $\theta_{0\%}$	···	0.0367		0.9300
Maximum lithium concentration $c_{s,max}$	mol m^{-3}	3.11e+04		4.92e+04
Electrode equilibrium potential E_{eq}	V vs Li/Li^+	14		14
Entropic coefficient $\partial E_{eq}/\partial T$	V K^{-1}	14		14
Transport				
Diffusion coefficient in solid D_s	m^2 s^{-1}	1.60e-14		1.00e-13
Diffusion coefficient activation energy E_{a,D_s}	J mol^{-1}	3.00e+04		3.00e+04
Diffusion coefficient in liquid D_l	m^2 s^{-1}		Eq. 26	
Electrical conductivity of the solid matrix σ	S m^{-1}	25.00		0.25
Kinetics				
Reference Reaction rate $k_{a/c}$	m s^{-1}	6.40e-09		1.68e-09
Reaction rate activation energy $E_{a,k}$	J mol^{-1}	6.80e+04		5.00e+04
Transfer coefficient $\alpha_{a/c}$	···	0.50		0.50

Table V. Thermal parameters assumed for the NMC-graphite cell.

Thermal Model		
Density ρ	kg m^{-3}	2354.42
Heat transfer coefficient h	W m^{-2} K^{-1}	6.00
Specific heat capacity C_p	J kg^{-1} K^{-1}	880.00
Surface area A	m^2	7.60e-02
Volume V	m^{-3}	4.20e-04

In contrast the Li-ion concentration in the electrode active material particle domain c_s is calculated by

$$\frac{\partial c_s(x, r, t)}{\partial t} = D_s\left(\frac{\partial^2 c_s(x, r, t)}{\partial r^2} + \frac{2}{r}\frac{\partial c_s(x, r, t)}{\partial r}\right) \quad [14]$$

where r is the radius of the active material particles and D_s is the diffusivity of the solid phase. The x-dimension is coupled to the r-dimension using a boundary condition that applies at the particle's surface. It therefore characterizes the diffusion of Li-ions from the electrolyte into the active material particles. The equation corresponding to the boundary condition is given by

$$\left.\frac{\partial c_s(x, r, t)}{\partial r}\right|_{r=r_p} = -\frac{1}{D_s} j_n(x, t) \quad [15]$$

where j_n is the ionic flux and D_s is the diffusivity at the surface of the idealized active material particles. Considering the diffusivity to be constant reduces the solid-phase diffusion to a linear problem. The equations can therefore be solved using a superposition technique.[34] The DUALFOIL model by default uses the Duhamel Superposition Integral (DSI) method. More detailed information about the DSI is given by Doyle et al.[34]

The Butler-Volmer equation relates the reaction-rate and thus the ionic pore-wall flux j_n to the spatial surface overpotential η. The pore-wall flux j_n is given by

$$j_n(x, t) = \frac{i_0(x, t)}{F}\left[\exp\left(\frac{\alpha_a F}{RT}\eta(x, t)\right) - \exp\left(\frac{\alpha_c F}{RT}\eta(x, t)\right)\right] \quad [16]$$

with the exchange current density i_0 that describes the reversibility of the reaction. It depends on the concentration of reactants and products[13] and on the reaction-temperature. In contrast the transfer coefficients α_a and α_c relate to how an applied potential gradient favors one reaction-direction over the other.

The surface overpotential η expresses the deviation of the electrode potential at the electrode-electrolyte interface from its thermodynamic equilibrium state E_{eq} and is given by

$$\eta(x, t) = \Phi_s(x, t) - \Phi_l(x, t) - E_{eq}(\theta(x, t)) \quad [17]$$

where $\Phi_s - \Phi_l$ is the potential of the electrode compared to an arbitrarily chosen location in the liquid phase. The electrode's equilibrium potential, respectively its open-circuit potential, depends on the stoichiometry θ and therefore on the electrode's state of charge (SOC).

The potential gradient in the electrolyte $\partial\Phi_l/\partial x$ is considered the driving force for the flow of current in the liquid phase. In accordance with the work of Newman[13] it is given by

$$\frac{\partial \Phi_l(x, t)}{\partial x} = -\frac{i_l(x, t)}{\kappa^{eff}} + \frac{2RT}{F}(1 - t_0^+)\left(1 + \frac{\partial \ln(f_\pm)}{\partial \ln(c_l)}\right)\frac{\partial}{\partial x}\ln(c_l) \quad [18]$$

with the effective conductivity κ^{eff} and the activity coefficient $\partial \ln(f_\pm)/\partial \ln(c_l)$ to be considered as functions of the Li-ion concentration in the electrolyte.

Journal of The Electrochemical Society, 2021 **168** 030535

The mathematical equation for the potential in the solid phase is directly derived by Ohm's law. In terms of the physiochemical P2D model, Ohm's law relates the current density in the solid phase i_s to the potential gradient in the solid phase. The spatial potential gradient in the electrode domain is therefore given by

$$\frac{\partial \Phi_s(x, t)}{\partial x} = -\frac{1}{\sigma^{eff}}[I(t) - i_l(x, t)] \quad [19]$$

with the electric conductivity of the porous composite electrode σ^{eff}, and the current density in the solid phase i_s calculated by $I(t) - i_l(x, t)$ according to Kirchhoff's law.

The current density in the electrolyte i_l can be derived by Faraday's law. It is therefore related to the pore-wall flux j_n and given by

$$\frac{\partial i_l(x, t)}{\partial x} = aFj_n(x, t) \quad [20]$$

where a specifies the electrode-electrolyte interfacial area that is calculated by the quotient of the particle's surface and its volume weighted with the volume fraction ε_l.

Since the current density in the electrode domain i_s is related to the potential gradient $\nabla\Phi_s$ by Ohm's law, one can derive the solid phase current density i_s by

$$i_s(x, t) = -\sigma^{eff}\frac{\partial \Phi_s(x, t)}{\partial x}. \quad [21]$$

Thermal Model

The thermal behavior of the Li-ion cell affects the electrochemical reactions substantially. In consequence, it is essential to take the heat generation and dissipation within the cell into account. Therefore, a simplified energy balance is applied that is derived from the work of Bernardi[35] and given by

$$\rho V C_p \frac{\partial T}{\partial t} = \left(U_0 - V - T\frac{\partial U_0}{\partial T}\right)I + hA(T - T_\infty) \quad [22]$$

where the irreversible heat generation is expressed by the term

$$(U_0 - V)I. \quad [23]$$

Here V is the cell voltage, U_0 is the OCV and I is the applied current. The reversible heat generation that is directly related to the cell's entropy is given by $TI(\partial U_0/\partial T)$. The parameters ρ, V, and C_p characterize the cell's density, its volume, and the specific heat capacity, respectively. The assumed values can be found in Table V.

The heat transfer between the Li-ion cell and the ambient air is given by the term

$$hA(T - T_\infty) \quad [24]$$

where h is the heat-transfer coefficient, A is the intersectional area, and T_∞ is the ambient air temperature.

The zero-dimensional thermal model is coupled to the physiochemical due to a temperature dependency of the diffusion coefficients in the electrode and the electrolyte D_s^{eff}, and D_l^{eff}, the reaction rate constants k_a, and k_c that contribute the exchange current density i_0, the ionic conductivity κ^{eff}, and the activity coefficient $\partial \ln(f_\pm)/\partial \ln(c_l)$.

The temperature dependencies of the solid phase diffusion coefficient D_s^{eff}, and the reaction rate constants k_a, and k_c are modeled by the general Arrhenius equation

$$\Psi = \Psi_{ref} \exp\left(\frac{E_{a,\Psi}}{R}\left(\frac{1}{T_{ref}} - \frac{1}{T}\right)\right). \quad [25]$$

In this context, Ψ specifies the temperature-dependent variable, and Ψ_{ref} is a reference value that was determined at a corresponding reference temperature T_{ref}. $E_{a,\Psi}$ is the activation energy of the process, and T is the temperature given by the zero-dimensional thermal model.

In contrast the temperature dependencies of the liquid phase diffusion coefficient D_l^{eff}, the ionic conductivity κ^{eff}, and the activity coefficient $\partial \ln(f_\pm)/\partial \ln(c_l)$ are given by the following equations:

$$D_l(c_l, T) = 10^{-4} \cdot 10^{-4.43 - 0.22\cdot 10^{-3}c_l - \frac{54}{T - 229 - 5\cdot 10^{-3}c_l}} \quad [26]$$

$$\begin{aligned}\kappa(c_l, T) = {} & c_l \cdot 10^{-4}(-10.5 + 0.074T - 6.96 \cdot 10^{-5}T^2 \\ & + 0.668 \cdot 10^{-3}c_l - 0.0178 \cdot 10^{-3}c_l T \\ & + 2.8 \cdot 10^{-8}c_l T^2 + 0.494 \cdot 10^{-6}c_l^2 \\ & - 8.86 \cdot 10^{-10}c_l T)^2\end{aligned} \quad [27]$$

$$\begin{aligned}\frac{\partial \ln(f_\pm)}{\partial \ln c_l} = {} & 0.601 - 0.24(c_l \cdot 10^{-3})^{0.5} + 0.982[1 \\ & - 0.0052(T - T_{ref})(c_l \cdot 10^{-3})^{1.5}] \\ & \cdot (1 - t_+)^{-1} - 1.\end{aligned} \quad [28]$$

Note that the parameters in Eqs. 26–28 are adopted from the literature.[36] Since the P2D model parametrization is beyond the scope of this study, it is referred to Valøen et al.[36] for more information about the parameter fitting.

