Machine Learning-based Online State-of-Health Estimation of Electric Vehicle Batteries

Artificial Intelligence Applied to Battery Management Systems

Alberto Barragán Moreno

Energy Technology, MCE4-1027, 2021-05

Master's Thesis

STUDENT REPORT

AALBORG UNIVERSITY

This report was written using Overleaf™ online LaTeX editor. Python v3.6.9 software was developed using Jupyter, along with packages Matplotlib v3.1.2, Numpy v1.18.1, Pandas v1.1.5, SciPy v1.4.1, Sklearn v0.24.1, TensorFlow v2.1.0. Some of the graphs and diagrams were created using Mathcha.io online editor and Inkscape. Front cover picture credit to Tennen-Gas, from Wikimedia Commons under cc-by-sa-3.0 license.

AALBORG UNIVERSITY
STUDENT REPORT

Energy Technology
Aalborg University
http://www.aau.dk

Title:
Machine Learning-based Online SoH Estimation of Electric Vehicle Batteries

Theme:
Mechatronics

Project Period:
Spring Semester 2021

Project Group:
MCE4-1027

Participant:
Alberto Barragán Moreno

Supervisor:
Erik Schaltz

Page Numbers: 88

Date of Completion:
May 28, 2021

Abstract:

With the increasing adoption of electrical vehicles (EV) by the general public, a lot of research is being conducted in Li-ion battery-related topics, having state-of-health (SoH) estimation a prominent role. In this work, machine-learning techniques are applied to estimate this parameter online in EV applications and in diverse scenarios. After a thorough analysis regarding cell ageing and the main factors influencing this process, a total of three approaches are developed: the first one is based on voltage measurements at fixed state-of-charge (SoC) levels, while the second one uses the charge gradients between certain voltage milestones as health indicators. The last method predicts capacity- and impedance-based SoH from a limited set of impedance measurements. The proposed approaches are tested for different chemistries and in various realistic scenarios to evaluate their performance and applicability. Great accuracy was obtained in all cases, with MAE as low as 0.4 % when making future predictions, 0.5 % when inferring at multiple temperatures, 0.4 % for diverse, realistic ageing, and 0.2 % for storage ageing. Thus, the methods constitute a powerful and viable alternative for online SoH estimation in real-world EV.

Preface

This thesis report was written by Alberto Barragán Moreno as only member of the MCE4-1027 group, belonging to the 4^{th} semester of the program *Master of Science in Energy Engineering, specialization in Mechatronic Control Engineering* at Aalborg University. This thesis extends across the period from February 1^{st}, 2021 until May 28^{th}, 2021. The title of the thesis is *Machine Learning-based Online State-of-Health Estimation of Electric Vehicle Batteries* and is under the general topic *Mechatronics*. The thesis was supervised by Associate Professor Erik Schaltz.

I would like to thank Erik for his wise and experienced guidance throughout this thesis, and to the members of the *E-Mobility and Industrial Drives* research group who carried out the experiments and allowed me to use the data for this project. I also want to thank my family for their love and support during all these years as a student, and my friends, who helped me overcome every challenge I had to face over the years.

Reading guide

This written report serves as a means to capture the theory, methods, assumptions and simulations necessary for the completion of this thesis. In order to attain a clear understanding of the whole project the report should be read in the presented order.

The report includes a table of contents displaying the chapters and sections with their corresponding starting page numbers. Nomenclature, as well as lists of both tables and figures are also part of this report. Elements such as tables, figures and equations are numbered for reference. A list of bibliography can be found at the end of the report. Citations to these references follow the IEEE style and are presented in order of appearance in the report.

This report is structured as follows: Chapter 1 introduces the topic of battery's state estimation and presents a state-of-the-art review of previous research. The chapter ends with the description of the methodology and the thesis statement. Chapter 2 explains background theory regarding Li-ion batteries and their most relevant characteristics, with a focus on degradation. Chapter 3 sets forth the basic concepts of artificial intelligence and neural networks. A state-of-health estimation method based on voltage measurements is designed and tested in Chapter 4, while a charge gradient-based approach is developed in Chapter 5. Lastly, an impedance-based state-of-health estimator is implemented in Chapter 6. Results are discussed in Chapter 7, conclusions are drawn in Chapter 8 and ideas for possible future works are presented in Chapter 9. Additionally, a thorough description of publicly available battery data sets is made in Appendix A.

Alberto Barragán Moreno
<abarra18@student.aau.dk>

Thesis summary

Over the last decade, a progressive adoption of electric vehicles by the general public has taken place, although nowadays they just constitute a tiny fraction of the sales worldwide. Even so, the field of battery technology is still in a high-paced development stage. This includes not only research on materials, form factors and arrangements, but also software aspects such as charge and health monitoring. At the same time, artificial intelligence applications have experienced an exponential growth, mostly due to increased availability of cheap powerful computing power and high-speed communications, as well as to the generation of massive amounts of data. This facilitates the deployment of this type of algorithms in tightly-constrained hardware, as it could be the case of the battery management system of an electric vehicle.

The main emphasis of this thesis is put on developing and evaluating new methodologies which take advantage of powerful, yet lightweight, intelligent algorithms in order to monitor the state of health of Li-ion batteries intended for usage in electric vehicles. Thus, after the current problems and state of the art had been discussed in Chapter 1, a study of the fundamental theory regarding Li-ion batteries was carried out in Chapter 2. It was composed of two parts: first, to review the main testing procedures used in this field, the information they provide and the way they are affected by e.g. ageing or temperature. Secondly, a thorough analysis of the main factors that have an impact on the way Li-ion cells deteriorate, including various usage conditions and chemistries. Since it was chosen to only use intelligent algorithms for this project, a basic explanation of their fundamental theory was presented in Chapter 3 in order to provide a knowledge base so that the developed methods are better understood. The first method was designed in Chapter 4. It was based on measuring the terminal voltage at a reduced amount of state-of-charge milestones. For practicality reasons, intermediate values were chosen. Then, in Chapter 5, another algorithm was proposed, but using the charge gradient of the battery as inputs to the neural networks. In this case, the measuring points were selected based on how this characteristic evolved throughout ageing in order to detect the most promising ones in terms of performance. Both types of algorithms were tested in scenarios of future, long-term predictions; estimation at multiple temperatures; and diversity of ageing procedures. The last approach was introduced in Chapter 6 from a completely different perspective: rather than using the measured voltage during charge, it received as inputs a very limited set of impedance measurements. In this case the algorithm was tested for different scenarios with cells which had undergone storage ageing in assorted conditions.

Analysis of the obtained results proved that all the proposed approaches were capable of delivering very low error rates when facing new situations, with a few exceptions in which the algorithms failed to learn the underlying behaviour. The diversity of methods and experiments allowed for a better comparison of which ones would be more appropriate for a given scenario.

The conclusion of this thesis is that artificial intelligence is a very powerful tool which can have plenty of applications in battery management systems, providing simplicity, great accuracy and generalisation capabilities. On the other hand, these algorithms require vast amounts of clean experimental data to learn a certain behaviour, which in the specific case of batteries may take several years to collect.

Contents

Nomenclature

Table 1: Description of symbols

Symbol	Description	Unit
b	Neuron's bias	[-]
C_i	i^{th}-branch polarization capacitance	F
e_i	i^{th}-sample estimation error	[-]
$f(\alpha)$	Neuron's activation function	[-]
$\mathbf{H}$	Enhancement nodes' output in BLS	[-]
I_t	Terminal current	A
J	Cost function of FC-FNN	[-]
MAE	Mean absolute error	[-]
ME	Maximum absolute error	[-]
N	Number of samples	[-]
$Q_{max,ch,aged}$	Maximum available charge capacity at aged stage	A h
$Q_{max,ch,pristine}$	Maximum available charge capacity at pristine stage	A h
$Q_{max,dis,aged}$	Maximum available discharge capacity at aged stage	A h
$Q_{max,dis,pristine}$	Maximum available discharge capacity at pristine stage	A h
R_i	i^{th}-branch polarization resistance	Ω
$R_{0,c}$	Charge-branch internal resistance	Ω
$R_{0,d}$	Discharge-branch internal resistance	Ω
RMSE	Root mean squared error	[-]
V_{OC}	Open-circuit voltage	V
V_t	Terminal voltage	V
$w^{x,n}_{n,1,m}$	Weight from input x_n to m^{th} neuron in 1^{st} hidden layer in FC-FNN	[-]
$w^{z,i-1,l}_{n,i,p}$	Weight from l^{th} neuron in $(i\text{-}1)^{th}$ hidden layer to p^{th} neuron in i^{th} hidden layer in FC-FNN	[-]
$w^{z,i,p}_{y,k}$	Weight from p^{th} neuron in i^{th} hidden layer to output y_k in FC-FNN	[-]
$w^{x,p}_{f,n,l}$	Weight from input x_p to feature mapping $f_{n,l}$ in BLS	[-]
$w^{z,n,l}_{e,m,s}$	Weight from feature mapping $f_{n,l}$ to enhancement node $e_{m,s}$ in BLS	[-]
$w^{z,n,l}_{y,q}$	Weight from feature mapping $f_{n,l}$ to output y_q in BLS	[-]
$w^{e,m,s}_{y,q}$	Weight from enhancement node $e_{m,s}$ to output y_q in BLS	[-]
$\mathbf{W}$	Weight matrix in BLS	[-]
x_i	i^{th} input to neural network	[-]
y_i	Real i^{th}-target value	[-]
$\hat{y}_i$	Estimated i^{th}-target value	[-]
$\mathbf{Z}$	Feature mappings' output in BLS	[-]
γ	Learning rate	[-]
δ	Hidden layer neuron's error	[-]
$\eta_{ch\text{-}dis,aged}$	Cell efficiency at aged stage	%

Continuation of Table 1

Symbol	**Description**	**Unit**
$\eta_{\text{ch-dis,pristine}}$	Cell efficiency at pristine stage	%
λ	Ridge-regression regularisation coefficient	[-]
μ_{res}	Residuals' mean value	[-]
$\sigma_{\frac{\Delta Q}{\Delta V}}$	Charge gradients' standard deviation	$\frac{\text{A h}}{\text{V}}$
$\sigma_{\mathfrak{Im}(Z)}$	Impedance's imaginary part's standard deviation	Ω
$\sigma_{\mathfrak{Re}(Z)}$	Impedance's real part's standard deviation	Ω
σ_{res}	Residuals' standard deviation	[-]

Table 2: Description of abbreviations

Abbreviation	Definition
AI	Artificial intelligence
ANN	Artificial neural network
BLS	Broad learning system
BMS	Battery management system
BoL	Beginning of life
CC	Constant current
CCCV	Constant-current, constant-voltage
CV	Constant voltage
DoD	Depth of discharge
DRBS	Discrete random binary sequence
DV	Differential voltage
ECM	Equivalent-circuit model
EIS	Electrochemical impedance spectroscopy
EoL	End of life
EQM	Electrochemical model
ESS	Energy-storage system
EV	Electric vehicle
FC-FNN	Fully-connected feed-forward neural network
FFT	Fast Fourier transform
FNN	Feed-forward neural network
FOM	Fractional-order model
FPGA	Field-programmable gate array
GPU	Graphics processing unit
HPPC	Hybrid-pulse power characterization
IC	Incremental capacity
ICEV	Internal combustion engine vehicle
LCO	Lithium-cobalt-oxide
LFP	Lithium-iron-phosphorus
LMO	Lithium-manganese-oxide
LSTM-RNN	Long short-term memory recurrent neural network
LTO	Lithium-titanium-oxide
ML	Machine learning
NCA	Nickel-cobalt-aluminium
NE	Negative electrode
NMC	Nickel-manganese-cobalt
OCV	Open-circuit voltage
PE	Positive electrode
PRBS	Pseudo-random binary sequence
PRTS	Pseudo-random ternary sequence
RC	Resistor-capacitor
ReLU	Rectified linear unit
RNN	Recurrent neural network

Continuation of Table 2

Abbreviation	Definition
RPT	Reference performance test
RVFLNN	Random-vector functional-link neural network
SEI	Solid-electrolyte inter-phase
SGD	Stochastic gradient descent
SoC	State of charge
SoH	State of health
SoH_Q	Capacity-based state of health
SoH_R	Resistance-based state of health
SoP	State of power
SoX	Battery states

List of Figures

List of Tables

Chapter 1

Introduction

In this chapter, a brief introduction to the topic of electric vehicles and battery technology is made, as well as an analysis of the problem this thesis aims at solving, followed by a state-of-the-art discussion. A project methodology is outlined next. The chapter ends with the thesis statement and the project limitations.

Contents

The transportation systems of the world are undergoing a deep transformation, moving from internal combustion engine vehicles (ICEV), which use petroleum- or gas-derivatives as fuel, to electric vehicles (EV) where the power is obtained from some kind of energy storage system (ESS). The most common ones nowadays are electrochemical batteries, hydrogen fuel cells and super-capacitors [1, pp. 237, 497]. This transition is mainly motivated by social and economic reasons. First, society is becoming increasingly concerned regarding the negative impact of pollution, which includes health risks and climate change. Then, fossil-fuel reserves become scarcer every day, which leads to rising prices until they become eventually depleted. Although driving performance is not usually listed as an argument for this transition, it has been demonstrated that EV have some advantages over traditional ICEV, such as more powerful acceleration thanks to the instantaneous torque availability in electric motors. On the other hand, driving range extension is still a significant obstacle to be overcome in order for EV to become mainstream. Since these technologies have only been thoroughly researched for less than two decades, there remain plenty of challenges and problems to be solved.

1.1 Problem analysis

One of the main problems when it comes to electric vehicles is the proper monitoring of the battery pack. These are very complex electrochemical systems and, thus, it is very difficult to derive accurate models which cover every possible operation scenario. Furthermore, a battery pack can be composed of hundreds of individual cells, which can present some differences between them due to manufacturing and operation conditions. Among many others, one of the main tasks of battery management systems (BMS) is precisely to try and keep all the cells as equalized as possible, which is a considerable challenge still under intensive research [2, p. 91], [1, p. 270].

There are three metrics of special interest in a battery, namely the state of charge (SoC), the state of health (SoH) and the state of power (SoP), referred to as SoX in general. The first one is related to the remaining electric energy in the cells, while the second one represents battery degradation through time and use; the SoP quantifies how much power the ESS can exchange under certain conditions. The SoX cannot be directly measured, therefore the estimation of these states is usually also carried out by the BMS based on the available measurements, typically voltage, current and temperature. Some of the challenges of SoX estimation are:

- *Sensitivity analysis*: before trying to model the behavior of an electrochemical cell, it is essential to analyse which factors and variables affect the system and how they do so. This implies extensive testing of battery cells under controlled laboratory conditions. Since some of the system's dynamics are very slow (in the order of days or months), this testing stage can last up to several years until a solid knowledge base is formed. Besides exogenous factors, such as temperature, endogenous ones also have a major influence, as is the case of the manufacturing process or the specific battery chemistry.

- *System modelling*: after the variables of interest have been identified, several types of models can be derived depending on the phenomenon which needs to be studied. A common remark is that there is usually a trade-off between accuracy and simplicity or performance. For example, electrochemical models can be very accurate in both short- and long-term, but they also require finding values for many parameters based on design and experimental data.

- *Online operation*: some of the states, especially SoC, can actually be measured under specific conditions. In this case, if experimental conditions can be tightly controlled and a high-accuracy current sensor is used, the remaining charge in the battery can be obtained by integrating the measured current, which is referred to as ampere or coulomb counting. However, in realistic driving scenarios, the conditions can change constantly and the quality of the measurements is considerably lower than the ones in a laboratory set-up. Thus, estimation algorithms should be designed in a way so that they can be applied in as diverse as possible scenarios.

These arguments support the idea that new methodologies are needed in order to improve the SoX estimation process in electric vehicles. Since SoH has a significant impact on the other two states, the focus will be put on the online estimation of the battery's degradation stage. This information can afterwards be used by the BMS to adapt the controllers, to provide more accurate range estimations and to avoid potentially hazardous situations for the battery pack. The next section discusses some of the relevant research that has been carried out in this field in recent years.

1.2 State-of-the-art review

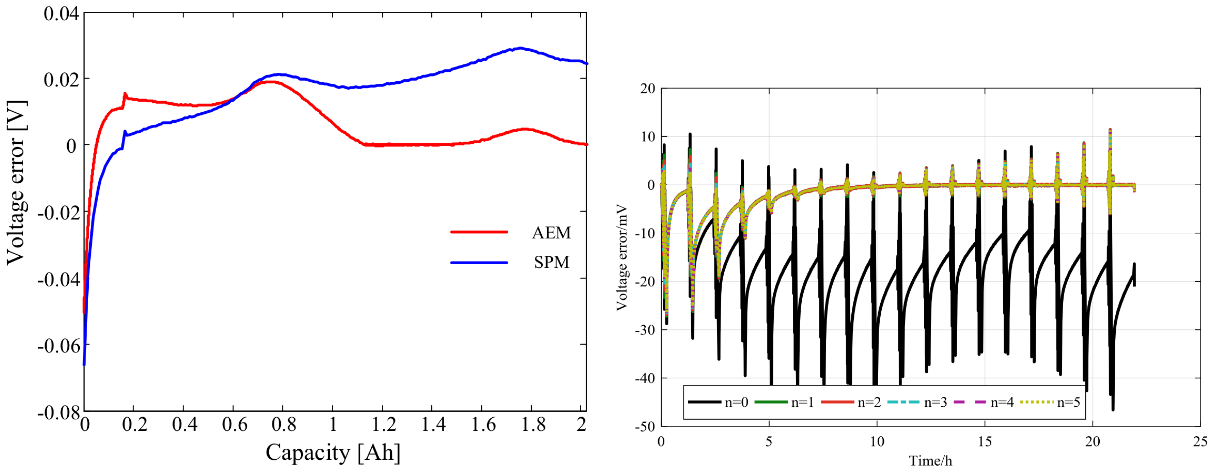
In the last two decades, battery-related research has experienced an exponential growth, in part due to the progressive adoption of, mainly, hybrid electric vehicles. This research includes different approaches for modelling batteries and several methods for SoH estimation.

1.2.1 Modelling

Over the years, researchers have developed several ways to describe and model the behaviour of electrochemical cells. One of the main goals is to try to estimate the open-circuit voltage (OCV) from the terminal voltage. According to [1, p. 258] and [3, p. 63], there are three prominent classes: the electrochemical models (EQM), which describe the physical-chemical reactions within the cells based on differential and algebraic equations; the equivalent-circuit models (ECM), which use resistor-capacitor (RC) networks and voltage sources to try and mimic the static and dynamic voltage characteristics of the system; and the fractional-order models (FOM), which replace the capacitors in the ECM by constant-phase elements and use fractional calculus to solve the equations in the time domain. A comparison between two simplified EQM-based SoC estimation algorithms was presented in [4]. Both of them showed great performance when tested on driving cycles in a laboratory setup, as shown in Figure 1.1a. Here it can be observed that both methods yield voltage errors under 3 % for most of the capacity range. The study serves as an example of how complex these models can be, even in their simplified versions. [5] proposed a solution to the modelling problem with two ECM: one for constant-current charging and one for dynamic operation. Each model, whose parameters are found by means of a least-squares optimization algorithm, contains two RC networks, so that both short- and long-term dynamics can be modelled. In [3, p. 81], the author compared the accuracy of several ECM with different amounts of RC units, being the one with a single RC network the best trade-off between complexity and accuracy, as observed in Figure 1.1b. It clearly shows that at least one RC branch is necessary to model the battery's behaviour. Adding more branches can increase fidelity at the cost of having to estimate more parameters. Even though results come from different sources, Figure 1.1 supports the idea that ECM can reach similar accuracy levels to EQM but with a much simpler structure. Yet another modification to the ECM, as employed in [1, p. 258] and [6], is to bifurcate the internal resistance branch of the model. This is motivated by the fact that there exists some hysteresis effect between charge and discharge processes. Furthermore, the efficiency can also differ slightly. An FOM-based SoC estimation method was proposed in [7]. The model's parameters were fitted from experimental data using a genetic algorithm and the estimation was done by means of an unscented Kalman filter. Good accuracy was proved when validated in several driving cycles. A common remark for these works is that none of the algorithms accounts for the effects of temperature nor ageing, which are known to have a significant influence in the battery's behaviour.

1.2.2 Impedance-based SoH estimation

Considering the cell as an electrical system, an impedance parameter can be defined as the ratio of voltage to current, both measured at the terminals. Battery impedance is commonly frequency-dependent, and experimental data shows that it is also affected by ageing, temperature and, to a lesser extent, SoC [8]. Thus, it can also be used to quantify the battery's

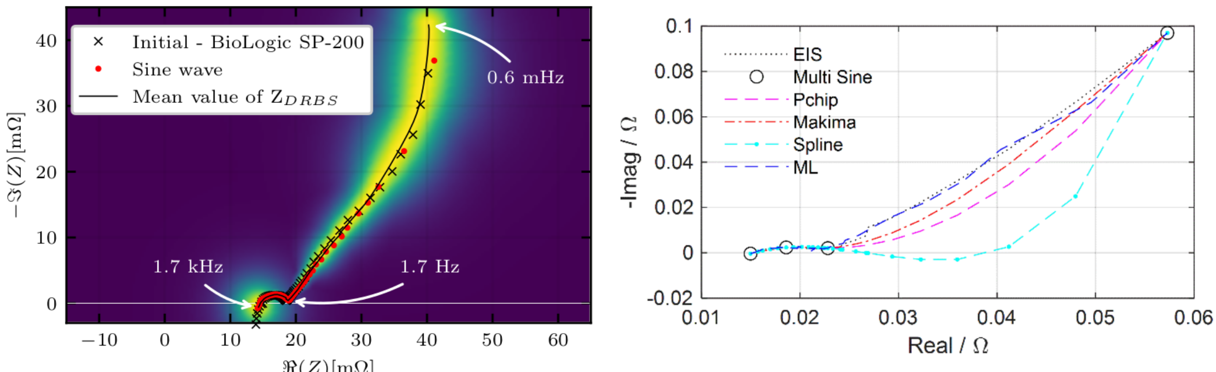


(a) Error comparison of the proposed reduced EQM [4]

(b) Error comparison between ECM of different orders [3]

Figure 1.1: Validation results of EQM and ECM models in [4] and [3], respectively

degradation stage. The most straightforward approach to measure the impedance is to inject low-amplitude, single-frequency sinusoidal current (or voltage) signals and measuring the voltage (or current) response. This technique is known as electrochemical impedance spectroscopy (EIS) and its main disadvantage is that it can take a long time to complete, since the interesting frequencies may be as low as 1 mHz. Therefore, a lot of focus has been put into alternative ways of measuring the impedance online, with discrete pseudo-random sequences having a prominent role. This is the case of [8], where a pseudo-random binary sequence (PRBS) was injected into the battery and the impedance spectrum was obtained by means of the fast Fourier transform (FFT). The paper also compares the performance of PRBS of different bit lengths and discusses some implementation aspects in EV. A similar work was done in [9], but the researchers used the continuous wavelet transform instead of the FFT. One advantage of this method is that it keeps both time and frequency information, rather than being limited to just the frequency domain. Figure 1.2a shows that this approach can provide impedance measurements on par with EIS, specially at mid- to high-frequency, but taking just 97 min to complete; the equivalent single-sine approach would take 54 h. The yellow shaded region represents the confidence interval of the impedance estimates. [10] proposed to use a pseudo-random ternary sequence (PRTS), which adds an extra excitation level to the PRBS, and obtained better results at lower frequencies. These works demonstrated that these methods can accurately estimate impedance while taking much less time than conventional EIS tests. Artificial intelligence (AI) has also been applied to this topic. For example, in [11] machine learning (ML) was used to obtain the full impedance spectrum based on a limited set of EIS measurements. When compared to other polynomial interpolation methods, the proposed approach performed considerably better, especially at low frequencies, as depicted in Figure 1.2b. It also shows that remarkable accuracy can be obtained by using very limited set of test points. Although none of the previous works explicitly mentions SoH as a result, the obtained impedance measurements can be directly used as health indicators and, thus, to estimate the state of health.



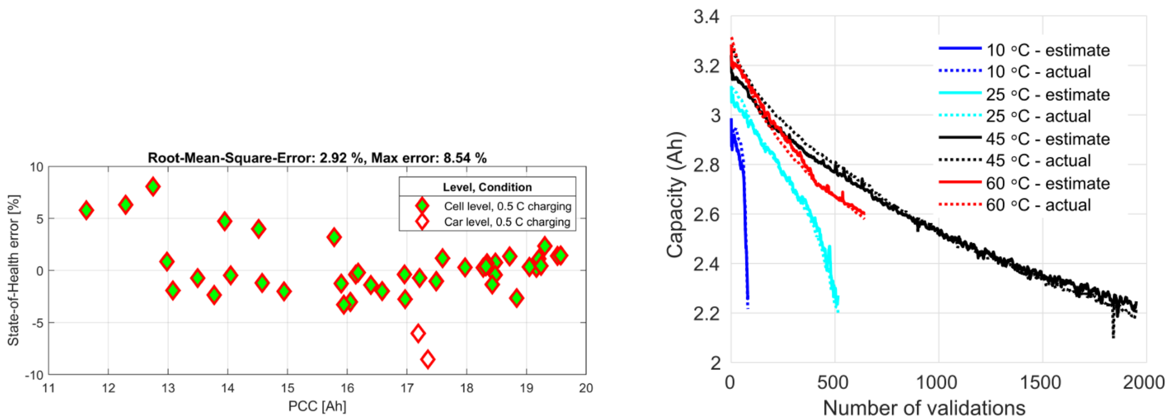
(a) Comparison between EIS and PRBS in [9]

(b) Comparison between ML and polynomial interpolation methods in [11]

Figure 1.2: Impedance spectrum measurement results from [9] and [11], respectively

1.2.3 Capacity-based SoH estimation

A state-of-the-art review of SoH estimation methods can be found in [12]. It considers both experimental and adaptive approaches, as well as degradation-based ones, such as incremental capacity (IC) and differential voltage (DV) curves analysis. A review of ML-based techniques was also presented in [13]. In [14] calendar ageing was studied by storing several Li-ion cells in different temperature and SoC conditions and conducting periodic EIS tests to evaluate degradation. It concluded that cells aged faster when stored at high temperatures and at intermediate charge levels. The same team proposed in [15] to estimate the SoH based on the variation of the charge and discharge curves for cells exposed to several cycling and calendar ageing conditions. They also conducted a detailed analysis in order to chose an appropriate charging interval as a trade-off between accuracy and charging time. Incremental capacity analysis was used in [16], [17] for SoH estimation. Based on experimental results of cells of different chemistries and ageing states, it proposed to focus the analysis on a reduced set of peaks and valleys, whose specific location was correlated with the battery's degradation. The method, whose performance is shown in Figure 1.3a as a function of partial charging capacity, presents good accuracy, although the error at car level rises above 5 %. In [18], an ECM was developed first based on EIS experiments, taking into account the effect of SoC but disregarding ageing and temperature. Then, a dual recurrent neural network (RNN) algorithm was designed in order to generate an SoH indicator based on both capacity and internal resistance degradation from voltage and temperature measurements, and SoC variations. The method, however, was only verified with the previously developed ECM and not with actual experimental data. [19] used a long short-term memory RNN (LSTM-RNN) to estimate SoH using voltage and current sequences during charging. The algorithm was trained and tested on cells which were aged using driving cycle-inspired data sets at different temperatures, as shown in Figure 1.3b. It can be observed that the algorithm can provide accurate estimations over a wide range of operating conditions. Lastly, [20] applied a Gaussian process regression model to the charging curves of a Li-ion battery. This statistical approach tries to estimate the SoH based on four features of the charging characteristic: constant-current (CC) and constant-voltage (CV) stages' duration, slope at the end of CC stage, and slope at the initial part of CC stage. Gray relational analysis is used to evaluate the relevance of each of these features for changes in SoH. High-accuracy results are obtained using a public battery-ageing data set, although temperature is not taken into consideration in this algorithm. In [21], deep neural networks were used to predict a cell's SoH and remaining useful life based on voltage, current and temperature measurements. It showed improved results over simpler structures, but only for SoH estimation. Researchers in [22] proposed a flexible method applicable during both charge and discharge stages. A novel health indicator was used by measuring the voltage increment over a fixed time interval. Then, SoH was estimated by means of an extreme-learning machine. One of the main advantages of this technique is that it is not based on fixed voltage or charge measuring points. Good accuracy was shown, with an average error of 0.5 %. A time series-based methodology was proposed in [23], where authors used LSTM-RNN to estimate the maximum available capacity of cells and battery packs. Transfer learning was also employed to fine-tune the parameters of the network. The charge duration between certain voltage levels was used as health indicator. Good results were obtained by the authors even when a limited amount of early stages were used for training. [24] proposed to combine several machine-learning algorithms to provide more accurate estimations, which were based on the energy increment between two specific points during charging. The algorithm showed performance levels on par with others based on a single neural network structure.



(a) SoH estimation error at cell and car level, from [16]

(b) Algorithm performance at different temperatures, from [19]

Figure 1.3: Results of proposed approaches in [16] and [19], respectively

Finally, it is worth mentioning that, motivated by the highly-complex, time-consuming battery testing processes, several institutions around the world have developed data sets based on comprehensive testing, which have then been published online for anyone to use. A detailed analysis of these sources can be found in appendix A. As a last remark, the works discussed throughout this section were carried out with cells of different chemistries and at different testing conditions, so it should be kept in mind that this severely conditions the results and comparisons.

1.3 Methodology

The state-of-the-art review showed that, even though plenty of approaches already exist for estimating an EV battery's SoH, there is still room for improvement in terms of accuracy, online execution or changing conditions. Furthermore, not a lot of attempts have yet been made at using machine learning for this purpose, although the good results of the aforementioned works imply that there is great potential for applying these techniques to BMS.
This thesis will therefore focus on facing some of these issues by, firstly, designing ML-based algorithms that can provide online SoH estimations and, secondly, testing them in diverse conditions which could occur in real-world applications. A key distinguishing feature of the approach presented in this report is to take as knowledge base the effects of several types of degradation, rather than the specific degradation processes themselves. At the same time, using AI eliminates the need for fitting complex models whose parameters are affected by several different factors and can only provide limited fidelity.

1.4 Thesis statement

Taking into account the information presented in the previous sections, the problem statement for this project is formulated as follows:

How can artificial intelligence techniques be applied to perform online estimation of a battery's state of health in electric vehicles?

1.4.1 Thesis limitations

Some limitations are established here in order to further set the scope of this thesis:

- Only Li-ion cells will be considered
- The devices under test will be individual cells, rather than entire battery packs
- No experimental validation will be conducted due to the Covid-19 pandemic situation, but data from real experiments will be used instead

Chapter 2

Fundamentals of Li-ion batteries

Li-ion batteries are very complex devices whose short- and long-term behaviour is conditioned by many factors. Therefore, prior to trying to develop any kind of model or algorithm, it is essential to understand the impact of these factors and the differences between them. To that end, a brief summary of the inner workings of batteries and the most common materials is given first. Then, several influences are investigated by analysing results of the main testing procedures. Lastly, a thorough study of ageing mechanisms and effects is made.

Contents

Before diving into the fundamentals of electrochemical batteries, some basic concepts must be defined first:

- **Beginning of life (BoL)**: time reference when a pristine cell has not yet been used and its properties remain as when it was manufactured.
- **Chemistry**: general term to denote the chemical species that constitute the various cell components, e.g. electrodes, electrolyte. It usually refers specifically to the cathode's material, since nowadays most batteries have carbon-based anodes and organic-solvent electrolytes.
- **Impedance**: from a linear point of view, ratio of voltage to current at a battery's terminals.
- **Coulombic / Faradaic efficiency**: difference in the amount of charge exchanged between charge and discharge processes. It is usually defined as the ratio of measured discharge capacity to charge capacity.
- **Cut-off voltage**: maximum and minimum voltages between which it is safe to operate the device, as specified by the manufacturer. Exceeding these limits can lead to hazardous situations, significant performance degradation or destruction of the cell.
- **C-rate**: current metric normalized to the cell's rated capacity. A 1 C current is defined as that capable of fully charging (or discharging) the cell in 1 h, for example 1 A for a 1 A h cell.
- **Depth of discharge (DoD)**: capacity range over which the battery is cycled, given as minimum and maximum SoC levels, or as an SoC increment.

- **End of life (EoL)**: time reference when a battery has been extensively used and it is considered unfit for the application. The threshold is usually defined as a 20 % drop in maximum available capacity with respect to the rated value, or a 100 % increase in internal resistance.
- **Maximum available capacity**: maximum capacity that can be charged (or discharged) from a cell at a specific moment of its operational life and under some specific conditions.
- **Rated capacity**: initial capacity of a cell before any degradation occurs, as provided by the manufacturer.
- **Remaining useful life**: estimation of the amount of time that a battery can still be used for in certain conditions before reaching the end-of-life threshold.
- **State of charge**: remaining-capacity metric computed as the ratio of charge level to maximum available capacity at a given moment.
- **State of health**: degradation metric which quantifies the degradation stage of the battery in terms of maximum available capacity (SoH_Q) or internal ohmic resistance (SoH_R).
- **State of power**: amount of power that the battery can exchange at a given moment and under certain conditions.
- **Specific energy / Energy density**: amount of energy that a cell can store per unit mass (specific energy) or volume (energy density).
- **Specific power / Power density**: amount of power that a cell can exchange per unit mass (specific power) or volume (power density).

