Otto-von-Guericke University Magdeburg

Department of Computer Science
Knowledge Management and Discovery Lab

Master Thesis

Transferability of Battery Cell Ageing Prediction Models

Author:
Maya Santhira Sekeran

April 24, 2020

Supervisor
Professor Dr. Myra Spiliopoulou
Department of Computer Science
Otto-von-Guericke University
Universitätsplatz 2
39106 Magdeburg, Germany

Santhira Sekeran, Maya:
Transferability of Battery Cell Ageing Prediction Models
Master Thesis, Otto-von-Guericke University
Magdeburg, 2020.

Declaration of Academic Integrity

I hereby declare that I have written the present work myself and did not use any sources or tools other than the ones indicated.

Datum: .....

(Signature)

Contents

Abstract

Electric vehicles are increasingly becoming the vehicle of choice in today's environmental conscious society and the heart of an electric vehicle is it's battery. Today, lithium-ion batteries are mainly used to power electric vehicles for its increased energy storage density and longevity. In order to estimate battery life, long and costly battery testing is required. Therefore, there is a need to investigate efficient ways that could reduce the amount of testing required by reusing existing knowledge of ageing patterns from different kinds of battery chemistry. This problem inspires the purpose of this thesis that comprises of two main goals. The first goal is to build a battery end of life prediction model using survival analysis. The second goal is to build a transfer learning method that can facilitate the reuse of a prediction model trained on one type of battery cell chemistry to predict the end of life for another type of battery cell chemistry.

Acknowledgements

The completion of this thesis would not have been possible without the support and encouragement from a number of individuals. I would first like to express my deepest gratitude to my university supervisor, Professor Dr. Myra Spiliopoulou who have been tirelessly supporting my work despite having to communicate remotely on most occasions. Her detailed feedback helped me to appreciate the process of scientific work and the need to be precise in explaining methods and concepts used in this thesis.

My sincere appreciation also goes to Dr. Koenraad van Woerden as my supervisor at AVL Software and Functions Regensburg who was instrumental in helping me to understand crucial mathematical concepts and reviewing my thesis drafts. My thanks also goes to the AVL Big Data team headed by Marek Hojgr for their support in ensuring that I had the necessary tools to facilitate the completion of this thesis.

The topic of this thesis was realised with the help of the Battery Ageing team based in AVL List Graz and I am ever grateful for the encouragement and support provided by Dr. Veronika Obersteiner, Dr. Thomas Traussnig and the rest of the team. I would also like to thank Matthias Scharrer of Virtual Vehicle who took the time to share his knowledge regarding the datasets used in this thesis.

I embarked on this thesis with little knowledge in survival analysis, transfer learning and battery ageing. Understanding these concepts was only possible with the help of numerous forum contributors and book authors. I would like to especially thank the following individuals in no particular order for their time and kindness to answer my questions: Professor Dr. Gregory Plett, Professor Dr. William Meeker, Dr. Stefan Gelissen, Pablo Espinoza, Dr. Peter Attia, Dr. Hu Xiasong, Professor Dr. Chandan Reddy, Dr. Michael Pecht, Professor Dr. Charles Bouveyron and Nabil Mahmoud.

Finally, I am deeply thankful to my family and friends for their constant encouragement and support to continue my studies and accompany me during challenging times of writing this thesis.

1 Introduction and Motivation

1.1 Motivation

The adoption of electric vehicles (EV) on the road is expected to be on an upward trend as demand is getting stronger for clean energy and environmentally friendly alternatives. In 2018 alone, global sales for EVs increased by 68% with active ownership, mainly in China and Europe [1]. The EV30@30 Campaign launched in 2017 established a goal to achieve 30% market share for electric vehicles out of the total vehicles (except two-wheelers) by 2030. This is also in line with the Paris Agreement to reduce carbon emission by at least 40% by 2030 compared to 1990 [1].

Many of these initiatives are policy led as user consumption is still limited by the price and safety factors of using an electric vehicle. The price of batteries is factored by the choice of chemistry, capacity, manufacturing capacity and charging speeds [1]. The chemical of choice now is lithium-ion that can be either lithium nickel manganese cobalt (NMC) or lithium nickel cobalt aluminum oxide (NCA), lithium manganese oxide (LMO) or lithium iron phosphate (LFP). The anode (negative electrode) is usually formed using graphite but lithium-titanite is also used (LTO) for heavy duty applications [1]. The NMC or NCA technologies are more commonly used in electric cars as they provide higher density energy compared to other chemistries. The LFP chemistry is mainly used in heavy duty EVs such as buses and has safer plus higher cycle performance.

According to the Global EV Outlook report 2018 [1], lithium-ion will still be the chemistry of choice for the next decade as even if there will be new battery designs by 2030, time is required to prepare the infrastructure needed till the point of adoption in the market.

One of the key areas in determining the robustness of a battery is its ageing process. This is because, as of now, battery manufacturers can only provide a battery lifetime warranty of between eight to ten years before a battery needs to be replaced. Since the cost of the battery itself makes up half of the price of an electric vehicle, more effort is needed to find ways to optimise and extend battery life. One of the key parameters needed for this purpose is by monitoring the battery state of health (SoH). Currently, the United States Council for Automotive Research (USCAR) has set an ambitious goal for EV manufacturers to provide a 15-year warranty on battery life by 2020 as the lifetime of cars are also expected to extend for more years.

Determining warranty periods for batteries requires obtaining battery life data from lab experiments conducted for a period between six months to two years. This duration is costly (could amount to several million dollars) and often needs to be repeated to observe battery cell ageing using different kinds of cell chemistry and charging conditions.

However, one of the key issues with analysing battery life data is the scarcity of capturing end of life points as most battery cells may have not experienced end of life during the experiment or observation period. This kind of incomplete data is called censored data [2]. In the case of battery life, this is due to the linear degradation that is observed in the beginning of a battery cell life and this pattern stays linear until the battery cell reaches 80% or 70% remaining capacity. This leads to using extrapolation functions to determine battery cell life and is often challenging to validate without the presence of actual battery life data. This begs the question on whether there are other approaches that can be used to handle censored observations such as survival analysis.