ORCID

Jacob C. Hamar https://orcid.org/0000-0002-0845-3667
Simon V. Erhard https://orcid.org/0000-0002-5029-7477
Christoph Zoerr https://orcid.org/000-0002-6361-2296
Andreas Jossen https://orcid.org/0000-0003-0964-1405

References

1. LLP Deloitte, Deloitte. New market. New entrants. New challenges. Battery Electric Vehicles (2019), https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/manufacturing/deloitte-uk-battery-electric-vehicles.pdf.
2. F. Leng, Z. Wei, C. M. Tan, and R. Yazami, "Heirarchical degradation processes in lithium-ion batteries during aging." *Electrochemica Acta*, **256**, 52 (2017).
3. J. Vetter, P. Novák, M. R. Wagner, C. Veit, K.-C. Möller, J. O. Besendhard, M. Winter, M. Wohlfahr-Mehrens, C. Volger, and A. Hammouche, "Ageing mechanisms in lithium-ion batteries." *Journal of Power Sources*, **147**, 269 (2005).
4. R. Bhattacharyya, B. Key, H. Chen, A. S. Best, A. F. Hollenkamp, and C. Grey, "In situ nmr observation of the formation of metallic lithium microstructures in lithium batteries." *Nat. Mater.*, **9**, 504 (2010).
5. C. Monroe and J. Newman, "Dendrite growth in lithium/polymer systems: A propagation model for liquid electrolytes under galvanostatic conditions." *J. Electrochem. Soc.*, **150**, A1377 (2003).
6. W. Mai, A. M. Colclasure, and K. Smith, "Model-instructed design of novel charging protocols for the extreme fast charging of lithium-ion batteries without lithium plating." *J. Electrochem. Soc.*, **167**, 1 (2020).
7. J. Newman and K. E. Thomas-Alyea, *Electrochemical Systems* (John Wiley and Sons, Inc., Hoboken, New Jersey) 3rd ed. (2004).
8. Z. Li, J. Huang, B. Y. Liaw, V. Metzler, and J. Zhang, "A review of lithium deposition in lithium-ion metal secondary batteries." *Journal of Power Sources*, **254**, 168 (2014).
9. P. Arora, R. E. White, and M. Doyle, *Capacity fade mechanisms and side reactions in lithium-ion batteries*, **145**(10), 3647 (1998).
10. D. R. Baker and M. W. Verbrugge, "Modeling overcharge at graphite electrodes: Plating and dissolution of lithium." *J. Electrochem. Soc.*, **167**, 1 (2020).
11. C. M. Doyle, "Design and Simulation of Lithium Rechargeable Batteries." *Dissertation* (University of California, Berkeley) (1995).
12. T. F. Fuller, "Simulation and optimization of the dual lithium ion insertion cell." *J. Electrochem. Soc.*, **141**, 1 (1994).
13. J. Newman and K. E. Thomas-Alyea, *Electrochemical Systems* (Wiley-Interscience, Hoboken, NJ) (2012), s.l.,3. aufl. edition.
14. J. Sturm, S. Ludwig, J. Zwirner, C. Ramirez-Garcia, B. Heinrich, M. F. Horsche, and A. Jossen, "Suitability of physiochemical models for embedded systems regarding a nickel-rich, silicon-graphite lithium-ion battery." *Journal of Power Sources*, **436**, 1 (2019).
15. D. Liu, X. Yin, Y. Song, W. Liu, and Y. Peng, "An on-line state of health estimation of lithium-ion battery using unscented particle filter." *IEEE Access*, **6**, 40990 (2018).

Journal of The Electrochemical Society, 2021 **168** 030535

16. R. Richardson, M. Osborne, and D. Howey, "Gaussian process regression for forcasting battery state of health." *Journal of Power Sources*, **357**, 209 (2017).
17. Y. Zhang, R. Xiong, H. He, and M. G. Pecht, "Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries." *IEEE Transactions on Vehicular Technology*, **67**, 5695 (2018).
18. A. G. Kashkooli, H. Fathiannasab, Z. Mao, and Z. Chen, "Application of artificial intelligence to state-of-charge and state-of-health estimation of calendar-aged lithium-ion pouch cells." *J. Electrochem. Soc.*, **166**, A605 (2019).
19. Kailong. Yi Li, Aoife M. Liu, Alana. Foley, Maitane. Zülke, and Elise. Berecibar, "Nanini-Maury, Joeri van Mierlo, and Harry E. Hoster. Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review." *Renew. Sustain. Energy Rev.*, **113**, 109254 (2019).
20. 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.*, **166**, A1893 (2019).
21. K. A. Severson et al., "Data-driven prediction of battery cycle life before capacity degradation." *Nat. Energy*, 1 (2019).
22. H. Ennifar, "Electrochemical State Estimation of a Lithium-Ion Battery: Using a Pseudo-2D Model and an Extended Kalman Filter." *Masterarbeit*, Technische Universität München, München (2017).
23. A. Tomaszewska et al., "Lithium-ion battery fast charging: A review." *eTransportation*, **1**, 100011 (2019).
24. B. Everitt and T. Hothorn, *An Introduction to Applied Multivariate Analysis with R* (Springer, New York, New York, NY) (2011).
25. J. O. Rawlings, S. G. Pantula, and D. A. Dickey, *Applied Regression A Research Tool* (Springer, Berlin) 2nd ed. (1998).
26. A. Raj, M. F. Rodrigues, and D. P. Abraham, "Rate-dependent aging resulting from fast charging li-ion cells." *J. Electrochem. Soc.*, **2020**, 1 (2020).
27. D. Anseán, M. Dubarry, A. Devie, B. Y. Liaw, V. M. Garcia, J. C. Viera, and M. González, "Fast charging technique for high power lifepo4 batteries: A mechanistic analysis of aging." *J. Electrochem. Soc.*, **321**, 201 (2016).
28. R Core Team, R: A language and environment for statistical computing (2019).
29. P. Bruce and A. Bruce, *Practical Statistics for Data Scientists: 50 Essential Concepts* (O'Reilly Media, Inc., Sebastopol, CA) (2017).
30. A. Burkov, *The Hundred Page Machine Learning Book* (Andriy Burkov, Online) (2019), http://themlbook.com/.
31. K. P. Murphy, *Machine Learning: A Probabilistic Perspective* (The MIT Press, Cambridge) (2012).
32. R. Klein, N. A. Chaturvedi, J. Christensen, J. Ahmed, R. Findeisen, and A. Kojic, "State estimation of a reduced electrochemical model of a lithium-ion battery." *IEEE Xplore, 2010 American Control Conference*, 6618 (2010).
33. V. Ramadesigan, P. W. C. Northrop, S. De, S. Santhanagopalan, R. D. Braatz, and V. R. Subramanian, "Modeling and simulation of lithium-ion batteries from a systems engineering perspective." *J. Electrochem. Soc.*, **159**, R31 (2012).
34. C. M. Doyle, T. F. Fuller, and J. Newman, "Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell." *J. Electrochem. Soc.*, **140**, 1526 (1993).
35. D. Bernardi., "A general energy balance for battery systems." *J. Electrochem. Soc.*, **132**, 5 (1985).
36. L. O. Valøen and J. N. Reimers, "Transport properties of lipf6-based li-ion battery electrolytes." *J. Electrochem. Soc.*, **152**, A882 (2005).

6 Summary and Conclusion

The combination of high-quality data sets, insights into battery behavior and the correct data-driven models has been effective at solving complex battery state estimation tasks. As the EV market continues to grow, the opportunity to improve state estimation functionality also grows. Harnessing the information contained in large data sets led to the development of machine learning, and supplements conventional battery state estimation approaches which often rely on complex differential equations or narrow equivalent circuit models.

One of the key enablers to this thesis was the large data set from the BMW i3 fleet. Applying statistical analysis to this data helps define the context of automotive operation. The distribution of highly relevant battery data from over 4000 vehicles driven under real-world conditions defines the extremes and the averages which can be anticipated for the majority of automotive battery aging scenarios. This information can be used as an input for battery design as well as for battery simulation. Two sub-sets from this data also shed light into the variation with the automotive context. When looking at the global temperatures, it was found that the higher latitudes experience a 17 K swing in average temperature due to the seasons, however, that variability can be ignored after as soon as 2.5 years in warmer countries (or 4.0 years in cooler regions) as the cumulative average settles. Another study looking into the difference between car sharing vehicles and privately owned vehicles confirmed the assumption that car sharing vehicles are more affected by cyclic aging as the average consumption is 18 % higher while also experiencing higher C-rates and lower average SOC. The ability to identify significant changes in the vehicle operation is also critical when monitoring battery health. The application of a changepoint detection algorithm on the data set highlights that even within a single vehicle, changes of up to 50 % in average discharge rate or 2 % in the average kilometers per year can be expected.

The insights into the variability within the real-world operation provided the impetus to investigate path dependent aging behavior in Chapter 3. Using a laboratory experiment where aging paths could be controlled, the assumption that path dependence could be ignored was tested. This experiment defined three path dependent conditions: cycling frequency, temperature and C-rate which were inspired by the more extreme operation spectrum seen during race operation. From the results, a path dependence was assumed under the temperature and C-rate conditions where deviations of up to 2 % in SOH were observed. A method of identifying aging mechanisms using the DVA curve helped confirm lithium plating as a significant aging mechanism for the profiles exhibiting a path dependence. A reversibility of plating due to higher discharge C-rate conditions seems to have also contributed to a path dependent aging behavior. Considering the large variation of SOH between complimentary paths, it is a strong recommendation of this thesis to consider the consistency of aging conditions, and further explore aging under dynamic conditions. The results from this paper support the efforts of a changepoint detection and clustering explored in Chapter 2 and Chapter 4, respectively, as they would help reduce variability and noise in the data.

Due to the difficulty in modeling health degradation under highly dynamic conditions, data-driven SOH estimation algorithms were also explored. Using the BMW i3 data as a basis– and the highly

accurate capacity measurements obtained from the offboard capacity tests– a novel machine learning approach using a neural network was compared with a more conventional semi-empirical method. The differences between the two approaches were minimized in part by combining the semi-empirical model with an unsupervised learning technique: K-means clustering, allowing for fitting of specific clusters of data exhibiting similar operation characteristics. The combined semi-empirical method provided an estimation error of 3.4 % in SOH and the neural network based approach an error of 3.0 % in SOH for over eight years of aging data. These results alone demonstrate the applicability of machine learning for supporting conventional and new SOH estimation techniques. Considering the results of Chapter 4 and the fact that few, if any, other algorithms have been tested on such a diverse data-set, the accuracy of these methods shows a high relevance for use in the automotive context.