2.1 Cell types and modelling

Internally, Li-ion batteries are electrochemical energy storage systems, that is, they store energy in chemical form and then transform it into electricity by means of a reduction-oxidation (redox) reaction, and vice versa. The process that takes place within a cell can be described in the general form of Equations (2.1) to (2.3), where rightwards indicates discharge and leftwards indicates charge.

$$\text{Positive electrode}: \quad \text{Metal-Oxide} + \text{Li}^+ + \text{e}^- \xleftrightarrow[\text{Charge (ox.)}]{\text{Discharge (red.)}} \text{Li-Metal-Oxide} \tag{2.1}$$

$$\text{Negative electrode}: \quad \text{LiC}_6 \xleftrightarrow[\text{Charge (red.)}]{\text{Discharge (ox.)}} \text{Li}^+ + \text{C}_6 + \text{e}^- \tag{2.2}$$

$$\text{Net cell reaction}: \quad \text{LiC}_6 + \text{Metal-Oxide} \xleftrightarrow[\text{Charge}]{\text{Discharge}} \text{Li-Metal-Oxide} + \text{C}_6 \tag{2.3}$$

Depending on the specific chemical species (i.e. metal oxide) taking part in the redox reaction, an associated reaction potential $\left(E^0\right)$ is obtained, which determines the cell's rated voltage. Oxidation always takes place at the anode, while reduction does so at the cathode. During discharge, electrons detach from the negative electrode (NE), acting as anode, and travel through the external load all the way to the positive electrode (PE), which acts as cathode, where they contribute to the formation of the metal salt. At the same time, anions/cations

flow through the electrolyte towards the NE/PE to donate/accept the transferred electrons, respectively. The opposite happens during charge, i.e. electrons flow externally into the negative electrode (cathode) as cations drift from PE to NE through the electrolyte for reduction, while anions move from NE to PE for oxidation.
In terms of materials, negative electrodes are most commonly built as a mix of graphite and solid lithium (LiC_6), although lithium-titanium-oxide salts ($Li_4Ti_5O_{12}$, LTO) have also been used. Many more possibilities exist for the positive electrode, being the most frequent ones lithium-cobalt-oxide ($LiCoO_2$, LCO), lithium-iron-phosphate ($LiFePO_4$, LFP) and lithium-manganese-oxide ($LiMn_2O_4$, LMO), as well as ternary salts such as lithium-nickel-cobalt-aluminium-oxide ($LiNi_{1-x-y}Co_xAl_yO_2$, NCA) and lithium-nickel-manganese-cobalt-oxide ($LiNi_{1-x-y}Mn_xCo_yO_2$, NMC). Each of them presents different characteristics and the choice of chemistry largely depends on the requirements for a specific application [1, p. 250], [3, p. 9], [25], [26]. A brief summary of their key properties is presented in Table 2.1.

Table 2.1: Characteristic comparison between Li-ion chemistries. Data from [25], [27], [28]

Chemistry	Specific capacity	Capacity density	Voltage	Advantage	Disadvantage
LCO	$150\,\mathrm{A\,h\,kg^{-1}}$	$550 \times 10^3\,\mathrm{A\,h\,m^{-3}}$	3.8 V	Low self-discharge	Lifespan
LFP	$165\,\mathrm{A\,h\,kg^{-1}}$	$589 \times 10^3\,\mathrm{A\,h\,m^{-3}}$	3.4 V	Lifespan	Specific energy
LMO	$120\,\mathrm{A\,h\,kg^{-1}}$	$596 \times 10^3\,\mathrm{A\,h\,m^{-3}}$	4.1 V	Specific power	Thermal stability
NCA	$190\,\mathrm{A\,h\,kg^{-1}}$	$700 \times 10^3\,\mathrm{A\,h\,m^{-3}}$	3.7 V	Energy and power density	Cost
NMC	$170\,\mathrm{A\,h\,kg^{-1}}$	$600 \times 10^3\,\mathrm{A\,h\,m^{-3}}$	3.7 V	Specific energy	Cost

As discussed in Section 1.2, there are several approaches for modelling the behaviour of Li-ion cells. Among them, ECM are the most commonly used and represent a good trade-off between accuracy and simplicity. Thus, only ECM are presented in this thesis to help understand the behaviour of electrochemical cells, although modelling is not part of this work. Figure 2.1 shows a general schematic of this type of models, where V_{OC} is the open-circuit voltage, R_0 represents the internal ohmic resistance, and $R_1, ..., R_n$, $C_1, ..., C_n$ are the polarization resistances and capacitances, respectively, and I_t and V_t are the terminal current and voltage. In literature, the most common order choices are n=0 (R_{int} model), n=1 (Thévenin model) and n=2 (dual polarization model). The role of R_0 is to quantify the static voltage change when current flows between the electrodes and is related to the resistance of the electrodes, the electrolyte and the connectors, while the R_iC_i networks model the dynamic behaviour of the cell, which may include several different time scales, and correspond to polarization and diffusion processes [29, p. 34], [3, p. 81].

This diagram represents an asymmetric model, that is, the charge and discharge paths present different internal resistance $\left(R_{0,c},\ R_{0,d}\right)$ and, furthermore, may also differ in efficiency.

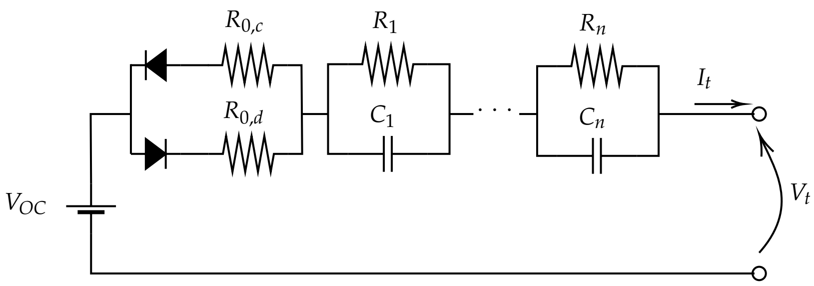


Figure 2.1: General diagram of ECM topology

It should also be noted that the voltage source V_{OC} is actually composed of two terms: an average term and a hysteresis one representing the offset between charge and discharge curves. Based on experimental data, it has been observed in previous research that all elements shown in Figure 2.1 can present some dependency with temperature, SoC and SoH [1, p. 258].

2.2 Standard testing procedures

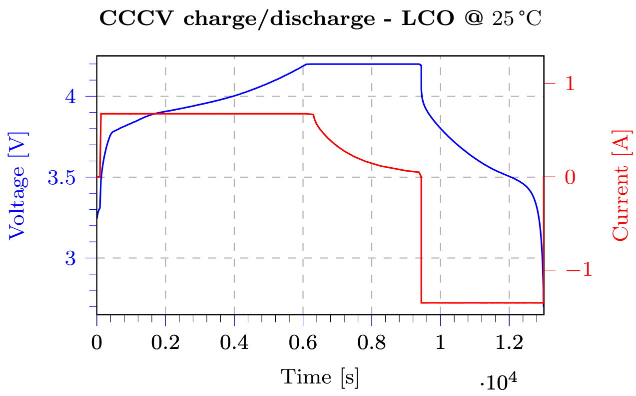
Over the years, several experiments have become industry standards to measure and observe the behaviour of batteries, as well as to gather relevant data to estimate the values of the different models' parameters. These are usually referred to as reference performance tests (RPT) [1, p. 263], [3, p. 33]. The experimental data used throughout this section corresponds to public data sets from *Battery Archive* [30], [31], *Sandia National Laboratory* [32]–[35], *Oxford University* [36], [37], *McMaster University* [38], and private data from *Aalborg University*. Some of these cells will be later on used for testing the proposed estimation algorithms.

2.2.1 Time-domain tests

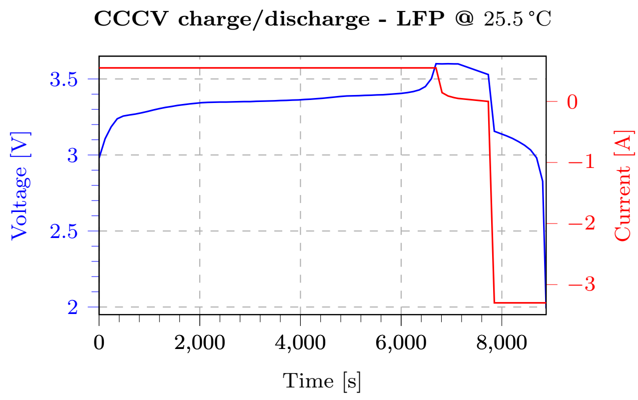
The goal of these tests is to analyse the short- and long-term response of the device by means of time-varying excitation signals. The duration of these tests can range from a few seconds to several days, depending on the phenomenon under study.

Constant-current, constant-voltage (CCCV)

The purpose of this test is to obtain the maximum available capacity of the cell at a given moment and under certain conditions, and it is also the most common charging method. It has two stages and it usually begins with a completely depleted battery. When charging, the first stage consists on injecting the battery with a constant C-rate (e.g. 0.5 C or 1 C) until the upper cut-off voltage is reached; then, in the second stage, the battery is charged at constant voltage as the current progressively decreases. Commonly, the test is considered to end whenever the current has dropped to around 0.05 C. After a resting period, a CC procedure with a negative current applied for discharging until the lower cut-off voltage is reached, followed by a CV stage. The charge levels at the end of the charging and discharge phases are taken as SoC = 100 % and SoC = 0 %, respectively. For more reliable results, the test may be repeated two or three times, and then the average is taken as the true value. Figure 2.2 shows the terminal voltage and current during CCCV for cells of five different chemistries. Note that the last CV discharge stage was not present in the data set and, thus, is not shown in the plots.

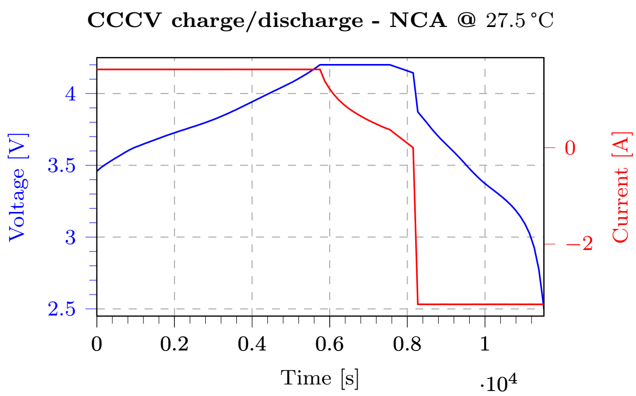


(a) LCO

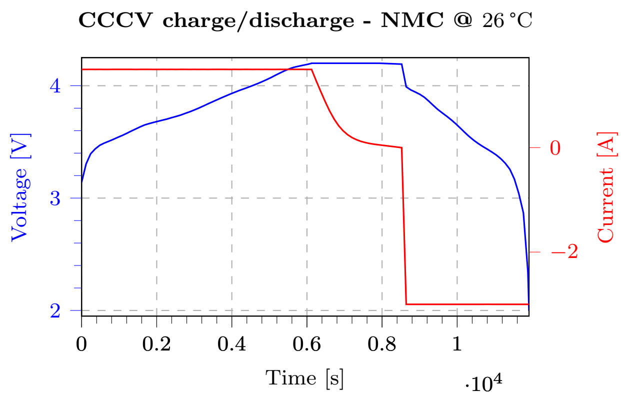


(b) LFP

Figure 2.2: CCCV tests. Data from [30], [31]

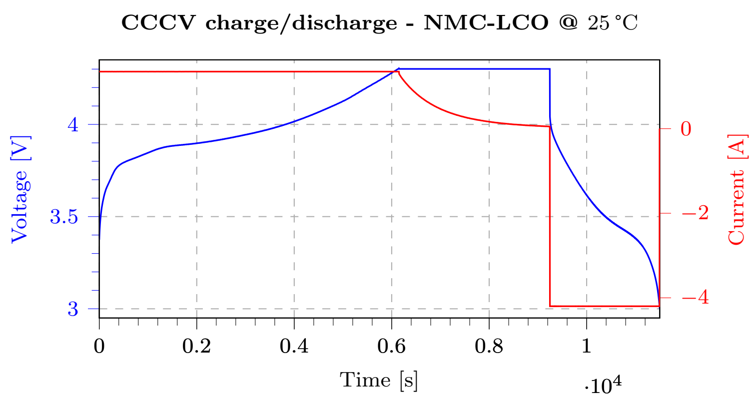


(c) NCA



(d) NMC

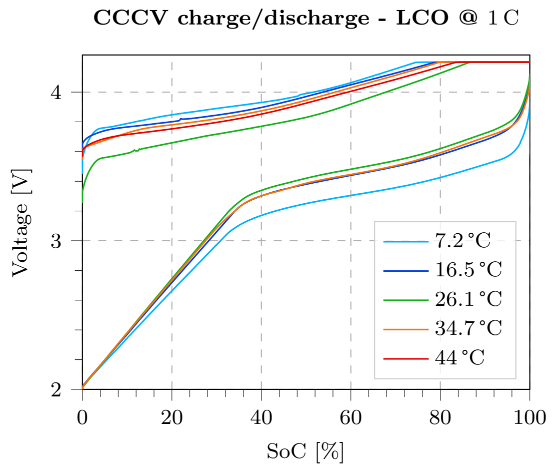
Figure 2.2: CCCV tests. Data from [30], [31] (cont.)



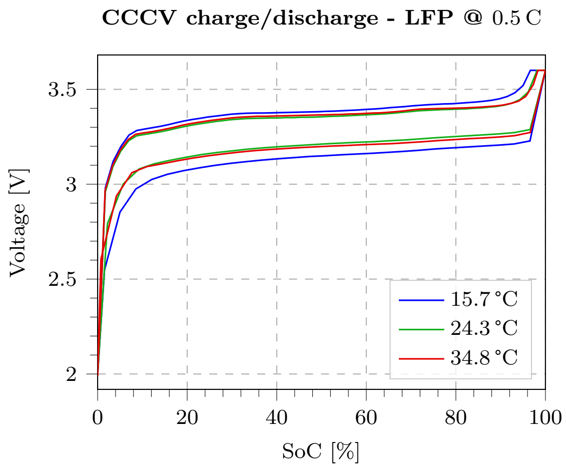
(e) NMC-LCO

Figure 2.2: CCCV tests. Data from [30], [31] (cont.)

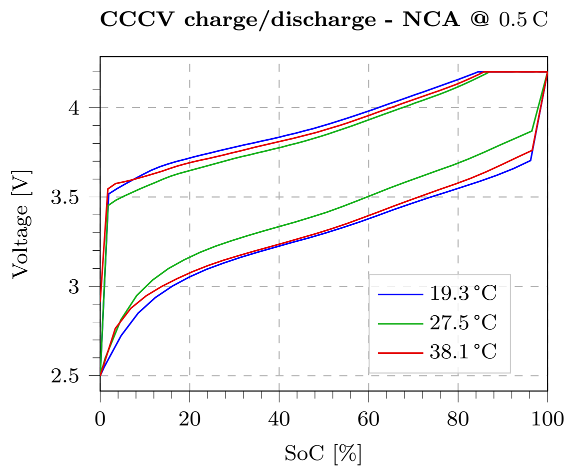
These plots show that the materials employed in a battery's electrodes have a significant impact on its behaviour. This is particularly clear when comparing the voltage during charging of the LFP and LCO cells, for example. The former has a very flat shape and a narrow voltage interval, while the latter is steeper and the voltage increment is larger. Another significant difference is that the LFP cell's CV stage is significantly shorter than the rest. Hysteresis effects and asymmetric efficiencies could also be observed by comparing each pair of consecutive charge and discharge cycles, although these are better studied by means of open-circuit voltage tests. Experimental data shows that the shape of the voltage curves over time depends on cell temperature, so several cells may need to be cycled to analyse the impact of this parameter, as shown in Figure 2.3. For each of the diagrams, the upper-half curves correspond to charge process (increasing SoC) and the lower-half ones to discharge (decreasing SoC). It can be observed that, in general, curves during charge displace upwards (equivalently downwards while discharging) as the temperature drifts away from 25 °C. This displacement is almost negligible in the case of LFP, while it is quite significant for NCA and NMC ones. Discharge data for the LCO unit at 44 °C was corrupt and is not shown here.



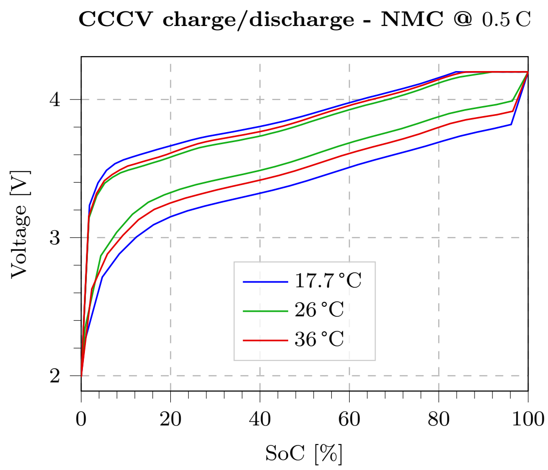
(a) LCO. Data from [32], [33]



(b) LFP. Data from [34], [35]



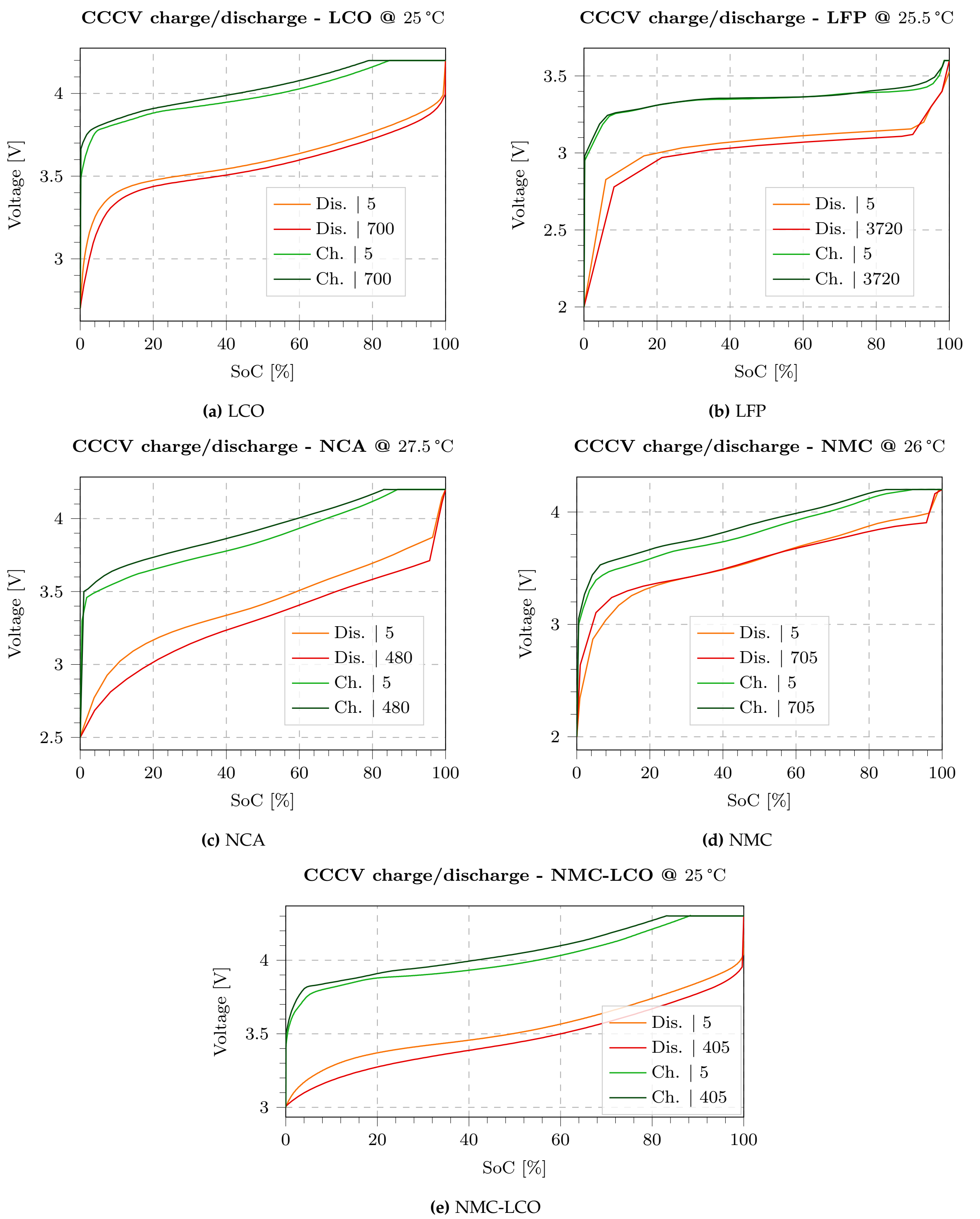
(c) NCA. Data from [34], [35]



(d) NMC. Data from [34], [35]

Figure 2.3: CCCV tests for cells at different temperatures

Lastly, Figure 2.4 shows how these CCCV voltage curves change as the battery undergoes more and more cycles, i.e. as degradation increases. All plots correspond to near-EoL conditions, that is, a maximum available capacity drop of around 20 % with respect to the rated value.



(a) LCO

(b) LFP

(c) NCA

(d) NMC

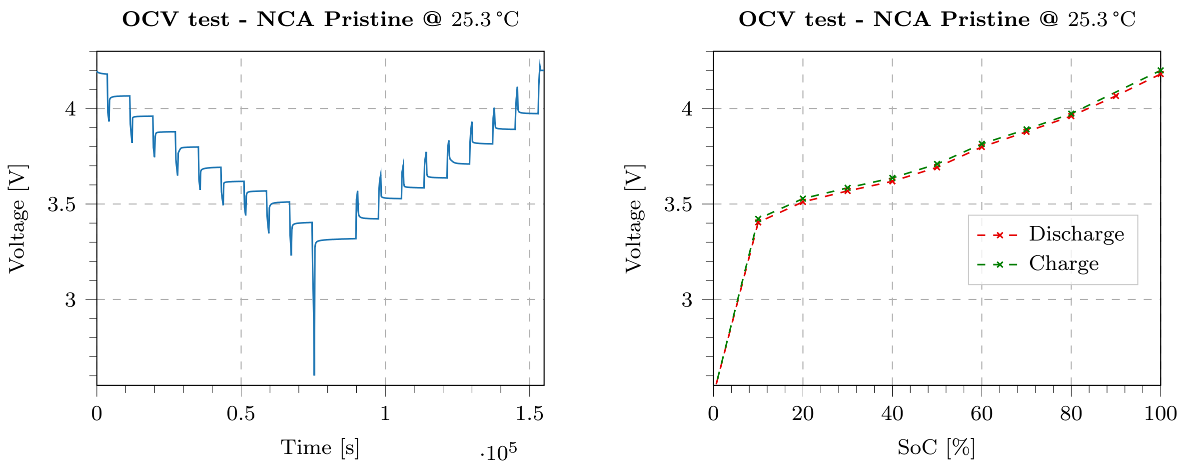
(e) NMC-LCO

Figure 2.4: CCCV tests for cells at different ageing stages. Data from [30], [31]

It is clear that terminal voltage while charging increases with the number of cycles for the same SoC level, and it decreases during discharge. This leads to reaching the cut-off voltages earlier and, thus, lower maximum capacity. Again, cell chemistry plays an important role in the effects of degradation. LFP's curves remain close together even when cells have been severely cycled, whereas the gaps are more significant in NCA units.
The charged (or discharged) capacity in CCCV experiments is usually measured by integrating the measured current (coulomb counting), but any other reliable method could be employed. If the charge capacity is greater than the discharge one, this would indicate that inefficiencies exist within the cell and, thus, an efficiency factor between charge and discharge could be computed and used when modeling, as described in Section 2.1.

Open-circuit voltage (OCV)

This test serves as a means to characterise the OCV-SoC relationship, that is, between open-circuit voltage and state of charge, under certain specific conditions. In this experiment, starting from a resting, fully depleted cell, a CCCV procedure is performed to charge the battery in fixed SoC increments, usually $\Delta\text{SoC} = 5\,\%$ or $10\,\%$. After every milestone is reached, the battery is left resting for a sufficiently long time, which may range from 1 h to 5 h, and then the measured voltage is recorded for the corresponding SoC level. The iterative process continues until the battery is fully charged. Similarly, the discharge phase follows the same approach, but starting with a rested, fully charged battery. Both processes are shown in Figure 2.5 for a pristine, NCA-type cell. As in the previous case, the charge and discharge curves are slightly different due to hysteresis, which is almost negligible according to Figure 2.5b. Since this experiment is usually performed prior to deriving an SoC estimation algorithm, the cell's SoC used as milestones are based on measurements, such as coulomb counting.



(a) Test time-series

(b) Charge and discharge comparison

Figure 2.5: OCV test of a pristine NCA cell. Data from [36], [37]

Due to the significantly long time required to conduct a complete OCV test, sometimes a pseudo-OCV test is performed instead. In this case, the cell is charged and discharged at a very low C-rate (around 0.05 C) so that the voltage drop due to resistive elements is low and the terminal voltage approaches the value of the open-circuit voltage. Therefore, this is also referred to as low-current OCV test and it is shown in Figure 2.6 for the same cell as in Figure 2.5, at both pristine and severely aged stages.

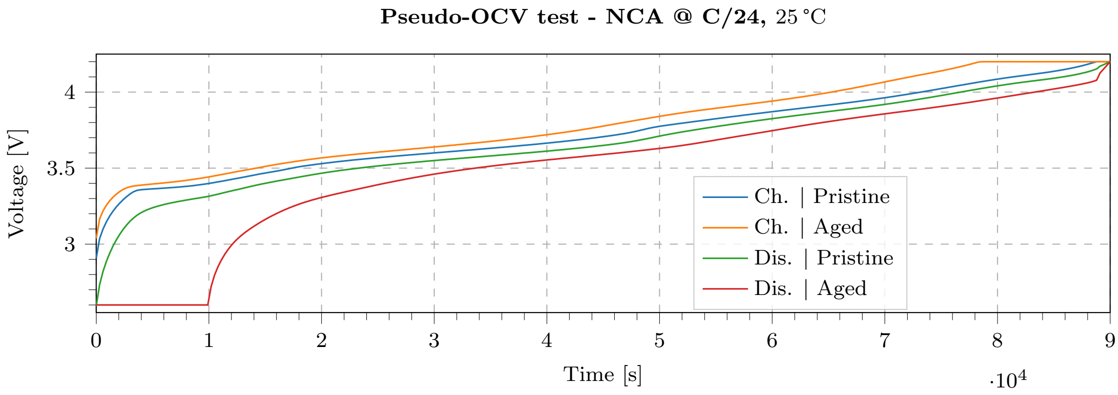


Figure 2.6: Pseudo-OCV test of an NCA cell at pristine and aged stages. Data from [36], [37]

The plot confirms the behaviour observed in Figure 2.4, with the voltage curves drifting away as damage accumulates. Based on these data it is also possible to study the difference between charge and discharge efficiencies, as well as the way they change as the cell degrades. For the specific case of Figure 2.6, these effects are computed in Equations (2.4) to (2.6), taking the maximum available charge capacity as reference. It can be observed that, even when the cell is halfway through its useful life, efficiency is still very close to 100 %.

$$\eta_{ch-dis,pristine} = \frac{Q_{max,dis,pristine}}{Q_{max,ch,pristine}} = \frac{3.0958}{3.0982} = 99.923\,\% \tag{2.4}$$

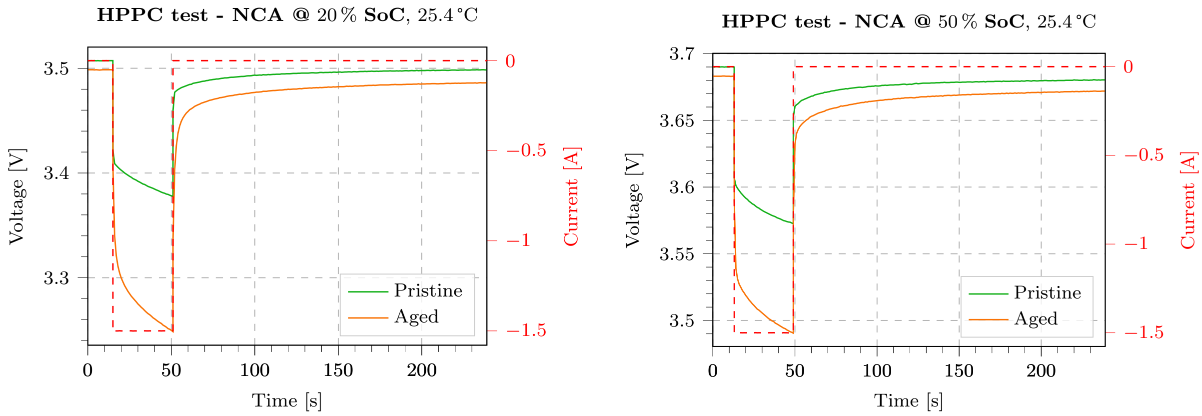
$$\eta_{ch-dis,aged} = \frac{Q_{max,dis,aged}}{Q_{max,ch,aged}} = \frac{2.7585}{2.791} = 98.836\,\% \tag{2.5}$$

$$SoH_Q = \frac{Q_{max,ch,aged}}{Q_{max,ch,pristine}} = \frac{2.791}{3.0982} = 90.085\,\% \tag{2.6}$$

Hybrid-pulse power characterisation (HPPC)

In this case, a sequence of current pulses are injected into, or withdrawn from, the cell. This allows for analysing the dynamics of the system and, thus, fitting the corresponding model parameters. If parameter variations are to be taken into consideration, the procedure would need to be repeated at different temperatures, degradation stages and charge levels. Figure 2.7 shows the HPPC tests for an NCA-type cell at three different SoC levels, both when the cell is pristine and severely aged.

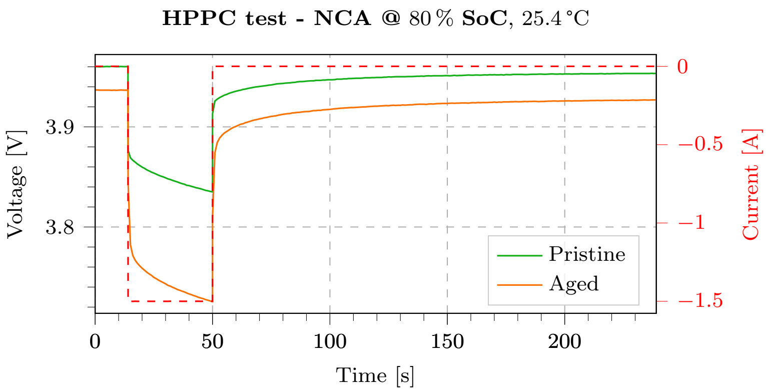
For a characterisation over the entire SoC range, the battery is first fully charged and allowed to rest until voltage equilibrium is reached. Then, the cell is excited with the sequence of current pulses, which are normally symmetric in polarity so that the SoC is not significantly modified during the test and to measure the response for both charge and discharge conditions. There is not a specific rule for the pulse amplitude, and they may range from 0.5 C to 10 C, and each of them may present a different amplitude. Similarly, pulse duration and time between pulses may also be freely chosen, as long as there is enough time for the cell's dynamics to manifest and settle. The battery is then discharged to the next SoC milestone, rested until equilibrium for e.g. 1 h and excited again with another sequence of pulses.



(a) 20 % SoC

(b) 50 % SoC

Figure 2.7: HPPC test at multiple SoC and SoH levels. Data from [36], [37]



(c) 80 % SoC

Figure 2.7: HPPC test at multiple SoC and SoH levels. Data from [36], [37] (cont.)