As mentioned above, acquiring battery life data is expensive. This is often a problem faced in other domains as well that has led to an active research area in transfer learning or domain adaptation. Transfer learning aims to reuse models that have been trained and have learned generalised patterns that can be used to improve the predictions for smaller datasets of a different or related domain. This step involves building a model on a source dataset that is then reused on another dataset referred as the target dataset.

In light of these research challenges, this thesis aims to tackle two problems. Firstly, to investigate the use of survival models to predict battery end of life with censored data. Secondly, is to implement a transfer learning approach that can possibly reduce the amount of battery life dataset required to predict battery end of life for different types of battery cell chemistry.

1.2 Aim of this thesis

Based on the challenges mentioned above, efforts are currently geared towards finding ways to reduce these long and costly battery cell experiment procedures. By having a battery end of life prediction model that can be generalised to predict the end of life of other battery chemistries, this will proof to be of great interest and value for battery manufacturers and OEMs. This thesis aims to address two goals as below:

- Firstly is to implement a data driven approach that can handle battery capacity degradation data while including cells that have not reached end of life. This goal mainly addresses the challenge of incomplete observations of failures or end of life in many experiments.

- Secondly, is to investigate transfer learning strategies that can be implemented to reuse knowledge gained from modeling one type of battery chemistry dataset (source dataset) that was built in the first goal to predict the battery end of life of another type of battery chemistry (target dataset).

1.3 Conceptual Architecture

A conceptual architecture or framework provides a solution for a problem without considering any constraints. This framework assumes that we have all the resources needed in order to solve this research problem. Fig.1.1 explains the conceptual architecture for this thesis.

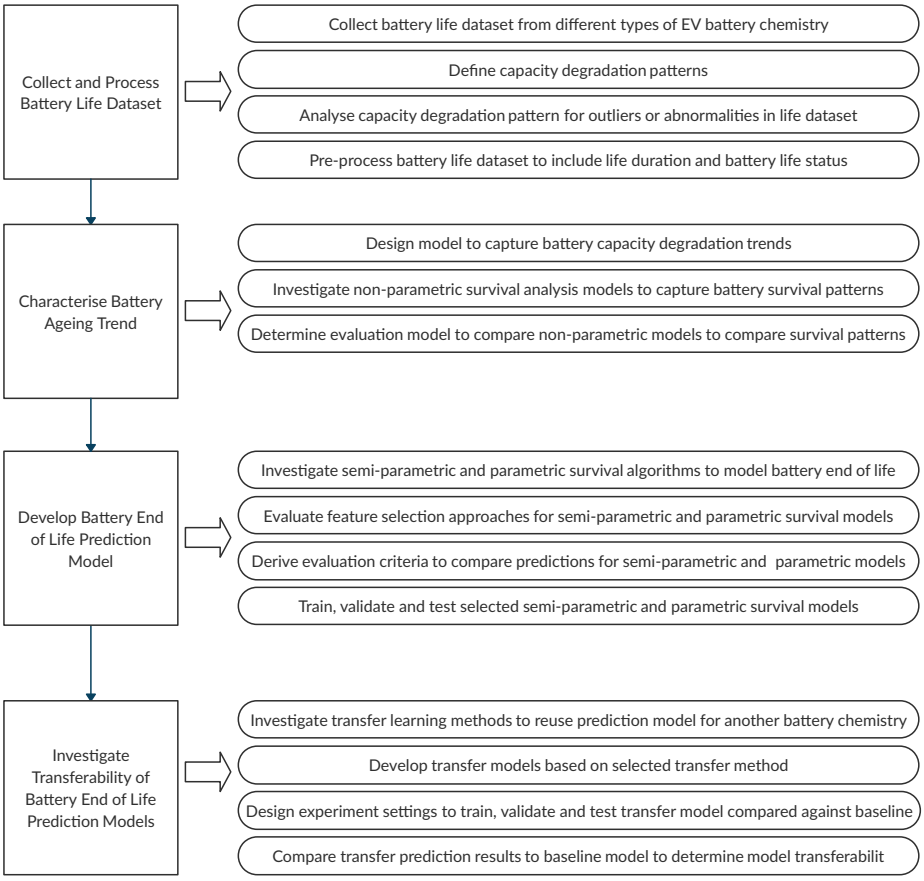


Figure 1.1: Conceptual Framework

Collect and Process Battery Life Dataset

Data collection and processing is the primary step needed in order to realise the goals of this thesis. The steps below explains the approaches used to address this requirement.

- Collect battery life dataset from different types of EV battery chemistry.

  The first requirement of this thesis is to be able to collect battery life dataset from different types of battery chemistry that are commonly used in electric vehicles.

- Define capacity degradation patterns.

  In this step, we define the expected capacity degradation patterns. This is addressed by collecting battery capacity measures from every reference test procedure and measure the rate of degradation against the nominal capacity.

  We use 80% as the end of life threshold to determine the point at which the battery cell is considered to have reached end of life.

- Analyse capacity degradation pattern for outliers or abnormalities in life dataset.

  Once these datasets are obtained, the next step is to capture the capacity of the battery over time and analyse battery capacity degradation trends. At this point, it is important to detect any possible abnormalities in the dataset and seek for domain experts opinion if the trend observed is as expected.

- Preprocess battery life dataset to include life duration and battery life status.

  As we are interested in investigating the use of survival analysis to model battery end of life, the next step is to preprocess the battery life dataset to be in the format that can be used to implement survival models. In this step, two additional variables are needed. The first variable provides the duration of life of every battery cell in the experiment. The second is a binary flag that indicates '1' if a battery cell capacity reaches 80% or lower or a '0' to indicate if the battery cell capacity is still above 80% at the end of the battery experiment. The variable is referred to as the "Status" of the battery cell where a '1' means that the battery cell has reached end of life and a '0' for being still usable by the end of the experiment period.

Characterise Battery Ageing Trend

Characterising battery ageing trend is important as a step to understand the battery life dataset collected from the battery ageing lab.

- Design model to capture battery capacity degradation trends.

This problem is addressed by investigating non-parametric survival models that can capture battery survival patterns and compare the survival patterns between different battery chemistry.

- Investigate non-parametric survival analysis models that can be used to capture survival patterns.

  This step provides preliminary analyses that can assist in testing the hypothesis of whether different types of battery chemistry share similarities in the way they age over time based on the dataset available.

- Determine evaluation model to compare non-parametric models which captures and compares survival patterns.