Finally, the use of data-driven algorithms was expanded in Chapter 5 to estimate the internal state of a battery, namely the anode potential. Using various regression techniques, an estimation accuracy within 15 mV of the simulated value was achieved. The most successful approach incorporated the machine learning method of random forest regression, which resulted in an accuracy of 2.6 mV. This approach is highly relevant for improving battery performance as manufactures continue to push the boundary of charging time, which is mainly limited by the onset of plating around the 0 V Li/Li^+ potential threshold, therefore knowing precisely where the anode potential is, allows for safe and effective battery use during fast charging. In addition to the model accuracy, the focus of this work was also on the applicability of data-driven methods for online estimation. By proposing two alternative empirical equations, an analysis on the trade-offs between the data and computational requirements and the accuracy was performed. This analysis showed that while the highly efficient random forest could produce a viable model using 100 Mb for 500 trees, linear and non-linear models could achieve an error under 15 mV using only a few kilobytes, which would be more acceptable in most battery management systems.

This thesis highlights the relevance of data-driven methods for battery state estimation. The key to unlocking the potential of the machine learning methods is the availability of data, and understanding the context for its application, which is why the focus on the data in Chapter 2 is so critical. As illustrated by Chapter 3, aging is a complex phenomenon, where understanding the specific aging is paramount to accurate modeling and estimation. But, when provided with adequate data, machine learning offers a powerful toolbox to improve battery state estimation, as confirmed in Chapter 4 and Chapter 5.

The methods in this thesis can be further improved by expanding the application of the data-driven methods to more areas within the battery state estimation context, such as SOC estimation and failure identification. More specifically, it would be of great interest to generate a large data-set containing more aging mechanism specific information, such as post-mortem measurements of SEI or lithium plating growth. The additional data could help isolate the potential coupling of aging mechanisms for improved modeling performance, but also provide more features for machine learning methods. Additionally, testing of the methods in a vehicle would clarify the applicability of these methods for onboard use. Many questions are still open regarding the measurement signal quality, data storage potential and system noise which can only be addressed on a system level. As more data-driven methods are adopted for solving more complex problems, the quality and specificity of the data collected will also have to be observed. Only by combining meaningful data with the appropriate methods can the desired outcome be achieved. Luckily, with the boom in EV production coinciding with the advent of digital cloud connectivity, the opportunities abound.

References

[1] International Energy Agency: *Global EV Outlook 2021: Accelerating ambitions despite the pandemic*. 2021. URL: https://iea.blob.core.windows.net/assets/ed5f4484-f556-4110-8c5c-4ede8bcba637/GlobalEVOutlook2021.pdf (see p. 1).

[2] David Jacobs: *The German Energiewende – History, Targets, Policies and Challenges*. In: *Renewable Energy Law and Policy Review* 3.4 (2012), pp. 223–233 (see p. 1).

[3] Celina Kacperski, Roberto Ulloa, Sonja Klingert, Benedikt Kirpes, and Florian Kutzner: *Impact of incentives for greener battery electric vehicle charging - a field experiment*. In: *Energy Policy* 161 (2022) (see p. 1).

[4] Ashley Nunes, Lucas Woodley, and Philip Rossetti: *Re-thinking procurement incentives for electric vehicles to achieve net-zero emissions*. In: *Nature Sustainability* 5 (2022), pp. 527–532 (see p. 1).

[5] Simon-Kucher & Partners: *The Global Sustainability Study*. In: *Strategy & Marketing Consultants* (2021). URL: https://www.simon-kucher.com/en/who-we-are/newsroom/recent-study-reveals-more-third-global-consumers-are-willing-pay-more#:~:text=Globally%2C%2085%20percent%20of%20people,meaningful%20generational%20differences%20in%20attitude. (see p. 1).

[6] Maximilian Fischer, Nicolaas Kramer, Inga Maurer, and Raschel Mickelson: *A turning point for the US auto dealers: the unstoppable electric car*. 2021. URL: https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/a-turning-point-for-us-auto-dealers-the-unstoppable-electric-car (see p. 1).

[7] Felix Kuhnert and Jörn Neuhausen: *Electric Vehicle Sales Review Q2: China is getting back on track as the global electric vehicle powerhouse*. 2021. URL: https://www.strategyand.pwc.com/de/en/industries/automotive/electric-vehicle-sales-review-q2.html#:~:text=BEV%20sales%20were%20up%20by,proof%20of%20continued%2C%20accelerating%20growth. (see p. 1).

[8] European Commission: *State aid: Commission approves 2.9 billion public support by twelve Member States for a second pan-European research and innovation project along the entire battery value chain*. Brussels, 2021. URL: https://ec.europa.eu/commission/presscorner/detail/en/ip_21_226 (visited on 2021) (see p. 1).

[9] Axel Schmidt, Teodoro Lio, Juergen Reers, and Andrea Regalia: *The Electric Vehicle: Why manufacturers need to rethink their ecosystems, the customer journey and the entire value chain*. 2021. URL: https://www.accenture.com/_acnmedia/PDF-143/Accenture-Electric-Vehicle-PoV-2021.pdf (see p. 2).

[10] Advanced Industries: *Monetizing car data: New service business opportunities to create new customer benefits*. Ed. by McKinsey & Company. www.mckenzie.com, 2016 (see p. 2).

[11] Harald Proff, Thomas Pottebaum, and Philipp Wolf: *Autonomous Driving: Moonshot Project with Quantum Leap from Hardware to Software AI Focus*. URL: https://www2.deloitte.com/content/dam/Deloitte/be/Documents/Deloitte_Autonomous-Driving.pdf (see p. 2).

[12] Jean-Baptiste Michel et al.: *Quantitative Analysis of Culture Using Millions of Digitalized Books*. In: *Science* (2010) (see p. 2).

[13] Journal of Power Sources. 2022. URL: https://www.journals.elsevier.com/journal-of-power-sources (see p. 2).

[14] Nature Portfolio. 2022. URL: https://www.nature.com/ (see p. 2).

[15] Tony Hey, Stewart Tansley, and Kristin Tolle: *The Fourth Paradigm: Data-Intensive Scientific Discovery*. 2009. ISBN: 978-0-9825442-0-4 (see p. 2).

[16] Francisco J. Montáns, Francisco Chinesta, Rafael Gómez-Bombarelli, and J. Nathan Kutz: *Data-driven modeling and learning in science and engineering*. In: *Data-Based Engineering Science and Technology* 347.11 (2019), pp. 845–855 (see p. 2).

[17] Mark Minevich: *The Automotive Industry and The Data Driven Approach*. In: *Forbes* (2013). URL: https://www.forbes.com/sites/markminevich/2020/07/13/the-automotive-industry-and-the-data-driven-approach/?sh=1cab893af9a5 (see p. 3).

[18] Nasim Arbabzadeh and Mohsen Jafari: *A Data-Driven Approach for Driving Safety Risk Prediction Using Driver Behavior and Roadway Information Data*. In: *IEEE Transactions on Intelligent Transportation Systems* 19.2 (2018). URL: 10.1109/TITS.2017.2700869 (see p. 3).

[19] Patrick Ebel, Julia Orlovska, Sebastian Hünemeyer, Casper Wickman, Andreas Vogelsang, and Rikard Sönerberg: *Automotive UX design and data-driven development: Narrowing the gap to support practitioners*. In: *Transportation Research Interdisciplinary Perspectives* 11.100455 (2021) (see p. 3).

[20] Andreas Theissler, Judith Pérez-Valázquez, Marcel Kettelgerdes, and Gordon Elger: *Predictive maintenance enabled by machine learning: Use cases and challenges in the automotive industry*. In: *Reliability Engineering and System Safety* 215.107864 (2021). URL: https://doi.org/10.1016/j.ress.2021.107864 (see p. 3).

[21] Yuejiu Zheng, Minggao Ouyang, Xuebing Han, Languang Lu, and Jianqiu Li: *Investigating the error sources of the online state of charge estimation methods for lithium-ion batteries in electric vehicles*. In: *Journal of Power Sources* 377 (2018), pp. 161–188 (see p. 3).

[22] Shrey Verma, Shubham Mishra, Ambar Gaur, Subhankar Chowdhury, Suhashree Mohapatra, Caurav Dwivedi, and Puneet Verma: *A comprehensive review on energy storage in hybrid electric vehicles*. In: *Journal of Traffic and Transportation Engineering* 8.5 (2021), pp. 621–637 (see p. 4).

[23] Guoxing Li: *Regulation Mass Transport Behavior for High-Performance Lithium Metal Batteries and Fast-charging Lithium-Ion Batteries*. In: *Advanced Energy Materials* 11.7 (2021). URL: https://doi.org/10.1002/aenm.202002891 (see p. 4).

[24] J. Nava-Avendano and J. Veilleux: *Plasma processes in the preparation of lithium-ion battery electrodes and separators*. In: *Journal of Physics D: Applied Physics* 50.16 (2017). URL: 10.1088/1361-6463/aa6245 (see p. 4).

[25] Juergen Garche and Klaus Brandt: *Electrochemical Power Sources: Fundamentals, Systems, and Applications: Li-Battery Safety*. Elsevier, 2018. ISBN: 978-0-444-63777-2. URL: https://doi.org/10.1016/C2015-0-00574-3 (see p. 4).

[26] J. Wu, Y. Cao, H. Zhao, J. Mao, and Z. Guo: *The critical role of carbon in marrying silicon and graphite anodes for high-energy lithium-ion batteries.* In: *Carbon Energy* 1.1 (2019), pp. 57–76. URL: DOI:%C2%A010.1002/cey2.2 (see p. 4).

[27] J. Asenbauer, T. Eisenmann, M. Kuenzel, A. Kazzazo, Z. Chen, and D. Bresser: *The success story of graphite as a lithium-ion anode material - fundamentals, remaining challenges and recent developments including silicon (oxide) composites.* In: *Sustainable Energy & Fules* 3.1 (2019), pp. 245–250. URL: DOI:%2010.1039/D0SE00175A (see p. 4).

[28] J. Sturm, A. Rheinfeld, I. Zilberman, F. B. Spingler, S. Kosch, F. Frie, and A. Jossen: *Modeling and simulation of inhomogeneities in a 18650 nickel-rich, silicon-graphite lithium-ion cell during fast charging.* In: *Journal of Power Sources* 412 (2019), pp. 204–223. URL: DOI:%2010.1016/j.jpowsour.2018.11.043. (see p. 5).

[29] T. D. Hatchard and J. R. Dahn: *In Situ XRD and Electrochemical Study of the Reaction of Lithium with Amorphous Silicon.* In: *Journal of The Electrochemical Society* 151.6 (2004). ISSN: 00134651. URL: DOI:%2010.1149/1.1739217. (see p. 5).