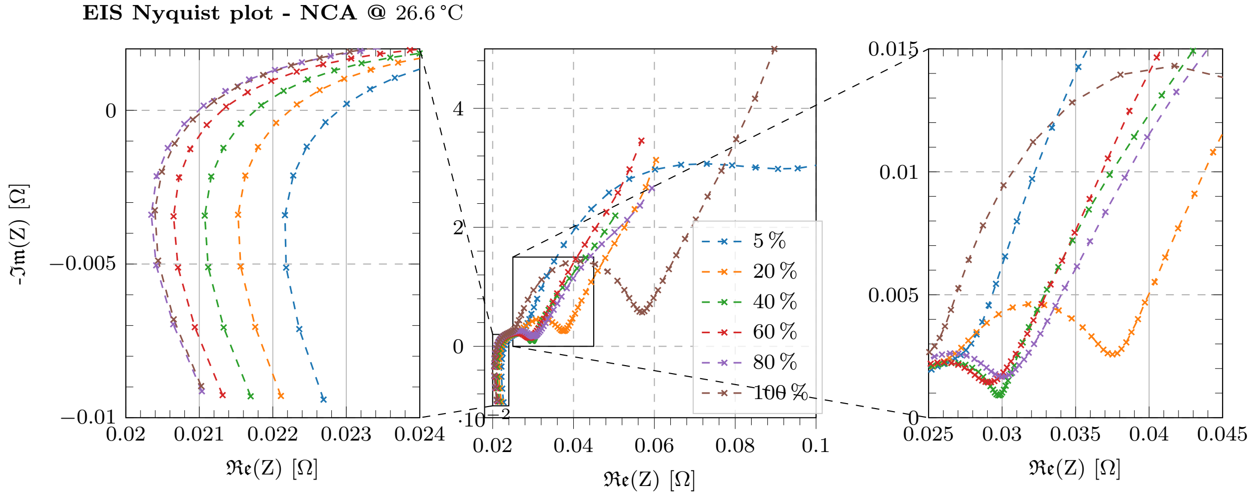
Once the experiment is concluded, terminal voltage and current measurements can be used to estimate ECM parameters, as well as the power capabilities of the cell. Two effects can be observed in the voltage response: a static voltage step, corresponding to the internal ohmic resistance (R_0), and a transient voltage change and recovery, related to the polarization networks ($R_i C_i$) in the ECM [29, p. 51], [3, pp. 38, 93]. It is clear how the voltage drop becomes larger as the cell undergoes more and more cycles, and also the transient seems to be slower. In terms of ECM parameters, this would correspond to increased R_0 value, and larger time constants in the RC networks.

2.2.2 Frequency-domain tests

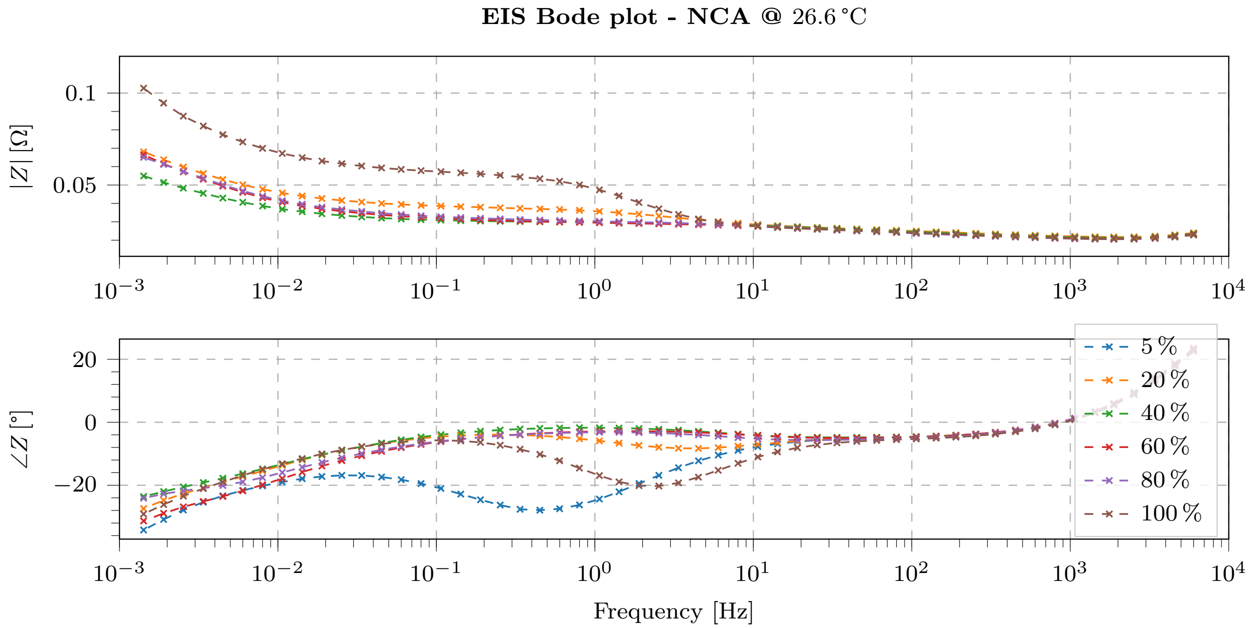
In order to further observe and analyse the frequency-dependent behaviour of the cell, experiments at several excitation frequencies need to be conducted [1, p. 267], [3, p. 42].

Electrochemical impedance spectroscopy (EIS)

The premise of this procedure is that, under the assumption that the cell can be regarded as a linear system at each operation point, a cell impedance can be defined as the ratio of terminal voltage to current. The linearisation point variables include temperature, SoH and SoC, which are the main factors affecting this parameter. The excitation signal can be either voltage (potentiostatic mode) or current (galvanostatic mode), and it consists on a sinusoidal signal of a certain frequency which may have a constant offset. The injected amplitude must be high enough to produce a measurable response, but low enough to keep the operating point unchanged. When conducting this experiment with laboratory equipment only one frequency is excited at a time, although multi-frequency approaches also exist for online applications, as discussed in Section 1.2. The frequency range of the excitation signals usually goes from a few mHz to several kHz. An example of the outcome of an EIS test is presented in Figure 2.8 for an NCA cell at several SoC levels. Another application of EIS experiments is to help computing the parameters of the ECM, as it has been shown in previous research [14], [1, p. 267], and also for the FOM [3]. The Nyquist plot, which represents the real and imaginary parts of the impedance's complex conjugate value, is shown in Figure 2.8a with a zoom in two areas of interest to help the analysis. Frequency-wise, the points in the lower-left corner correspond to high-frequency inputs, while the ones in the top-right zone proceed from low-frequency excitation. This is more clearly seen in the Bode plot in Figure 2.8b. In the leftmost plot of Figure 2.8a an evident tendency in the curves is observed, which is that the real part decreases for higher values of SoC. On the other hand, the rightmost plot shows how the differences between SoC levels are much more significant. In this segment it can be observed how the valley moves to the right (higher real part) as the SoC drifts away from 50 %, although the pattern is not symmetric.
Similar observations can be made from Figure 2.8b. At low frequency (top-right zone in the Nyquist plot) both the impedance's magnitude and phase reach their maximum absolute values, whereas at high frequency (bottom-left area in the Nyquist plot) the magnitude reaches its lowest value while the phase increases with the opposite sign. In the mid-frequency range the impedance approaches the 0° line, although never crossing it back. It is also clear that SoC has a much more sensible influence on impedance at low- to mid-frequencies, and becomes negligible beyond 10 Hz.



(a) Impedance Nyquist plot



(b) Impedance Bode plot

Figure 2.8: EIS for NCA cell at various SoC levels. Data from [38]

Figure 2.9 portrays the EIS data for an NMC cell which has only undergone storage ageing (see C1 in Table 6.1). Again, the Nyquist plot shows that differences at both low and high frequencies appear as the cell ages. In the former, it is clear that the real part of the impedance grows as damage accumulates. Meanwhile, the valley at middle frequencies seems to displace rightwards as time passes, although the tendency here is not as clear as in the previous case. In the Bode plot in Figure 2.9b it can be observed how the magnitude of the impedance increases with ageing, specially at high frequencies, which is coherent with the previous analysis.

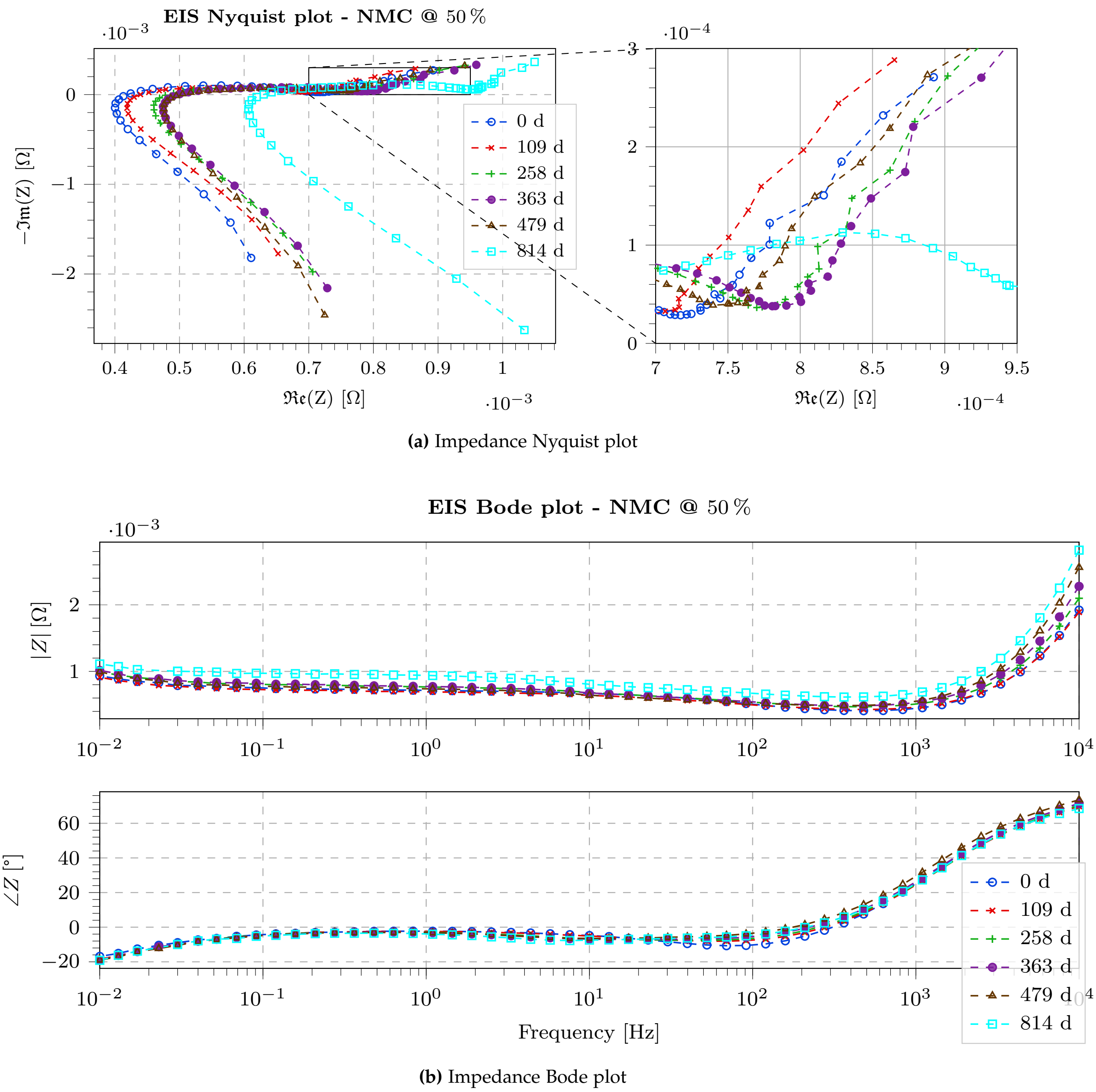


Figure 2.9: EIS for NMC cell at various ageing stages. Own data (see Chapter 6)

Lastly, Figure 2.10 shows the impact of temperature on the impedance of an NMC-type cell. It can be observed how, as the temperature decreases, the real part of the impedance increases, which leads to higher losses when charging or discharging the battery. As a final remark, it is interesting to see that chemistry also determines the shape of the impedance spectrum, in this case between NCA and NMC cells.

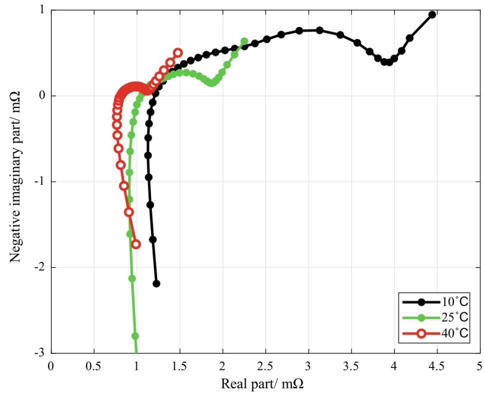
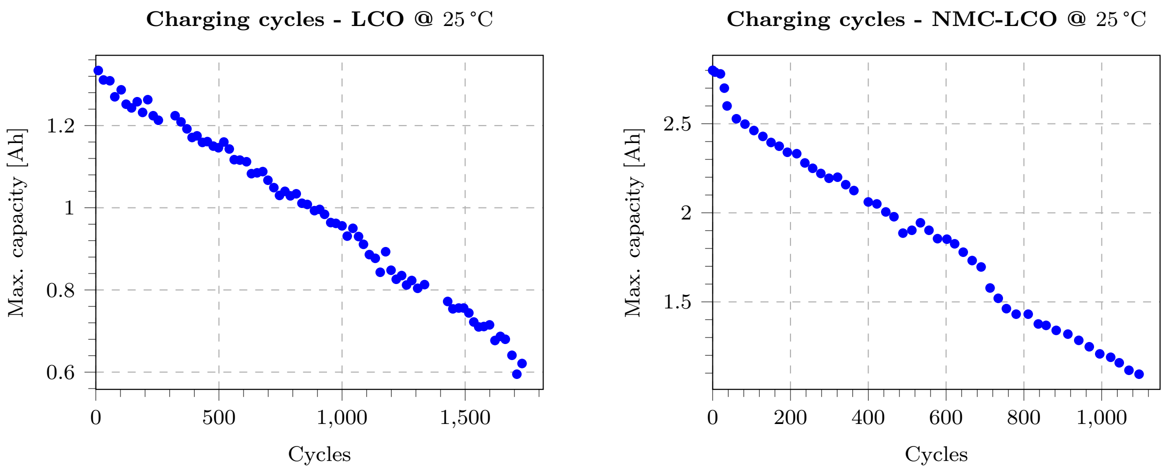


Figure 2.10: Influence of temperature on NMC-type cell impedance. Replicated from [3, p. 52]

2.3 Ageing of Li-ion cells

As electrochemical cells are used, they accumulate damage and their internal structure changes due to e.g. solid-electrolyte inter-phase (SEI) film formation, material adsorption at the electrodes or loss of active material [12]. Among the factors which can influence the degradation process of a Li-ion battery the most relevant are C-rate, temperature, DoD and storage conditions. These are analysed in the following. Although these data may not quite represent realistic usage, they allow for studying each of the aforementioned factors independently. It should also be considered that each set of data corresponds to a unique individual cell, so it would be expected to have some slight behaviour differences due to manufacturing and not to the processes themselves. There is one general comment which is applicable to any of the ageing diagrams shown next, which is that usually two different stages can be distinguished: in the first one the capacity decreases very fast in just a few cycles, while in the second one the reduction becomes nearly linear and the slope is smaller. The former can be linked to the consumption of active material during the formation of the SEI film, and the latter corresponds to the scenario after this layer is completely formed [3, p. 55].



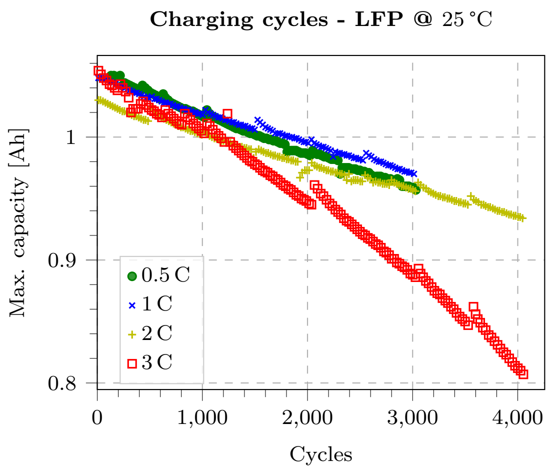
(a) LCO **(b)** NMC-LCO

Figure 2.11: Maximum capacity degradation. Data from [30], [31]

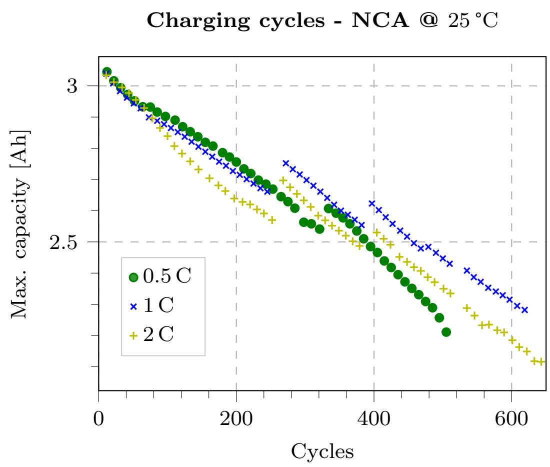
Since only constant-temperature and -current tests were available for LCO and NMC-LCO units, their ageing characteristics are shown in Figure 2.11. These two types of cell present a quite linear degradation with the number of cycles, thus the first stage lasts for only a few cycles.

2.3.1 C-rate's impact on cell degradation

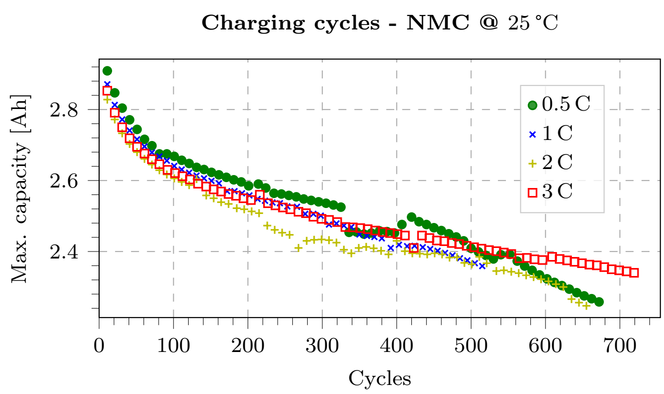
The current withdrawn from an EV battery is directly linked to the power demand of the vehicle, which will be higher whenever the motor needs to produce more torque for accelerating or climbing a slope [39]. On the other hand, the current injected into the battery comes from charging most times and, therefore, it depends on the capabilities of the charging station. Figure 2.12 shows the evolution of the maximum available capacity of three different types of Li-ion cells for several discharge C-rates. The charge C-rate is 0.5 C for all of them, since it was the only one available in the data set. The plots show that this parameter has a clear effect on capacity degradation and that some differences exist between chemistries, being the most significant the large amount of cycles that the LFP cell can withstand compared to the other two. This is in agreement with the analysis of Figure 2.4.



(a) LFP



(b) NCA



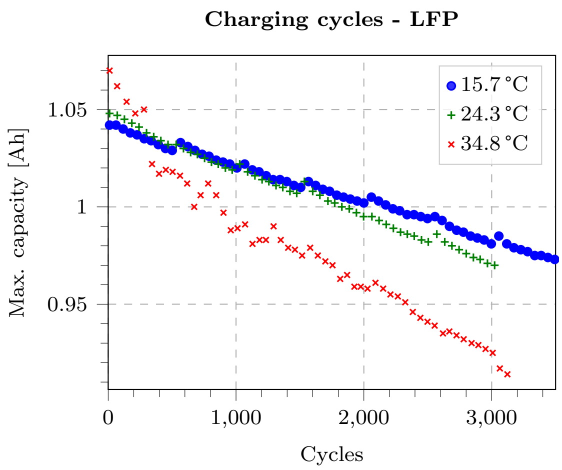
(c) NMC

Figure 2.12: Maximum capacity degradation for different discharge C-rate. Data from [30], [31]

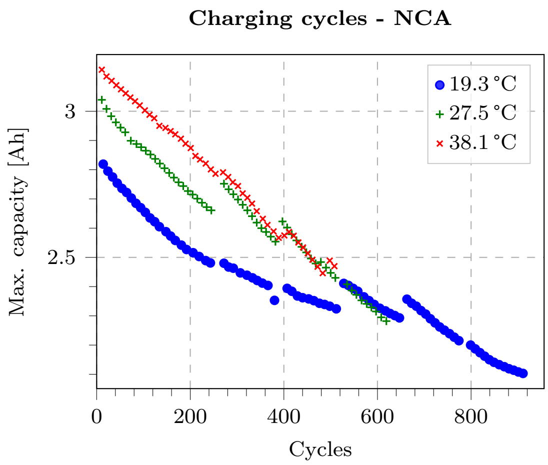
For the LFP unit in Figure 2.12a, deterioration is lowest at 0.5 C and 1 C, while degradation at 2 C seems worse because of lower initial capacity. In terms of slope all three scenarios are very similar. The main difference occurs for 3 C, whose slope is much steeper than the rest and, thus, the cell loses capacity much faster. Identical observations can be made for the NCA unit in Figure 2.12b, although data for 3 C was not available. Lastly, Figure 2.12c shows that the NMC device is the most immune to the discharge current's magnitude and all experiments have alike degradation slopes. It is also in this case in which the two stages are best distinguished, the transition happening at around 100 cycles.

2.3.2 Temperature's impact on cell degradation

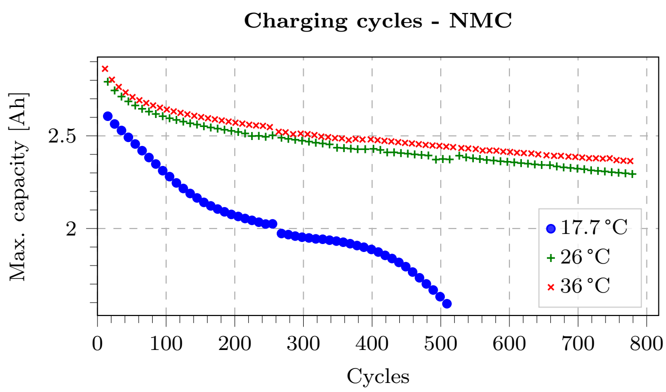
Electrochemical reactions taking place within the Li-ion cells are affected by the ambient and internal temperatures, so it would be expected for this parameter to have an influence on ageing. Figure 2.13 depicts the available charge capacity degradation as a function of the number of cycles for three different temperatures, which are kept nearly constant throughout both charge and discharge processes. Specific information regarding the cells and the cycling conditions is presented in Table 4.4.



(a) LFP



(b) NCA



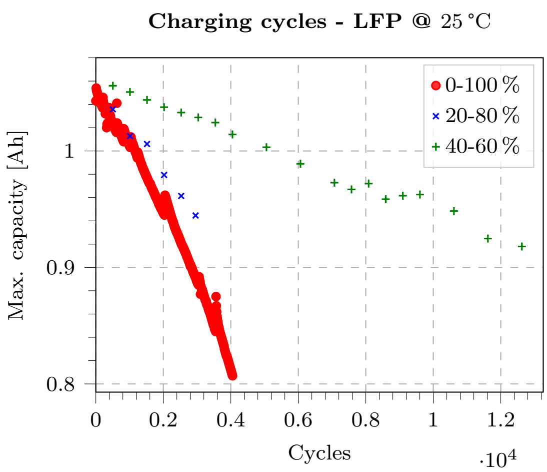
(c) NMC

Figure 2.13: Maximum capacity degradation for different temperatures. Data from [30], [31]

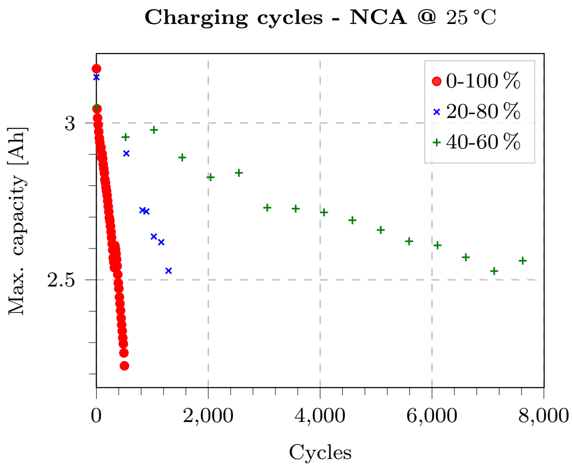
It must be pointed out that the temperatures presented here constitute a quite narrow range and, therefore, similar experiments with a wider interval (e.g. from −20 °C to 50 °C) should be conducted and analysed to get a better grasp of the influence of temperature in active operation. It can be observed in Figure 2.13a that the LFP unit's capacity decrease is faster the higher the cycling temperature is, with a very significant difference between 15.7 °C and 34.8 °C, while the initial charge capacity is slightly higher at 34.8 °C. For the NCA cell (see Figure 2.13b), degradation is also worst at the highest temperature. Nevertheless, initial maximum capacity is significantly higher at 38.1 °C, which indicates that the battery cannot hold as much charge at low temperatures as it can when it is warmer. According to Figure 2.13c, degradation and initial maximum capacity in the NMC-type device are worse when the temperature decreases. Although similar performance is obtained at 26 °C and 36 °C, the capacity loss at 17.7 °C is much higher and it even shows an entirely different pattern.

2.3.3 DoD's impact on cell degradation

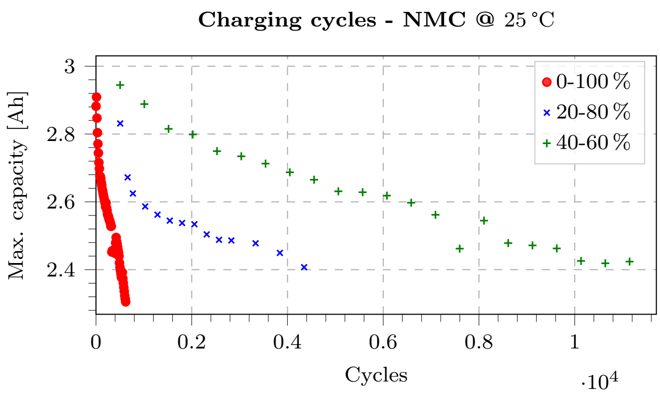
The amount of charge used and refilled when driving an EV also has a severe influence on how capacity decreases throughout the cell's lifetime, as it would be expected for deterioration to become worse as the DoD approaches 0-100 %. Figure 2.14 shows the capacity degradation for cells following CCCV cycling with three different DoD intervals.



(a) LFP



(b) NCA



(c) NMC

Figure 2.14: Maximum capacity degradation for different DoD. Data from [30], [31]

The plots confirm the hypothesis that capacity loss occurs faster whenever the battery is used in wider SoC ranges. Note that much fewer data points were recorded for 40-60 % and 20-80 % since, in order to find the maximum available capacity, a full charge and discharge cycle must be completed, which modifies the DoD-based ageing procedure. Figure 2.14a shows that, for the LFP cell, degradation with full cycles is just slightly worse than with 20-80 % DoD, but it is much more accentuated than with 40-60 %. For example, approximately the same capacity drop happens at 1000 cycles with full-range CCCV and at 5000 cycles for the narrowest DoD procedure. For the NCA cell, according to Figure 2.14b, more differences exist between the two widest intervals, following the same tendency as before. In this case, equivalent degradation happens at 320 cycles for 0-100 %, 1300 cycles for 20-80 % and 6500 cycles for 40-60 % DoD. From Figure 2.14c it can be observed that the degradation pattern in the NMC unit resembles that of the NCA one. A similar capacity decrease occurs at around 400, 2300 and 8600 cycles for the widest to the narrowest cycling intervals, respectively. Furthermore, the slope difference between early and late stages becomes smoother when the battery is operated in small intervals.

2.3.4 Storage conditions' influence on capacity loss

Since EV batteries may spend a significant amount of time in an idle state, it is worth analysing whether storage conditions affect the way the device degrades. For this purpose, data from NMC-type cells stored in different conditions (see Table 6.1) were recorded and are presented in Figure 2.15.

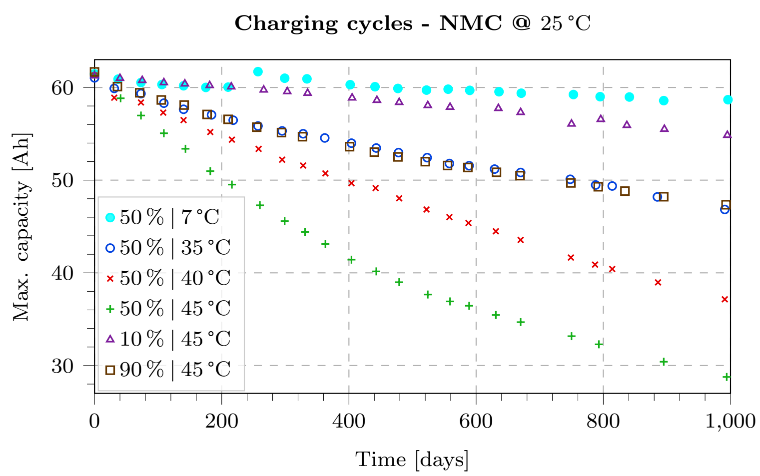


Figure 2.15: Maximum capacity degradation for different storage ageing conditions. Own data (see Chapter 6)

The main observation is that, for the same SoC level, degradation becomes worse as the ambient temperature is increased. On the other hand, for constant temperature, extreme SoC levels seem to be more beneficial than intermediate ones and, among the extremes, low ones preserve the cell's health better than high ones. The capacity drop between the best (C6, cyan dots) and worst (C3, green crosses) ageing procedures is quite significant, retaining 58.7 A h and 28.8 A h after 1000 days, respectively. In terms of capacity loss, these correspond to 93.2 % and 45.7 % of the initial capacity.

2.3.5 Operation mode's impact on cell degradation

The final ageing factor accounted for in this thesis is hybrid ageing, that is, the combination of cycling and storage periods, which aims at better representing the real usage of an EV battery. Data from four groups undergoing different procedures are recorded. The specific details are presented in Table 4.6 and the results are depicted in Figure 2.16 for one cell in each of the groups. It is clear from the plots that the specific type of hybrid ageing does not seem to have a huge impact on capacity degradation, as all four groups present a similar patterns. The capacity decay does not show a clear border between early and late stages, which was also observed for the NCA-type cell in the previous analysis.

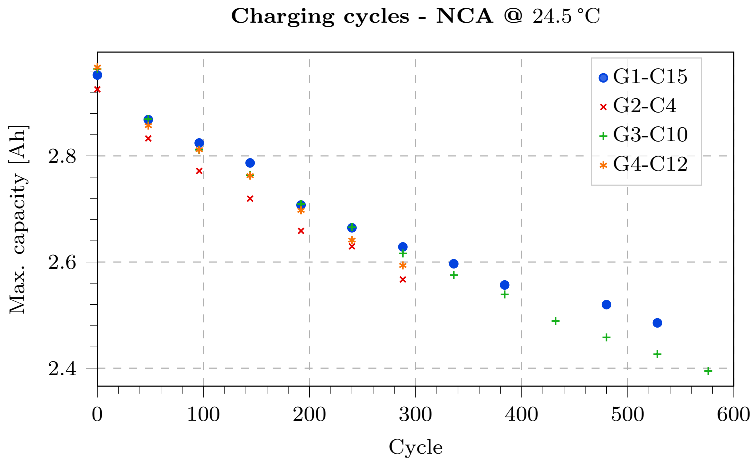


Figure 2.16: Maximum capacity degradation for different hybrid ageing procedures. Data from [36], [37]

After analysing the main factors affecting the ageing of Li-ion cells some conclusions can be drawn. In general, in order to maximize the lifetime of these devices, they should be operated at relatively low C-rate (see Figure 2.12) and narrow DoD intervals (see figure 2.14). However, when it comes to temperature, LFP and NCA cells degrade less at low temperature (at least up to 15 °C), while NMC cells keep higher capacity when cycled at high temperature (at least up to 36 °C), as it is shown in Figure 2.13. Analysis of calendar ageing data proved that Li-ion cells, or at least NMC-type ones, can lose a large amount of available capacity depending on how they are stored. Among the several options considered, the best one was to keep the battery at half charge (50 % SoC) and very low temperature (7 °C). Finally, it was observed that, when it comes to hybrid ageing, differences between procedures are not significant. A direct comparison between purely-cycling and hybrid ageing could not be made since data belonged to different data sets, although a brief comparison between Figures 2.13b and 2.16 points towards the hypothesis that full-range and hybrid cycling do not lead to significant differences in the pattern of capacity decrease. Once the fundamental background theory of Li-ion ESS has been set forth, the next chapter will do the same for the topic of neural networks.

Chapter 3

Fundamentals of Artificial Neural Networks

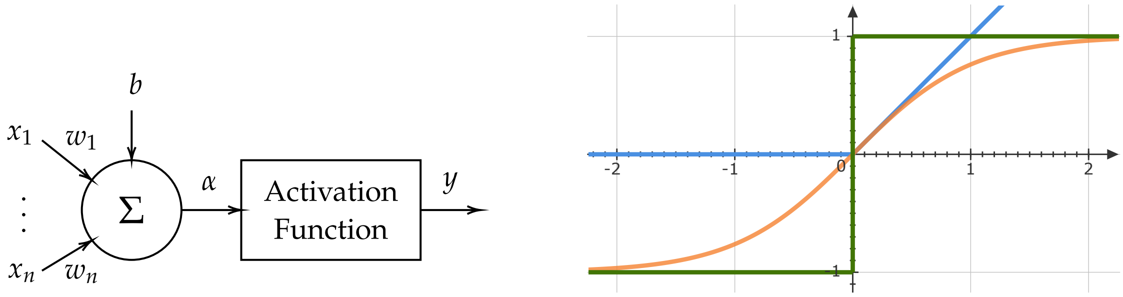
This chapter provides a theoretical background to the topic of artificial neural networks, with a focus on the two specific structures implemented later on in this thesis, namely fully-connected feed-forward neural networks and broad learning systems.