  We determine the evaluation model to compare the different non-parametric approaches that can be used by examining the approaches available and selecting a model that can provide a statistical comparison of survival patterns and includes a way to test the hypothesis of similarities between battery chemistry capacity degradation trends.

Develop Battery End of Life Prediction Model

To the best of our knowledge and at the point of writing this thesis, there were no research available that points to a specific survival model that can be considered to model battery end of life. Therefore, this step is crucial to ensure all possible survival algorithms are considered.

- Investigate semi-parametric and parametric survival algorithms to model battery end of life.

  Here, we investigate semi- parametric and parametric survival algorithms to model battery end of life. This problem is addressed by implementing a semi-parametric approach namely the Cox-PH model. The parametric method is addressed by testing several parametric shapes such as Weibull, log-normal and log-logistic that can best model the survival patterns and to also consider domain application when selecting which parametric survival model to use.

- Evaluate feature selection approaches for semi-parametric and parametric survival models.

Once a suitable survival algorithm is selected, we evaluate feature selection approaches that can derive or suggest several models to be validated. These approaches include considering backward selection, forward selection and exhaustive search and evaluated using evaluation methods such as Mallow's Cp, Adjusted R-squared or Bayesian Information Criterion (BIC).

- Derive evaluation criteria to compare predictions for semi-parametric and parametric models.

  Here, we determine the evaluation criteria to compare the prediction results for the selected semi-parametric and parametric models. One approach is by using root mean squared error (RMSE). The prediction model that produces the lowest RMSE is selected as the preferred prediction model.

- Train, validate and test semi-parametric and parametric survival models.

  Depending on battery life dataset availability, the last step in this component is addressed by splitting the dataset into train, validation and test datasets to determine model performance and select a model that can provide the best prediction results.

Develop Transfer Learning Model to Predict End of Life of Different Battery Chemistry

In this component, we address the fundamental purpose of the thesis which is to examine the transferability of battery end of life prediction model.

- Investigate transfer learning methods to reuse prediction model for other battery chemistry.

  To determine which transfer learning method is most suitable to test the transferability of the battery end of life prediction model developed in the previous step, we investigate the different transfer learning methods. This step is done by scouring transfer learning survey papers and relevant research.

- Develop transfer model based on selected transfer method.

Based on the selected transfer method, we consider several transfer models to capture knowledge gained from the existing prediction model that predicts the end of life of one type of battery chemistry to be reused to predict the end of life for a battery with a different type of battery cell chemistry.

- Design experiment settings to train, validate and test transfer models to be compared against baseline.

  We then validate these models by designing several experiment scenarios. These models are then compared against the baseline model which is the prediction model that does not involve any knowledge transfer.

- Compare transfer prediction results to baseline model to determine model transferability.

  The final step is to compare the prediction results and to determine if the hypothesis of transferability of battery end of life prediction models can be accepted or rejected. This is done by analysing the RMSE results that shows the difference between the true end of life and predicted end of life.

1.4 Thesis Workflow

As an overview, the workflow for this thesis is as shown in Fig. 1.2.

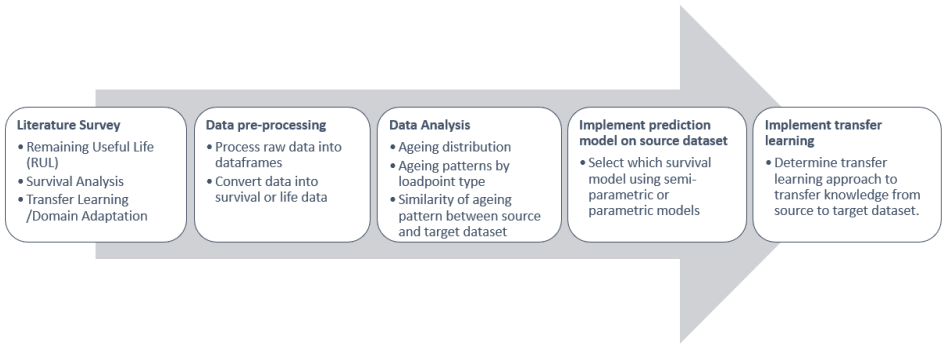


Figure 1.2: Thesis Workflow

Each of the components in the workflow above is further broken down into sub-tasks as follows:

1. Literature Review and critical assessment
    - Research existing studies and methods used in predicting battery end of life.
    - Research on survival analysis as an approach used in battery life data and it's applicability for this use case.
    - Research on transfer learning strategies that can be adopted especially for regression problems.
2. Data preprocessing
    - Extract capacity values from reference test procedures for each battery cell that were experimented under different load settings. This step is done twice for the two different types of battery chemistry used.
    - Generate survival data format by calculating final recorded capacity value against nominal capacity. This results in the number of days the battery cells were experimented and whether each battery cell has reached the end of life threshold of 80%. Label '1' is assigned for battery cells that have reached end of life and '0' for battery cells that have not.
3. Data analysis
    - Investigate capacity degradation for the source dataset and the target dataset to answer questions below:
        - Are there significant differences between the ageing trend of different types of battery cell chemistry?
        - Are there significant differences of battery cell ageing based on identified ageing influences in the source dataset and target dataset? (e.g. impact of high and low temperatures, swing in state of charge, etc.)
        - Does the survival trend follow a specific distribution pattern? (e.g. Weibull, log-normal, log-logistic) by way of using non-parametric methods.
4. Data modeling for end of life prediction
    - Investigate application of semi-parametric and parametric models for battery life datasets.

- Determine distribution type to be used for parametric modeling.
- Feature selection for source dataset by way of using probabilistic approaches to generate a list of models for end of life prediction.
- Model selection by way of using re-sampling approaches specifically bootstrapping. Model with the lowest RMSE is selected.

5. Data modeling for transfer learning

   - Investigate transfer learning strategy by way of addressing the questions below:
     - What will be transferred from the source model to the target model?
     - How will the knowledge gained from the prediction model built for the source dataset be reused for the target dataset?
   - Design and execute experiments to facilitate knowledge transfer between the source model and the target model.
   - Evaluate model transfer by way of comparing average root mean squared error from every experiment.

1.5 Structure of this thesis

In Chapter 1, we discuss the motivation along with the conceptual framework and thesis workflow that was used as a guide for the realisation of this thesis.