[30] Newman, J. and Thomas-Alyea, K. E.: *Electrochemical Systems: Third Edition.* Wiley-Interscience, 2004 (see pp. 5–8, 11).

[31] R. Schmuch, R. Wagner, G. Hörpel, T. Placke, and M. Winter: *Performance and cost of materials for lithium-based rechargeable automotive batteries.* In: *Nature Energy 3.4* 3.4 (2018), p. 267.278. URL: DOI:%2010.1038/s41560-018-0107-2 (see p. 5).

[32] D. Andre, S.-J. Kim, P. L., S. F. Lux, f. Maglia, and O. Paschos, B. S.: *Future generations of cathode materials: An automotive industry perspective.* In: *Journal of Material Chemistry* 3.13 (2015), pp. 6709–6732 (see p. 5).

[33] N. Nitta, F. Wu, J. T. Lee, and G. Yushin: *Li-ion battery materials: Present and future.* In: *Materials Today* 18.5 (2015), pp. 252–264 (see p. 5).

[34] F. Wu, J. Maier, and Y. Yu: *Guidelines and trends for next-generation rechargeable lithium and lithium-ion batteries.* In: *Chemical Society Reviews* 49.5 (2020), pp. 1569–1614. URL: DOI:%2010.1039/c7cs00863e. (see p. 5).

[35] H. Li, M. Cormier, N. Zhang, J. Inglis, J. Li, and J. R. Dahn: *Is Cobalt Needed in Ni-Rich Positive Electrode Materials for Lithium-Ion Batteries?* In: *Journal of The Electrochemical Society* 166.4 (2019), A429–A439. ISSN: 00134651. URL: DOI:%2010.1149/2.1381902jes (see p. 5).

[36] J. Morris: *Tesla's Shift to Cobalt-Free Batteries is its Most Important Move Yet.* In: *Forbes* (2020). URL: https://www.forbes.com/sites/jamesmorris/2020/07/11/teslas-shift-to-cobalt-free-batteries-is-its-most-important-move-yet/#5740958546b4 (see p. 5).

[37] D. Andre: *Lithium-ion batteries and applications: Batteries.* London: Artech House, 2020. ISBN: 9781630817688 (see p. 5).

[38] D. Andre, M. Meiler, K. Steiner, C. Wimmer, T. Soczka-Guth, and D. U. Sauer: *Characterization of high-power lithium-ion batteries by electrochemical impedance spectroscopy. I. Experimental investigation.* In: *Journal of Power Sources* 196.12 (2011), pp. 5334–5351 (see p. 6).

[39] S. V. Erhard: *Multi-dimensional electrochemical-thermal modeling of lithium-ion batteries.* PhD Thesis, 2017 (see pp. 6, 10).

[40] E. Sarasketa-Zabala, E. Martinez-Lasema, M. Berecibar, I. Gandiaga, and L. M. Rodriguez-Martinez, I. Villarreal: *Realistic lifetime prediction approach for Li-ion batteries*. In: *Applied Energy* 2016.162 (2016), pp. 839–852. URL: http://dx.doi.org/10.1016/j.apenergy.2015.10.115 (see pp. 7, 12, 45, 59).

[41] Maik Naumann, Michael Schimpe, Peter Keil, Holger C. Hesse, and Andreas Jossen: *Analysis and modeling of calendar aging of a commercial $LiFePO_4$/graphite cell*. In: *Journal of Energy Sources* 2018.17 (2018), pp. 153–169. URL: https://doi.org/10.1016/j.est.2018.01.019 (see pp. 7, 45).

[42] E. Sarasketa-Zabala, I. Gandiaga, L.M. Rodriguez-Martinez, and I. Villarreal: *Calendar ageing analysis of a LiFePO4/graphite cell with dynamic model validations: Towards realistic lifetime predictions*. In: *Journal of Power Sources* 2014.272 (2014), pp. 45–57. URL: http://dx.doi.org/10.1016/j.jpowsour.2014.08.051 (see p. 7).

[43] Peter Keil, Simon F. Schuster, Jörn Wilhelm, Julian Travi, Andreas Hauser, Ralph C. Karl, and Andreas Jossen: *Calendar Aging of Lithium-Ion Batteries*. In: *Journal of The Electrochemical Society* 2016.163 (2016), A1872–A1880. ISSN: 00134651. URL: 10.1149/2.0411609jes (see pp. 7, 8, 36, 45).

[44] Vetter, J. et al.: *Ageing mechanisms in lithium-ion batteries*. In: *Journal of Power Sources* 147 (2005), pp. 269–281 (see pp. 7, 74).

[45] V. A. Agubra and J. W. Fergus: *The formation and stability of the solid electrolyte interface on the graphite anode*. In: *Journal of Power Sources* 268 (2014), pp. 153–162 (see p. 7).

[46] S. Watanabe, M. Kinoshita, and K. Nakura: *Capacity fade of $LiNi_{(1-x-y)}Co_xAl_yO_2$ cathode for lithium-ion batteries during accelerated calendar and cycle life test: comparison analysis between $LiNi_{(1-x-y)}Co_xAl_yO_2$ and $LiCoO_2$ cathodes in cylindrical lithium-ion cells*. In: *Journal of Power Sources* 247 (2014), pp. 620–625 (see pp. 7, 46).

[47] Arora, P., White, R. E., and Doyle, M.: *Capacity Fade Mechanisms and Side Reactions in Lithium-Ion Batteries*. In: *Journal of The Electrochemical Society* 146 (1999). ISSN: 00134651. URL: 10.1149/1.1838857 (see p. 7).

[48] N. Legrand, B. Knosp, P. Desprez, F. Lapicque, and S. Rael: *Physical characterization of the charging process of a Li-ion battery and prediciton of Li plating by electrochemical modelling*. In: *Journal of Energy Sources* 245 (2014), pp. 297–302 (see p. 7).

[49] Raj, A., Rodrigues, M. F., and Abraham, D. P.: *Rate-Dependent Aging Resulting from Fast Charging Li-Ion Cells*. In: *Journal of The Electrochemical Society* 2020.167 (2020), pp. 1–8. ISSN: 00134651 (see pp. 7, 74).

[50] Anseán, D., Dubarry, M., Devie, A., Liaw, B. Y., V. M. Garcia J. C. Viera, and González, M.: *Fast charging technique for high power LiFePO4 batteries: A mechanistic analysis of aging*. In: *Journal of The Electrochemical Society* 2016.321 (2016), pp. 201–209. ISSN: 00134651 (see pp. 7, 74).

[51] Carlos Pastor-Fernández, Kotub Uddin, Gael H. Chouchelamane, W. Dhammika Widanage, and James 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*. In: *Journal of Power Sources* 360 (2017), pp. 301–318. URL: http://dx.doi.org/10.1016/j.jpowsour.2017.03.042 (see pp. 8, 45, 46).

[52] Matthieu Dubarry, Cyril Truchot, and Bor Yann Liaw: *Synthesize battery degradation modes via a diagnostic and prognostic model*. In: *Journal of Power Sources* 2012.219 (2012), pp. 204–216. URL: http://dx.doi.org/10.1016/j.jpowsour.2012.07.016 (see pp. 8, 45).

[53] K. Uddin, S. Perera, W. Widanage, L. Somerville, and J. Marco: *Characterising lithium-ion battery degradation through the identificaiton and tracking of electrochemical battery model parameters*. In: *Batteries* 2.2 (2016), p. 13. URL: http://dx.doi.org/10.3390/batteries2020013 (see p. 8).

[54] Doyle, C. M.: *Design and Simulation of Lithium Rechargeable Batteries*. Dissertation. Berkeley: University of California, 1995 (see pp. 8, 11, 74).

[55] J. Newman and W. Tiedemann: *Porous-electrode theory with battery applications*. In: *AIChE Journal* 21.1 (1975), pp. 25–41 (see pp. 8, 11).

[56] K. Thomas, J. Newman, and R. Darling: *Mathematical modeling of lithium batteries*. In: *Advances in lithium-ion batteries* (2002) (see pp. 8, 11).

[57] J. Newman, K. E. Thomas, H. Hafezi, and D. R. Wheeler: *Modeling of lithium-ion batteries*. In: *Journal of Power Sources* 119-121 (2003), pp. 838–843 (see pp. 8, 11).

[58] Y. Zheng, M. Ouyang, M. Han, L. Lu, and J. Liu: *Investigating the error sources of the online state of charge estimation methods for lithium-ion batteries in electric vehicles*. In: *Journal of Power Sources* 377 (2018), pp. 161–188 (see p. 8).

[59] Johannes Schmalstieg, Stefan Käbitz, Madeleine Ecker, and Dirk Uwe Sauer: *A holistic aging model for Li(NiMnCo)O_2 based 18650 lithium-ion batteries*. In: *Journal of Power Sources* 2014.257 (2014), pp. 325–334 (see pp. 8, 12, 45, 59).

[60] C. H. Chen, J. Liu, and K. Amine: *Symmetric cell approach and impedance spectroscopy of high power lithium-ion batteries*. In: *Journal of Power Sources* 2 (2001), pp. 321–328. URL: https://doi.org/10.1016/S0378-7753(00)00666-2 (see p. 8).

[61] Rudolph G. Jungst, Ganesan Nagasubramanian, Herbert L Case, Bor Yann Liaw, Angel Urbina, Thomas L Paez, and Daniel H Doughty: *Accelerated calendar and pulse life analysis of lithium-ion cells*. In: *Journal of Power Sources* (2003), pp. 870–873. URL: https://doi.org/10.1016/S0378-7753(03)00193-9 (see p. 8).

[62] Andreas Jossen: *Fundamentals of battery dynamics*. In: *Journal of Power Sources* 154.2 (2005), pp. 530–538 (see p. 9).

[63] Kurzweil, P.: *Elektrochemische Speicher: Superkondensatoren, Batterien, Elektrolyse-Wasserstoff, rechtliche Rahmenbedingungen*. Spinger, 2018 (see p. 9).

[64] W. Waag, C. Fleischer, and D. U. Sauer: *Adaptive on-line prediciton of the available power of lithium-ion batteries*. In: *Applied Energy* 242 (2013), pp. 548–559 (see p. 9).