Contents

Artificial intelligence is the knowledge field related to the design and development of computational systems which, rather than just being programmed to perform a fixed, specific task, have the ability to learn a certain ground truth, understand a context and adapt their operation to achieve a certain goal. Although an explosive expansion of AI has taken place over the last two decades, its principles date back to the first half of the 20^{th} century [40]. Some of the most important areas within this field include fuzzy logic [41], machine learning [42] and deep learning [43]. Despite being based on similar general concepts, each of them presents unique characteristics which make them better fit for different applications, such as autonomous driving, system modeling and control, computer vision or natural-language processing [44, p. 46], [43, p. 443]. A key aspect of these algorithms is that, just like human knowledge, are based on heuristics, that is, they try to learn from previous experiences and then apply the attained knowledge to new scenarios.

3.1 Feed-forward neural networks

Artificial neural networks (ANN) are computational models inspired by biological brains. As such, their fundamental building block is the artificial neuron, which tries to mimic the behaviour of the individual neurons in the nervous system. A neural network is then formed by stacking several neurons at the same depth, forming a *layer*, and several layers consecutively within the network, forming a sequence of *hidden layers*. A key milestone in the history of ANN was the formal proof that these are universal function approximators, meaning that any arbitrarily complex function can be accurately approximated by an ANN [45]. Figure 3.1a shows the functional diagram of a single neuron, where $\boldsymbol{x} = (x_1, ..., x_n)$ are the inputs to the neuron, $\boldsymbol{w} = (w_1, ..., w_n)$ are the input weights, b is the bias term and y is the neuron's output. The mathematical expression describing an individual neuron is given by Equation (3.1), where $f(\alpha)$ is the *activation function*. This simple expression shows that ANN fundamentally perform linear algebra operations, which makes them very efficient when implemented in hardware.



(a) Artificial neuron

(b) Non-linear activation functions

Figure 3.1: Diagram of a basic neuron

$$y = f(\alpha) = f\left(\sum_{i=1}^{n} w_i\, x_i + b\right) = f\left(\boldsymbol{w}\boldsymbol{x}^T + b\right) \tag{3.1}$$

The design and choice of the activation function is one of the most important aspects when designing AI systems, since it will determine the complexity of the operations that the algorithm can accomplish. Systems without activation function or with threshold-like ones are only able to produce linear discrimination functions at the output and, thus, they cannot carry out non-linear tasks. On the other hand, using non-linear activation functions extends the capabilities of the network to more complex problems. The most commonly used ones are sigmoid-like (such as the hyperbolic tangent), although alternative types have gained a lot of prominence over the years (e.g. the rectified linear unit function, shortened ReLU) [44, p. 11]. For reasons that will become evident later, a major requirement for designing activation functions is that their derivatives are simple and easy to compute. These derivatives should also have values nor too large nor too close to 0 in order to avoid the exploding- and vanishing-gradient problems, respectively [44, p. 129]. Graphical representations of *sign* (green), *tanh* (orange) and *ReLU* (blue) activation functions are shown in Figure 3.1b, and their mathematical expressions and derivatives are given by Equations (3.2) and (3.3).

$$f_{sign}(\alpha) = \begin{cases} 1, & \text{if } \alpha > 0 \\ -1, & \text{otherwise} \end{cases} \quad f_{tanh}(\alpha) = \frac{e^{2\alpha} - 1}{e^{2\alpha} + 1} \quad f_{ReLU}(\alpha) = \begin{cases} \alpha, & \text{if } \alpha > 0 \\ 0, & \text{otherwise} \end{cases} \tag{3.2}$$

$$f'_{sign}(\alpha) = 0 \quad f'_{tanh}(\alpha) = 1 - \left(\frac{e^{2\alpha} - 1}{e^{2\alpha} + 1}\right)^2 \quad f'_{ReLU}(\alpha) = \begin{cases} 1, & \text{if } \alpha > 0 \\ 0, & \text{otherwise} \end{cases} \tag{3.3}$$

However, the introduction of non-linear functions is not enough to solve non-linear problems. Neural networks consisting only on a single layer of neurons can only solve problems which are linearly separable, such as logic AND and OR functions. In order to enhance the capabilities of the network so that it can solve non-linear discriminant functions, such as logic XOR, multiple layers with non-linear activation functions need to be stacked together [44, p. 32]. The number of hidden layers, the amount of neurons in each layer and the activation function will determine the level of complexity of the tasks that the ANN has to learn, also referred to as *network's capacity*.

Figure 3.2 shows a generic diagram of a fully-connected feed-forward neural network (FC-FNN) with p inputs, i hidden layers, m and s nodes in the first and last hidden layers,

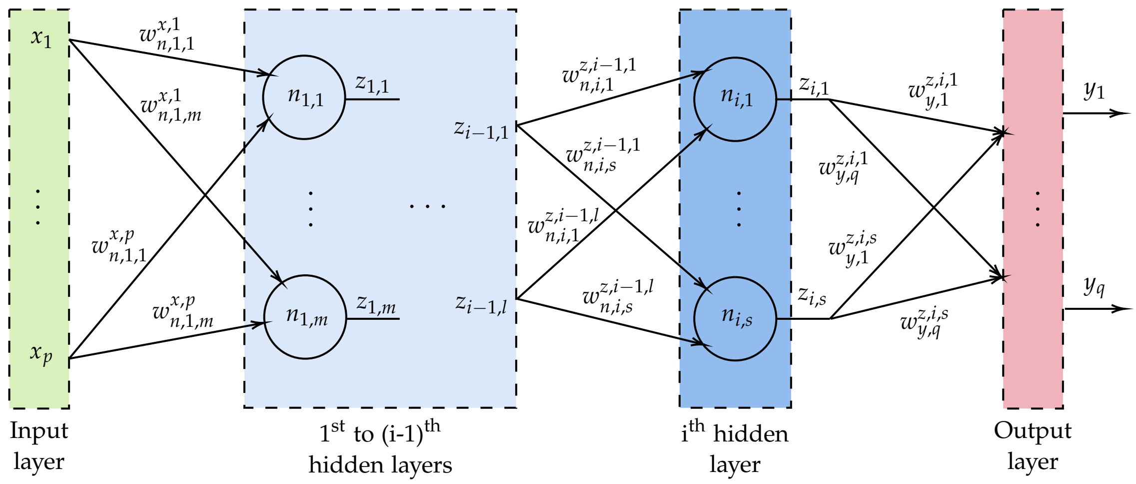


Figure 3.2: Diagram of a fully-connected feed-forward neural network

respectively, and q outputs. A description of the symbols is presented in Table 3.1. For the remainder of this report, the nomenclature employed to describe an FC-FNN structure will be $FNN\,(n_1 - ... - n_i,\ AF)$, where n_1 is the number of neurons in the first hidden layer, n_i is the amount of neurons in the last one and AF denotes the activation function employed in all neurons. Note that each circle represents a complete neuron as depicted in Figure 3.1a, that is, including bias, summation and activation function. Each hidden layer may have a different number of neurons, and different activation functions may be set for each neuron. In an FC-FNN, each neuron within a layer takes as inputs all the outputs from the preceding layer, multiplied by the corresponding weights and added to the bias. A quite important advantage of ANN can be inferred from Figure 3.2: at each layer many calculations are performed independently and, therefore, could be executed in parallel to speed up the process. This is the reason why they are normally deployed in concurrent-computing hardware, such as graphics processing units (GPU) or field-programmable gate array (FPGA) boards [44, p. 156].

Table 3.1: Description of variables of FC-FNN in Figure 3.2

Symbol	Description
$\boldsymbol{x} = (x_1, ..., x_p)$	Input vector
$\boldsymbol{y} = (y_1, ..., y_q)$	Output vector
$w^{x,p}_{n,1,m}$	Weight from input x_p to m^{th} neuron in 1st hidden layer
$w^{z,i-1,l}_{n,i,s}$	Weight from l^{th} neuron in (i-1)th hidden layer to s^{th} neuron in i^{th} hidden layer
$w^{z,i,s}_{y,q}$	Weight from s^{th} neuron in i^{th} hidden layer to output y_q

The cornerstone in AI applications is to find a way to train the system so that it learns the behaviour that the developer has designed. In the case of FC-FNN, this process consists on setting appropriate values for the weights and biases within the network. Due to the complexity of the structure and the task to be solved, there is not an analytical expression to compute such weights in a single step. Instead, an iterative numerical optimization process is carried out trying to minimize a loss function. This training process consists on three stages, which together constitute an *epoch*:

1. *Forward pass*: first, input samples (features) are fed to the FC-FNN and the output values (labels) are computed. Initially, the weight matrices are randomly initialised, unless more advanced techniques (e.g. transfer learning) are implemented [42, p. 630]. The derivative of the loss function with respect to the outputs is also computed.

2. *Backward pass*: for the second stage, the desired output values (targets) for these inputs are used for computing the output error of the network, which is usually taken as the loss function. Then, the error is propagated layer by layer from the outputs to the inputs to calculate the gradient of the loss function with respect to the parameters for each neuron in a process referred to as *back-propagation*. The error at a neuron within a hidden layer is obtained by multiplying the errors in the next layer's neurons by the connecting weights.

3. *Parameter update*: finally, the gradient of the loss function is employed to update the values of the weights and biases of the network. Normally, the chosen method for solving the optimization problem is based on the steepest-descent method, that is, the gradient of the loss function with respect to the weights is computed first and, then, the parameters are updated in the opposite direction of the one that yields the largest gradient [44, p. 21].

For a certain weight $w_{i,j}$ connecting the i^{th} neuron in layer L-1 and the j^{th} neuron in layer L, the partial derivatives of the cost function J are computed by means of the chain rule as in Equation (3.4), where α_j is the internal state of the j^{th} neuron, that is, before the activation function is applied, k and l are the indices of the neurons in layers L-1 and L, respectively, and δ is the output error of a neuron in a hidden layer.

$$\frac{\partial J}{\partial w_{i,j}} = \frac{\partial J}{\partial f\left(\alpha_j\right)} \frac{\partial f\left(\alpha_j\right)}{\partial \alpha_j} \frac{\partial \alpha_j}{\partial w_{i,j}} \tag{3.4}$$

$$\frac{\partial \alpha_j}{\partial w_{i,j}} = \frac{\partial}{\partial w_{i,j}} \left(\sum_{k=1}^{m} w_{k,j}\alpha_k + b_j \right) = \frac{\partial}{\partial w_{i,j}} \left(w_{i,j} f\left(\alpha_i\right) \right) = f\left(\alpha_i\right) \tag{3.5}$$

$$\frac{\partial f\left(\alpha_j\right)}{\partial \alpha_j} = \begin{cases} 1, & \text{if } \alpha_j > 0 \\ 0, & \text{otherwise} \end{cases} \tag{3.6}$$

$$\frac{\partial J}{\partial f\left(\alpha_j\right)} = \frac{\partial}{\partial f\left(\alpha_j\right)} \left(y_j^* - f\left(\alpha_j\right) \right)^2 = 2 \left(f\left(\alpha_j\right) - y_j^* \right) \quad \text{[Output layer]} \tag{3.7}$$

$$= \sum_{l=1}^{p} w_{i,l}\delta_l \quad \text{[Hidden layer]} \tag{3.8}$$

During back-propagation, the task is to compute each of the elements in the right-hand side of Equation (3.4) by applying Equations (3.5) to (3.8). Note that, for demonstration purposes, ReLU as activation function and mean squared error as loss function are chosen. The process for computing the gradient with respect to the bias terms is analogous and only Equation (3.5) needs to be modified. By repeating this set of operations from the output layer all the way to the inputs, the complete gradient of the loss function can be obtained collecting all the partial derivatives as in Equation (3.9).

$$\nabla J\left(\boldsymbol{w}, \boldsymbol{b}\right) = \left(\frac{\partial J}{\partial w_1}, \ldots, \frac{\partial J}{\partial w_n}, \frac{\partial J}{\partial b_1}, \ldots, \frac{\partial J}{\partial b_m}\right) \tag{3.9}$$

Lastly, the weights and biases are updated by applying the steepest-descent algorithm in the form of Equation (3.10), where $\gamma \in (0, 1)$ is the learning rate.

$$\boldsymbol{w}\left[k+1\right] = \boldsymbol{w}\left[k\right] - \gamma\, \nabla J\left(\boldsymbol{w}\right)\left[k\right] \qquad \boldsymbol{b}\left[k+1\right] = \boldsymbol{b}\left[k\right] - \gamma\, \nabla J\left(\boldsymbol{b}\right)\left[k\right] \tag{3.10}$$

Another key aspect when developing intelligent algorithms is the appropriate choice of a learning strategy. For example, as opposed to weights and biases, the learning rate (γ) is a parameter which is not updated during training (hyperparameter) and, thus, the problem of choosing an suitable value arises. Furthermore, having a constant learning rate throughout the entire learning process may result in poor performance. A better approach is to set a large initial value so that the algorithm quickly approaches an optimal solution, and then reduce it to keep the system from oscillating or diverging from it. Likewise, better results may be obtained if some history of the cost function's gradient is used rather than just the latest value. These techniques are known as *learning rate decay* and *gradient momentum*, respectively, and they are part of optimizer algorithms which help improve the performance of the neural network. Some popular examples are RMSProp and Adam [44, p. 134].

One of the main problems that AI applications may face during training is *overfitting*, which means that the algorithm tends to memorise the input-output training data, rather than learning the underlying characteristics of the process [44, p. 25]. It is usually caused by too-low amount or diversity of samples in the training data set. This phenomenon can be detected during training if, when feeding the network with data from the test data set, the training error keeps decreasing but the test error starts increasing. In other words, a system experiencing overfitting will show very high accuracy for the training data set, but poor performance when tasked with new, unseen data. Fortunately, many techniques have been proposed to avoid this situation, such as early stopping, batch training and neuron drop-out. When using early stopping, a certain limit is set to the number of consecutive epochs for which the validation error does not decrease, so that the training process is halted whenever the system stops improving and before overfitting occurs [44, pp. 27, 192]. Batch training means that, rather than updating the weights every time a new set of inputs is used, the optimization routine is only executed after a certain number of randomly selected samples have been processed. Thus, the average error of the entire batch is used to update the weights instead of the individual errors. This helps the system have a more robust progression towards the optima, since the effect of outlier values is mitigated by the good ones in the same batch, and can reduce computational cost. In this case, the steepest-descent method becomes the stochastic gradient descent (SGD) [44, p. 121]. Lastly, as its name indicates, neuron drop-out consists on randomly deactivating some neurons in each layer throughout the entire network. This way, the algorithm becomes better at generalising and the influence of individual paths is reduced [44, p. 188].

3.2 Broad learning system

The broad learning system (BLS) is a novel ML algorithm which was first introduced in [46], [47] and has been further extended and applied to various engineering problems [48]–[50]. It is an evolution of a sub-class of FNN referred to as random-vector functional-link neural networks (RVFLNN), whose main feature is that the output is directly connected to both a hidden layer and the input. Figure 3.3 shows a generic diagram of a BLS algorithm with p inputs, q outputs, n feature-mappings windows, k feature maps per window, and m enhancement nodes. Some of the weights' labels have been omitted and their lines dashed to improve readability. The diagram's symbols are explained in Table 3.2. For the remainder of this report, the nomenclature employed to describe a BLS structure will be $BLS\left(n_{fea} - n_{win} - n_{enhan},\ AF\right)$, where n_{fea} and n_{win} are the number of features per window and the amount of feature windows in the feature-mapping layer, n_{enhan} is the amount of enhancement nodes, and AF denotes the activation function employed in all neurons.

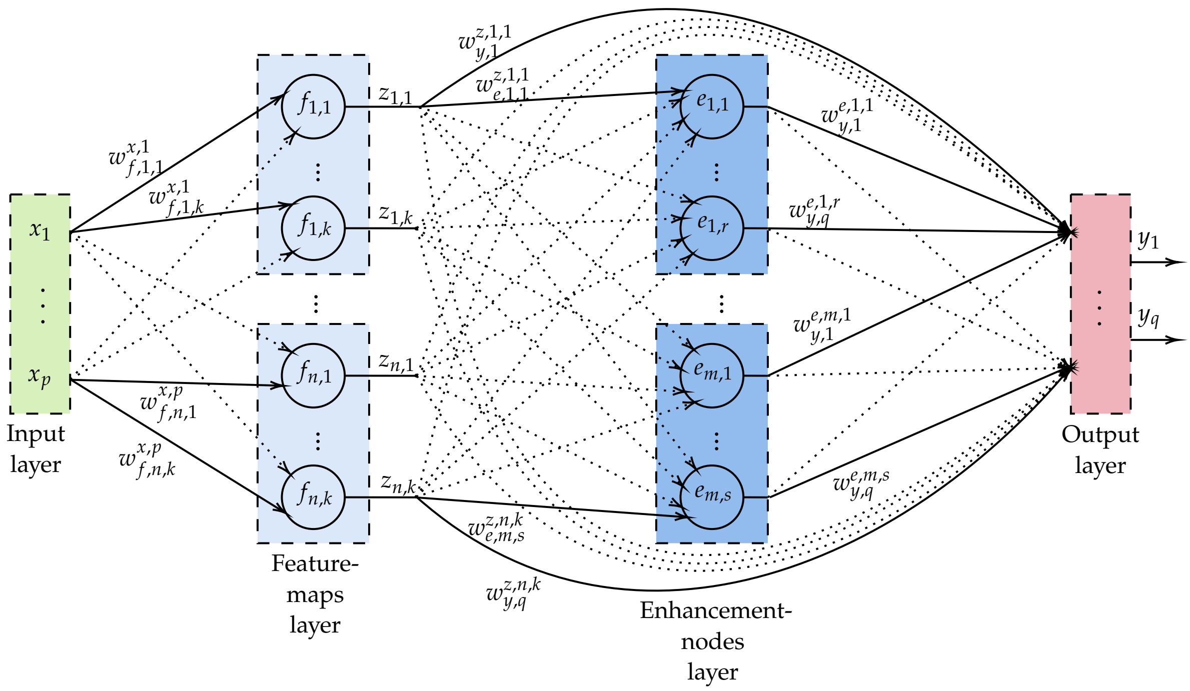


Figure 3.3: Structure of a BLS. Inspired from [46]

One of the improvements of BLS over RVFLNN is that the inputs are not directly connected to the output layer. Instead, a hidden layer of feature mappings is inserted between them. The purpose of this modification is to introduce a first stage in which some features are already extracted from the input data, which is a very common practice in deep-learning

Table 3.2: Description of variables of BLS in Figure 3.3

Symbol	Description
$\boldsymbol{x} = (x_1, ..., x_p)$	Input vector
$\boldsymbol{y} = (y_1, ..., y_q)$	Output vector
$w_{f,n,k}^{x,p}$	Weight from input x_p to feature mapping $f_{n,k}$
$w_{e,m,s}^{z,n,k}$	Weight from feature mapping $f_{n,k}$ to enhancement node $e_{m,s}$
$w_{y,q}^{z,n,k}$	Weight from feature mapping $f_{n,k}$ to output y_q
$w_{y,q}^{e,m,s}$	Weight from enhancement node $e_{m,s}$ to output y_q

algorithms. This is done by means of a sparse autoencoder [43, p. 502], that is, a compact neural network trained to replicate the input information at its output. Evidently, a perfect copy of the inputs would not provide any benefits and, therefore, the system has to be constrained somehow. When done properly, an autoencoder is capable of transferring the information from the inputs and, at the same time, extracting relevant features from them. Another useful application of these elements is dimensionality reduction in problems with a large number of variables. In Figure 3.3, the sparse autoencoder is simply represented by the weights between the input layer and the feature mappings.

The nodes in the feature-maps layer take as inputs the outputs of the sparse autoencoder, and their outputs are directly fed to both the enhancement nodes and the output layer. All these nodes are also basic neurons with a set of weights, bias and activation function as depicted in Figure 3.1a. The operations carried out between the inputs and the feature mappings, between the feature mappings and the enhancement nodes, and between these and the output are shown in Equations (3.11) to (3.13), respectively, where $\boldsymbol{H}$ denotes the output of the enhancement nodes.

$$\boldsymbol{Z} = f\left(\boldsymbol{W}_f^x \boldsymbol{x} + \boldsymbol{B}_f^x\right) \tag{3.11}$$

$$\boldsymbol{H} = f\left(\boldsymbol{W}_e^z \boldsymbol{Z} + \boldsymbol{B}_e^z\right) \tag{3.12}$$

$$\boldsymbol{Y} = \left[\boldsymbol{Z} | \boldsymbol{H}\right] \left[\boldsymbol{W}_y^z | \boldsymbol{W}_y^e\right]^T = \boldsymbol{A}\,\boldsymbol{W} \tag{3.13}$$

Although the weights in the BLS could be computed by means of SGD-based optimization, the original authors took advantage of the very flat and compact structure of the network and used matrix-inversion methods instead. More precisely, they decided to apply a ridge-regression algorithm (also known as norm-regularization), which is a very efficient way of approximating a matrix's pseudo-inverse. The optimization problem is thus reduced to solving Equation (3.14), where λ is a positive regularisation coefficient and $\boldsymbol{I}$ is the identity matrix. Consequently, rather than back-propagating the output error to compute the gradient of some loss function with respect to each parameter, the entire weight matrix is optimised at once. This makes training the BLS significantly faster than an equivalent FC-FNN.

$$\boldsymbol{W} = \left(\lambda \boldsymbol{I} + \boldsymbol{A}\boldsymbol{A}^T\right)^{-1} \boldsymbol{A}^T \boldsymbol{Y} \tag{3.14}$$

One of the features that make BLS particularly interesting is its incremental learning capabilities. This means that the structure can be augmented, by increasing the number of either feature maps or enhancement nodes, without the need of training the entire network from scratch. Instead, only the new parts of the network are taken into account to update the system's weights. This can largely reduce both computational power requirements and training time when working with large data sets or a very high amount of parameters. In the original work, algorithms were derived to update the network whenever its structure was modified or new training samples were used. Furthermore, since machine learning models have a tendency to exponentially grow in size as the problems become more complex, a structure-simplification method was also proposed based on singular-value decomposition.

The two architectures presented in this chapter are just two examples in the ever-broadening field of artificial intelligence. More advanced algorithms exist for other types of tasks, such as recurrent neural networks for time-series data [44, p. 271] or convolutional neural networks for data with a grid-like topology [43, p. 330]. Neither is the case of this thesis. The FC-FNN structure was chosen because it is the most basic one and can perform well in a wide variety of applications, while BLS was used since it is still a recent development and comparative studies with other architectures are necessary to determine its potential.
The field of AI is very extensive and new strategies, algorithms and applications are constantly being developed. Thus, only the very basic and strictly necessary concepts have been discussed in this chapter. For a more in-depth study of the topics, the reader is referred to dedicated references [41]–[44], [51]. After the fundamental knowledge regarding batteries and neural networks has been presented, the following chapters are focused on the design, development and testing of intelligent algorithms for SoH estimation.

Chapter 4

SoC-based State-of-Health Estimation

In this chapter, an intelligent algorithm for SoH_Q estimation based on voltage measurements at certain SoC milestones during charging is derived and evaluated, taking into consideration different chemistries, temperatures and ageing procedures.

Contents

It was shown in the previous chapters that two generally accepted definitions of state of health exist: the first one is related to the maximum available capacity of a battery at a certain moment and under certain conditions, while the second one is based on the change in battery impedance as the device degrades. Furthermore, these definitions also allow for establishing end-of-life thresholds, that is, some criteria such that, when met, the battery is considered to have lost significant performance capabilities and cannot be properly used anymore. In the first case, EoL is usually defined as the point at which the maximum available capacity is 80 % of the nominal capacity of a pristine unit. In the case of impedance the definition becomes more complex, since it largely depends on the excitation frequency. There is however a common agreement to use the internal resistance value, that is, the one with null complex part. Thus, the cell is said to have reached EoL whenever this value becomes 200 % of the original one. Both definitions are shown in Equation (4.1). Practically, each of these definitions is linked to a specific application: for those cases where capacity is most relevant, such as battery electric vehicles, the capacity-based definition is regularly used, while the impedance-based definition is the one used for applications where power delivery is more important, such as hybrid electric vehicles. It is important to note that, as pointed out in Chapter 2, both capacity and impedance depend on the cell's conditions and, therefore, these must also be included in the previous definitions. Most authors use 25 °C as the reference temperature and 100 % SoC, as well as some specific C-rate.

$$SoH_Q = \frac{Q_{max}}{Q_{rated}} \cdot 100\,\% \qquad\qquad SoH_R = \frac{R_0}{R_{0,pristine}} \cdot 100\,\% \tag{4.1}$$

It was shown in Section 1.2 that many methodologies exist for $\mathrm{SoH_Q}$ estimation, each of them studying a different phenomenon, such as CCCV curve displacement or IC and DV trajectory variations. For this thesis, the proposed approaches here and those in Chapter 5 are based on the voltage-capacity relationship during charging, since this process is considerably more controllable than discharge, which makes the methods more applicable and reliable. As a consequence, these algorithms cannot be applied in real time while driving, and they have to wait until the vehicle is stationed at a charger. This is considered acceptable, since the $\mathrm{SoH_Q}$

is a slow-varying parameter and does not need to be continuously estimated.
Two AI-based estimators are designed and tested for this algorithm, namely a fully-connected feed-forward neural network and a broad learning system (see Chapter 3). For all the FC-FNN algorithms employed throughout this thesis the configuration in Table 4.1 is used for training, unless noted otherwise. Many FC-FNN structures in terms of number of hidden layers, number of neurons per layer or activation functions are tested for each of the approaches, and the one with best performance in the test data set is selected. For the BLS algorithms, the structures are generated by applying a grid-search iterative procedure to the three parameters, i.e. number of feature windows, number of features per window and number of enhancement nodes, and the model is selected as in the FC-FNN case. For each case, the FC-FNN and the BLS are trained and tested on exactly the same samples to make comparisons easier.

Table 4.1: FC-FNN configuration parameters

Early stopping	Batch size	Optimizer	Epochs	Loss	Train / Validation split
50 epochs	5 %, min. 5 samples	Adam	2500	MSE	70 % / 30 %

The performance metrics used for comparison and model selection are the root mean squared error (RMSE), the mean absolute error (MAE) and the maximum absolute error (ME), as well as the residuals' mean (μ_{res}) and standard deviation (σ_{res}), all of them normalized. The formulas for these metrics are as given by Equations (4.2) and (4.3), where $\hat{y}_i, y_i$ are the estimated and real output values of the ith sample, respectively, and N is the total number of samples.

$$e_i = \frac{\hat{y}_i - y_i}{y_i} \qquad \mu_{res} = \frac{1}{N}\sum_{i=0}^{N-1} e_i \qquad \sigma_{res} = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1}\left(e_i - \mu_{res}\right)^2} \tag{4.2}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1} e_i^2} \qquad MAE = \frac{1}{N}\sum_{i=0}^{N-1} |e_i| \qquad ME = \max(|e_i|) \tag{4.3}$$

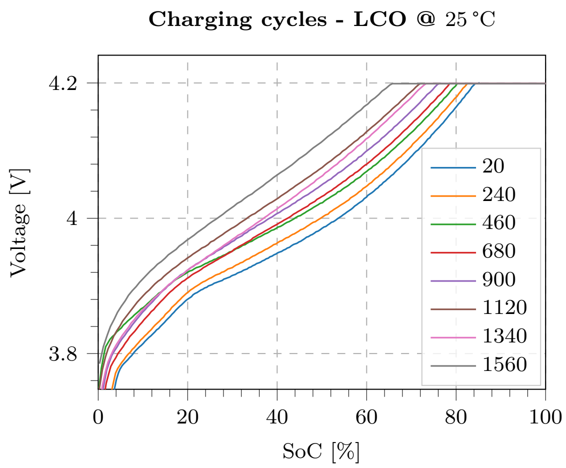
4.1 Approach A1: terminal voltage at SoC milestones with incomplete data

In this first approach it is proposed to measure the battery's terminal voltage at certain SoC milestones while charging. These milestones could be chosen according to one of two criteria: either analysing which ones present bigger differences as the cell ages, or based on real applicability considerations. The former one could lead to better results, but it may be less practical if the most significant differences happen at extreme capacity milestones which may not be reached during regular charging processes. Here the latter is chosen in order to guarantee good applicability, taking 40 %, 50 % and 60 % SoC as milestones. This algorithm is applied to several cells aged by means of CCCV full charge-discharge cycles trying to keep a constant temperature of 25 °C. All cells' data are part of *Battery Archive*'s data set [30], [31] (see Appendix A) and the cycling conditions are presented in Table 4.2.

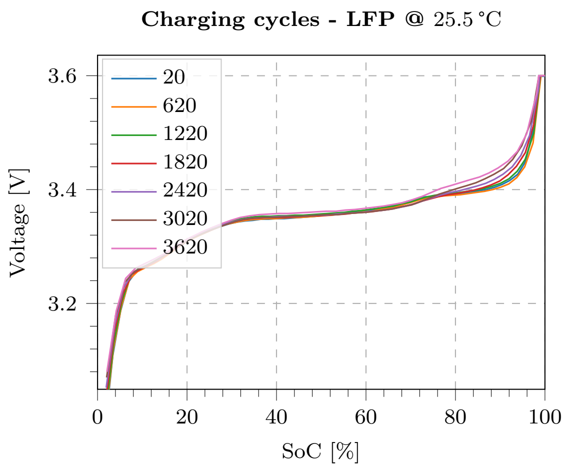
Table 4.2: Cells used for approach A1. Data from [30], [31]

Chemistry	C-rate (ch/dis)	Cut-off voltage	Rated capacity	Temperature	ID
LCO	0.5 C / 0.5 C	2.7 V / 4.2 V	1.35 A h	25 °C	e
LFP	0.5 C / 3 C	2 V / 3.6 V	1.1 A h	25.5 °C	d
NCA	0.5 C / 1 C	2.5 V / 4.2 V	3.2 A h	27.5 °C	a
NMC	0.5 C / 1 C	2 V / 4.2 V	3 A h	26 °C	b
NMC-LCO	0.5 C / 1.5 C	2.8 V / 4.3 V	2.8 A h	25 °C	a

The algorithm operates as follows: as the battery is being charged at constant current, terminal voltage is measured and registered whenever one of the SoC milestones is reached. Then, all three measurements are fed to the SoH_Q estimator, which returns the estimated maximum available capacity of the battery. This algorithm assumes that a reliable SoC estimator is available and that temperature does not significantly change while charging.
The charging curves for all cells are shown in Figure 4.1 for some of the cycles in the time series. In general, it can be observed that, as the cell ages, the slope becomes steeper and the constant-current zone limit is reached faster, thus the maximum available capacity decreases. This behaviour is specially clear in the case of LCO and NMC-LCO cells (see Figures 4.1a and 4.1e), and the distance between the curves is higher at high SoC than at low SoC, while NCA and NMC experience the opposite behaviour and the difference is less significant. LFP cells present nearly flat CCCV curves and, therefore, the voltage trajectory barely changes as the cell ages. Based on these observations, the algorithms would be expected to perform best on NCA and NMC cells, but poorly on LFP cells, since the curves are further apart from each other. For LCO and NMC-LCO cells great performance would not be expected, since these types, when severely aged, may reach the CV region at 60 % SoC. Thus, cells at different degradation stages would show the same voltage at this milestone.