In Chapter 2, we discuss lithium-ion batteries and how batteries work in general. This is followed by investigating battery ageing factors and their effects on battery end of life. A description of the datasets used for this thesis and initial data analysis on capacity degradation trends are discussed in this chapter as well.

Chapter 3 discusses the underpinnings of battery end of life predictions and critical assessment on each of the approach studied.

Chapter 4 discusses transfer learning strategies and related work in this domain.

Chapter 5 is based on the critical assessments conducted in Chapter 3. This chapter discusses the approach used in this thesis to build a battery end of life prediction model for the source dataset. Here, we also discuss the transfer learning strategy that is most applicable to this use case based on the assessment done in Chapter 4.

This is followed by Chapter 6 that discusses the prediction model building process for the source dataset and the results obtained.

Chapter 7 shows the implementation of the transfer learning strategy discussed in Chapter 4 and the results obtained.

Chapter 8 provides a conclusion and suggestions for future work.

Some terms that are used interchangeably in this thesis are as follows:

- Covariate, variable, feature, factor, regressor
- Dataset A1 (source dataset) is also known as ALICe1 and dataset A2 (target dataset) is also known as ALICe2
- Transfer learning, domain adaptation, covariate shift
- Cell ageing, battery end of life, battery ageing

2
Materials

This chapter aims to discuss the theoretical aspects of the workings of a battery. This is followed by understanding of the different parameters that influences battery ageing. We then describe the battery life datasets that were used and provide a definition of the capacity degradation trends observed.

2.1 Lithium-ion Battery

The lithium-ion battery technology was developed by a team of scientists, John B. Goodenough, M. Stanley Whittingham and Akira Yoshino in the seventies and eighties earning the Nobel Prize in Chemistry in 2019 [3]. Their work was instrumental in introducing rechargeable lithium-ion batteries that powers most appliances today.
There are different types of lithium-ions; Lithium Cobalt Oxide (LCO), Lithium Iron Phosphate (LFP), Lithium Manganese Oxide (LMO), Lithium Nickel Manganese Cobalt Oxide (NMC), Lithium Nickel Cobalt Aluminium Oxide (NCA), Lithium Titanate Oxide (LTO)[4]. The formation of these different types of lithium-ion materials allowed to stabilise the highly reactive lithium metal while at the same time maintaining its potential charging capacity. Details on these discoveries can be found in [3].

There are a few reasons for the common use of lithium-ion in electric car batteries. Lithium-ion contains higher energy densities and high power per mass compared to lead-acid or nickel-metal hybrid batteries that allows for the weight of the batteries to be significantly reduced. It also has long cycle life and is highly safe with fire resistance capabilities that is important to avoid thermal runaway situations.

The benefits of using the lithium-ion technology makes it a popular choice for electric vehicles as studied in [5]. Lithium-ion batteries are also considered to have the best charge to weight ratio solution and gradual loss of maximum capacity even with repeated recharging that results in longer lifespans.

However, lithium-ion batteries are also highly influenced by factors such as temperature that affects its efficiency over time. The ageing process of lithium-ion batteries is a very active research area due the external and internal factors that can affect battery life. By having a better understanding on these ageing influences, steps can be taken to optimise and lengthen battery usage.

2.1.1 Battery Cell Design Overview

Components of a battery cell consists of the positive electrode (cathode), negative electrode (anode), separator (mechanically and electrical isolation of two sides of battery cell), electrolyte, and current collectors (metal foils that hold the electrode powder together) as in Fig. 2.1[6].

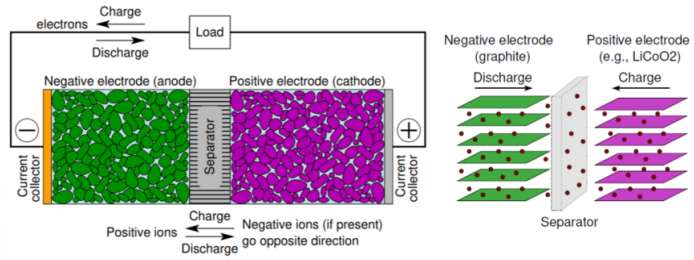


Figure 2.1: Lithium-ion Battery Cell Design [6]

A charged battery cell holds electrochemical potential energy. Electrochemical potential refers to the tendency of a metal to release electrons. Among various chemicals, lithium-ion holds the highest capability to lose these electrons. In a battery cell, both the positive and negative electrode holds potential energy.

In a non-rechargeable or what is called as primary battery cell, the negative electrode holds the potential energy that initiates a chemical process to release electrons into the external circuit and these electrons moves towards the cathode end of the battery. The positive electrode receives the

electrons from the negative electrode through the external circuit. As this process is occurring, electricity is generated to power appliances. Over time, a process called oxidation occurs in the negative electrode as electrons gets depleted and a reduction process occurs in the positive electrode as the acceptance of negatively charged electrons neutralises the positively charged electrons. This eventually leads to the end of the battery life when battery is no longer able to provide a potential difference to the external circuit.

In the case of rechargeable or secondary battery cell, the design of the chemical reaction allows for a reversible process to occur. This process is facilitated by injecting energy into the cell from the external potential and reforms the chemical composition of the electrons before it was discharged. At this state, the battery is said to have recharged and this process continues where every charge discharge of the battery is called a cycle.

While the cycling of the battery takes place multiple times, other chemical processes occurs in the battery cell that causes the degradation of the battery over time. These are called parasitic chemical reactions which causes materials in the battery to degrade. The causes of these parasitic chemical reactions can be due to high voltage or abuse from exposure to very high temperatures. There could also be mechanical processes causing stresses and strains that cause the materials in the battery to degrade. So the life of a secondary battery cell is not primarily determined by its main chemical reaction, rather the life is determined by how quickly these other degradation processes occur [7].

Recharging the battery requires care as overcharging a battery cell means that we are forcing energy into the cell where the cell is no longer able to accept the charge in a safe way. This causes irreversible damage to the cell, and in some cases, even fires or explosions [7].

During discharge, cations move from the negative electrode to the positive electrode and anions move from positive electrode to the negative electrode.