[65] I. Baghdadi, O. Briat, J. Y. Deletage, P. Gyan, and J. M. Vinassa: *Lithium battery aging model based on Dakin's degradation approach*. In: *Journal of Power Sources* 325 (2016), pp. 273–285 (see p. 9).

[66] Reinhold Koch: *On-line Electrochemical Impedence Spectroscopy for Lithium-Ion Battery Systems; Estimation, Compensation and Avoidance of Measurement Deviations: Dissertation*. Media TUM, 2021. URL: https://mediatum.ub.tum.de/doc/1343447/1343447.pdf (see p. 9).

[67] Sankhadeep Sarkar, S. Zohra Halim, Mahmoud M. El-Halwagi, and Faisal I. Khan: *Electrochemical models: methods and applications for safer lithium-ion battery operation*. In: *Journal of The Electrochemical Society* 169.10 (2022). ISSN: 00134651 (see p. 9).

[68] Venkatasailanathan Ramadesigan, Paul W. C. Northrop, Sumitava De, Shriiram Santhanagopalan, Richard D. Braatz, and Venkat R. Subramanian: *Modeling and Simulation of Lithium-Ion Batteries from a Systems Engineering Perspective*. In: *Journal of The Electrochemical Society* 159.3 (2012). ISSN: 00134651. URL: DOI%2010.1149/2.018203jes (see p. 9).

[69] Marc Doyle, Thomas F. Fuller, and John Newman: *Modeling of Galvanostatic Charge and Discharge of the Lithium/Polymet/Insertion Cell*. In: *Journal of The Electrochemical Society* 140.6 (1993). ISSN: 00134651. URL: DOI%2010.1149/1.2221597 (see p. 9).

[70] Ali Jokar, Barzin Rajabloo, Martin Desilets, and Marcel Lacroix: *Review of simplified Pseudo-two-Dimensional models of lithium-ion batteries*. In: *Journal of Power Sources* 327 (2016), pp. 44–55 (see p. 9).

[71] John Newman and Nitash P. Balsara: *Electrochemical Systems*. Wiley, 2020. ISBN: 978-1-119-51460-2 (see pp. 10–12).

[72] Jacob C. Hamar, Simon V. Erhard, Christoph Zoerr, and Andreas Jossen: *Anode Potential Estimation in Lithium-Ion Batteries Using Data-Driven Models for Online Applications*. In: *Journal of The Electrochemical Society* 168 (2021). ISSN: 00134651 (see pp. 11, 23).

[73] Johannes Joerg Sturm: *State-Estimation of Lithium-Ion Batteries using Physiochemical Models and Experimental Characterization Techniques*. MediaTUM, 2021. URL: https://mediatum.ub.tum.de/1574895 (see p. 12).

[74] Sturm, J., Ennifar, H., Erhard, S. V., Rheinfeld, A., Kosch, S., and Jossen, A.: *State estimation of lithium-ion cells using a physicochemical model based extended Kalman filter*. In: *Applied Energy* 223 (2018), pp. 103–123 (see p. 12).

[75] Jonas Keil and Andreas Jossen: *Electrochemical Modeling of Linear and Nonlinear Aging of Lithium-Ion Cells*. In: *Journal of The Electrochemical Society* 167 (2020). ISSN: 00134651. URL: https://iopscience.iop.org/article/10.1149/1945-7111/aba44f/pdf (see p. 12).

[76] Martin Petit, Eric Prada, and Valerie Sauvant-Moynot: *Development of an empirical aging model for Li-ion batteries and applicaiton to assess the impact of Vehicle-to-Grid strategies on battery lifetime*. In: *Applied Energy* 172 (2016), pp. 398–407 (see p. 12).

[77] Severin Lukas Hahn, Mathias Storch, Rammanathan Swaminathan, Bjoern Obry, Jochen Bandlow, and Kai Peter Birke: *Quantitative validation of calendar aging models for lithium-ion batteries*. In: *Journal of Power Sources* 400 (2018), pp. 402–414. URL: https://doi.org/10.1016/j.jpowsour.2018.08.019 (see pp. 12, 13, 59).

[78] Svante Arrhenius: *Über die Dissociationswärme und den Einfluss der Temperatur auf den Dissociationsgrad der Elektrolyte*. In: *Zeitschrift für Physikalische Chemie* 4 (1889), pp. 96–116 (see p. 13).

[79] Julius Tafel: *Über die Polarisation bei kathodischer Wasserstoffentwicklung*. In: *Zeitschrift für Physikalische Chemie* 50A (1905), p. 641 (see p. 13).

[80] Carlos Vidal, Pawel Malysz, Phillip Kollmeyer, and Ali Emadi: *Machine Learning Applied to Electrified Vehicle Battery State of Charge and State of Health Estimation: State-of-the-Art*. In: *IEEE Access* 2020.Volume 8 (2020), pp. 52796–52814 (see pp. 13, 59).

[81] Man-Fai Ng, Jin Zhao, Qingyu Yan, Gareth J. Conduit, and Zhi Wei Seh: *Predicting the state of charge and health of batteries using data-driven machine learning*. In: *Nature Machine Intelligence* 2 (2020), pp. 161–170. URL: http://doi.org/10.1038/s42256-020-0156-7 (see p. 13).

[82] Andriy Burkov: *The Hundred-Page Machine Learning Book*. 2019. URL: http://themlbook.com/ (see pp. 14, 16, 17, 19–21, 60, 74).

[83] Francois Chollet: *Deep Learning with Python*. Manning Publication, 2018 (see pp. 15, 21, 60).

[84] F. Pedregosa et al.: *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research. 2011. URL: https://scikit-learn.org/stable/about.html#citing-scikit-learn (see p. 16).

[85] Harshit Sethy, Amit Patel, and Vineet Padmanabhan: *Real Time Strategy Games: A Reinforcement Learning Approach*. In: *Procedia Comupter Science* 54 (2015), pp. 257–264 (see p. 19).

[86] Yimo Yan, Andy H. F. Chow, Chin Pang Ho, Yong-Hong Kuo, Qihao Wu, and Chengshuo Ying: *Reinforcement learning for logistics and supply chain management: Methodologies, state of the art, and future opportunities*. In: *Transportation Research Part E: Logistics and Transportation Review* 162 (2022) (see p. 19).

[87] W. Zhang, D. Yang, H. Peng, W. Wu, W. Quan, H. Zhang, and X. Shen: *Deep Reinforcement Learning Based Resource Management for DNN Inference in Industrial IoT*. In: *IEEE Transactions on Vehicular Technology* 70.8 (2021), pp. 7605–7618. ISSN: 0018-9545 (see p. 19).

[88] Kevin P. Murphy: *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012. ISBN: 978-0-262-01802-9 (see pp. 19, 74).

[89] W. Shen and R. Xiong: *Advanced Battery Management Technologies for Electric Vehicles*. United Kingdom: John Wiley & Sons, 2019. ISBN: 9781119481676 (see p. 21).

[90] 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 batteries*. In: *Applied Energy* 300 (2021), p. 117346. URL: 10.1016/j.apenergy.2021.117346 (see p. 22).

[91] Goncalo dos Reis, Calum Strange, Mohit Yadav, and Shawn Li: *Lithium-ion battery data and where to find it*. In: *Energy and AI* 5.100081 (2021). URL: https://doi.org/10.1016/j.egyai.2021.100081 (see pp. 22, 25).

[92] Dawn McIntosh: *Li-ion Battery Aging Datasets*. 2010. URL: https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository. (see p. 22).

[93] Robert R. Richardson, Michael A. Osborne, and David A. Howey: *Gaussian process regression for forecasting battery state of health*. In: *Journal of Power Sources* 2017.357 (2017), pp. 209–219 (see pp. 22, 59).

[94] E.F. Hogge, B. M. Bole, S. L. Vasquez, J. R. Celaya, T. H. Strom, and B. L. Hill: *Adaption of an Electrochemistry-based Li-ion Battery Model to Account for Deterioration Observed Under Randomized Use*. In: *SGT Inc.* (2014) (see p. 22).

[95] B. E. Olivares, M. A. C. Munoz, M. E. Orchard, and J. F. Silva: *Particle-filtering-based prognosis framework for energy storage devices with a statistical characterization of state-of-health regeneration phenomena*. In: *IEEE Trans Instrm Meas* 62.2 (2012), pp. 364–376 (see p. 22).

[96] J. Zhou, D. Liu, Y. Peng, and X. Peng: *Dynamic battery remaining useful life estimation: an online data-driven approach*. In: *IEEE international instrumentation and measurement technology conference proceedings* (2012), pp. 2196–2199 (see p. 22).

[97] B. Saha and K. Goebel: *Model adaption for prognostics in a particle filtering framework*. In: *International Journal of Prognostic Health Management* 2 (2011), p. 61 (see p. 22).

[98] Sandia National Lab: *Data for degradation of commercial lithium-ion cells as a function of chemistry and cycling conditions*. 2020. URL: https://www.batteryarchive.org/snl_study.html. (see p. 22).

[99] J. Govindarajan: *Lifecycle ageing tests on commercial 18650 Li ion cell @ 10°C, and 0°C: Everlasting: Electric Vehicle Enhanded Range, Lifetime And Safety Through INGenious battery management*. 2021. URL: 10.4121/14377295.v1 (see p. 22).

[100] D. Howey: *Oxford Battery team data and code*. 2011. URL: http://howey.eng.ox.ac.uk/data-and-code/ (see p. 22).

[101] D. Gun, H. Perez, and S. Moura: *Berkeley: eCal fast charging test data*. 2015. URL: https://datadryad.org/stash/dataset/doi:10.6078/D1MS3X. (see p. 22).

[102] G. Steinbuss, B. Rzepka, S. Bischof, T. Blank, and K. Böhm: *Frequent observations from a battery system with subunits*. 2019. URL: DOI:%2010.5445/IR/1000094469 (see p. 22).

[103] D. Roman, S. Saxena, V. Robu, M. Pecht, and D. Flynn: *Machine learning pipeline for battery state of health estimation*. In: *Nature Machine Intelligence* (2021). URL: DOI:%2010.1038/s42256-021-00312-3 (see p. 22).