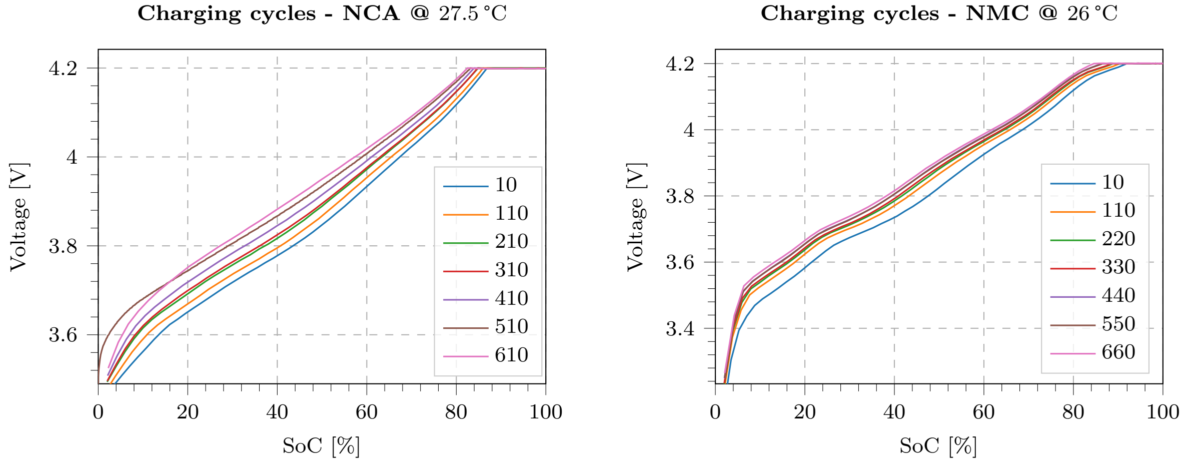


(a) LCO



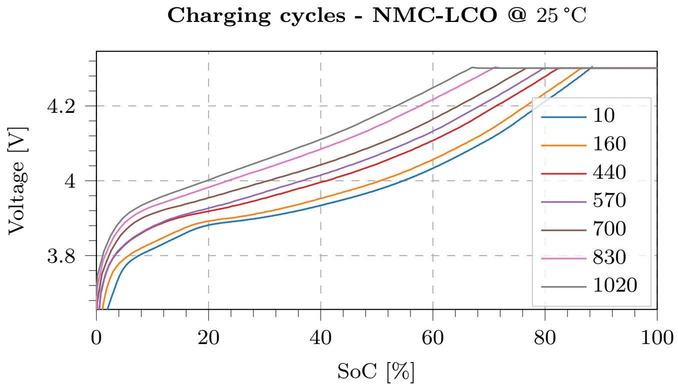
(b) LFP

Figure 4.1: CCCV charge curves evolution with ageing



(c) NCA **(d)** NMC

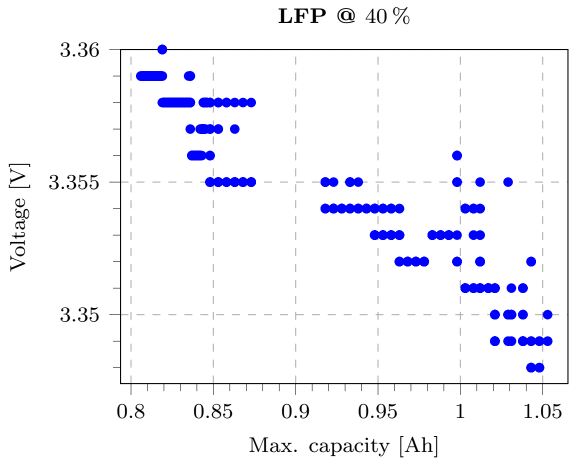
Figure 4.1: CCCV charge curves evolution with ageing (cont.)



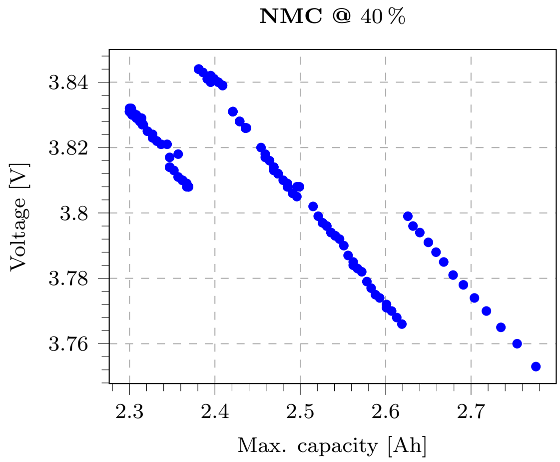
(e) NMC-LCO

Figure 4.1: CCCV charge curves evolution with ageing (cont.)

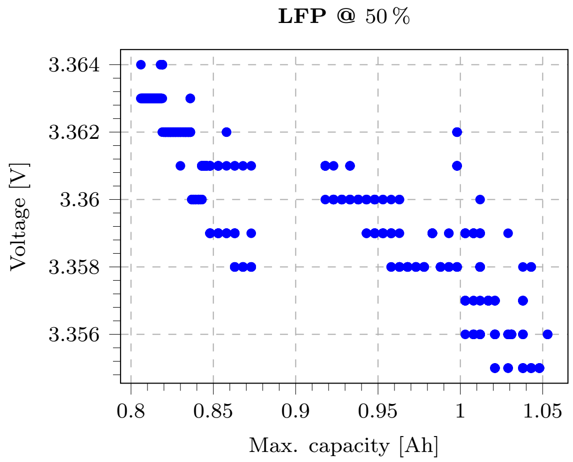
Figure 4.2 shows the evolution of the voltage at the three chosen SoC milestones as the cell degrades. Only LFP and NMC types are shown for comparison. The amount of points has been reduced in order to improve readability. As expected from the analysis of Figure 4.1, terminal voltage increases at all three points as the maximum available capacity decreases. Besides the clear general tendency, it should be noted how, in the LFP cell, different levels of maximum capacity have the same terminal voltage, which is a natural consequence of the voltage curves staying almost unchanged throughout ageing (see Figure 4.1b). This could pose a significant challenge for the estimator, because the same input values would be linked to different output ones. The plots also give insight regarding the magnitude of voltage change with ageing. In the case of the NMC units, the variation is around 100 mV across the entire range, while for the LFP ones it is just 10 mV. The fact that this voltage variation is very low could imply that the intelligent algorithm has difficulties learning the behaviour. Furthermore, this also means that very accurate and noise-proof voltage sensors would be required for this application, since just a small deviation could lead to significant capacity estimation errors.



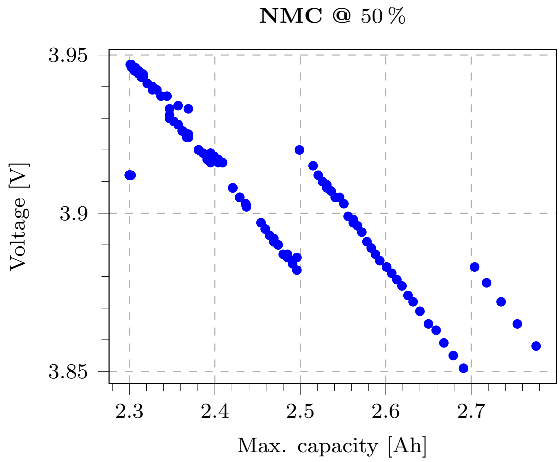
(a) Terminal voltage evolution at 40 % SoC. LFP cell



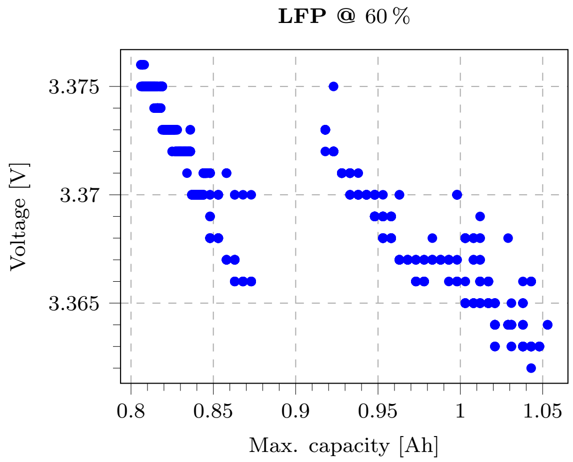
(b) Terminal voltage evolution at 40 % SoC. NMC cell



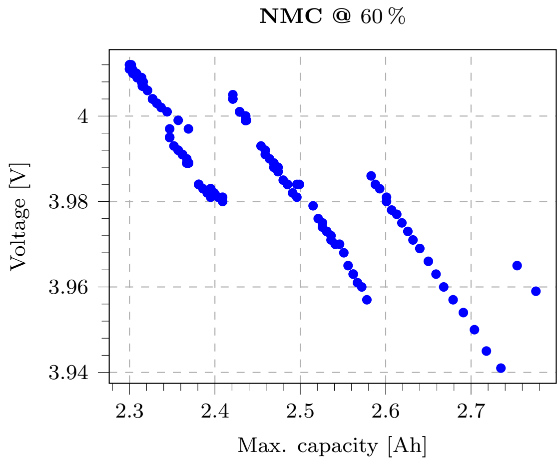
(c) Terminal voltage evolution at 50 % SoC. LFP cell



(d) Terminal voltage evolution at 50 % SoC. NMC cell



(e) Terminal voltage evolution at 60 % SoC. LFP cell



(f) Terminal voltage evolution at 60 % SoC. NMC cell

Figure 4.2: Terminal voltage evolution with ageing of cells for approach A1

To further test the capabilities of the algorithm, three scenarios are considered: one in which 90 % of the entire data set is available for training, another one in which the first 60 % of the cycles can be used and, lastly, one in which only the earliest 25 % is available. This aims at exploring the performance of the methods in the case that reference cells have not yet been cycled until EoL but the application needs to be deployed. The results obtained when applying this algorithm to the testing data set are summarised in Table 4.3.

Table 4.3: Results for approach A1

Cell	Structure	Data split	RMSE	MAE	ME	μ_{res}	σ_{res}
LCO	BLS(1-6-3, tanh)	90 %	5.15 %	3.45 %	13.8 %	1.6 %	4.9 %
	BLS(2-5-4, tanh)	60 %	10.1 %	8.8 %	30.2 %	8.6 %	5.3 %
	BLS(2-1-7, tanh)	25 %	8.03 %	6.43 %	38.5 %	3.4 %	7.3 %
	FNN(30-30-30, ReLU)	90 %	15.8 %	15.8 %	17.3 %	16 %	0.9 %
	FNN(30-30-30, ReLU)	60 %	17.7 %	15.2 %	43.2 %	15 %	9 %
	FNN(30-30-30, ReLU)	25 %	30.1 %	25.1 %	75.6 %	25 %	17 %
LFP	BLS(1-2-2, tanh)	90 %	3.14 %	2.25 %	18.2 %	0.03 %	3.1 %
	BLS(1-2-2, tanh)	60 %	9.15 %	7.58 %	39 %	7.3 %	5.6 %
	BLS(1-4-2, tanh)	25 %	9.7 %	7.8 %	56 %	−6.1 %	7.6 %
	FNN(20-20-20-10, ReLU)	90 %	6.13 %	4.82 %	14.7 %	0.5 %	6.1 %
	FNN(20-20-20-10, ReLU)	60 %	16.2 %	15.6 %	22.6 %	16 %	4.4 %
	FNN(20-20-20-10, ReLU)	25 %	14.5 %	12.2 %	27.8 %	12 %	8 %
NCA	BLS(1-1-12, tanh)	90 %	0.85 %	0.7 %	3.29 %	−0.1 %	0.9 %
	BLS(2-14-14, tanh)	60 %	1.63 %	1.43 %	4.9 %	−0.3 %	1.6 %
	BLS(2-2-12, tanh)	25 %	1.89 %	1.63 %	6.18 %	1.1 %	1.6 %
	FNN(100-100-100, ReLU)	90 %	1.56 %	1.27 %	2.71 %	0.6 %	1.5 %
	FNN(100-100-100, ReLU)	60 %	2 %	1.37 %	4.32 %	0.9 %	1.8 %
	FNN(100-100-100, ReLU)	25 %	17.1 %	15.2 %	31 %	15 %	7.9 %
NMC	BLS(1-4-2, tanh)	90 %	0.45 %	0.38 %	0.84 %	−0.03 %	0.45 %
	BLS(1-2-2, tanh)	60 %	1.05 %	0.88 %	2.9 %	0.8 %	0.7 %
	BLS(6-2-3, tanh)	25 %	1.24 %	0.98 %	4.2 %	0.3 %	1.2 %
	FNN(50-50-50-30, ReLU)	90 %	1.13 %	0.66 %	3.27 %	0.7 %	0.9 %
	FNN(50-200-200-30, ReLU)	60 %	2.52 %	2.41 %	3.69 %	1.8 %	1.8 %
	FNN(50-200-200-30, ReLU)	25 %	6.85 %	6.51 %	11.5 %	6.5 %	2.1 %
NMC-LCO	BLS(3-4-1, tanh)	90 %	5.55 %	4.04 %	15.1 %	1.2 %	5.4 %
	BLS(7-5-1, tanh)	60 %	6.7 %	4.31 %	31.2 %	2.8 %	6.1 %
	BLS(3-1-8, tanh)	25 %	4.68 %	2.9 %	33.9 %	0.9 %	4.6 %
	FNN(20-20-20, ReLU)	90 %	9.16 %	6.19 %	40.3 %	−3.6 %	8.4 %
	FNN(20-20-20, ReLU)	60 %	25 %	23 %	44 %	23 %	9.9 %
	FNN(20-20-20, ReLU)	25 %	9.55 %	6.7 %	44.7 %	−5.8 %	7.6 %

It is evident that, in general, having more data available for training leads to better prediction capabilities. It can be observed that, indeed, the best results correspond to the NCA and NMC chemistries, which show good accuracy even when only early data are available, although some overfitting was present when the FC-FNN was trained with only 25 % of the samples. Both algorithms experienced overfitting for the LFP cell even when almost the entire data set was used for training. Similarly, relatively large errors were obtained for the LCO and NMC-LCO units. Using the regularization techniques as discussed in Section 3.1 did not improve their performance. These results are as expected based on the previous analysis. Lastly, the BLS was more accurate than the FC-FNN in all cases.

4.2 Approach A2: terminal voltage at SoC milestones with multiple temperatures

In the previous approach it was assumed that the batteries were always cycled at the same temperature. However, this may not be the case in a real scenario and, since this parameter has a significant impact on the voltage-capacity curves (see Chapter 2), the SoH_Q estimator should be capable of performing well even when the charging temperature is different from the one used during training. Therefore, the approach presented here aims at taking temperature into consideration in order to predict the cell's capacity. In this case, the cells have been cycled at three temperatures, as shown in Table 4.4. The data used here come from *Battery Archive*'s data set [30], [31]. This algorithm also takes as inputs the measured terminal voltages at 40 %, 50 % and 60 % SoC, and produces an estimation of the maximum available capacity as output, which corresponds to the temperature at which the voltage has been measured. The algorithm assumes constant-current charging and constant temperature for each individual cycle. Data from cycling at approximately 15 °C and 35 °C are used for training, while data at 25 °C are used for testing. The entire time series for each temperature is used for training or testing.

Table 4.4: Cells used for approach A2. Data from [30], [31]

Chemistry	C-rate (ch/dis)	Cut-off voltage	Rated capacity	Temperature	ID
LFP	0.5 C / 1 C	2 V / 3.6 V	1.1 A h	15.7 °C	b
	0.5 C / 1 C	2 V / 3.6 V	1.1 A h	24.3 °C	c
	0.5 C / 1 C	2 V / 3.6 V	1.1 A h	34.8 °C	a
NCA	0.5 C / 1 C	2.5 V / 4.2 V	3.2 A h	19.3 °C	b
	0.5 C / 1 C	2.5 V / 4.2 V	3.2 A h	27.5 °C	a
	0.5 C / 1 C	2.5 V / 4.2 V	3.2 A h	38.1 °C	a
NMC	0.5 C / 1 C	2 V / 4.2 V	3 A h	17.7 °C	a
	0.5 C / 1 C	2 V / 4.2 V	3 A h	26 °C	b
	0.5 C / 1 C	2 V / 4.2 V	3 A h	36 °C	c

Charging cycles for pristine and severely aged LFP and NMC cells are shown in Figure 4.3 for all three temperatures in the data set. It can be observed that voltage levels are higher as the temperature drifts away from 25 °C. When it comes to ageing, both types show the same tendency: voltage increases with ageing at any temperature. This is in accordance with the analysis elaborated in Sections 2.2 and 2.3.

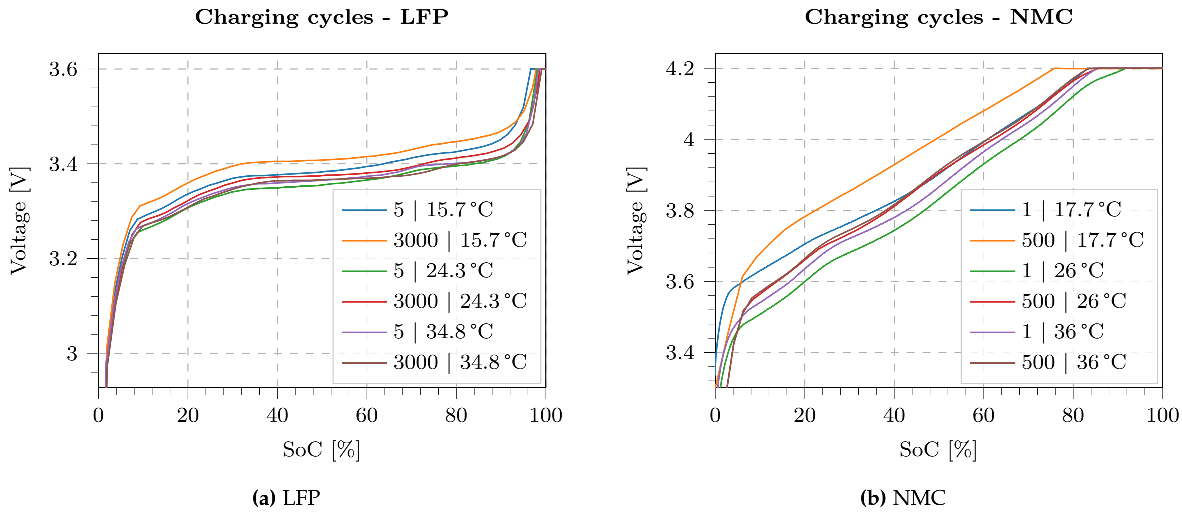
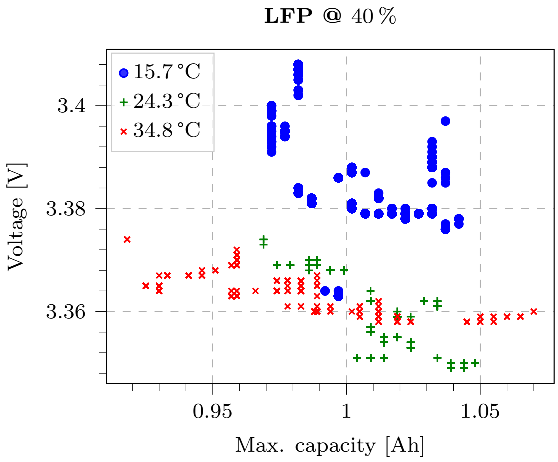
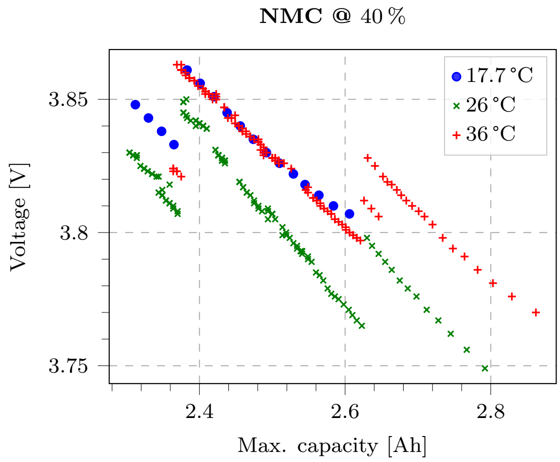


Figure 4.3: CCCV charge curves evolution with ageing and temperature for cells used in approach A2

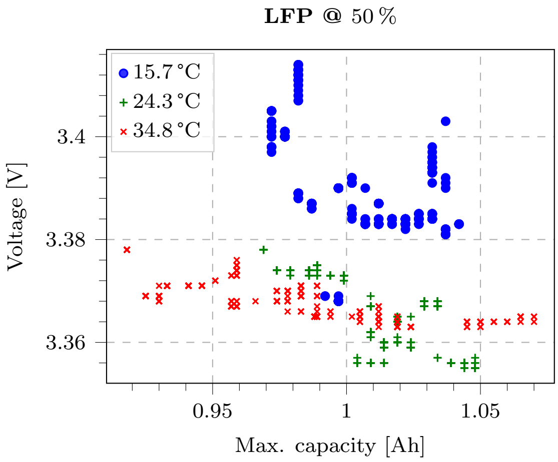
Voltage variation at the chosen SoC milestones for all considered temperatures is shown in Figure 4.4 for both LFP and NMC cells. The number of points has been reduced to improve readability. It can be observed that the patterns of the two cells are very different. The NMC cell has a more consistent behaviour, with a voltage increase as the maximum capacity diminishes as previously noted. The effect of temperature is that the voltage is higher as the temperature drifts away from 25 °C, in accordance with Figure 4.3b. This is congruent throughout all three SoC milestones. Even though there is a clear tendency regarding both temperature and ageing, the superposition of points from two different temperatures could make the learning process more difficult. On the contrary, the LFP cell only shows significant differences at 15.7 °C and the voltage variation at any temperature is quite small, also in accordance with Figure 4.3a. The lack of large differences and the limited diversity of data available could make it very challenging for the algorithms to learn the underlying behaviour of the LFP cell. Thus, it would be expected to obtain superior performance on NMC cells than on LFP ones.



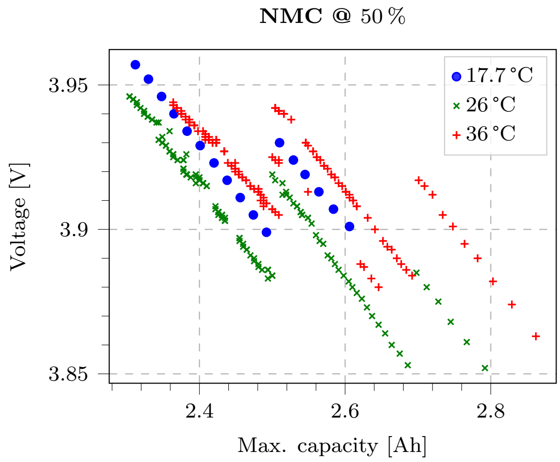
(a) Terminal voltage evolution at 40 % SoC. LFP cell



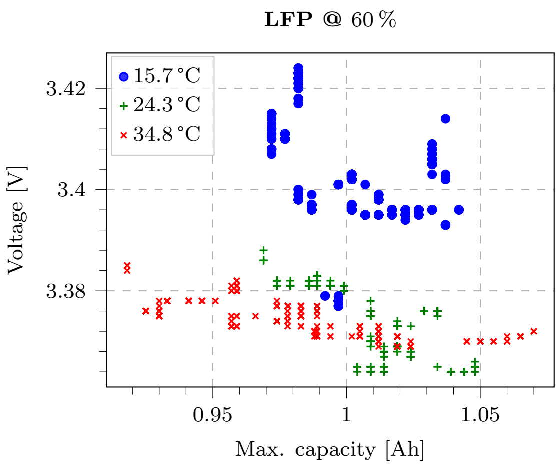
(b) Terminal voltage evolution at 40 % SoC. NMC cell



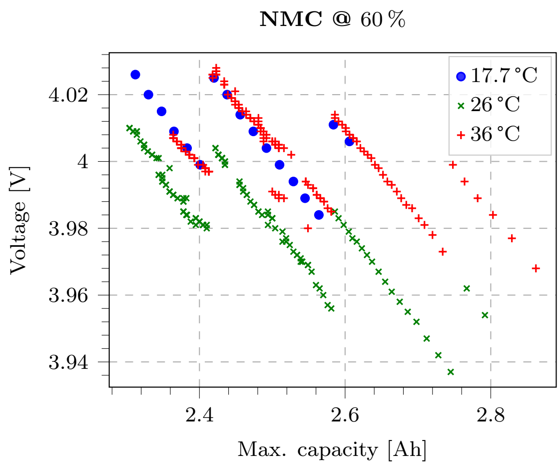
(c) Terminal voltage evolution at 50 % SoC. LFP cell



(d) Terminal voltage evolution at 50 % SoC. NMC cell



(e) Terminal voltage evolution at 60 % SoC. LFP cell



(f) Terminal voltage evolution at 60 % SoC. NMC cell

Figure 4.4: Terminal voltage evolution with ageing of cells for approach A2

The results are presented in Table 4.5. According to the performance metrics, only the BLS algorithm provides great accuracy for all cells. Nonetheless, a visual inspection of the individual results confirmed that, for the LFP unit, both BLS and FC-FNN had a tendency to overfit the training data. Some overfitting was also observed in the FC-FNN tests with NCA and NMC cells. This is caused by the low variety of samples in the data set, as having only two temperatures for training does not seem enough for learning the influence of this parameter. Furthermore, the fact that the LFP's curves barely change due to ageing and temperature made the algorithm prone to overfitting the training data. Similarly to approach A1, the BLS results are considerably better than the FC-FNN ones in all scenarios.

Table 4.5: Results of approach A2

Cell	Structure	RMSE	MAE	ME	μ_{res}	σ_{res}
LFP	BLS(12-4-1, tanh)	1.05 %	0.8 %	6.46 %	−0.04 %	1.1 %
	FNN(30-30-30, ReLU)	2.31 %	1.95 %	5.24 %	−1.2 %	2 %
NCA	BLS(1-4-4, tanh)	1.77 %	1.38 %	5.56 %	−0.5 %	1.7 %
	FNN(30-30-30, ReLU)	6.69 %	5.8 %	12.4 %	−5.6 %	3.6 %
NMC	BLS(1-2-2, tanh)	1.39 %	1.08 %	3.38 %	0.58 %	1.3 %
	FNN(30-30-30, ReLU)	5.32 %	4.37 %	13 %	1.4 %	5.1 %

4.3 Approach A3: terminal voltage at SoC milestones with realistic ageing

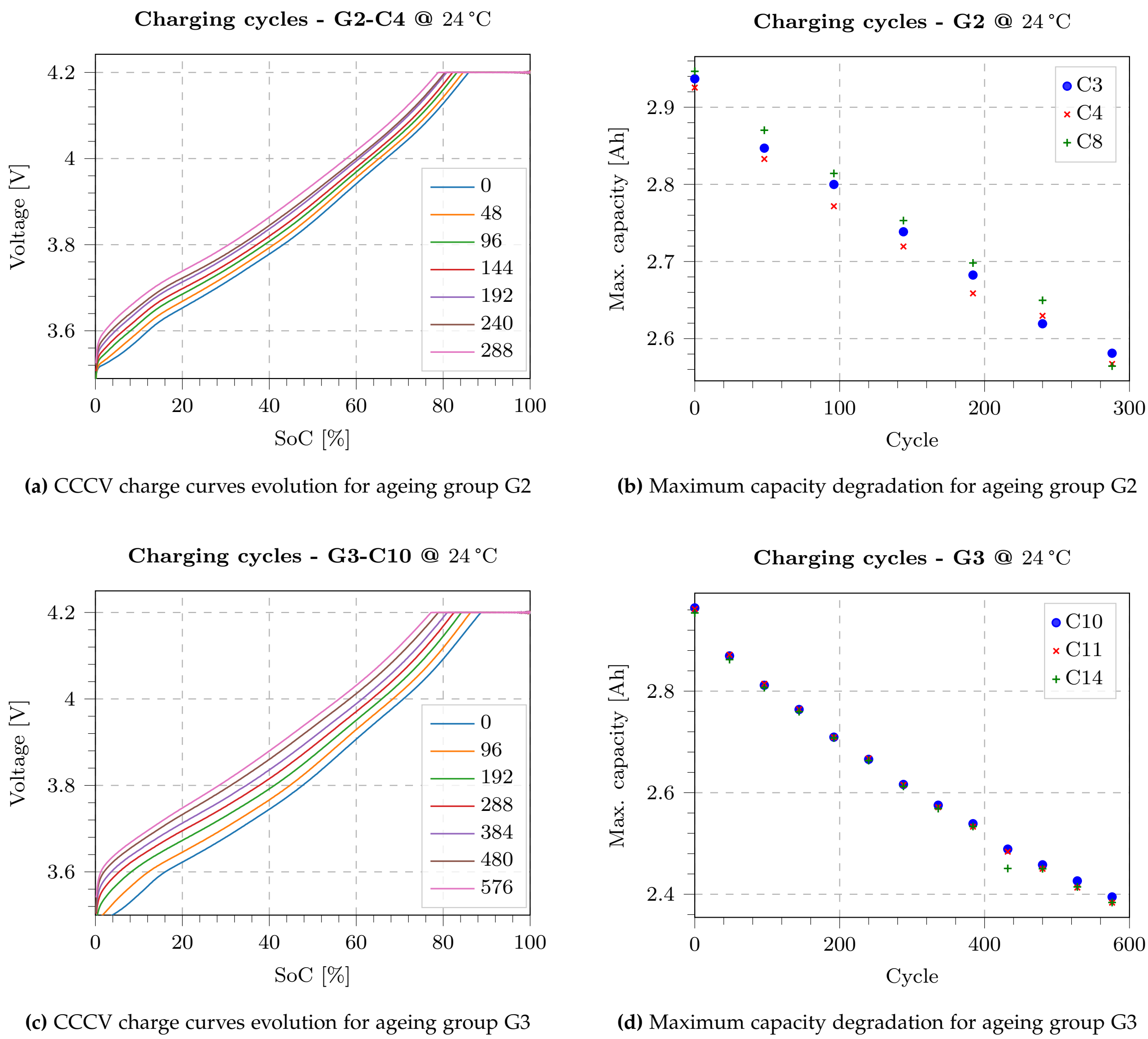
In all the previous approaches the cells had been aged by means of full CCCV charge-discharge cycles, which do not represent how batteries are actually used in EV applications. In order to explore the capabilities of the proposed approaches under realistic ageing conditions, the same algorithm is applied to cells from *Oxford University*'s data set [36], [37] (see Appendix A). The cells are NCA-type with the same characteristics presented in Table 4.2. Each iteration of the realistic ageing procedure consisted on two stages: a series of constant-current cycles followed by some idle days in storage, as described in Table 4.6. The first one represents the active usage of the battery while driving in a real application, while the second one represents calendar ageing when the vehicle is not being used. As it would be the case of an actual EV, the time allocated for calendar ageing is considerably larger than that for cycling. Furthermore, reference performance tests were conducted periodically to characterise the degradation of the units. All aforementioned tests were carried out at a constant chamber temperature of 24 °C. The whole process was repeated until the cells reached EoL conditions.

Similarly to approaches A1 and A2, this algorithm takes as inputs the measured voltages at 40 %, 50 % and 60 % SoC, and its output is the estimated maximum available capacity of the cell. Figures 4.5a and 4.5c show the voltage curves during charging corresponding to the CCCV reference performance tests of cell C4 in ageing group G2, and cell C10 in ageing group G3, respectively. It is clear that the curves displace upwards as deterioration increases. Capacity degradation for all cells in groups G2 and G3 is displayed in Figures 4.5b and 4.5d, respectively. It can be observed that all cells belonging to the same group experience nearly

Table 4.6: Ageing groups of cells used for approach A3. Data from [36], [37]

Group	Cells	Cycling ageing	Calendar ageing
G1	C9, C15, C20	1 day at 0.5 C	5 days at 90 % SoC
G2	C3, C4, C8	1 day at 0.25 C	5 days at 90 % SoC
G3	C10, C11, C14	2 days at 0.5 C	10 days at 90 % SoC
G4	C12, C18, C19	2 days at 0.25 C	10 days at 90 % SoC

identical degradation, which is almost linear with the number of cycles. The same applies to cells that have been aged in different ways. This is a very important feature, since cell-to-cell differences in the any of the data sets could lead to wrong estimations as the algorithm would be affected by a phenomenon which should not exist. On an additional note, the features observed in Figure 4.5 are coherent with those of similar cells in other data sets explored previously.



(a) CCCV charge curves evolution for ageing group G2

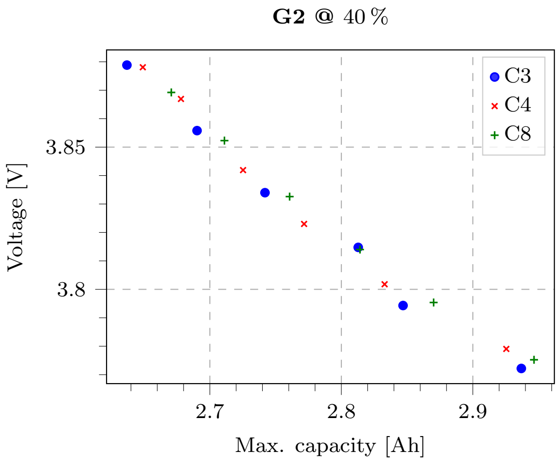
(b) Maximum capacity degradation for ageing group G2

(c) CCCV charge curves evolution for ageing group G3

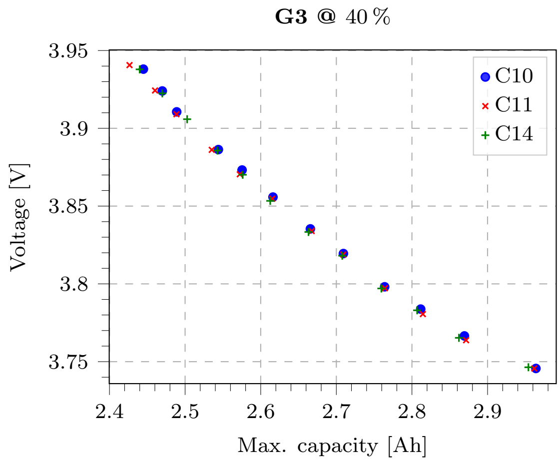
(d) Maximum capacity degradation for ageing group G3

Figure 4.5: Capacity degradation analysis of cells for approach A3

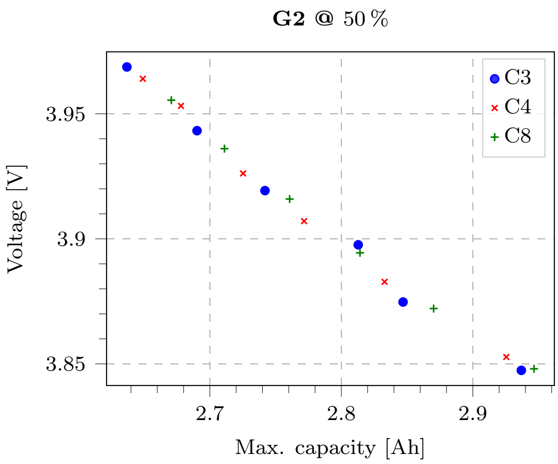
The evolution of the measured voltage at the SoC milestones for all the cells in ageing groups G2 and G3 is shown in Figure 4.6.