2.1.2 Battery Pack Topology

This section illustrates how battery cells forms the design of the battery pack that is build from combining multiple battery cells in several mod-

Functions of Negative Electrode	**Functions of Positive Electrode**
Discharge - gives up electrons Oxidation is Loss (OIL)	Discharge - accepts electrons Resistance is Gained (RIG)
Charge - accepts electrons from external circuit, (RIG)	Charge - gives up electrons to external circuit, (OIL)

Table 2.1: Functions of Negative and Positive Electrode

ules. However, it is important to note that the scope of this thesis will only concentrate on battery ageing at the cell level.

Electrical power is computed by multiplying current and voltage:

$$P = I \cdot V \tag{2.1}$$

P = Power
I = Current
V = Voltage

Battery pack engineers decide the topology of the battery pack by considering the design requirements to achieve maximum power. The question to address will be the voltage range and the peak current.

As per Kirchhoff's current and voltage law, high power battery packs can deliver high voltage, high current or both. For high voltage packs, we stack the cells in series while for high current packs, the cells are stacked in parallel but usually the design is a combination of both series and parallel. If the cells are configured in series, then the battery voltage is the sum of individual cell voltages. For example with AVL's E-Coupe car that has a 800V battery, the sum of the cell voltages would approximate this value. Battery capacity equals to the individual cell capacity.

$$v_{pack} = N_s \cdot v_{cell} \tag{2.2}$$

v_{pack} = pack voltage
N_s = number of cells in series
v_{cell} = cell voltage

Example: If we have 3 batteries of 2V each with 20Ah capacity, then the battery energy capacity will be 2V · 3 · 20Ah = 120Wh.

If the cells are configured in parallel, then the battery voltage is equal to the cell voltage. Battery capacity equals to the sum of all cell capacities as calculated in the formula below:

$$i_{pack} = N_p \cdot i_{cell} \tag{2.3}$$

i_{pack} = pack current
N_p = number of cells in parallel
i_{cell} = cell current

Example: If we have 5 batteries of 2V each with 20Ah capacity, then the battery energy capacity will be 20Ah · 5 · 2V = 200Wh.

The total energy storage capacity is calculated using the formula below:

$$E = V_{nominal} \cdot Ah_{nominal} \tag{2.4}$$

E = Energy storage capacity
$V_{nominal}$ = Nominal voltage
$Ah_{nominal}$ = Nominal capacity

Battery packs consists of multiple cells and is managed and controlled by the Battery Management System (BMS). The BMS is an embedded system (purpose built electronics plus processing to enable a specific application). Below are formulas to calculate the battery power and energy capacities:

$$E_{pack} = n_s \cdot n_p \cdot q_{cell} \cdot v_{cell} \tag{2.5}$$

$$P_{pack} = n_s \cdot n_p \cdot i_{cell} \cdot v_{cell} \tag{2.6}$$

E_{pack} = pack energy
n_s = number of cells configured in series
n_p = number of cells configured in parallel
q_{cell} = cell resistance
i_{cell} = cell current

v_{cell} = cell voltage

An estimate of energy is important for EVs where it is the fundamental input to calculate vehicle range while an estimate of power is important for HEVs where power estimates shows the rate at which the energy can be moved from the battery pack to the wheels or vice versa without exceeding cell or electronic design limits. To meet energy and power requirements, n_s and n_p needs to fulfill some criteria but the individual values are flexible.

High voltage batteries require careful cell matching especially when using heavy loads in cold temperatures. As the cells are connected as a string, the possibility of a failure will occur when one cell fails. Usually there will be a switch in large packs to bypass the failed cell to allow current to continue to flow although with lower voltage [8].

High voltage batteries would normally have multiple modules so that when a cell fails in one module, only the affected module can be replaced but a slight cell imbalance will occur when replaced since the new cell will have higher voltage [8].

2.1.3 Causes and Effect of Cell Ageing

Causes of Cell Ageing

There are many factors that can contribute towards how a battery cell ages and Fig. 2.2 gives an overview of these causes based on different categories. Essentially, battery cell ageing occurs at the chemical and mechanical level and is dependent on the chemical composition of the electrodes. The ageing of the positive and negative electrodes of the battery cell also differs and a detailed explanation on the ageing process can be found in [9].

At the first charge of a battery cell, a layer referred as the Solid Electrolyte Interface (SEI) forms as a protective agent against corrosion but this layer is highly unstable [9]. Overtime this layer grows at the anode part of the battery that causes capacity fading and power loss. A high SoC can increase the frequency of the expansion of the SEI through some chemical reaction and therefore justifies the need for SoC measurements to be included in battery ageing models.

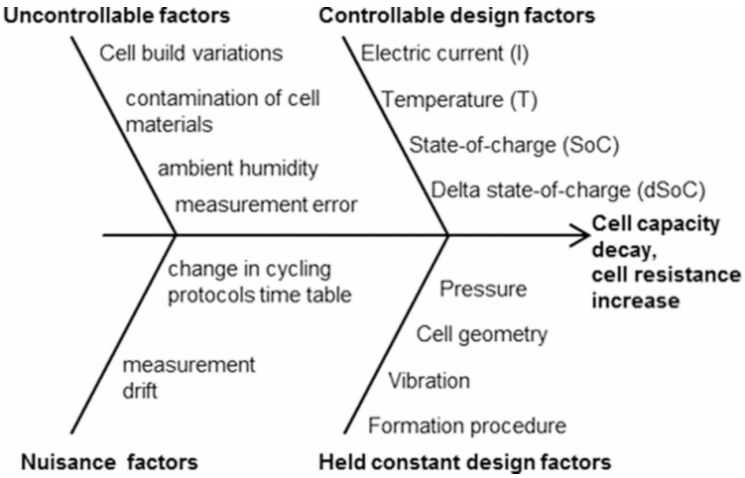


Figure 2.2: Fishbone diagram of ageing influencing factors [10]

When a battery is in use, it goes through cycles of charge and discharge. The cyclic ageing is therefore dependent on the utilisation of the battery, temperature and the current requirements to charge and discharge the battery. This explains the need to consider temperature, average discharge and charge current as measurable parameters for battery ageing models.

As mentioned in [9], the SoC swing or delta SoC represents the state of charge variation during a cycle. Through various experiments, it was found that as the SoC swing becomes larger, this causes high loss in battery power as well. This is attributed to the degradation in the positive electrode and the SEI development due to high charge and discharge [9].