[104] J. Tian and R. Xiong w. Shen: *State-of-Health Estimation Based on Differential Temperature for Lithium Ion Batteries*. In: *IEEE Transactions on Power Electronics* 35.10 (2020), pp. 10363–10373. URL: 10.1109/TPEL.2020.2978493 (see pp. 22, 60).

[105] Z. Chen, M. sun, X. Shu, R. Xiao, and J. Shen: *Online State of Health Estimation for Lithium-Ion Batteries Based on Support Vector Machines*. In: *Applied Sciences* 8.6 (2018), p. 20. URL: 10.3390/app8060925 (see pp. 22, 60).

[106] T. Oji, Y. Zhou, S. Ci, F. Kang, X. Chen, and X. Liu: *Data-Driven Methods for Battery SOH Estimation: Survey and a Critical Analysis*. In: *IEEE Access* 9 (2021), pp. 126903–126916. URL: 10.1109/ACCESS.2021.3111927 (see pp. 22, 60).

[107] 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*. In: *Journal of Power Sources* 482 (2021), p. 228863. URL: 10.1016/j.jpowsour.2020.228863 (see pp. 22, 60).

[108] G. Zou, Z. Yan, C. Zhang, and L. Song: *Transfer learning with CNN-LSTM model for capacity prediction of lithium-ion batteries under small sample*. In: *Journal of Physics: Conference Series* 2258.1 (2022). URL: 10.1088/1742-6596/2258/1/012042 (see p. 22).

[109] J. C. Alvarez Anton, P. J. Garcia Nieto, and C. Blanco Viejo and J. A. Vilan Vilan: *Support Vector Machines Used to Estimate the Battery State of Charge*. In: *IEEE Transactions on Power Electronics* Band 12 Nr. 12 (2013), pp. 5919–5926 (see p. 22).

[110] G. O. Sahinoglu, M. Pajovic, Z. Sahinoglu, Y. Wang, P.V. Orlik, and T. Wada: *Battery State-of-Charge Estimation Based on Regular/Recurrent Gaussian Process Regression*. In: *IEEE Transactions on Industrial Electronics* Band 65 Nr. 5 (2018), pp. 4311–4321 (see p. 22).

[111] E. Chemali, P. J. Kollmeyer, M. Preindl, and A. Emadi: *State-of-charge estimation of Li-ion batteries using deep neural networks: a machine learning approach*. In: *Journal of Power Sources* 400 (2018), pp. 242–255 (see p. 22).

[112] C. Li, F. Xiao, and Y. Fan: *An Approach to State of Charge Estimatoin of Lithium-ion Batteries Based on Recurrent Neural Networks with Gated Recurrent Unit*. In: *Energies* 12.19 (2019), p. 1592 (see p. 22).

[113] Y. Pan, X. Feng, M. Zhang, X. Han, L. Lu, and M. Ouyang: *Internal short circuit detection for lithium-ion battery pack with parallel-series hybrid connections*. In: *J. Clean. Prod.* 255 (2020), p. 120277 (see p. 22).

[114] B.-R. Chen, M.R. Kunz, T. T., and E.J. Dufek: *A machine learning framework for early detection of lithium plating combining multiple physics-based electrochemical signatures*. In: *Cell Rep Phys Sci* 2.3 (2021), p. 100352 (see p. 22).

[115] Y. Shang, K. Liu, N. Cui, N. Wang, K. Li, and C. Zhang: *A compact resonant switched-capacitor heater for lithium-ion battery self-heating at low temperatures*. In: *IEEE Transactions on Power Electronics* 35.7 (2019) (see p. 22).

[116] E. Chemali, P. J. Kollmeyer, M. Preindl, R. Ahmed, and A. Emadi: *Long Short-Term Memory Networks for Accurate State-of-Charge Estimation of Li-ion Batteries*. In: *IEEE Transactions on Industrial Electronics* 65.8 (2018), pp. 6730–6739 (see p. 22).

[117] F. Yang, W. Li, and Q. Miao: *State-of-charge estimation of lithium-ion batteries based on gated recurrent neural networks*. In: *Energy* 175 (2019), pp. 66–75 (see p. 22).

[118] Q. Gong, P. Wang, and Z. Cheng: *A Data-Driven Model Framework Based on Deep Learning for Estimating the States of Lithium-Ion Batteries*. In: *Journal of The Electrochemical Society* 196.3 (2022). ISSN: 00134651. URL: 10.1149/1945-7111/ac5bac (see pp. 22, 60).

[119] Y. Choi, S. Ryu, K. Park, and H. Kim: *Machine Learning-Based Lithium-Ion Battery Capacity Estimation Exploiting Multi-Channel Charging Profiles*. In: *IEEE Access* 7 (2019). URL: 10.1109/ACCESS.2019.2920932 (see pp. 22, 60).

[120] Jacob C. Hamar, Magdalena Stuckenberger, Johannes Sturm, Julius Schmitt, Marcel Rogge, Simon V. Erhard, and Andreas Jossen: *Investigating the path dependent aging behavior of nickel cobalt aluminum oxide batteries during high C-rate cycling conditions*. In: *Journal of The Electrochemical Society* (2024). ISSN: 00134651. URL: 10.1149/1945-7111/ad2952 (see p. 23).

[121] Jacob C. Hamar, Simon V. Erhard, Angelo Canesso, Jonas Kohlschmidt, Nicolas Olivain, and Andreas Jossen: *State-of-health estimation using a neural network trained on vehicle data*. In: *Journal of Power Sources* 512 (2021). URL: https://doi.org/10.1016/j.jpowsour.2021.230493 (see p. 23).

[122] BMW AG: *The i3*. 2021. URL: https://www.bmw.de/de/neufahrzeuge/bmw-i/i3/2021/bmw-i3-ueberblick.html (see p. 25).

[123] BMW AG. 2022. URL: https://www.bmw.com/de/index.html (see p. 25).

[124] Michael Taylor: *Last Call for BMW's Pioneering i3 Electric Car as Production Ends*. In: *Forbes* (2021). URL: https://www.forbes.com/sites/michaeltaylor/2021/06/29/last-call-for-bmws-pioneering-i3-electric-car-as-production-ends/ (see p. 25).

[125] Jiuchun Jiang and Caiping Zhang: *Fundamentals and Applications of Lithium-Ion Batteries in Electric Drive Vehicles*. Wiley Online Library: John Wiley Sons Singapore Pte Ltd., 2015. ISBN: 9781118414781. URL: 10.1002/9781118414798 (see p. 26).

[126] V. Pop, H.J. Bergveld, D. Danilov, P.P.L. Regtien, and P.H.L. Notten: *Battery Management Systems: Accurate State-of-Charge Indication for Battery-Powered Applications*. Springer Science+Business Media B. V, 2008. ISBN: 978-1-4020-6945-1 (see p. 26).

[127] Hossam A. Gabbar, Ahmed M. Othman, and Muhammad R. Abdussami: *Review of Battery Management Systems (BMS) Development and Industrial Standards*. In: *technologies* 2021.9 (2021), pp. 1–23. URL: https://doi.org/10.3390/technologies9020028 (see p. 26).

[128] John W. Tukey: *Exploratory Data Analysis*. Addison-Wesley Publishing Company, 1977. ISBN: 0-201-07616-0 (see pp. 27, 28).

[129] Peter Bruce and Andrew Bruce: *Practical Statistics for Data Scientists: 50 Essential Concepts*. Vol. 2017. O'Reilly, 2017 (see pp. 27, 28, 74).

[130] United States Department of Transportation. 2018. URL: https://www.fhwa.dot.gov/ohim/onh00/bar8.htm (see p. 35).

[131] Elena Paffumi, Michele De Gennaro, and Giorgio Martini: *European-wide study on big data for supporting road transport policy*. In: *Case Studies on Transport Policy* 6 (2018), pp. 785–802. URL: https://doi.org/10.1016/j.cstp.2018.10.001 (see p. 35).

[132] Jack Ewing: *With i3 Electric Car, BMW Tries to Ease Range Anxiety*. 2013. URL: https://www.nytimes.com/2013/07/16/business/with-i3-electric-car-bmw-tries-to-ease-range-anxiety.html#:~:text=Extensive%20trial%20projects%20with%20electrified,range%20is%20about%20120%20miles. (see p. 35).

[133] Abby Brown, Stephen Lommele, Alexis Schayowitz, and Wmily Klotz: *Electric Vehicle Charging Infrastructure Trends from the Alternative Fueling Station Locator: First Quarter 2020*. In: *Technical Report* NREL/TP-5400-77508 (2020). URL: https://www.nrel.gov/docs/fy20osti/77508.pdf (see p. 35).

[134] Wenzel Prochazka, Gudrun Pregartner, and Martin Cifrain: *Design-of-Experiment and Statistical Modeling of a Large Scale Aging Experiment for Two Popular Lithium Ion Cell Chemistries*. In: *Journal of The Electrochemical Society* 2013.160 (2013), A1039–A1051. ISSN: 00134651. URL: 10.1149/2.003308jes (see p. 36).

[135] Thomas Waldmann, Marcel Wilka, Michael Kasper, Meike Fleischhammer, and Margret Wohlfahrt-Mehrens: *Temperature dependent ageing mechanisms in lithium-ion batteries - a post-mortem study*. In: *Journal of Power Sources* 2014.262 (2014), pp. 129–135. URL: http://dx.doi.org/10.1016/j.jpowsour.2014.03.112 (see p. 36).

[136] Marius Bauer, Clemens Guenther, Michael Kasper, Mathias Petzl, and Michael A. Danzer: *Discrimination of degradation process in lithium-ion cells based on the sensitivity of aging indicators towards capacity loss*. In: *Journal of Power Sources* 2015.283 (2015), pp. 494–504. URL: http://dx.doi.org/10.1016/j.jpowsour.2015.02.130 (see pp. 36, 45).

[137] Andrew Price: *BMW Launches "DriveNow" Car-Sharing System in Munich*. 2011. URL: https://www.good.is/articles/bmw-launches-drivenow-car-sharing-system-in-munich (see p. 39).

[138] Patrick Dillenberger: *5000 Times Around the World*. 2021. URL: https://www.share-now.com/de/en/press-release-electric-record/ (see p. 39).