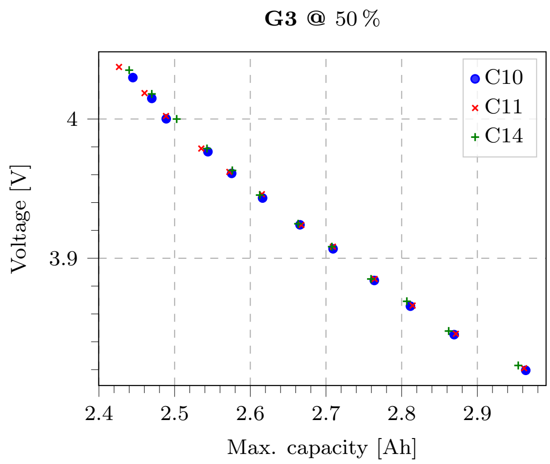
(a) Terminal voltage evolution at 40 % SoC. Group G2



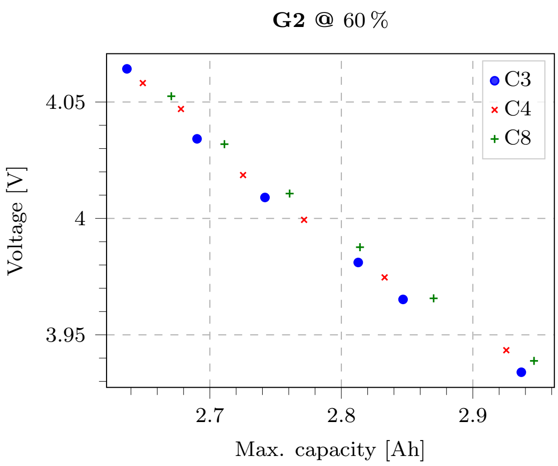
(b) Terminal voltage evolution at 40 % SoC. Group G3



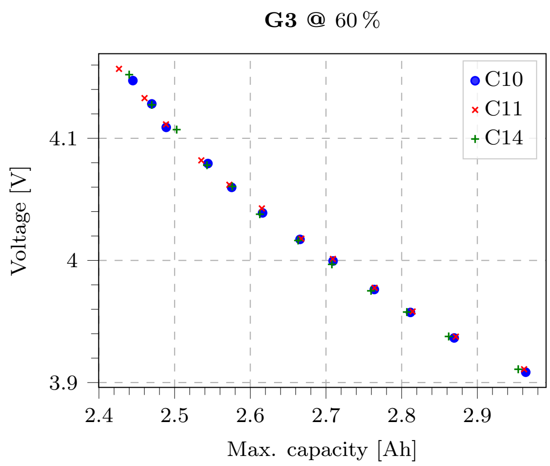
(c) Terminal voltage evolution at 50 % SoC. Group G2



(d) Terminal voltage evolution at 50 % SoC. Group G3



(e) Terminal voltage evolution at 60 % SoC. Group G2



(f) Terminal voltage evolution at 60 % SoC. Group G3

Figure 4.6: Terminal voltage evolution with ageing of cells for approach A3

As expected, there exists an almost linear relationship between capacity decay and voltage increase that all cells align with, and cell-to-cell differences are negligible. This situation indicates that a relatively simple pattern exists, so it would be expected to obtain good accuracy when trying to predict the cell's maximum capacity. However, some differences in voltage levels for the same maximum capacity are present between the two groups, which may impact the algorithms' performance given the low amount of data available.

Two different scenarios are used to verify the performance of this approach. First, two groups are used during training and the remaining two are used for testing. Each of the training groups (G2 and G3) corresponds to one of the shorter and longer ageing procedures and to the lowest and highest C-rate, so that the algorithm could learn both types of behaviour during training and then it would be tested on similar scenarios. For the second one, the algorithm is trained only on one of the groups (G3) and then tested on the rest. This represents a more realistic situation, since it would be unlikely to have experimental data for each and every possible scenario. The set-up for both cases and the obtained results are presented in Table 4.7.

Table 4.7: Results of approach A3

Training	Testing	Structure	RMSE	MAE	ME	μ_{res}	σ_{res}
G2, G3	G1, G4	BLS(1-5-1, tanh)	0.53 %	0.44 %	1.16 %	0 %	0.5 %
		FNN(30-300-30-30, ReLU)	0.95 %	0.74 %	2.26 %	0 %	0.9 %
G3	G1, G2, G4	BLS(1-6-8, tanh)	1.67 %	1.33 %	3.78 %	−0.8 %	1.4 %
		FNN(30-300-30-30, ReLU)	1.87 %	1.48 %	3.83 %	−1.2 %	1.4 %

It is clear that both the BLS and the FC-FNN can provide highly-accurate predictions regarding SoH_Q even when the algorithm is facing new situations on which it has not been trained, although performance slightly worsens when only one of the groups is taken as the training data set. Results would probably improve if more data, both from more cells and more frequent RPT, were available. In this case the BLS and the FC-FNN are on par in terms of performance, nevertheless the former contains less parameters, and thus requires lower computing power, which makes it more appealing for real-world implementations.

Chapter 5

Charge gradient-based State-of-Health Estimation

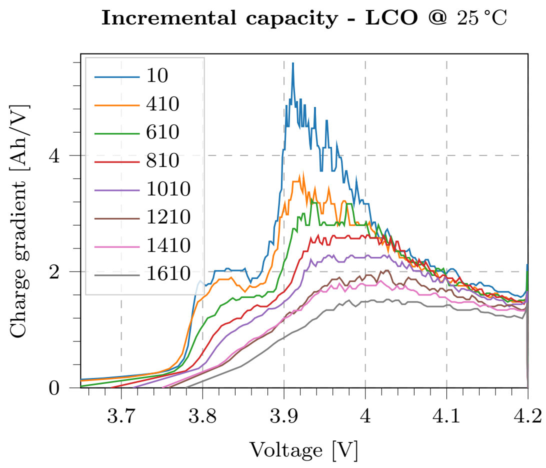
A novel, intelligent algorithm for SoH_Q estimation is designed in this chapter based on charge gradient data. As before, three approaches are implemented and tested in scenarios with different chemistries, temperatures and ageing procedures.

Contents

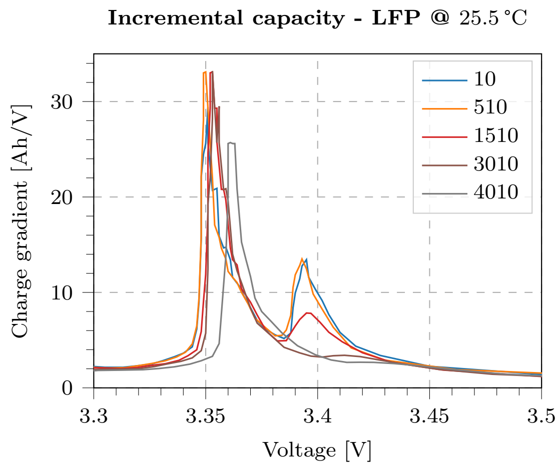
Although very effective, the three previous approaches presented some drawbacks, mainly the need for an accurate SoC estimator in order to detect the milestones for making the voltage measurements. Nevertheless, this can be avoided by reversing the strategy: rather than measuring the voltage at some specific SoC points, the charge increment between a sew specific voltage points is measured instead. Therefore, this algorithm requires accurate charge measurements, by means of e.g. coulomb counting, to obtain the increments between voltage milestones, as opposed to an absolute SoC estimation. This procedure is inspired by other IC-based methods, and these curves are analysed to decide the most favourable voltage milestones for each type of cell chemistry.

5.1 Approach A4: charge gradient at voltage milestones with incomplete data

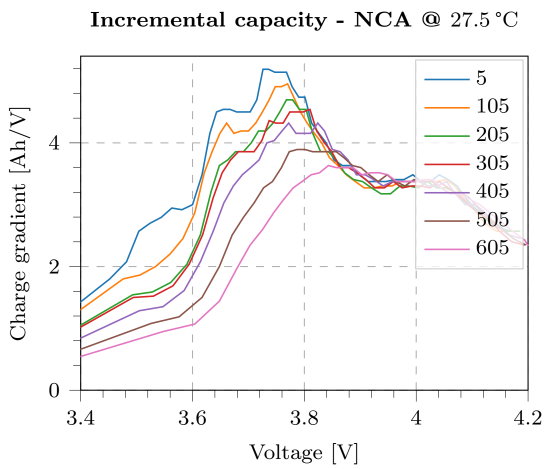
The first test is the same as in A1, that is, trying to make long-term SoH_Q predictions based on a reduced amount of data. The cells used for this approach are the same as in Section 4.1 in order to better compare the results between both methods. Cells' information is presented in Table 4.2 and their IC curves are shown in Figure 5.1 for some of the cycles in the time series. Note that these have been smoothed with a moving average of window size 5 samples. The most relevant feature portrayed in these plots is the existence of peaks and valleys at different points of the charging process, as well as the way they displace and change throughout degradation. This pattern has been used in previous research for SoH_Q estimation [16]. Nonetheless, the approach proposed here simplifies the peak/valley-tracking ones, since it removes the need for implementing a zone-detection algorithm and reduces the processing and memory requirements. Moreover, it relaxes the applicability constraints because it does not require to register the entire charging process.



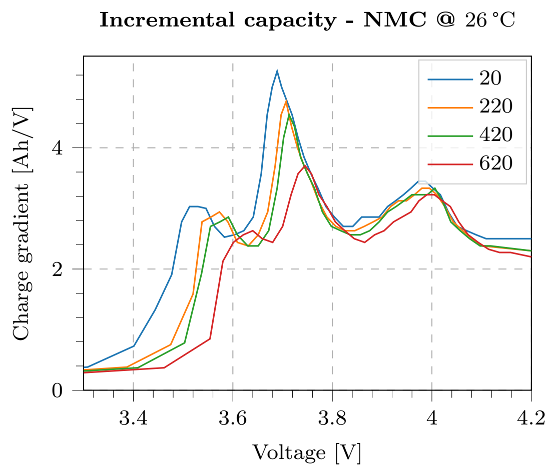
(a) LCO



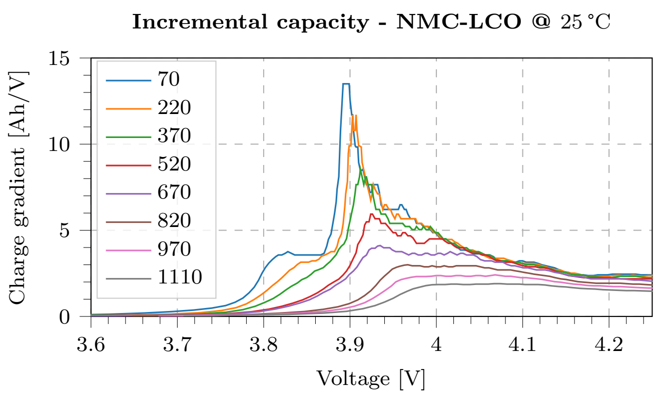
(b) LFP



(c) NCA



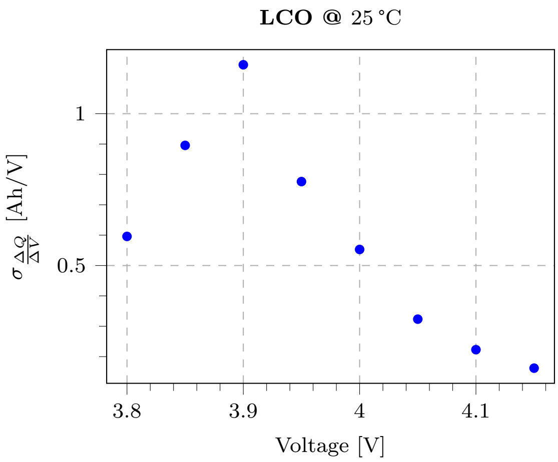
(d) NMC



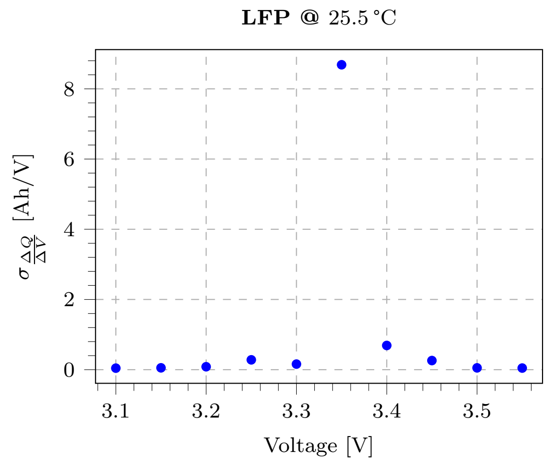
(e) NMC-LCO

Figure 5.1: IC curves evolution with ageing of cells for approach A4

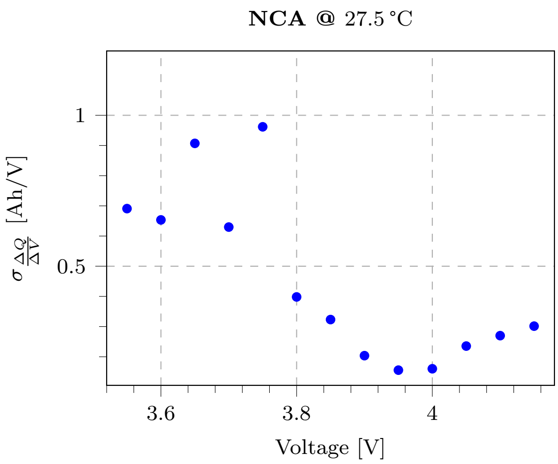
These plots show that all five chemistries follow a similar tendency, that is, the curves displace right- and downwards as degradation becomes larger, and the peaks and valleys become smoother. Displacement is predominant in NCA and NMC cells, while attenuation is more significant in LCO and NMC-LCO ones. The charge gradient in the LFP unit is not severely affected by degradation, except for the attenuation of the second peak at 3.4 V. The interesting feature for this approach is not the magnitude of the charge gradient at a certain voltage, but the magnitude of the variation of such gradient as the cell ages. Thus, the preferred voltage milestones would be those for which the gradient of a pristine cell is very different from that of a severely aged one. In order to further analyse this characteristic, the standard deviation of the charge gradient at 50 mV increments is shown in Figure 5.2. This provides a clearer overview of how the charge gradient varies as the cell degrades. LFP- and NMC-LCO-type cells follow a similar pattern with one voltage level significantly larger than the rest, although in the former case the rest of the values are very close to 0. The peaks are in the expected positions after observing Figures 5.1b and 5.1e. The LCO unit also has an unique peak, but the standard deviations at the surrounding voltage milestones are not far away from it, in accordance with Figure 5.1a. The most complex patterns are those of the NCA and NMC cells. For the former, the distribution is as expected from Figure 5.1c, i.e. peaking between 3.6 V and 3.8 V, and low values thereafter. For the NMC unit, the peaks at 3.5 V and 3.7 V match the ones in Figure 5.1d, although the difference in magnitude between all points is quite small. Based on the analysis of the standard deviation for the different chemistries, it would be expected to obtain good performance when working with the LFP cell, and poor with the NMC one.



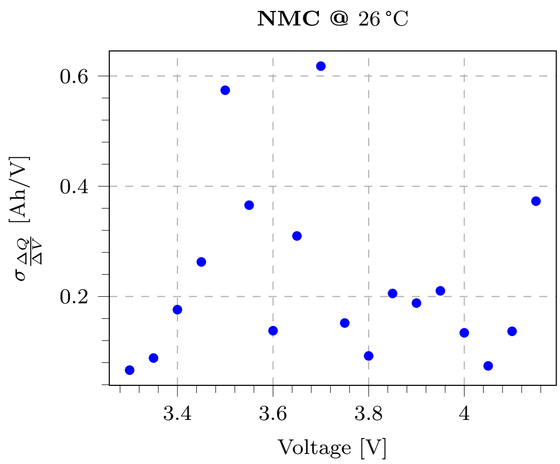
(a) LCO



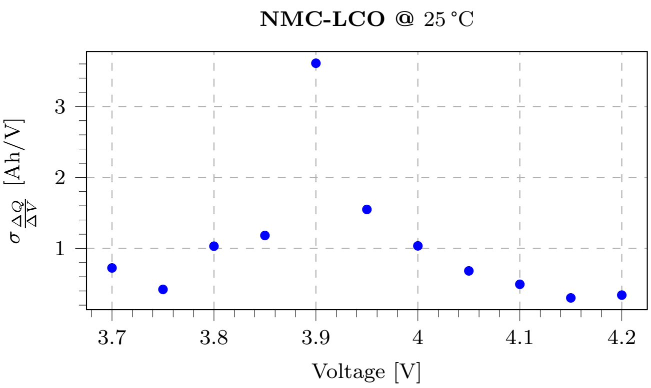
(b) LFP



(c) NCA



(d) NMC



(e) NMC-LCO

Figure 5.2: Standard deviation of charge gradient for cells for approach A4

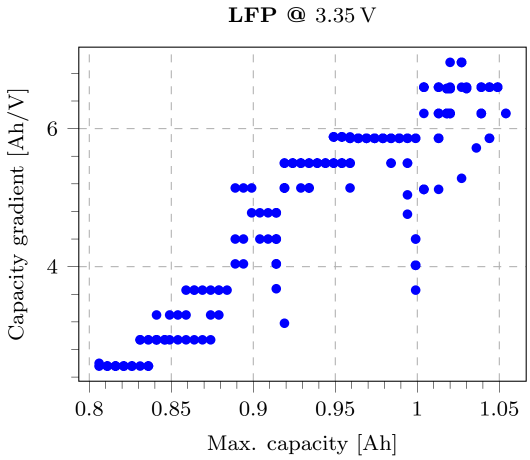
After considering the aforementioned remarks, the chosen voltage milestones for each chemistry are shown in Table 5.1. The OCV-SoC curves were not available in the data set, so it is not possible to find the equivalent SoC at each voltage milestone. The main drawback of this approach is that, in order to reach the most useful voltage milestones, the charging range might need to be extended or it may correspond to not very practical ones, which reduces its applicability. The first value in the list is taken as reference for the rest. Then, the gradient is computed as in Equation (5.1) at each milestone, where Q_i, V_i are the charge and terminal voltage measured at the i$^{\text{th}}$ milestone, respectively, and Q_{ref}, V_{ref} are the charge and voltage measured at the reference milestone. Once all points have been reached, the values are fed as inputs to the AI algorithms. As previously mentioned, the charge increment may be computed by means of any method deemed reliable.

$$x_i = \frac{\Delta Q_i}{\Delta V_i} = \frac{Q_i - Q_{ref}}{V_i - V_{ref}} \tag{5.1}$$

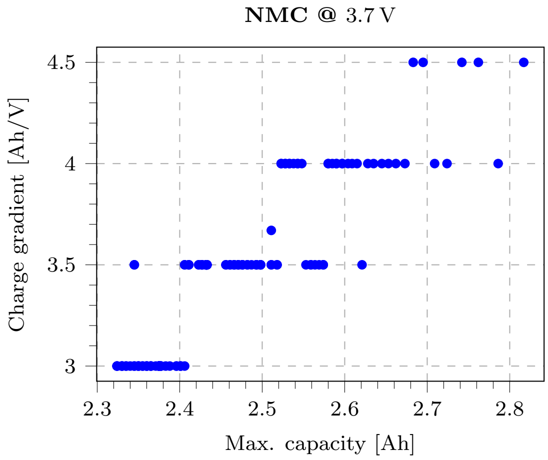
Table 5.1: Voltage milestones for approach A4

Cell	**Voltage milestones**
LCO	3.85 V / 3.9 V / 3.95 V / 4 V
LFP	3.3 V / 3.35 V / 3.4 V / 3.45 V
NCA	3.6 V / 3.7 V / 3.8 V
NMC	3.6 V / 3.7 V
NMC-LCO	3.9 V / 3.95 V / 4 V / 4.05 V

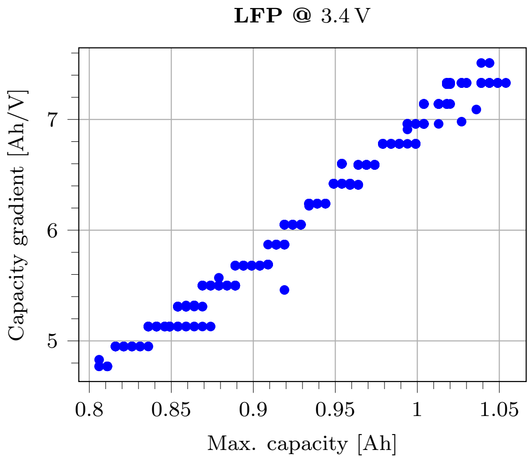
The evolution of the charge capacity gradients with ageing at the chosen voltage milestones are shown in Figure 5.3 for LFP and NMC cells. The LFP clearly follows the expected behaviour based on Figure 5.1b, that is, it decreases with the maximum available capacity. The plot also shows that some noise still remains in the measurements, since there are some significant drops which deviate from the underlying trend. It should be noted that, specially at 3.35 V, the same gradient is obtained for different degradation levels, which could have a negative impact in the algorithms' performance. On the other hand, the NMC cell only presents a clear tendency at 3.7 V and, even so, the same problem of equal gradient for distinct maximum capacity occurs. This is expected from Figures 5.1d and 5.2d, seeing as how the IC curves are close together and have similar slope after this point. This reinforces the expectation of achieving far superior performance for the LFP cell than for the NMC one.



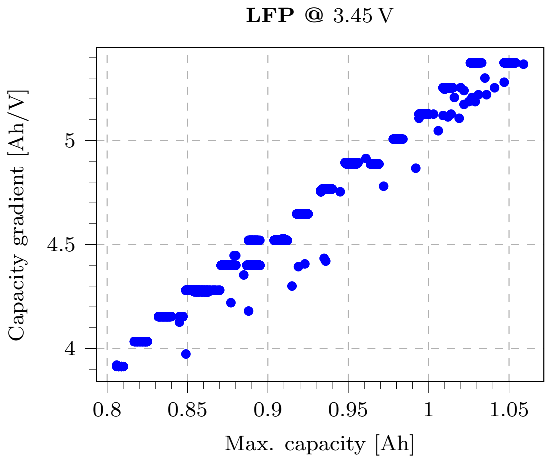
(a) Charge gradient evolution at 3.35 V. LFP cell



(b) Charge gradient evolution at 3.7 V. NMC cell



(c) Charge gradient evolution at 3.4 V. LFP cell



(d) Charge gradient evolution at 3.45 V. LFP cell

Figure 5.3: Charge gradient evolution with ageing of cells for approach A4

The results obtained when testing this approach are summarised in Table 5.2 for both BLS and FC-FNN estimators. They show that the algorithms are capable of providing highly-accurate capacity estimations when a large part of the data set is used for training. As expected, performance for the LFP cell was very high in all cases, although an individualized observation of the algorithms' predictions showed that they had a tendency to overfit the training data set as the amount of training samples was reduced. This is accentuated by the phenomena observed in Figure 5.3. A similar problem occurred with FC-FNN trying to predict the LCO cell's SoH_Q. In the case of the NMC one, this already happened when training on 90 % of the time series, which was expected since only one valid voltage milestone was found, while the BLS kept a fair amount of error until trained on only 25 % of the samples. Accuracy for the NCA device was good in general and only decayed when a very limited part of the data set was used for training. Results were also positive for the NMC-LCO unit, although the error increased significantly when using the earliest 60 % samples. As in previous cases, the BLS was more accurate than the FC-FNN.

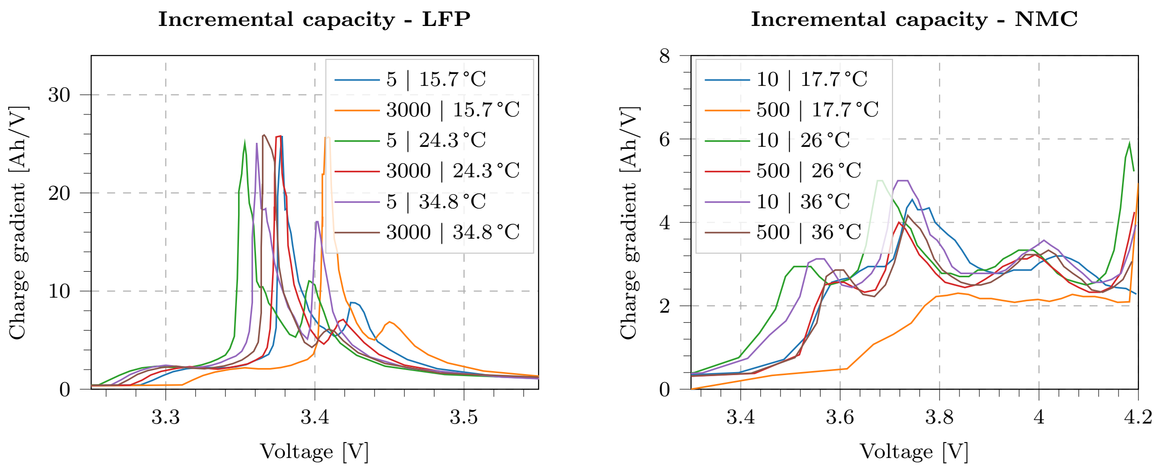
Table 5.2: Results for approach A4

Cell	Structure	Data split	RMSE	MAE	ME	μ_{res}	σ_{res}
LCO	BLS(2-12-8, tanh)	90 %	1.84 %	1.39 %	8 %	0.3 %	1.8 %
	BLS(1-7-6, tanh)	60 %	2.22 %	1.81 %	12.1 %	1.1 %	1.9 %
	BLS(3-8-12, tanh)	25 %	1.78 %	1.24 %	19 %	−0.2 %	1.8 %
	FNN(10-10-10-10, ReLU)	90 %	2.9 %	2.3 %	13.1 %	1.8 %	2.3 %
	FNN(10-10-10-10, ReLU)	60 %	8.51 %	7.24 %	22.1 %	7.2 %	4.5 %
	FNN(10-10-10-10, ReLU)	25 %	12.3 %	8.88 %	36.4 %	8.9 %	8.5 %
LFP	BLS(6-9-6, tanh)	90 %	0.51 %	0.36 %	3.81 %	0.2 %	0.5 %
	BLS(7-12-8, tanh)	60 %	0.91 %	0.67 %	6.9 %	−0.3 %	0.9 %
	BLS(4-12-3, tanh)	25 %	3.44 %	2.91 %	9.67 %	−2.2 %	2.6 %
	FNN(10-100-10-10, ReLU)	90 %	0.92 %	0.8 %	3.38 %	0.8 %	0.5 %
	FNN(10-100-10-10, ReLU)	60 %	0.97 %	0.74 %	5.95 %	−0.4 %	0.9 %
	FNN(10-100-10-10, ReLU)	25 %	3.71 %	3.19 %	9.95 %	−3.1 %	2 %
NCA	BLS(12-12-1, tanh)	90 %	0.85 %	0.6 %	2.36 %	0.1 %	0.9 %
	BLS(7-1-2, tanh)	60 %	1.37 %	1.06 %	3.69 %	0.6 %	1.3 %
	BLS(2-1-1, tanh)	25 %	4.02 %	3.66 %	8 %	3.7 %	1.7 %
	FNN(20-50-50-20, ReLU)	90 %	1.35 %	1.13 %	2.63 %	−0.3 %	1.3 %
	FNN(20-50-50-20, ReLU)	60 %	4.3 %	3.56 %	9.36 %	−3 %	3 %
	FNN(20-50-50-20, ReLU)	25 %	2.62 %	2.18 %	6.36 %	2 %	1.7 %
NMC	BLS(1-1-1, tanh)	90 %	0.33 %	0.22 %	1.73 %	0 %	0.3 %
	BLS(1-2-1, tanh)	60 %	2.02 %	1.6 %	6.55 %	−0.9 %	1.8 %
	BLS(1-4-9, tanh)	25 %	3.53 %	2.94 %	9.65 %	1.1 %	3.3 %
	FNN(10-50-10, ReLU)	90 %	2.07 %	2.06 %	2.8 %	2.1 %	0.2 %
	FNN(10-50-10, ReLU)	60 %	1.65 %	1.25 %	5.65 %	0.8 %	1.4 %
	FNN(10-50-10, ReLU)	25 %	4.72 %	4.47 %	9.24 %	4.5 %	1.5 %
NMC-LCO	BLS(4-7-12, tanh)	90 %	1.49 %	1.24 %	3.69 %	−0.6 %	1.4 %
	BLS(1-1-1, tanh)	60 %	8.36 %	7.45 %	33.1 %	−4.9 %	6.8 %
	BLS(11-1-1, tanh)	25 %	9.83 %	8.59 %	46.3 %	5.2 %	8.4 %
	FNN(10-100-100-20, ReLU)	90 %	3.17 %	2.57 %	6.35 %	2.5 %	2 %
	FNN(10-100-100-20, ReLU)	60 %	24.9 %	23.2 %	42.5 %	23 %	9 %
	FNN(10-100-100-20, ReLU)	25 %	11 %	8.85 %	27.6 %	8.8 %	6.6 %

5.2 Approach A5: charge gradient at voltage milestones with multiple temperatures

Following the same path as with the approaches in Chapter 4, after the charge gradient method has been tested for incomplete data sets in approach A4, the next step is to apply it to scenarios where temperature cannot be considered constant. The cells are the same as for approach A2 and their characteristics are presented in Table 4.4. Figure 5.4 shows the incremental capacity curves for LFP and NMC cells at early and late degradation stages for each one of the available temperatures after being smoothed with a moving average of width

5 samples. For both of them, for the same ageing stage, it can be observed how the curves displace to the right as temperature drifts away from 25 °C. Then, as previously mentioned, degradation pushes the curves to the right and, in the case of the LFP cell, attenuates the second peak. Meanwhile, all peaks become smaller in the NMC one, with a very significant degradation at 17.7 °C, which is coherent with the analysis in Section 2.3



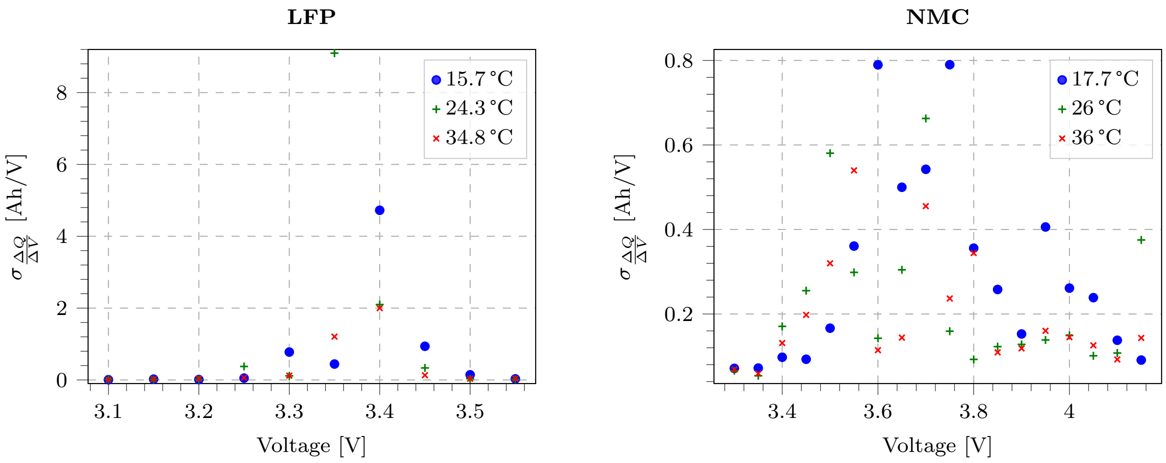
(a) LFP **(b)** NMC

Figure 5.4: Charge gradient evolution with ageing at multiple temperatures of cells for approach A5

Figure 5.5 shows the standard deviation of the charge capacity gradients as the cell ages and at the three available temperatures. The LFP cell maintains a coherent behaviour across all temperatures and presents only peaks at 3.35 V and 3.4 V at 24.3 °C and 15.7 °C, respectively. These match the position of the main peaks in Figure 5.4a. The differences are more noticeable in the NMC case, specially when it comes to the peaks at 3.6 V and 3.75 V, which are only present at 17.7 °C. For the other temperatures, the major differences are located at 3.5 V and 3.7 V for 26 °C, and at 3.55 V and 3.7 V for 36 °C, which is also coherent with the curves in Figure 5.4b. As before, the lack of a clear pattern or large differences may lead to bad estimation results when testing the algorithms. Based on these observations, the voltage milestones are chosen as those in Table 5.3.