Other causes of battery cell ageing depends on the cell chemistry type, cell design, usage and environmental conditions. Although cross dependencies among these factors exists, it is not possible to consider all ageing influencing factors at the same time. Therefore one of the key activities in researching and understanding battery ageing involves selecting relevant and important ageing factors to be tested usually in a controlled lab environment. Other causes of battery cell ageing:

- Failure causes: cell design faults, poorly controlled manufacturing process, ageing, uncontrolled operations and abuse [7].

- Cell performance usually degrades due to unwanted chemical reactions, physical changes and active chemicals. This process is irreversible [7].

- The effect of these internal chemical reactions causes batteries to form effects such as corrosion, crystal formation, dendrite growth, chemical loss through evaporation, passivation, short circuit cell and cracking of the electrode or electrolyte [7].

For the purpose of modeling cell ageing prediction using data driven methods, the observable or controlled parameters as they are measurable compared to uncontrollable factors that are difficult to be quantified.

Effects of Battery Ageing

Effects of battery ageing include reduced capacity, increased internal resistance or increased cell discharge [7].

The reduced capacity effect is usually used to observe battery life degradation. This is because with internal resistance, there is a tendency for this value to remain consistent until the battery reaches end of life and increased cell discharge is not a commonly used parameter for observing battery degradation.

2.2 Description of Dataset

2.2.1 Data Collection

The datasets used in this thesis were collected from a battery life testing project with the goal of understanding cell ageing influences and to build statistical models that can describe cell ageing. The datasets from this project comprised of battery cell life from two battery cell chemistries namely the NMC and NCA cell chemistry type. Therefore, we will focus on understanding the ageing trends for these available battery cell life datasets.

As depicted in Fig. 2.2, seven controllable factors were selected; temperature in °C (Temp), state of charge in percentage (SoC), state of charge

swing in percentage (DSoC), peak discharge current in Amps (PDC), average discharge current in Amps (ADC), frequency pulse in Hz (Freq) and charge current in Amps (CC).

A design of experiment (DoE) task was executed for this purpose to form statistical foundation to determine the experiment configurations that should be used for testing. A range is determined for each ageing influence that later forms the factor levels for experiment configurations. The combination of factor levels defines a load point configuration for each cell and is tested by way of replication. Each loadpoint configuration was repeated for three or four cells. To measure the capacity of the cell over time, a Reference Test Procedure (RTP) is conducted every three or four weeks where this procedure usually lasts for about two days and the capacity of the battery cell is estimated. More information on the DoE can be found regarding DoE in [10].

Experiments were conducted in a lab environment for two kinds of battery cell chemistry. The first battery cell testing was executed in 2013 until end of 2015, labelled as the ALICe1 dataset while the second test experiment was held in 2016 until mid 2018, labelled as the ALICe2 dataset. From now on these datasets will be referred to as the ALICe1 dataset and ALICe2 dataset.

In each of these datasets, the cells were divided between small battery cells (18650) and large battery cells (PHEV2) that are used in commercial vehicles. Essentially, there are four datasets, two sets for the 18650 type battery cells and the other two sets for the PHEV2 type battery cells. The reason why these datasets need to be treated separately is even though the cells share the same chemistry type, the geometry within these cells differ. In this thesis, statistical models will be built using the ALICe1 and ALICe2 18650 battery cells only.

The loadpoint configuration used for ALICe1 and ALICe2 are as shown in Table 2.2 and Table 2.3

		ALICe1						
Factor	Unit	Min	Max	Factor Levels				
Temperature (Temp)	° C	-10	40	-10	5	20	40	
Charge Current (CC)	A	0.2	2.4	0.2	0.8	2.4		
Average Discharge Current (ADC)	A	0	8	Z	0.2	1	4	8
Peak Discharge Current (PDC)	A	0.2	14	0.2	3	8	10	14
Frequency (Freq)	Hz	3.33e-04	0.5	3.33e-04	0.03	0.1	0.5	
State of Charge (SoC)	%	15	95	15	25	55	85	95
Delta State of Charge (DSoC)	%	0.01	80	0.01	2.5	15	50	80

Table 2.2: Load Point Configurations for ALICe1

		ALICe2									
Factor	Unit	Min	Max	Factor Levels							
Temperature (Temp)	°C	0	45	0	20	30	45				
Charge Current (CC)	A	0	0.5	0	0.05	0.5					
Average Discharge Current (ADC)	A	0	2	0	0.2	0.3	0.75	0.9	1	1.5	2
Peak Discharge Current (PDC)	A	0	3	0.2	1.5	3					
Frequency (Freq)	Hz	0	0.5	0	3.0e-04	3.33e-04	0.2	0.25	0.5		
State of Charge (SoC)	%	15	95	15	25	40	55	70	85	95	
Delta State of Charge (DSoC)	%	0	80	0	2.5	20	30	50	80		

Table 2.3: Load Point Configurations for ALICe2

In earlier works [11], a multiple linear regression model (MLR) was implemented using the controllable ageing influence factors as depicted in Fig. 2.2 as features for the model. Two types of battery cell datasets were used for the experiments. The small type of batteries (18650) were used to parameterise the model. This model was then used to model the behaviours of the larger cell type (PHEV2) that are used in commercial vehicles. The smaller cells were used for economical purposes and are assumed to depict similar ageing behaviours as the larger cells.

The combination of these seven factors defines the loadpoint (LP). In the beginning the linear regression model used several different factor combinations as shown in the equation below from [11]. Here, β represents the coefficient value for each factor while ε is the estimated error between the observed and predicted battery cell end of life:

$$f(EoL) = \beta_0 + \beta_1 Temp + \beta_2 CC + \beta_3 ADC + \beta_4 PDC + \beta_5 Freq + \beta_6 SoC + \beta_7 dSoC + \beta_8 (Temp)^2 + \beta_9 ADC^2 + \beta_{10} SoC^2 + \beta_{11} dSoC^2 + \beta_{12} Temp^3 +$$

$\beta_{13}Temp \cdot CC + \beta_{14}Temp \cdot PDC + \beta_{15}CC \cdot PDC + \beta_{16}CC \cdot dSoC + \beta_{17}PDC \cdot Freq + \beta_{18}PDC \cdot SoC + \beta_{19}PDC \cdot dSoC + \beta_{20}Freq \cdot dSoC + \beta_{21}SoC \cdot dSoC + \beta_{22}Temp \cdot Freq + \beta_{23}Temp \cdot SoC + \beta_{24}Freq \cdot SoC + \varepsilon$

The next step was to identify significant regressors that can describe the relationship between battery cell usage and ageing where in the end, the model used between six to ten regressors that showed comparable results to the extrapolated end of life. In this study [11], the battery cell capacity degradation threshold was set to be at 70% but most of the battery cells still did not reach this end of life threshold at the end of the study. This required the end of life of these battery cells to be extrapolated for predictions. Some battery cells also had to be taken out from the study due to defects found and these cells were replaced with newer battery cells to maximise the experiment space available.