[139] Rebecca Killick and Idris A. Eckley: *An R Package for Changepoint Analysis*. In: *Journal of Statistical Software* 58.3 (2013), pp. 1–19. URL: https://doi.org/10.18637/jss.v058.i03 (see p. 42).

[140] Michael J. Evans and Jeffery S. Rosenthal: *Probability and Statistics: The Science of Uncertainty*. Toronto, 2009. ISBN: 978-1429224628 (see p. 42).

[141] Auger, I. E. and Lawrence C.E.: *Algorithms for the Optimal Identifiaciton of Segment Neighborhoods*. In: *Bulletin of Mathematical Biology* 51.1 (1989), pp. 39–54 (see p. 43).

[142] Edwards, A. and Cavalli-Sforza, L.: *A Method for Cluster Analysis*. In: *Biometrics* 21.2 (1965), pp. 362–375 (see p. 43).

[143] Leng, F., Wei, Z., Tan, C. M., and Yazami, R.: *Hierarchical degradation processes in lithium-ion batteries during aging*. In: *Electrochemica Acta* 256 (2017), pp. 52–62 (see pp. 45, 74).

[144] Jiangong Zhu et al.: *Investigation of lithium-ion battery degradation mechanisms by combining differential voltage analysis and alternating current impedance*. In: *Journal of Power Sources* 448.227575 (2020), pp. 1–12. URL: https://doi.org/j.jpowsour.2019.227575 (see p. 45).

[145] S.Sankarasubramanian and B. Krishnamurthy: *A capacity fade model for lithium-ion batteries including diffusion and kinetics*. In: *Electrochemica Acta* 70 (2012), pp. 248–254. URL: https://doi.org/10.1016/j.electacta.2012.03.063 (see p. 45).

[146] Kevin L. Gering, Sergly V. Sazhin, David K. Jamison, Christopher J. Michelbacher, Bor Yann Liaw, Mathhiew Dubarry, and Mikael Cugnet: *Investigation of path dependence in commercial lithium-ion cells chosen for plug-in hybrid vehicle duty cycle protocols*. In: *Journal of Power Sources* 2011.196 (2011), pp. 3395–3403. URL: doi:10.1016/j.jpowsour.2010.05.058 (see p. 45).

[147] Laisuo Su, Jianbo Zhang, Jun Huang, Hao Ge, Zhe Li, Fengchao Xie, and Bor Yann Liaw: *Path dependence of lithium ion cells aging under storage conditions*. In: *Journal of Power Sources* 2016.315 (2016), pp. 35–46. URL: http://dx.doi.org/10.1016/j.jpowsour.2016.03.043 (see p. 45).

[148] Fridolin Roeder and Srivastan Ramasubramanian: *A Review and Perspective on Path Dependencey in Batteries*. In: *Energy Techonolgy* 11 (2022). URL: 10.1002/ente.202200627 (see p. 45).

[149] Alexander Karger, Leo Wildfeuer, Deniz Ayguel, Arpit Maheshwari, Jan P. Singer, and Andreas Jossen: *Modeling capacity fade of lithium-ion batteries during dynamic cycling considering path depenedence*. In: *Journal of Energy Storage* 52.Part A (2022). URL: https://doi.org/10.1016/j.est.2022.104718 (see p. 45).

[150] M. Schimpe, M.E. von Kuepach, M. Naumann, H. C. Hesse, K. Smith, and A. Jossen: *Comprehensive Modeling of Temperature-Dependent Degradation Mechanisms in Lithium Iron Phosphate Batteries*. In: *Journal of The Electrochemical Society* 2018.165 (2018), A181–A193. ISSN: 00134651. URL: 10.1149/2.1181714jes (see pp. 45, 59).

[151] Matthieu Dubarry, Cyril Truchot, Bor Yann Liaw, Kevin Gering, Sergiy Sazhin, David Jamison, and Christopher Michelbacher: *Evaluation of commercial lithium-ion cells based on composite positive electrode for plug-in hybrid electric vehicle applications. Part II. Degradation mechanism under 2 C cycle aging*. In: *Journal of Power Sources* 196 (2011), pp. 10336–10343. URL: doi:10.1016/j.jpowsour.2011.08.078 (see p. 45).

[152] Matthieu Dubarry, M. Berecibar, A. Devie, D. Anseán, N. Omar, and I. Villarreal: *State of health battery estimator enabling degradation diagnosis: Model and algorithm description*. In: *Journal of Power Sources* 2017.360 (2017), pp. 59–69. URL: http://dx.doi.org/10.1016/j.jpowsour.2017.05.121 (see p. 45).

[153] M. Dubarry, v. Svoboda, R. Hwu, and B. Y. Liaw: *Incremental Capacity Analysis and Close-to-Equilibrium OCV Measurements to Quantify Capacity Fade in Commercial Rechargeable Lithium Batteries*. In: *Electrochemical and Solid-State Letters* 9 (2006), A454–A457. URL: 10.1149/1.2221767 (see p. 45).

[154] Kohei Honkura, Hidetoshi Honbo, Yoshimasa Koishikawa, and Tatsuo Horiba: *State Analysis of Lithium-Ion Batteries Using Discharge Curves*. In: *ECS Transactions* 13.19 (2008), pp. 61–73. URL: 10.1149/1.3018750 (see p. 45).

[155] Arnaud Devie and Matthieu Dubarry: *Durability and reliability of Electric Vehicle Batteries under Electric Utility Grid Operations. Part 1: Cell-to-Cell Variations and Preliminary Testing*. In: *Batteries* 2.28 (2016), pp. 1–13. URL: https://10.3390/batteries2030028 (see p. 45).

[156] Ira Bloom, Andrew N. Jansen, Daniel P. Abraham, Jamie Knuth, Scott A. Jones, Vincent S. Battaglia, and Gary L. Henriksen: *Differential voltage analyses of high-power, lithium-ion cells: 1. Technique and application*. In: *Journal of Power Sources* 139 (2005), pp. 295–303. URL: 10.1016/j.jpowsour.2004.07.021 (see p. 45).

[157] Peter Keil and Andreas Jossen: *Calendar Aging of NCA Lithium-Ion Batteries Investigated by Differential Voltage Analysis and Coulomb Tracking*. In: *Journal of The Electrochemical Society* 164.1 (2017), A6066–A6074. ISSN: 00134651. URL: 10.1149/2.0091701jes (see p. 45).

[158] Julius Schmitt, Markus Schindler, and Andreas Jossen: *Change in the half-cell open-circuit potential curves of silicon-graphite and nickel-rich lithium nickel manganese cobalt oxide during cycle aging*. In: *Journal of Power Sources* 506 (2021). URL: https://doi.org/10.1016/j.jpowsour.2021.230240 (see p. 45).

[159] S. Watanabe, M. Kinoshita, and K. Nakura: *Capacity fade of $LiNi_{(1-x-y)}Co_xAl_yO_2$ cathode for lithium-ion batteries during accelerated calendar and cycle life test: Comparison of the surface changes on cathode during long term storage testing of high energy density cylindrical lithium-ion cells*. In: *Journal of Power Sources* 196.16 (2011), pp. 6906–6910. URL: http://dx.doi.org/j.jpowsour.2010.12.028 (see p. 46).

[160] S. Watanabe, M. Kinoshita, T. Hosokawa, K. Morigaki, and K. Nakura: *Capacity fade of $LiNi_{(1-x-y)}Co_xAl_yO_2$ cathode for lithium-ion batteries during accelerated calendar and cycle life test: surface analysis of $LiNi_{(1-x-y)}Co_xAl_yO_2$ cathode after cycle tests in restricted depth of discharge ranges during long term storage test*. In: *Journal of Power Sources* 258 (2014), pp. 210–217. URL: http://dx.doi.org/j.jpowsour.2014.02.018 (see p. 46).

[161] S. Watanabe, M. Kinoshita, T. Hosokawa, K. Morigaki, and K. Nakura: *Capacity fade of $LiNi_{(1-x-y)}Co_xAl_yO_2$ cathode for lithium-ion batteries during accelerated calendar and cycle life test: Effect of depth of discharge in charge-discharge cycling on the suppression of the micro-crack generation of $LiNi_{(1-x-y)}Co_xAl_yO_2$*. In: *Journal of Power Sources* 260 (2014), pp. 50–56. URL: http://dx.doi.org/j.jpowsour.2014.02.103 (see p. 46).

[162] M. Broussely, P. Biensan, F. Bonhomme, P. Blanchard, S. Herreyre, K. Nechev, and R. F. Staniewicz: *Main aging mechanisms Li ion batteries*. In: *Journal of Power Sources* 146.1-2 (2005), pp. 90–96. URL: http://dx.doi.org/j.jpowsour.2005.03.172 (see p. 46).

[163] S. Bourlot, P. Blanchard, and S. Robert: *Investigation of aging mechanisms of high power Li-ion cells used for hybrid electric vehicles*. In: *Journal of Power Sources* 196.16 (2011), pp. 6841–6846. URL: http://dx.doi.org/j.jpowsour.2010.09.103 (see p. 46).

[164] Ira Bloom, Jon Christophersen, and Kevin Gering: *Differential voltage analyses of high-power lithium-ion cells: 2. Applications*. In: *Journal of Power Sources* 139 (2005), pp. 304–313. URL: 10.1016/j.jpowsour.2004.07.022 (see p. 46).

[165] Ian D. Campbell, Mohamed Marzook, Monica Marinescu, and Gregory J. Offer: *How Observable is Lithium Plating? Differential Voltage Analysis to Identify and Quantify Lithium Plating Following Fast Charging of Cold Lithium-Ion Batteries*. In: *Journal of The Electrochemical Society* 166.4 (2019), A725–A739. ISSN: 00134651. URL: 10.1149/2.0821904jes (see p. 46).

[166] Christoph R. Birkl, Matthew R. Roberts, Euan McTurk, Peter G. bruce, and David A. Howey: *Degradation diganostics for lithium ion cells*. In: *Journal of Power Sources* 341 (2017), pp. 373–386. URL: https://doi.org/10.1016/j.jpowsour.2016.12.011 (see p. 46).

[167] Johannes Sieg, Jochen Bandlow, Tim Mitsch, Daniel Dragicevic, Torben Marterna, Bernd Spier, Heiko Witzenhausen, Madeleine Ecker, and Dirk Uwe Sauer: *Fast charging of an electric vehicle lithium-ion battery at the limit of the lithium deposition process*. In: *Journal of Power Sources* 427 (2019), pp. 260–270 (see p. 46).