Table 5.3: Voltage milestones for approach A5

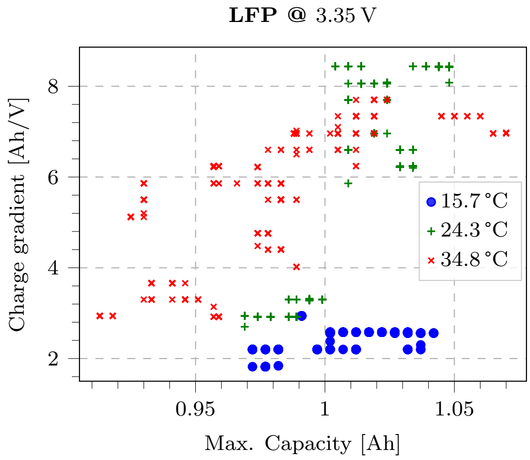
Cell	Voltage milestones
LFP	3.3 V / 3.35 V / 3.4 V / 3.45 V
NCA	3.6 V / 3.7 V / 3.8 V
NMC	3.6 V / 3.7 V / 3.8 V



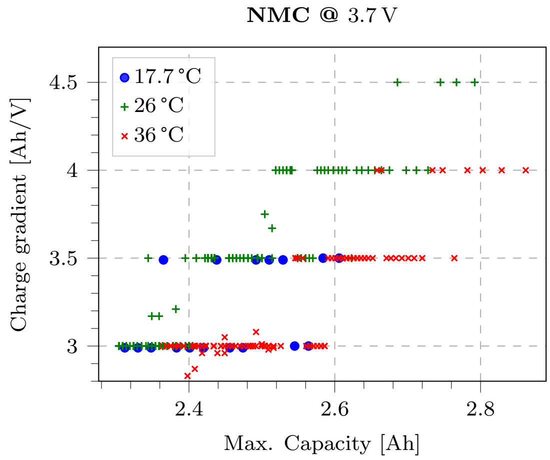
(a) LFP **(b)** NMC

Figure 5.5: Standard deviation of charge gradient through ageing at multiple temperatures of cells for approach A5

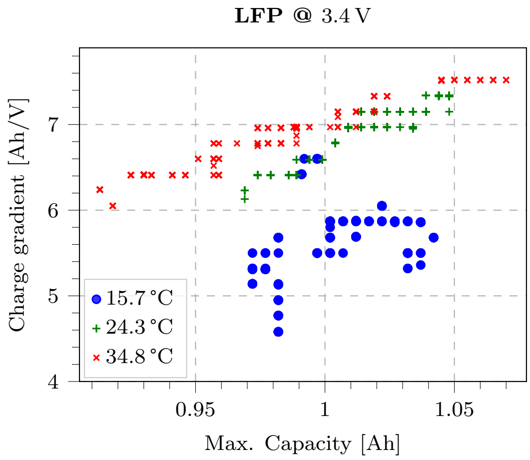
The evolution of the charge gradients at the chosen milestones is shown in Figure 5.6. As before, the LFP cell presents a consistent behaviour across the three temperatures, which follow the expected trend of decreasing gradient with ageing. This change is less pronounced at 15.7 °C, as it could be inferred from the previous analysis. The differences between temperatures are most significant at 3.4 V, while they overlap each other considerably at 3.45 V. For the NMC cell, the gradient shows a very clear general tendency at any temperature, although the fact that the points for all three temperatures overlap with each other may make it difficult for the AI systems to learn and make accurate predictions in changing conditions. Thus, it would be expected to obtain considerably better results for the LFP cell than for the NMC one.



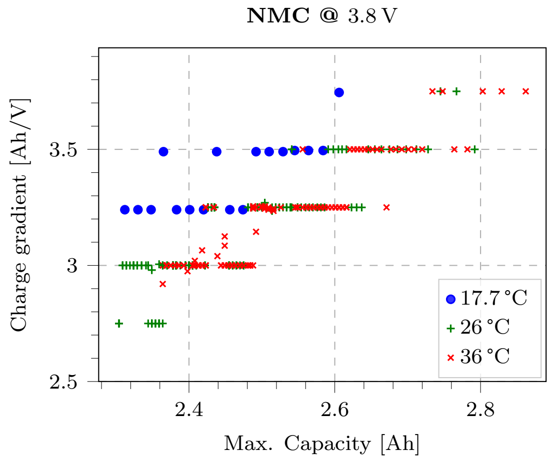
(a) Charge gradient evolution at 3.35 V. LFP cell



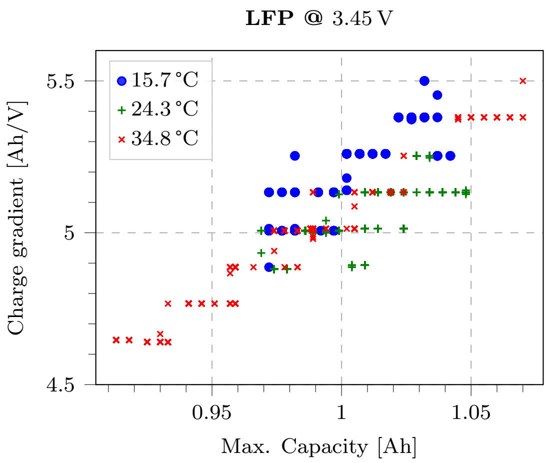
(b) Charge gradient evolution at 3.7 V. NMC cell



(c) Charge gradient evolution at 3.4 V. LFP cell



(d) Charge gradient evolution at 3.8 V. NMC cell



(e) Charge gradient evolution at 3.45 V. LFP cell

Figure 5.6: Charge gradient evolution with ageing at multiple temperatures of cells for approach A5

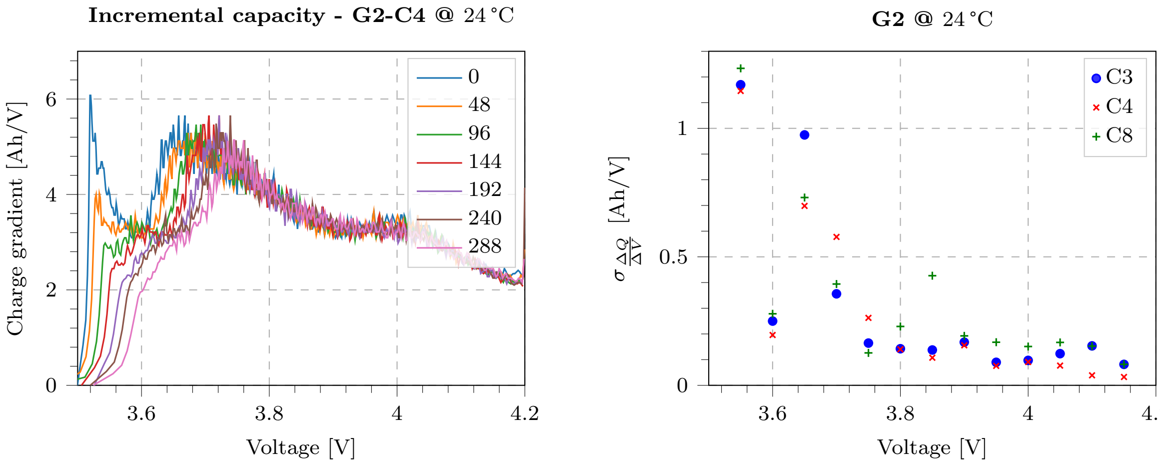
The algorithms were trained on data at 15 °C and 35 °C, then tested at 25 °C. Results for this approach are presented in Table 5.4. The best performance was obtained by the BLS in all three scenarios. Overall accuracy was high, having the LFP cell the best results. Nevertheless, the FC-FNN algorithm showed some overfitting for the NCA and NMC tests when predictions were analysed individually. These results match the expectations, being the higher errors related to the fact that the charge gradient did not show significant differences between temperatures in the analysis. Then again, trying to learn the impact of temperature by using only two experiments is very complicated, specially if this influence is not linear or monotonic.

Table 5.4: Results of approach A5

Cell	Structure	RMSE	MAE	ME	μ_{res}	σ_{res}
LFP	BLS(3-1-4, tanh)	0.64 %	0.49 %	1.81 %	−0.1 %	0.6 %
	FNN(300-300-30, ReLU)	1.51 %	1.34 %	3.36 %	−1 %	1.2 %
NCA	BLS(8-12-4, tanh)	2.16 %	1.76 %	5.96 %	0.1 %	2.2 %
	FNN(300-300-30, ReLU)	3.23 %	2.59 %	9 %	2 %	2.5 %
NMC	BLS(9-10-2, tanh)	1.93 %	1.65 %	5.08 %	−0.3 %	1.9 %
	FNN(300-300-30, ReLU)	4.13 %	3.42 %	8.58 %	−3.3 %	2.5 %

5.3 Approach A6: charge gradient at voltage milestones with realistic ageing

For this last approach, the intelligent algorithm is tasked with predicting the maximum available capacity based on the measured charge increments at fixed voltage milestones, similar to approaches A4 and A5, but using the hybrid ageing data set (see Table 4.6). Figure 5.7a shows the IC curves for one of the cells, smoothed with a moving average of window size 2 samples. This plot has a higher voltage resolution than the one obtained from the previous data set for a similar cell (see Figure 5.1c), however there is also slightly more noise in the measurements. The tendency of the curves is to attenuate the first peak and to displace the second one rightwards as the number of cycles increases, although beyond 3.8 V all curves overlap. This is in accordance with the standard deviation analysis of the charge gradient in Figure 5.7b, where the points with highest dispersion are 3.55 V and 3.65 V. It can also be observed how all three cells in the same group present similar evolution of the charge gradient, which remarks that cell-to-cell differences are negligible.

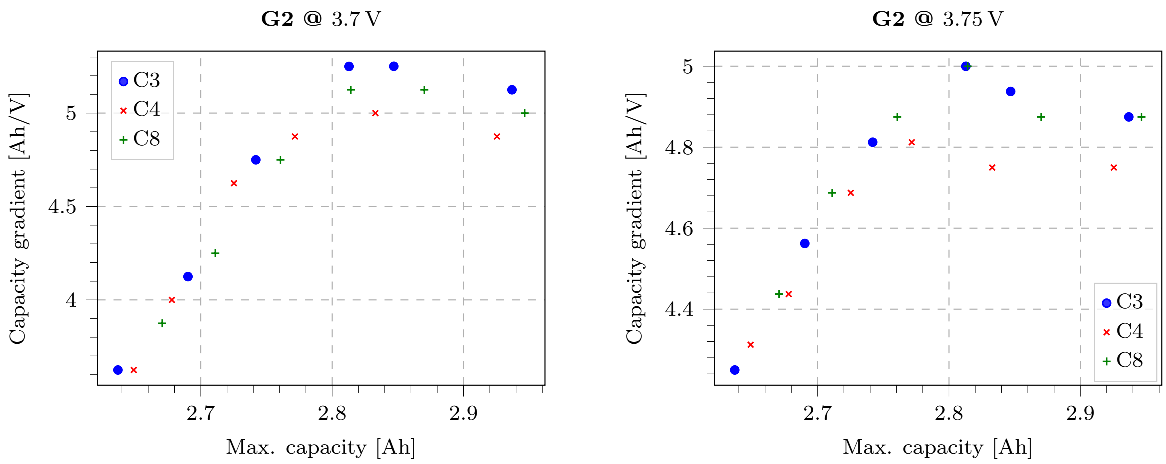


(a) IC curves through ageing

(b) Standard deviation of $\Delta Q/\Delta V$ through ageing

Figure 5.7: IC curves analysis of cell for approach A6

Based on these plots, the reference voltage milestone is taken as 3.65 V, and the measuring ones are 3.7 V and 3.75 V, which correspond to 48 %, 58 % and 62 % SoC, respectively, for a pristine cell (see Figure 2.5). This means a total range of approximately $\Delta\text{SoC} = 14\,\%$, which is shorter than the ones used in Chapter 4. The evolution of the charge gradient at each of the milestones is shown in Figure 5.8 for all cells in group G2. At both 3.7 V and 3.75 V, it decreases as the device loses capacity, with a steeper slope below 2.8 A h. However, the gap between a pristine and an aged cell is more significant at the first milestone. This behaviour is in accordance with what was depicted in Figure 5.7a. Once again it becomes clear that no relevant cell-to-cell differences exist since all of them follow a similar pattern.



(a) Charge gradient evolution at 3.7 V

(b) Charge gradient evolution at 3.75 V

Figure 5.8: Charge gradient evolution with ageing of cells for approach A6

The same two scenarios as for approach A3 are used here, that is, a first scenario where data from two types of ageing are used for training, and a second one where only one of the groups is used for training and the other three are reserved for testing the algorithms. The results of this approach are presented on Table 5.5. It is clear that both algorithms perform quite well in any of the two designed scenarios. It is also interesting to note that similar accuracy was achieved in any of them, which points towards the hypothesis that the specific conditions in which the cells are aged do not have a severe influence on the effects of degradation, at least when it comes to hybrid ageing. As a final remark, it should also be taken into consideration that this data set contained a very small amount of data, which definitely had a negative impact on performance.

Table 5.5: Results of approach A6

Training	Testing	Structure	RMSE	MAE	ME	μ_{res}	σ_{res}
G2, G3	G1, G4	BLS(1-2-4, tanh)	3.02 %	2.5 %	5.82 %	−0.4 %	3 %
		FNN(20-100-20-10, ReLU)	3.53 %	3.16 %	5.88 %	1.1 %	3.4 %
G3	G1, G2, G4	BLS(5-1-5, tanh)	2.79 %	2.34 %	5.66 %	−0.3 %	2.8 %
		FNN(20-100-20-10, ReLU)	3.2 %	2.67 %	6.28 %	0.5 %	3.2 %

Chapter 6

Impedance-based State-of-Health Estimation

In this chapter, an intelligent method to simultaneously estimate SoH_Q and SoH_R based on a very small amount of impedance measurements is developed and then tested on cells undergoing storage ageing.

Contents

One of the main drawbacks of the previously presented approaches is that they require the vehicle to be stationed at a charger and that the charging process covers the pre-established SoC or voltage milestones. In real-world applications, this may not happen each and every time the battery is being recharged. For those scenarios in which a faster, more flexible method for monitoring the battery's health is required, an estimation algorithm based on impedance measurements is developed here. The advantages over the previous methods are that it can be executed online, and that the amount of time required to perform the measurements can be drastically reduced. Furthermore, the algorithm is capable of providing estimations of degradation for both the battery's maximum available capacity ($\mathrm{SoH_Q}$) as well as its internal ohmic resistance ($\mathrm{SoH_R}$).

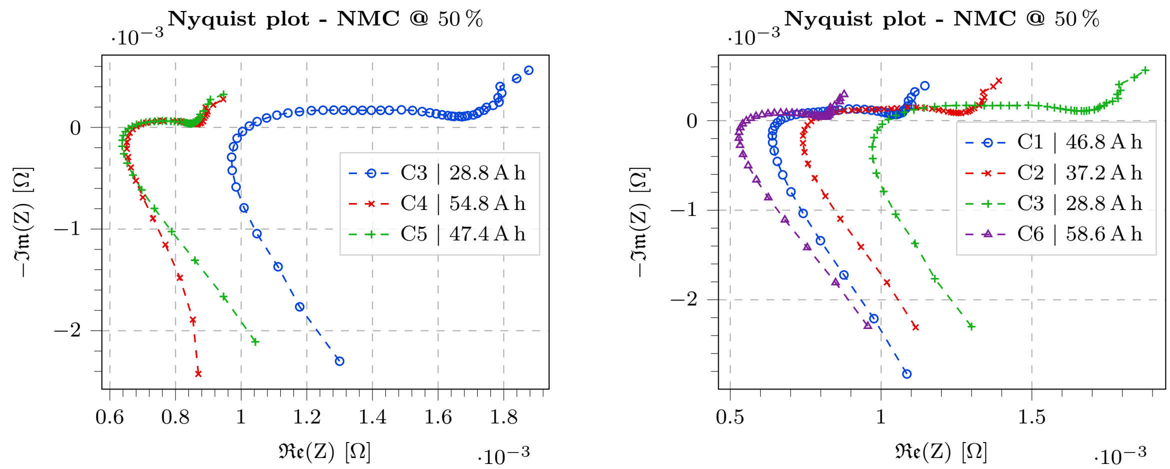
6.1 Approach A7: impedance measurements at limited frequency points

In order to design and test this method, data from cells aged in different storage conditions were used, as presented in Table 6.1. Unfortunately, these data are part of *Aalborg University E-Mobility and Industrial Drives* research group and are not publicly available, although papers describing and using the data set have been published in the past [14], [16]. This arrangement allows for analysing the impact of both SoC level and ambient temperature while in storage. The cells were taken out to undergo reference performance tests approximately every month over a total period of 33 months. These included, among others, EIS and CCCV charge and discharge experiments, all of them conducted trying to keep a constant temperature of 25 °C. The impedance measurements were conducted at 10 %, 50 % and 90 % SoC, and it has been shown by the original authors that this factor does not have a significant influence and, thus, just measurements at 50 % can be used for simplification [14]. For this reason, combined with he fact that, in realistic scenarios, it would be expected to be at 50 % SoC more often than at the rest, only these measurements will be used in the remainder of the section. However, it would be possible to develop similar approaches for any given SoC level. On the other hand, the lack of EIS measurements at multiple temperatures makes it impossible to quantify the impact of this parameter on impedance, so the approach presented here is only valid for estimation at the same temperature as in the original data set. This is not ideal, since it was shown in Section 2.2.2 that temperature has a sensible effect on cell impedance and should be included in the design of experiments and algorithms.

Table 6.1: Storage conditions and characteristics of cells used for approach A7 [14], [16]

Cell	C1	C2	C3	C4	C5	C6	**Chemistry**	NMC
SoC	50 %	50 %	50 %	10 %	90 %	50 %	**Cut-off voltage**	3 V / 4.125 V
Temperature	35 °C	40 °C	45 °C	45 °C	45 °C	7 °C	**Rated capacity**	63 A h

Based on the collected data, it is possible to analyse how cells stored at different temperatures and SoC levels age. A capacity degradation-focused analysis of these data was previously done in Section 2.3.4, while impedance is the variable of interest in this section. Figure 6.1 shows the Nyquist plots of severely-aged cells (around 33 months) which were stored at the same temperature and multiple SoC levels (see Figure 6.1a), and at the same SoC and multiple temperatures (see Figure 6.1b). It can be observed that, when it comes to SoC, degradation is much more significant at intermediate levels (C3) than when the battery is almost entirely depleted (C4) or charged (C5), while the differences are minimal at the extremes. For cells stored at different temperatures, capacity loss is worse for those which underwent higher temperatures (C3). Furthermore, the plots also show that, in general, cell degradation leads to increased impedance's real part at all frequencies or, in other words, a rightwards displacement of the Nyquist plots. These observations are coherent with those discussed in Section 2.3.4.

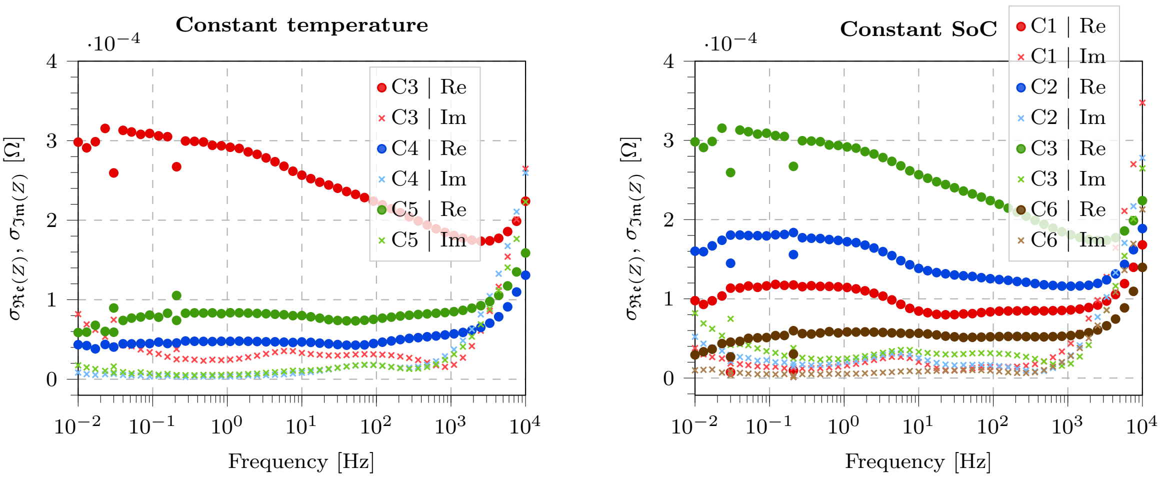


(a) SoC influence (45 °C)

(b) Temperature influence (50 % SoC)

Figure 6.1: Effects of long-term storage ageing at constant SoC and constant temperature

The choice of frequency points at which to measure the cell's impedance is a critical part of this approach. Using low-frequency excitation signals would largely extend the process' duration, while high-frequency ones would increase the magnitude of the inductive effects of the wires and could make noise-filtering more difficult [1, p. 267]. At the same time, the amount of points should be kept as small as possible so that the estimation can be carried out faster and the computing requirements are low. Moreover, it is necessary to analyse if the effects of degradation are more significant at some frequencies than others, as points with large variations would be preferred. This is done by computing the standard deviation of the complex impedance at each frequency throughout the entire ageing process of each of the cells, as shown in Figure 6.2. It is clear that the imaginary part does not significantly change with ageing except at very high frequencies, no matter the ageing procedure. In terms of real-part variation, it is larger for the cell stored at 50 % (C3) when looking at the constant-temperature data, while it is higher for the one kept at 45 °C (C3) when comparing data at the same SoC. These observations match the ones outlined before. Based on these considerations it was chosen to use the following frequencies as milestones: 10 Hz, 20 Hz, 50 Hz, 100 Hz, 200 Hz, 500 Hz and 1 kHz. It must be noted that these are approximated values, since they do not perfectly match the ones present in the data set. Exactness is not a critical aspect, as long as the frequencies used during training are the same ones as in the real application. Impedance may be measured by means of a regular EIS test or, to save time, with any of the random sequence-based methods discussed in Section 1.2.



(a) SoC influence (45 °C)

(b) Temperature influence (50 % SoC)

Figure 6.2: Standard deviation of impedance spectrum with ageing

In order to validate the proposed approach, the intelligent algorithms were tested on two scenarios. In the first one, data from four cells were used for training and the rest for testing. The training data selection here is based on using different temperatures and SoC levels, giving priority to the first ones since they seemed to have a bigger impact on ageing and more diverse situations were considered. This aims at verifying if, by only using data from a very limited set of conditions, it is possible to predict the SoH of cells stored in different ones. In the second one, the algorithms were trained on early degradation data and then they were tasked with predicting the SoH at late stages. Thus, this one investigates how accurate estimations are when only early data are available. There were however problems with the testing equipment and data between 17 and 24 months were missing. The training configuration for the intelligent algorithms is presented in Table 6.2 and the obtained results are summarised in Table 6.3.

Table 6.2: FC-FNN configuration parameters for approach A7

Early stopping	Batch size	Optimizer	Epochs	Loss	Train / Validation split
250 epochs	5 %, min. 5 samples	Adam	5000	MSE	70 % / 30 %

Table 6.3: Results of approach A7

Train	Test	Structure	Variable	RMSE	MAE	ME	μ_{res}	σ_{res}
C1-C4	C5-C6	BLS(16-19-1, tanh)	SoH_Q	1.2 %	1 %	2.7 %	0.6 %	1.1 %
0-33 M[1]	0-33 M[1]		SoH_R	0.3 %	0.2 %	0.5 %	−0.1 %	0.2 %
		FNN(200-200-200-	SoH_Q	2.2 %	1.9 %	3.2 %	1.1 %	1.9 %
		200-30, ReLU)	SoH_R	0.5 %	0.4 %	0.7 %	0.4 %	0.7 %
C1-C6	C1-C6	BLS(10-18-3, tanh)	SoH_Q	1.1 %	0.9 %	1.8 %	−0.3 %	1 %
0-26 M[1]	27-33 M[1]		SoH_R	0.3 %	0.3 %	0.5 %	0 %	0.3 %
		FNN(100-100-100-	SoH_Q	2.1 %	1.6 %	3.8 %	−1.1 %	1.8 %
		100-30, ReLU)	SoH_R	0.9 %	0.8 %	1.4 %	−0.8 %	0.3 %

Results show that the algorithms can accurately estimate both SoH_Q and SoH_R, specially the latter one, in any of the scenarios. During testing it was observed that modifying the algorithm so that it only estimated SoH_Q led to even higher accuracy, specially for the FC-FNN-based methods. This is due to the fact that this one takes into account the MSE of both outputs to update the weights during training and, since the resistance's error proved to decrease much faster than the capacity one, the training ended when the latter was still relatively high. A possible solution to this problem would be to choose the excitation frequencies so that they are close to the cross-over point, and then performing some form of interpolation to obtain the SoH_R. In the specific case of the NMC cells employed in this approach, the intersection with $\mathfrak{Im}(Z) = 0$ occurred in the range from 200 Hz to 400 Hz (see Figure 2.9b) and, therefore, there would be no need for additional frequency measurements. Another possibility would be to reconstruct the entire impedance spectrum using machine learning, as shown in Section 1.2.
It was also observed that similar accuracy was obtained when using data at 10 % and 90 % SoC, which implies that, potentially, the algorithm could be applied to any given charge level. Lastly, data scarcity definitely had a negative impact on the system's performance, as a significant part of the time series was not available. As a final remark, no specific methodology has been established for impedance measurements since many fast and reliable approaches have already been proposed in literature, as discussed in Section 1.2.

This was the final method developed in this thesis and, thus, the practical aspects of this report come to an end here. The next chapter contains a discussion regarding methodologies and results.

[1]Data from months 17 to 24 were not available.

Chapter 7

Discussion

Throughout this report a total of seven machine learning-based approaches for SoH estimation have been proposed. They can be split into three types, each of them based on different considerations and with different requisites:

1. The first one (A1, A2, A3; see Chapter 4) is SoC-based, that is, terminal voltage is measured at certain SoC levels and then these data are used as inputs for the estimator. These SoC milestones were chosen based on applicability rather than potential performance. In other words, rather than analysing which SoC levels would show more differences with ageing, intermediate values (40 %, 50 % and 60 %) were chosen instead, as it was considered that these would be more frequent in real EV applications. No specific SoC estimation algorithm was designed for this approach, but plenty of reliable methods are available in the recent literature [13], [52]–[54].

2. The second one (A4, A5, A6; see Chapter 5) is voltage-based, which means that charge increments are measured between some pre-established voltage points. Then, charge gradients are computed based on these increments and used as inputs for the intelligent algorithms. In this case, an individualised analysis of each chemistry was performed in order to find those voltage milestones at which the charge gradient showed a larger variation as the cell deteriorated. When choosing the voltage milestones, the ones further away from the cut-off voltages were preferred to try and make the method more applicable. One of the main advantages of this approach was that it did not require the existence of an SoC estimator, but only accurate charge measurements.

3. The last type (A7; see Chapter 6) is based on impedance spectrum and attempts to estimate both SoH_Q and SoH_R from a very limited amount of frequency measurements. The improvement of this method over the previous two is that it can truly be executed online, rather than demanding the car to be connected to a charging station. On the other hand, the measurement process becomes slightly more complex, since injecting a series of excitation signals at a number of frequencies is necessary for computing the cell's impedance. Moreover, the intelligent algorithms were trained to work with measurements at a certain SoC and temperature and, thus, the vehicle would need to be in some specific conditions prior to conducting the SoH estimation. As with the second type, the selection of frequencies was based on a variation analysis through degradation and, at the same time, very low and high ones were avoided in order to reduce execution time and inductive effects from the wires, respectively. No specific method for signal injection was derived, although random sequence-based ones [8]–[10] are recommended over single-frequency excitation so that impedance measurements are made faster.

In order to test the accuracy of the methods proposed in Chapters 4 to 6, several different scenarios were considered. The aim of these experiments was to verify if the AI-based methods were capable of providing accurate estimations when facing new, demanding conditions. The best results for each scenario are summarised in Table 7.1.

1. First (A1, A4; see Sections 4.1 and 5.1), a situation in which the reference cells have not yet been cycled until EoL in the laboratory, as a way of simulating an incomplete lifetime data set. The algorithms were trained on progressively early stages of the data set and then tested on the later ones. Good all-round accuracy was obtained for most of the cells in both approaches, with the exception of e.g. the LFP cell in A1, which showed overfitting problems. This was caused by the voltage curves being very close to each other throughout ageing. In most cases, training the algorithms on increasingly earlier stages led to performance deterioration, which is a natural consequence of the reduction in training samples and the increased differences with the test data set.

2. Secondly (A2, A5; see Sections 4.2 and 5.2), the case where temperature is taken into consideration as a parameter conditioning cell behaviour. The estimators were trained on two temperatures and tested for prediction at a different one. Here, for example, it was observed how, for the LFP cell, the results of the SoC-based method were not good, but they were the best when using the charge gradient-based one. This supports the idea that each of the methods may be a better option depending on the battery's specific chemistry. The employed data set, however, was quite limited in terms of temperature diversity as only three levels were available. This severely conditioned the algorithms' capabilities, since they had to learn the general effect of temperature based on just two of them and, as observed in Chapter 2, this relationship can be rather complex. It would be interesting to conduct these experiments over a wider range and to evaluate again the goodness of the proposed approaches.

3. For the third one (A3, A6; see Sections 4.3 and 5.3), data from realistic hybrid ageing procedures were used to assess the accuracy of both types of approaches. Out of a total of four available ageing groups, the algorithms were trained on either one or two of them, and then tested on the other ones. The good overall results corroborated that the intelligent algorithms were capable of estimating the SoH_Q even when tasked with cells aged in different ways. It was observed how the SoC-based approach was significantly better than the voltage-based one, and that performance did not largely worsen if only one of the ageing groups was used for training. A key limiting factor in this case was the relative scarcity of data, as RPT were only conducted once every many cycles.

4. Since it required a different type of measurements, approach A7 (see Section 6.1) could not be applied to the aforementioned scenarios. Instead, two types of situations with incomplete data were investigated: one in which just a few cases of storage ageing were available, in order to explore whether the approach could be trained on limited scenarios and then used for general predictions, and another one in which only data from early degradation stages was used for training, to verify if the algorithms could make good predictions for future events. Quite low error was obtained in both of them, with no signs of overfitting. Predictions when training and testing at other SoC levels were also accurate, which implies that this approach could potentially be applied to any SoC. Nevertheless, the reduced size of the data set and, more importantly, the absence of

several months' data definitely burdened the ability of the methods to provide accurate long-term predictions.