2.2.2 Data Analysis

The characteristics of the ALICe datasets used is as shown in Table 2.4 below:

Dataset	Battery Type	Chemistry Type	No of Cells	Start Date	End Date	Duration	Cells Reaching EoL	Censored Cells
ALICe1	18650	NMC	122	07-08-2013	25-11-2015	840	37	85
ALICe2	18650	NCA	100	28-11-2016	07-05-2018	525	38	62

Table 2.4: ALICe1 and ALICe2 Dataset Description

The capacity of the battery cells were collected at reference test procedure points and Fig. 2.3 and Fig. 2.4 shows the capacity degradation trends for these two datasets over time.

It can be observed that the battery cells from ALICe1 showed a more linear ageing trend with some battery cells experiencing more rapid ageing than the others. This can explained through the different loadpoint configurations based on the temperature and state of charge values.

Similarly with the ALICe2 dataset, it shows more variability in the ageing of the different battery cells. There are also many occurrences of the battery cells with capacities increasing and dropping over the time.

Upon verification, the oscillations observed could result from temperature conditioning and as tracing of experiment influences is not trivial, it is difficult to determine other reasons for the trend observed. There is also the

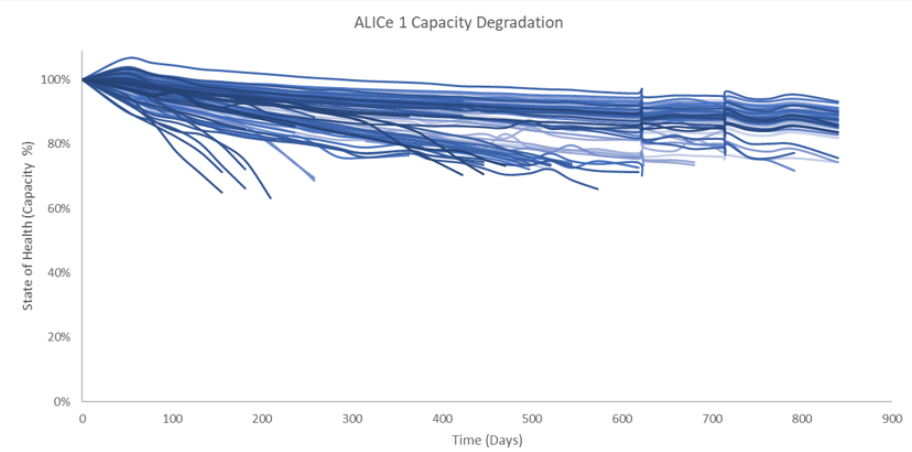


Figure 2.3: ALICe1 Capacity Degradation Trend

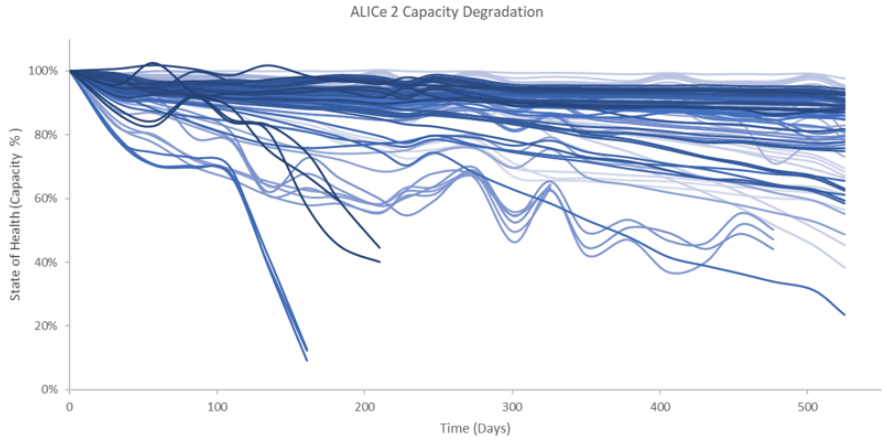


Figure 2.4: ALICe2 Capacity Degradation Trend

possibility of capacity spikes when the battery cell goes into a rest period where some chemical reaction occurs during battery usage and chemical products accumulate near the electrodes. In this situation, the battery goes into a rest period to melt these chemicals and this eventually could lead to self-charging scenario.

The variability of the capacity degradation trend between ALICe1 and ALICe2 datasets is due to the kind of battery cells that were used. For ALICe1, commercial 18650 battery cells were used while for ALICe2 tailored cylindrical battery cells were used instead.

From the information gathered on the capacity degradation over time, the data preprocessing step involves taking the last capacity measurement over the nominal capacity to determine the remaining capacity left in the battery cell at the final reference test procedure. If the remaining capacity percentage is at 80% or less, the battery cell is considered to have reached end of life and the status is labelled as '1', otherwise the cell is still active and is labelled as '0' to indicate that the cell has not reached end of life or what is called censored cell.

3

Underpinnings on Battery Life Prediction

This chapter aims to investigate studies that have focused on predicting battery life. Studies on remaining useful life focus on predicting how long a battery can still operate before reaching its end of life threshold and is usually given by number of remaining cycles left. In predicting battery end of life, the outcome refers to a specific cycle number or time when the battery is expected to reach its end of life threshold, for example, predicting when the state of health (SoH) reaches 80%. This information is particularly useful for battery manufacturers to determine a more accurate approximation of the battery lifespan.