[168] Simon F. Schuster, Tobias Bach, Elena Fleder, Jana Müller, Martin Brand, Gerhard Sextl, and Andreas Jossen: *Nonlinear aging characteristics of lithium-ion cells under different operation conditions*. In: *Journal of Energy Sources* 1 (2015), pp. 44–53 (see p. 46).

[169] A. Brand, L. Allen, M. Altman, M. Hlava, and J. Scott: *Beyond authorship: Attribution, contribution, collaboration and credit*. In: *Learned Publishinig* 28 (2015), pp. 151–155 (see pp. 46, 60, 74).

[170] Xiao-Sheng Si, Wenbin Wang, Chang-Hua Hu, and Dong-Hua Zhou: *Remaining useful life estimation- A review on the statistical data driven approaches*. In: *European Journal of Operational Research* 2011.213 (2011), pp. 1–14. URL: doi:10.1016/j.ejor.2010.11.018 (see p. 59).

[171] Wladislaw Waag, Christian Fleischer, and Dirk Uwe Sauer: *Critical review of the methods for monitoring of lithium-ion batteries in electric and hybrid vehicles*. In: *Journal of Power Sources* 2014.258 (2014), pp. 321–339. URL: http://dx.doi.org/10.1016/j.jpowsour.2014.02.064 (see p. 59).

[172] International Energy Agency, ed.: *Germany 2020: Energy Policy Review*. www.iea.org/t&c/, 2020 (see p. 59).

[173] M. Berecibar, I. Gandiaga, I. Villarreal, N. Omar, J. Van Mierla, and P. Van den Bossche: *Critical review of state of health estimation methods of Li-ion batteries for real applications*. In: *Renewable and Sustainable Energy Reviews* 2016.56 (2016), pp. 572–587 (see p. 59).

[174] Phillip Kollmeyer: *Panasonic 18650PF Li-ion Battery Data*. 2018. URL: 10.17632/wykht8y7tg.1 (see p. 59).

[175] Yi Li, Kailong Liu, Aoife M. Foley, Alana Zülke, Maitane Berecibar, Elise Nanini-Maury, Joeri Van Mierlo, and Harry E. Hoster: *Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review*. In: *Renewable and Sustainable Energy Reviews* 2019.113 (2019), pp. 1–18 (see p. 59).

[176] MacQueen, J.: "Some methods for classification and analysis of multivariate observations." In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. Ed. by MSP. Berkeley, Calif.: University of California Press, 1967, pp. 281–297. URL: https://projecteuclid.org/euclid.bsmsp/1200512992 (see p. 59).

[177] Arthur, D. and Vassilvitskii, S.: "K-Means++: The Advantages of Careful Seeding." In: *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. Ed. by ACM-SIAM. SODA '07. USA: Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035. ISBN: 9780898716245 (see p. 59).

[178] Kristen A. Severson et al.: *Data-driven prediction of battery cycle life before capacity degradation*. In: *Nature Energy* 2019.Vol. 4 (2019), pp. 383–391. URL: https://doi.org/10.1038/s41560-019-0356-8 (see pp. 59, 74).

[179] Python Software Foundation: *Python*. 2020. URL: http://www.python.org (see p. 60).

[180] Francois Chollet: *Keras*. 2020. URL: https://keras.io/ (see p. 60).

[181] M. Fleischhammer, T. Waldmann, G. Bisle, B.I: Hogg, and M. Wohlfahrt-Mehrens: *Interaction of cyclic ageing at high-rate and low temperatures and safety in lithium-ion batteries*. In: *Journal of Power Sources* 274 (2015), pp. 432–439. URL: DOI:10.1016/j.jpowsour.2014.08.135 (see p. 73).

[182] Yangying Zhu et al.: *Fast lithium growth and short circuit induced by localized-temperature hotspots in lithium batteries*. In: *Nature Communications* 10.2067 (2019). URL: https://doi.org/10.1038/s41467-019-09924-1 (see p. 73).

[183] Anna Tomaszewska et al.: *Lithium-ion battery fast charging: A review*. In: *eTransportation* 1 (2019), pp. 1–22. ISSN: 25901168 (see p. 73).

[184] UK, B.: *New Market. New entrants. New challenges. Battery Electric Vehicles*. In: *Bloomberg* (2019), pp. 1–20 (see p. 74).

[185] Bhattacharyya, R., Key, B., Chen, H., Best, A. S., Hollenkamp, A. F., and Grey, C.: *In Situ NMR observation of the formation of metallic lithium microstructures in lithium batteries*. In: *Nature Materials* 9 (2010), pp. 504–510 (see p. 74).

[186] Monroe, C. and Newman, J.: *Dendrite Growth in Lithium/Polymer Systems: A Propagation Model for Liquid Electrolytes under Galvanostatic Conditions*. In: *Journal of The Electrochemical Society* 150 (2003), A1377–A1384. ISSN: 00134651 (see p. 74).

[187] L. O. Valoen and J. N. Reimers: *Transport properties of $LiFePO_6$-based li-ion battery electrolytes*. In: *Journal of The Electrochemical Society* 152 (2005), A882–A891. ISSN: 00134651 (see p. 74).

[188] Klein, R., Chaturvedi, N. A., Christensen, J., Ahmed, J., Findeisen, R., and Kojic, A.: *State estimation of a reduced electrochemical model of a lithium-ion battery*. In: *IEEE Xplore* 2010 American Control Conference (2010), pp. 6618–6623 (see p. 74).

[189] Fuller, T. F.: *Simulation and Optimization of the Dual Lithium Ion Insertion Cell*. In: *Journal of The Electrochemical Society* 141.1 (1994), p. 1. ISSN: 00134651 (see p. 74).

[190] Newman, J. and Thomas-Alyea, K. E.: *Electrochemical Systems*. 3. Aufl. s.l.: Wiley-Interscience, 2012. ISBN: 0-471-47756-7 (see p. 74).

[191] Sturm, J., Ludwig, S., Zwirner, J., Ramirez-Garcia, C., Heinrich, B., Horsche, M. F., and Jossen, A.: *Suitability of physiochemical models for embedded systems regarding a nickel-rich, silicon-graphite lithium-ion battery*. In: *Journal of Power Sources* 436 (2019), pp. 1–17 (see p. 74).

[192] Tomaszewska, A. et al.: *Lithium-ion battery fast charging; A review*. In: *eTransportation* 1 (2019), pp. 1–28. ISSN: 25901168 (see p. 74).

[193] Mai, W., Colclasure, A. M., and Smith, K.: *Model-Instructed Design of Novel Charging Protocols for the Extreme Fast Charging of Lithium-Ion Batteries Without Lithium Plating*. In: *Journal of The Electrochemical Society* 167 (2020), pp. 1–16. ISSN: 00134651 (see p. 74).

[194] C. M. Doyle, T. F. Fuller, and J. Newman: *Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell*. In: *Journal of The Electrochemical Society* 140.6 (1993), pp. 1526–1533. ISSN: 00134651 (see p. 74).

[195] Liu, D., Yin, X., Song, Y., Liu, W., and Peng, Y.: *An On-line State of Health Estimation of Lithium-Ion Battery Using Unscented Particle Filter*. In: *IEEE Access* Volume 6 (2018), pp. 40990–41001 (see p. 74).

[196] Richardson, R., Osborne, M., and Howey, D.: *Gaussian process regression for forecasting battery state of health*. In: *Journal of Power Sources* 357 (2017), pp. 209–219 (see p. 74).

[197] Zhang, Y., Xiong, R., He, H., and Pecht, M. G.: *Long Short-Term Memory Recurrent Neural Network for Remaining Useful Life Prediction of Lithium-Ion Batteries*. In: *IEEE Transactions on Vehicular Technology* 67.7 (2018), pp. 5695–5705. ISSN: 0018-9545. DOI: 10.1109/TVT.2018.2805189 (see p. 74).

[198] Kashkooli, A. G., Fathiannasab, H., Mao, Z., and Chen, Z.: *Application of Artificial Intelligence to State-of-Charge and State-of-Health Estimation of Calendar-Aged Lithium-Ion Pouch Cells*. In: *Journal of The Electrochemical Society* 166.4 (2019), A605–A615. ISSN: 00134651 (see p. 74).

[199] Li, Y., Liu, K., Foley, A. M., Zülke, A., Berecibar, M., Nanini-Maury, E., van Mierlo, J., and Hoster, H. E.: *Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review*. In: *Renewable and Sustainable Energy Reviews* 113 (2019), p. 109254. DOI: 10.1016/j.rser.2019.109254 (see p. 74).

[200] Lin, X.: *Real-Time Prediction of Anode Potential in Li-Ion Batteries Using Long Short-Term Neural Networks for Lithium Plating Prevention*. In: *Journal of The Electrochemical Society* 166 (2019), A1893–A1904. ISSN: 00134651 (see p. 74).

[201] V. Ramadesigan, P. W. C. Northrop, S. De, S. Santhanagopalan, R. D. Braatz, and V. R. Subramanian: *Modeling and simulation of lithium-ion batteries from a systems engineering perspective*. In: *Journal of The Electrochemical Society* 159.3 (2012), pp. 31–45. ISSN: 00134651 (see p. 74).

[202] D. Bernardi, E. Pawlikowski, and J. Newman: *A General Energy Balance for Battery Systems*. In: *Journal of The Electrochemical Society* 132 (1985), pp. 5–12. ISSN: 00134651. URL: 10.1149/1.2113792 (see p. 74).

[203] Ennifar, H.: *Electrochemical State Estimation of a Lithium-Ion Battery: Using a Pseudo-2D Model and an Extended Kalman Filter*. Masterarbeit. München: Technische Universität München, 2017 (see p. 74).

[204] Rawlings, J. O., Pantula, S. G., and Dickey, D. A.: *Applied Regression A Research Tool: Second Edition*. Springer, 1998 (see p. 74).

[205] Everitt, B. and Hothorn, T.: *An Introduction to Applied Multivariate Analysis with R*. New York, NY: Springer New York, 2011. ISBN: 978-1-4419-9649-7. DOI: 10.1007/978-1-4419-9650-3 (see p. 74).

List of Figures

List of Tables