Table 7.1: Summary of results of SoH estimation algorithms

Case	Structure	RMSE	MAE	ME	μ_{res}	σ_{res}
A1, LCO, Train: 90 %	BLS(1-6-3, tanh)	5.2 %	3.5 %	13.8 %	1.6 %	4.9 %
A1, LFP, Train: 90 %	BLS(1-2-2, tanh)	3.1 %	2.3 %	18.2 %	0 %	3.1 %
A1, NCA, Train: 90 %	BLS(1-1-12, tanh)	0.9 %	0.7 %	3.3 %	−0.1 %	0.9 %
A1, NMC, Train: 90 %	BLS(1-4-2, tanh)	0.5 %	0.4 %	0.8 %	0 %	0.5 %
A1, NMC-LCO, Train: 25 %	BLS(3-1-8, tanh)	4.7 %	2.9 %	33.9 %	0.9 %	4.6 %
A2, LFP, Train: 15 °C, 35 °C	BLS(12-4-1, tanh)	1.1 %	0.8 %	6.5 %	0 %	1.1 %
A2, NCA, Train: 15 °C, 35 °C	BLS(1-4-4, tanh)	1.8 %	1.4 %	5.6 %	−0.5 %	1.7 %
A2, NMC, Train: 15 °C, 35 °C	BLS(1-2-2, tanh)	1.4 %	1.1 %	3.4 %	0.6 %	1.3 %
A3, Train: G2, G3	BLS(1-5-1, tanh)	0.5 %	0.4 %	1.2 %	0 %	0.5 %
A4, LCO, Train: 90 %	BLS(2-12-8, tanh)	1.8 %	1.4 %	8 %	0.3 %	1.8 %
A4, LFP, Train: 90 %	BLS(6-9-6, tanh)	0.5 %	0.4 %	3.8 %	0.2 %	0.5 %
A4, NCA, Train: 90 %	BLS(12-12-1, tanh)	0.9 %	0.6 %	2.4 %	0.1 %	0.9 %
A4, NMC, Train: 90 %	BLS(1-1-1, tanh)	0.3 %	0.2 %	1.7 %	0 %	0.3 %
A4, NMC-LCO, Train: 90 %	BLS(4-7-12, tanh)	1.5 %	1.2 %	3.7 %	−0.6 %	1.4 %
A5, LFP, Train: 15 °C, 35 °C	BLS(3-1-4, tanh)	0.6 %	0.5 %	1.8 %	−0.1 %	0.6 %
A5, NCA, Train: 15 °C, 35 °C	BLS(8-12-4, tanh)	2.2 %	1.8 %	6 %	0.1 %	2.2 %
A5, NMC, Train: 15 °C, 35 °C	BLS(9-10-2, tanh)	1.9 %	1.7 %	5.1 %	−0.3 %	1.9 %
A6, Train: G3	BLS(5-1-5, tanh)	2.8 %	2.3 %	5.7 %	−0.3 %	2.8 %
A7, Train: C1-C4, 0-33 M	BLS(16-19-1, tanh)	1.2 %	1 %	2.7 %	0.6 %	1.1 %
		0.3 %	0.2 %	0.5 %	−0.1 %	0.2 %
A7, Train: C1-C6, 0-26 M	BLS(10-18-3, tanh)	1.1 %	0.9 %	1.8 %	−0.3 %	1 %
		0.3 %	0.3 %	0.5 %	0 %	0.3 %

While designing the tests, it was decided to make them quite demanding in order to find how powerful the algorithms were in challenging situations. Although it has not been stated in this report, the performance on the validation data sets, which are extracted from the same group as the training data, was always significantly better than that obtained on the test data set. This implies that, even though the proposed methods were good in new situations, they were even better when operating in situations similar to those they were trained on. On a similar consideration, it was chosen to keep the data as raw as possible in order to have a closer situation to real-world conditions. Thus, only evident outliers or erroneous measurements were taken out of the data sets. Most likely, a more thorough pre-processing would have considerably improved the results, although in that case the problem of distinguishing if information of only noise is being removed arises. The same applies to the amount of available data for training, which was very reduced for some of the approaches.

As a final remark, the BLS showed higher performance than the FC-FNN in all cases, which proves its potential as an ML algorithm. This, combined with the fact that it contained less parameters and that the training process was much faster, are very relevant arguments to choose this strategy when deploying it in real-world applications.

To provide some context for the previous results, Table 7.2 summarises some of the best SoH estimation methods found in Section 1.2. This should not be taken as a direct comparison of the algorithms, since the data, the pre-processing and the implementation methodologies differ for each of them. This shows that all seven approaches proposed in this thesis can provide similar or superior accuracy than already existing methods, both data- and model-based. Furthermore, the algorithms in this project have been tested for more diverse cell types, demonstrating good applicability for all of them, as well as for combined, more realistic ageing procedures.

Table 7.2: Summary of capacity-based SoH estimation methods from state-of-the-art review

Method	Observations	Results
SVM [55]	Single chemistry. Multiple temperatures. CCCV and driving-cycle ageing. No long-term degradation	MSE: 0.4 %
LSTM-RNN [19]	Single chemistry. Multiple temperatures. CCCV and driving-cycle ageing. Long-term degradation.	MAE: 2.5 %
Peak/valley tracking in IC curves [16]	LMO and NMC chemistries. Calendar ageing at multiple temperatures and SoC levels.	MSE: 1.8 %
Partial charging [15]	NMC chemistry. Calendar and cycling ageing at multiple temperatures and SoC levels.	MSE: 2.3 %

The approaches proposed throughout this thesis should not be considered as a final solution, since plenty of improvements are still possible and necessary, as discussed in Chapter 9. These affect both the methodologies themselves and the validation experiments. The next chapter draws some relevant conclusions regarding the entire work presented in this report.

Chapter 8

Conclusion

As stated in Section 1.4, the purpose of this thesis was to investigate if machine learning algorithms could be applied for online SoH estimation of batteries in electric vehicles. For this purpose, several strategies were designed based on different considerations and types of measurements. Then, they were evaluated in different scenarios trying to cover several diverse ageing procedures and cell chemistries. The results proved that, indeed, all the algorithms can accurately estimate a Li-ion cell's maximum available capacity even when tasked with challenging situations.

This thesis makes several contributions to the topic of SoH estimation of Li-ion batteries. First, a very thorough analysis of the effects of several factors related to cell degradation was done in Chapter 2, along with a study of the main reference performance tests and the way they are affected by the experiment's conditions. Then, three completely novel approaches were developed in Chapters 4 to 6 based on various effects of cell degradation, which had been identified in the previous analysis. All methods can be applied online, but without the need of completing an entire charging cycle nor wide-range, high-resolution impedance spectrum measurements. Then, it was shown how these methods were capable of estimating a cell's maximum available capacity when making long-term predictions based only on early data, when operating at temperatures other than the ones used in the laboratory tests, and when applied to cells which had undergone different types of ageing.

Lastly, this work is one of the few to consider several types of Li-ion cells, it constitutes the first attempt at applying the novel BLS algorithm to battery management systems, and also the first time that the majority of publicly available data sets are explored and compared in depth.

Chapter 9

Future work

Battery pack validation The proposed approaches have been applied to individual Li-ion cells, rather than large battery packs. According to some literature in Section 1.2, battery-pack curves may present some differences with those of single cells. It would therefore be necessary to evaluate the performance of the algorithms for a real EV battery.

Large-scale data set In order to develop algorithms which can operate in very diverse scenarios it would be necessary to build a reliable data set which contains RPT data from several units of different chemistries, cycled at different C-rates, temperatures, DoD, but also realistic driving cycles and calendar ageing in different conditions. For example, no data were available for LMO nor LTO cells, and some methods could only be tested on a single type.

Multi-temperature SoH_Q interpolation Some of the proposed approaches have limited applicability in terms of ambient conditions. For example, methods proposed in Chapters 4 and 5 provide an estimate of the maximum available capacity at a certain temperature, but cannot extrapolate it to a different one. However, this could be possible if consecutive RPT were conducted in different conditions at certain points of a cell's ageing process.

Realistic SoC milestones If real data from user habits were available, these could be used to choose the most common SoC interval for the approaches A1 and A4 to further increase applicability. Another option would be to obtain several models with different SoC milestones and then using the one corresponding to the specific case. Nevertheless, this could largely increase the computation requirements of the algorithm.

BLS smart structure selection The BLS algorithms for this application did not require a lot of nodes and, therefore, a grid-search procedure was efficient enough. If, however, the problem had been more complex and required many nodes, it would have been interesting to develop some intelligent or adaptive method to iterate between promising structures in order to save training time.

Publicly available data Due to their intrinsic characteristics, Li-ion-cells testing for degradation is a process that may take up to several years and require costly equipment. To favour the advancement of this field, research institutions and companies should consider making their data sets publicly available so that more researchers can use them to develop better methodologies.

Test for different DoD or C-rate Although the algorithms have been extensively tested for different ageing conditions and procedures, they have not been applied to other relevant factors such as multiple DoD or C-rate. Since the algorithms learn the effects of degradation rather than its causes, it would be expected for them to work well on these situations too.

References

[1] A. Emadi, *Advanced Electric Drive Vehicles*. CRC Press, 2015, ISBN: 978-1-4665-9770-9.

[2] S. S. Williamson, *Energy Management Strategies for Electric and Plug-in Hybrid Electric Vehicles*. Springer New York, 2013, pp. 1–6, ISBN: 978-1-4614-7711-2. DOI: 10.1007/978-1-4614-7711-2_1. [Online]. Available: https://doi.org/10.1007/978-1-4614-7711-2_1.

[3] R. Xiong, *Battery Management Algorithm for Electric Vehicles*, 1st ed. Spinger Singapore, 2020, ISBN: 978-981-15-0248-4. DOI: 10.1007/978-981-15-0248-4.

[4] C. Lin, A. Tang, and J. Xing, "Evaluation of Electrochemical Models Based Battery State-of-charge Estimation Approaches for Electric Vehicles," *Applied Energy*, vol. 207, pp. 394–404, 2017, Transformative Innovations for a Sustainable Future – Part II, ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2017.05.109. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S030626191730613X.

[5] J. Yang, B. Xia, Y. Shang, W. Huang, and C. Mi, "Improved Battery Parameter Estimation Method Considering Operating Scenarios for HEV/EV Applications," *Energies*, vol. 10, p. 5, Dec. 2016. DOI: 10.3390/en10010005.

[6] P. Malysz, J. Ye, R. Gu, H. Yang, and A. Emadi, "Battery State-of-Power Peak Current Calculation and Verification Using an Asymmetric Parameter Equivalent Circuit Model," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 6, pp. 4512–4522, Jun. 2016, ISSN: 1939-9359. DOI: 10.1109/TVT.2015.2443975.

[7] H. Mu, R. Xiong, H. Zheng, Y. Chang, and Z. Chen, "A Novel Fractional Order Model Based State-of-charge Estimation Method for Lithium-ion Battery," *Applied Energy*, vol. 207, pp. 384–393, 2017, Transformative Innovations for a Sustainable Future – Part II, ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2017.07.003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261917308668.

[8] Z. Geng, T. Thiringer, Y. Olofsson, J. Groot, and M. West, "On-board Impedance Diagnostics Method of Li-ion Traction Batteries Using Pseudo-Random Binary Sequences," in *2018 20th European Conference on Power Electronics and Applications (EPE'18 ECCE Europe)*, Sep. 2018, P.1–P.9.

[9] G. Nusev, D. Juričić, M. Gaberšček, J. Moškon, and P. Boškoski, "Fast Impedance Measurement of Li-ion Battery Using Discrete Random Binary Excitation and Wavelet Transform," *IEEE Access*, pp. 1–1, 2021, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3058368.

[10] J. Sihvo, D. Stroe, T. Messo, and T. Roinila, "Fast Approach for Battery Impedance Identification Using Pseudo-Random Sequence Signals," *IEEE Transactions on Power Electronics*, vol. 35, no. 3, pp. 2548–2557, Mar. 2020, ISSN: 1941-0107. DOI: 10.1109/TPEL.2019.2924286.

[11] C. Chen, G. Yesilbas, A. Lenz, O. Schneider, and A. C. Knoll, "Machine Learning Approach for Full Impedance Spectrum Study of Li-ion Battery," in *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*, Oct. 2020, pp. 3747–3752. DOI: 10.1109/IECON43393.2020.9254622.

[12] M. Berecibar, I. Gandiaga, I. Villarreal, N. Omar, J. Van Mierlo, and P. Van den Bossche, "Critical Review of State of Health Estimation Methods of Li-ion Batteries for Real Applications," *Renewable and Sustainable Energy Reviews*, vol. 56, pp. 572–587, 2016, ISSN: 1364-0321. DOI: https://doi.org/10.1016/j.rser.2015.11.042. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032115013076.

[13] C. Vidal, P. Malysz, P. Kollmeyer, and A. Emadi, "Machine Learning Applied to Electrified Vehicle Battery State of Charge and State of Health Estimation: State-of-the-Art," *IEEE Access*, vol. 8, pp. 52 796–52 814, 2020. DOI: 10.1109/ACCESS.2020.2980961.

[14] A. Gismero, D. I. Stroe, and E. Schaltz, "Calendar Aging Lifetime Model for NMC-based Lithium-ion Batteries Based on EIS Measurements," English, pp. 1–8, 2019. DOI: 10.1109/EVER.2019.8813635. [Online]. Available: https://sfx.aub.aau.dk/sfxaub?sid=pureportal&doi=10.1109/EVER.2019.8813635.

[15] E. Schaltz, D. Stroe, K. Nørregaard, B. Johnsen, and A. Christensen, "Partial Charging Method for Lithium-Ion Battery State-of-Health Estimation," in *2019 Fourteenth International Conference on Ecological Vehicles and Renewable Energies (EVER)*, May 2019, pp. 1–5. DOI: 10.1109/EVER.2019.8813645.

[16] E. Schaltz, D. I. Stroe, K. Nrregaard, L. S. Ingvardsen, and A. Christensen, "Incremental capacity analysis applied on electric vehicles for battery state-of-health estimation," *IEEE Transactions on Industry Applications*, pp. 1–1, 2021, ISSN: 1939-9367. DOI: 10.1109/TIA.2021.3052454.

[17] D. Stroe and E. Schaltz, "Lithium-Ion Battery State-of-Health Estimation Using the Incremental Capacity Analysis Technique," *IEEE Transactions on Industry Applications*, vol. 56, no. 1, pp. 678–685, Jan. 2020, ISSN: 1939-9367. DOI: 10.1109/TIA.2019.2955396.

[18] A. Eddahech, O. Briat, N. Bertrand, J.-Y. Delétage, and J.-M. Vinassa, "Behavior and State-of-health Monitoring of Li-ion Batteries Using Impedance Spectroscopy and Recurrent Neural Networks," *International Journal of Electrical Power & Energy Systems*, vol. 42, no. 1, pp. 487–494, 2012, ISSN: 0142-0615. DOI: https://doi.org/10.1016/j.ijepes.2012.04.050. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0142061512001779.

[19] G. You, S. Park, and D. Oh, "Diagnosis of electric vehicle batteries using recurrent neural networks," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 6, pp. 4885–4893, Jun. 2017, ISSN: 1557-9948. DOI: 10.1109/TIE.2017.2674593.

[20] D. Yang, X. Zhang, R. Pan, Y. Wang, and Z. Chen, "A Novel Gaussian Process Regression Model for State-of-health Estimation of Lithium-ion Battery Using Charging Curve," *Journal of Power Sources*, vol. 384, pp. 387–395, 2018, ISSN: 0378-7753. DOI: https://doi.org/10.1016/j.jpowsour.2018.03.015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378775318302398.

[21] P. Khumprom and N. Yodo, "Data-driven prognostic model of li-ion battery with deep learning algorithm," in *2019 Annual Reliability and Maintainability Symposium (RAMS)*, 2019, pp. 1–6. DOI: 10.1109/RAMS.2019.8769016.

[22] W. Liu, Y. Xu, and X. Feng, "A hierarchical and flexible data-driven method for online state-of-health estimation of li-ion battery," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 14 739–14 748, 2020. DOI: 10.1109/TVT.2020.3037088.

[23] X. Shu, J. Shen, G. Li, Y. Zhang, Z. Chen, and Y. Liu, "A flexible state of health prediction scheme for lithium-ion battery packs with long short-term memory network and transfer learning," *IEEE Transactions on Transportation Electrification*, pp. 1–1, 2021. DOI: 10.1109/TTE.2021.3074638.

[24] G. Li, Q. Chu, Z. Wang, and L. Li, "State of health prediction for battery based on ensemble learning," in *2021 International Conference on Electronics, Circuits and Information Engineering (ECIE)*, 2021, pp. 16–19. DOI: 10.1109/ECIE52353.2021.00011.

[25] M. Armand, P. Axmann, D. Bresser, M. Copley, K. Edström, C. Ekberg, D. Guyomard, B. Lestriez, P. Novák, M. Petranikova, W. Porcher, S. Trabesinger, M. Wohlfahrt-Mehrens, and H. Zhang, "Lithium-ion Batteries – Current State of the Art and Anticipated Developments," *Journal of Power Sources*, vol. 479, p. 228 708, 2020, ISSN: 0378-7753. DOI: https://doi.org/10.1016/j.jpowsour.2020.228708. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378775320310120.

[26] A. Masias, "Lithium-ion battery design for transportation," in *Behaviour of Lithium-Ion Batteries in Electric Vehicles: Battery Health, Performance, Safety, and Cost*, G. Pistoia and B. Liaw, Eds. Springer International Publishing, 2018, pp. 1–33, ISBN: 978-3-319-69950-9. DOI: 10.1007/978-3-319-69950-9_1. [Online]. Available: https://doi.org/10.1007/978-3-319-69950-9_1.

[27] M. A. Hannan, M. M. Hoque, A. Hussain, Y. Yusof, and P. J. Ker, "State-of-the-Art and Energy Management System of Lithium-Ion Batteries in Electric Vehicle Applications: Issues and Recommendations," *IEEE Access*, vol. 6, pp. 19 362–19 378, 2018, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2018.2817655.

[28] N. Nitta, F. Wu, J. T. Lee, and G. Yushin, "Li-ion Battery Materials: Present and Future," *Materials Today*, vol. 18, no. 5, pp. 252–264, 2015, ISSN: 1369-7021. DOI: https://doi.org/10.1016/j.mattod.2014.10.040. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1369702114004118.

[29] R. Xiong and S. Weixiang, *Advanced Battery Management Technologies for Electric Vehicles*, 1st ed. John Wiley & Sons, 2018, ISBN: 9781119481652. DOI: 10.1002/9781119481652.

[30] U.S. Department of Energy, *Battery Archive*, https://batteryarchive.org/index.html, [Accessed 19/02/2021].

[31] V. De Angelis, Y. Preger, and B. R. Chalamala, *Battery Lifecycle Framework: A Flexible Repository and Visualization Tool for Battery Data from Materials Development to Field Implementation*, Mar. 2021. DOI: 10.1149/osf.io/h7c24. [Online]. Available: ecsarxiv.org/h7c24.

[32] Sandia National Laboratories, *Battery Cell Testing Data Archive*, https://www.sandia.gov/energystoragesafety-ssl/research-development/research-data-repository/, [Accessed 19/02/2021], 2017.

[33] H. Barkholtz, A. Fresquez, B. Chalamala, and S. Ferreira, "A Database for Comparative Electrochemical Performance of Commercial 18650-Format Lithium-Ion Cells," *Journal of The Electrochemical Society*, vol. 164, no. 12, A2697–A2706, 2017. DOI: 10.1149/2.1701712jes. [Online]. Available: https://doi.org/10.1149/2.1701712jes.

[34] Sandia National Laboratories, *Battery Cell Aging Data Archive*, https://batteryarchive.org/snl_study.html, [Accessed 19/02/2021], 2020.

[35] Y. Preger, H. Barkholtz, A. Fresquez, D. Campbell, B. Juba, J. Romàn-Kustas, S. Ferreira, and B. Chalamala, "Degradation of Commercial Lithium-Ion Cells as a Function of Chemistry and Cycling Conditions," *Journal of The Electrochemical Society*, vol. 167, no. 12, p. 120 532, Sep. 2020. DOI: 10.1149/1945-7111/abae37. [Online]. Available: https://doi.org/10.1149/1945-7111/abae37.

[36] T. Raj, *Path Dependent Battery Degradation Dataset Part 1*, https://ora.ox.ac.uk/objects/uuid:de62b5d2-6154-426d-bcbb-30253ddb7d1e, [Accessed 19/02/2021], 2020.

[37] T. Raj, A. Wang, C. Monroe, and D. Howey, "Investigation of Path Dependent Degradation in Lithium-Ion Batteries," *Batteries and Supercaps*, vol. 3, no. 12, pp. 1377–13 851, 2020.

[38] P. Kollmeyer, *Panasonic 18650PF Li-ion Battery Data*, http://dx.doi.org/10.17632/wykht8y7tg.1, [Accessed 19/2/2021], 2018.

[39] A. Barragán Moreno, "Optimal Control of the Ecoracer Powertrain: Design and evaluation of efficient and robust control algorithms," 2020. [Online]. Available: https://projekter.aau.dk/projekter/en/studentthesis/optimal-control-of-the-ecoracer-powertrain(66ea9a6f-6f5e-4465-9c07-a56c37cc7ba0).html.

[40] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biology*, vol. 52, no. 1, pp. 99–115, 1990, ISSN: 0092-8240. DOI: https://doi.org/10.1016/S0092-8240(05)80006-0. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0092824005800060.

[41] D. Driankov, P. W. Eklund, and A. L. Ralescu, "Fuzzy logic and fuzzy control," Springer Berlin Heidelberg, 1994, ISBN: 978-3-540-48602-2.

[42] K. P. Murphy, *Probabilistic Machine Learning: An introduction*. MIT Press, 2021. [Online]. Available: probml.ai.

[43] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[44] C. C. Aggarwal, *Neural Networks and Deep Learning: A Textbook*. Springer International Publishing, 2018, ISBN: 978-3-319-94463-0. DOI: 10.1007/978-3-319-94463-0_1. [Online]. Available: https://doi.org/10.1007/978-3-319-94463-0_1.

[45] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989, ISSN: 0893-6080. DOI: https://doi.org/10.1016/0893-6080(89)90020-8. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0893608089900208.

[46] C. L. P. Chen and Z. Liu, "Broad learning system: A new learning paradigm and system without going deep," in *2017 32nd Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, 2017, pp. 1271–1276. DOI: 10.1109/YAC.2017.7967609.

[47] ——, "Broad learning system: An effective and efficient incremental learning system without the need for deep architecture," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 1, pp. 10–24, 2018. DOI: 10.1109/TNNLS.2017.2716952.

[48] M. Xu, M. Han, C. L. P. Chen, and T. Qiu, "Recurrent broad learning systems for time series prediction," *IEEE Transactions on Cybernetics*, vol. 50, no. 4, pp. 1405–1417, 2020. DOI: 10.1109/TCYB.2018.2863020.

[49] H. Huang, T. Zhang, C. Yang, and C. L. P. Chen, "Motor learning and generalization using broad learning adaptive neural control," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 10, pp. 8608–8617, 2020. DOI: 10.1109/TIE.2019.2950853.

[50] R. Mao, R. Cui, and C. L. P. Chen, "Broad learning with reinforcement learning signal feedback: Theory and applications," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–13, 2021. DOI: 10.1109/TNNLS.2020.3047941.

[51] G. Rebala, A. Ravi, and S. Churiwala, *An Introduction to Machine Learning*. Springer International Publishing, 2019, ISBN: 978-3-030-15729-6. DOI: 10.1007/978-3-030-15729-6_1. [Online]. Available: https://doi.org/10.1007/978-3-030-15729-6_1.

[52] R. Zhang, B. Xia, B. Li, L. Cao, Y. Lai, W. Zheng, H. Wang, and W. Wang, "State of the Art of Lithium-Ion Battery SOC Estimation for Electrical Vehicles," *Energies*, vol. 11, no. 7, 2018, ISSN: 1996-1073. DOI: 10.3390/en11071820. [Online]. Available: https://www.mdpi.com/1996-1073/11/7/1820.

[53] M. Hannan, M. Lipu, A. Hussain, and A. Mohamed, "A Review of Lithium-ion Battery State of Charge Estimation and Management System in Electric Vehicle Applications: Challenges and Recommendations," *Renewable and Sustainable Energy Reviews*, vol. 78, pp. 834–854, 2017, ISSN: 1364-0321. DOI: https://doi.org/10.1016/j.rser.2017.05.001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032117306275.

[54] W. He, N. Williard, C. Chen, and M. Pecht, "State of Charge Estimation for Li-ion Batteries Using Neural Network Modeling and Unscented Kalman Filter-based Error Cancellation," *International Journal of Electrical Power & Energy Systems*, vol. 62, pp. 783–791, 2014, ISSN: 0142-0615. DOI: https://doi.org/10.1016/j.ijepes.2014.04.059. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0142061514002646.

[55] V. Klass, M. Behm, and G. Lindbergh, "A Support Vector Machine-based State-of-Health Estimation Method for Lithium-ion Batteries Under Electric Vehicle Operation," *Journal of Power Sources*, vol. 270, pp. 262–272, 2014, ISSN: 0378-7753. DOI: https://doi.org/10.1016/j.jpowsour.2014.07.116. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378775314011707.

[56] B. Saha and K. Goebel, *'Battery Data Set', NASA Ames Prognostics Data Repository*, https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#battery, [Accessed 19/02/2021], 2007.

[57] B. Bole, C. Kulkarni, and M. Daigle, *'Randomized Battery Usage Data Set', NASA Ames Prognostics Data Repository*, https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#batteryrnddischarge, [Accessed 19/02/2021], 2014.

[58] ——, "Adaptation of an Electrochemistry-based Li-ion Battery Model to Account for Deterioration Observed Under Randomized Use," in *Annual Conference of the Prognostics and Health Management Society*, 2014.

[59] Y. Preger, H. Barkholtz, A. Fresquez, D. Campbell, B. Juba, J. Romàn-Kustas, S. Ferreira, and B. Chalamala, "Degradation of Commercial Lithium-Ion Cells as a Function of Chemistry and Cycling Conditions," *Journal of the Electrochemical Society*, vol. 167, Aug. 2020. DOI: 10.1149/1945-7111/abae37.

[60] P. Kollmeyer, C. Vidal, M. Naguib, and M. Skells, *LG 18650HG2 Li-ion Battery Data and Example Deep Neural Network xEV SOC Estimator Script*, http://dx.doi.org/10.17632/cp3473x7xv.3, [Accessed 19/2/2021], 2020.

[61] G. Nusev, *DRBS Excitacion Data for Impedance Measurement*, https://repo.ijs.si/gnusev/supplementary_material.git, [Accessed 19/2/2021], 2020.

[62] G. Nusev, D. Juričić, M. Gaberšček, J. Moškon, and P. Boškoski, "Fast Impedance Measurement of Li-ion Battery Using Discrete Random Binary Excitation and Wavelet Transform," *IEEE Access*, 2021. DOI: 10.1109/ACCESS.2021.3058368.

[63] CALCE Battery Research Group. University of Maryland, *CALCE Battery Dataset*, https://web.calce.umd.edu/batteries/data.htm, [Accessed 19/02/2021].

[64] Severson et al., "Data-driven Prediction of Battery Cycle Life Before Capacity Degradation," *Nature Energy*, vol. 4, 383–391, 2019, [Accessed 19/2/2021]. [Online]. Available: https://data.matr.io/1/projects/5c48dd2bc625d700019f3204.

[65] Attia et al., "Closed-loop Optimization of Extreme Fast Charging for Batteries Using Machine Learning," *Nature*, [Accessed 19/2/2021]. [Online]. Available: https://data.matr.io/1/projects/5d80e633f405260001c0b60a.

[66] C. Birkl, *Oxford Degradation Dataset 1*, https://ora.ox.ac.uk/objects/uuid:03ba4b01-cfed-46d3-9b1a-7d4a7bdf6fac, [Accessed 19/02/2021], 2017.

[67] C. Birkl, "Diagnosis and Prognosis of Degradation in Lithium-ion Batteries," Ph.D. dissertation, University of Oxford, 2017.

[68] A. Devie, G. Baure, and M. Dubarry, "Intrinsic Variability in the Degradation of a Batch of Commercial 18650 Lithium-Ion Cells," *Energies*, vol. 11, no. 5, 2018, ISSN: 1996-1073. DOI: 10.3390/en11051031. [Online]. Available: https://www.mdpi.com/1996-1073/11/5/1031.

[69] Advanced Energy Storage and Applications. Beijing Institute of Technology, *Battery datasets*, http://en.aesa.net.cn/About.aspx?ClassID=37, [Accessed 01/04/2021], 2020.

Appendix A

Summary of publicly available data sets

Table A.1: Description of publicly available data sets

Organization	Devices	Content	Link
National Aeronautics and Space Administration	Li-ion. 35 units	Charge/discharge cycles at several temperatures and with different current profiles. EIS experiments after every half-cycle to quantify cell aging. Includes terminal voltage and current, impedance, temperature, charger/load voltage and current, and discharge capacity measurements as MAT files.	[56]
	18650 Li-ion. 24 units	Charge/discharge cycles using random-walk-like currents, generated using different probability distributions. Tests are conducted at different temperatures. Reference CCCV and pulsed-current cycles every several iterations to evaluate SoH and dynamics degradation. Includes voltage, current and temperature data as MAT and R files.	[57], [58]
Sandia National Laboratories	18650 LCO, LFP, NCA, NMC. 6 units each	CCCV charge/discharge cycles at several temperatures. EIS experiments at every iteration to quantify cell degradation. Some tests are conducted in abusive conditions, that is, disregarding recommended operation conditions. Includes voltage, current, capacity, energy and temperature measurements.	[32], [33]
	18650 LFP, NCA, NMC	CCCV charge/discharge cycles until reaching EoL criteria under different temperature, DoD and discharge current conditions. Includes voltage, current, temperature, capacity and energy measurements as CSV files.	[34], [35], [59]
McMaster University, University of Wisconsin-Madison	18650 NCA	Combination of multi-level HPPC and driving cycles, with EIS experiments at certain SoC milestones, as well as reference capacity tests. The procedure is repeated at different temperatures, both constant and varying. Includes voltage, current, impedance, capacity, energy, power and temperature measurements as MAT and CSV files.	[38]
	18650 NMC	Combination of multi-level HPPC and mixed driving cycles. Reference charging tests are conducted between cycles. The procedure is repeated at a wide temperature range. Includes voltage, current, capacity, energy, power and temperature measurements as MAT and CSV files.	[60]

Continuation of Table A.1

Organization	Devices	Description	Link
Jožef Stefan Institute	18650 NMC	Impedance measurement by means of DRBS excitation, with EIS data as reference. Includes voltage, current, frequency and EIS measurements as MAT files.	[61], [62]
University of Maryland	18650 LFP, NMC. Pouch LCO. Prismatic LCO	Several types of experiments: low-current and incremental-current OCV modelling, driving cycle tests, CCCV lifetime cycling and storage life tests with impedance measurement. Most of the tests are repeated at different temperatures. Data is available as XLS files.	[63]
Battery Archive (U.S. Department of Energy)	18650 LCO, LFP, NCA, NMC. Pouch LCO. Prismatic LCO.	Compilation of data from several research institutions. Focused on life cycling of cells at different conditions (temperature, DoD, SoC range, current profile). Includes voltage, current, temperature, capacity and energy measurements as CSV files.	[30], [31]
Stanford University, Massachusetts Institute of Technology	18650 LFP, 124 units	Several fast-charging experiments at constant temperature for profile comparison. Internal resistance measured by means of HPPC. Includes voltage, current, charge and resistance measurements as MAT and CSV files.	[64]
	18650 LFP, 137 units	Several fast-charging experiments at constant temperature for profile comparison and optimization. Internal resistance measured by means of HPPC. Includes voltage, current, charge and resistance measurements as MAT and CSV files.	[65]
University of Oxford	Pouch LCO, 8 units	CCCV charging followed by Artemis-profile driving cycle at constant temperature. Periodic characterization procedures in order to assess long-term degradation effects. Includes voltage, current, charge and temperature data as MAT files.	[66], [67]
	18650 NCA, 12 units	Long-term degradation tests following different combinations of cycling and calendar ageing at constant temperature. Reference performance tests including EIS and HPPC performed periodically to characterise cell degradation. Includes voltage, current, temperature, capacity and impedance data as MAT files.	[36], [37]

Continuation of Table A.1

Organization	Devices	Description	Link
Hawaii Natural Energy Institute	18650 NMC-LCO, 14 units	CCCV charging and discharging tests at constant temperature from BoL to EoL. Includes voltage, current, temperature, capacity and energy data as CSV files.	[30], [68]
Beijing Institute of Technology	LFP; NMC; NCA; LiPb; LTO; LMO	Periodic characterization tests, including maximum capacity, HPPC and OCV, every 100 iterations of an ageing procedure. Tests are conducted at several temperatures. For some of the chemistries, EIS and driving cycle tests are also available. More focused on battery packs than in individual cells. Data is available upon request to the organization.	[29], [69]

Table A.2: Content in publicly available data sets

data set	CCCV	OCV	HPPC	EIS	DC [1]	CA [2]	MT [3]	MD [4]	MC [5]
[56]	✓	X	X	✓	X	X	✓	X	✓
[57], [58]	✓	X	✓	X	X	X	✓	X	✓
[32], [33]	✓	X	X	✓	X	X	✓	X	X
[34], [35]	✓	X	X	X	X	X	✓	✓	✓
[38]	✓	X	✓	✓	✓	X	✓	X	X
[60]	✓	X	✓	X	✓	X	X	✓	X
[61], [62]	X	X	X	✓	X	X	X	X	X
[63]	✓	✓	✓	✓	✓	✓	✓	✓	✓
[30], [31]	✓	X	X	X	X	X	✓	✓	✓
[64]	✓	X	✓	X	X	X	X	X	X
[65]	✓	X	✓	X	X	X	X	X	X
[66], [67]	✓	X	X	X	✓	X	X	X	X
[36], [37]	✓	✓	✓	✓	X	✓	X	X	X
[30], [68]	✓	X	X	X	X	X	X	X	X
[29], [69]	✓	✓	✓	✓	✓	X	✓	✓	✓

[1] Driving cycles
[2] Calendar ageing
[3] Multiple temperatures
[4] Multiple DoD
[5] Multiple C-rate

This page intentionally left blank

STUDENT REPORT

AALBORG UNIVERSITY