Method and Reference	Data Source	Data Size	Approach Summary
Dempster–Shafer theory (DST) and Bayesian Monte Carlo (BMC), 2011 [12]	CALCE, University of Maryland	4 batteries, 18 and 250 cycles	• RUL prediction from early point • Used two exponential functions to model capacity degradation • Initial model parameter selection based on DST used to get model parameters with highest degree of belief • BMC used to update model parameters based on new measurement • with tuned parameters obtained from BMC, capacity fade extrapolated to provide RUL prediction • Showed that BMC can provide more accurate parameter estimation compared to Extended Kalman Filtering (EKF)
Adaptive Bath-tub Function (ABF),2013 [13]	CALCE, University of Maryland	four batteries 300 cycles	• ABF is used to model the normalised battery cycle to predict RUL based on historical data • Artificial fish swarm algorithm (AFSA) applied to optimise parameter determination of ABF curves
Particle Swarm Optimisation (PSO), 2013 [14]	CALCE,University of Maryland	4 batteries, 80, 95, 100, 126 cycles	• Used statistical time series (Autoregressive (AR)) model to predict near future values • No uniform criteria to determine AR model order • Used RMSE to determine model order • Applied PSO to search for optimal AR model order • Improved accuracy by making AR model order adaptable to change

Unscented Particle Filter (UPF),2013 [15]	CALCE, University of Maryland	4 batteries, cycle 15 and 32	• System state is represented by a p.d.f that is approximated from a set of particles that represent values from an unknown state space • UPF uses UKF to generate proposal distribution to obtain proposal distribution then use standard PF algorithm to get the final results • Exponential growth model used to describe regression process • Depends on accuracy of estimated parameters
Relevance Vector Machine (RVM), 2013 [16]	CALCE, University of Maryland	3 batteries 104, 128, 146, 157, 170, 183 cycles (randomly selected)	• RVM used to find representative training vectors containing cycles of relevance vectors • The conditional three parameter capacity degradation model used to fit the predictive values on the cycles of the relevance vectors • Three conditional parameter model is extrapolated to a failure threshold to estimate RUL
Naive Bayes,2014 [17]	NASA Dataset	12 batteries 18 and 32 cycles (Type A), 250 cycles(Type B)	• Considers different operating conditions • Results comparable to SVM method
Support Vector Regression (SVR),2014 [18]	NASA Dataset	40,60 and 80 cycles	• Developed two models: non-iterative prediction model based on flexible SVR and iterative multi-step prediction model based on SVR • Used current, voltage, time, battery temperature, ambient temperature and correlation between current, voltage and operation time as features • Non-iterative method used to consider regional variation in capacity degradation • Multi-step method used to iteratively predict by using part of current status to predict next state • Flexible SVR suitable for long term while multi-step suitable for short term predictions

Wiener Process with Measurement Error(WPME), 2014 [19]	NASA Dataset	4 batteries 68 cycles	• Used truncated normal distribution (TND) to estimate degradation state and obtain exact and closed form RUL • Maximum Likelihood Estimation (MLE) applied to improve estimation efficiency • Earlier works used Monte Carlo approach to estimate p.d.f which is time consuming • Wiener process introduced to solve this problem • Measurement error considered for possible contamination in CM data
Unscented Kalman Filter (UKF), 2015 [20]	CALCE, University of Maryland	4 batteries, 100 and 160 cycles	• Two step prediction process • Start with UKF to produce sequential estimation of system state • Apply RVR to predict RUL • Applicable for short term capacity prediction • Accuracy increases when more samples are used for training
Gaussian Process Mixture (GPM),2016 [21]	NASA Dataset	4 batteries, 60 and 80 cycles	• A hard cut Expected Maximisation(EM) algorithm used to estimate parameters • GPM is applied to deal with multi-modality observed in the capacity degradation curves • Results showed SVM performing slightly better than GPM but SVM does not produce confidence intervals

Spherical Cubature Particle Filter (SCPF),2016 [22]	NA	26 batteries	• State space model is first constructed to assess capacity degradation • SCPF applied to solve state space model • Adapts a spherical cubature integration based KF to provide importance function of a standard PF • Once state space model is determined, it is then extrapolated to a specified failure threshold to infer RUL • Used exponential model to describe battery degradation trend
Recurrent Neural Network (RNN), 2018 [23]	Online data from real applications offline data from experiments	6 cells with 2 offline cells and 4 online cells	• Uses gaussian distribution of the observed capacity degradation as input • One input layer fed into LSTM layer with 50 neurons and added another layer with 100 neurons • Ends with fully connected layer with one neuron • Uses dropout method to deal with overfitting issue and claims to perform better than SVM approach

The analysis of these papers are based on comparisons made between the groups of techniques surveyed in [24]. Therefore the selected papers are just a subset of many other studies on predicting battery remaining useful life.

Papers that used the dataset provided by CALCE, University of Maryland preceded an initial study done by He et al. The approach in [13] proposed a computational intelligence technique but no comparisons in terms of model performance were made against He et al.'s work.

Preceding [13], Long et al. proposed a Particle Swarm Optimisation technique to optimise results from using the Autoregressive Moving Average (ARMA) model approach. This model was compared against traditional ARMA models and a study on Extended Kalman Filtering (EKF) in [25] where the ARMA approach is said to outperform EKF. The study by [15] proposed a particle based model, the Unscented Particle Filter approach by borrowing the ideas from Particle Filters and Unscented Kalman Filtering where the prediction error rate is claimed to be lower than using PF algorithms that suffer from particle degeneracy problem.

In [16], a Relevance Vector Machine (RVM) was applied where it claims to perform better than the ARMA and EKF approaches studied before. The study also mentions the downside of [12] where initial parameters need to be determined using historical samples while using the RVM approach this step can be ignored by using randomly selected parameters. The parameters using the RVM model can be determined by using the conditional three parameter capacity degradation method against the sum of exponential models used in [12] that makes it difficult to identify the unique parameter values.

In [17], a Naive Bayes (NB) model is applied on the publicly available dataset from NASA and compared against SVM based approaches. It was able to show that using NB method can achieve comparable results to SVM but it also mentions that the NB approach loses accuracy power at the critical point when the capacity of the battery reaches the specified threshold compared to SVM that has better predictive power at this phase. The Support Vector Regression approach proposed in [18] compares two SVR models developed in this study and considers temperature as an input that was not considered in the RVM study by [16]. There were no com-

parisons made in terms of model performance against other approaches except for the difference in features used. The Wiener process method proposed in [19] aims to optimise model computational performance since the probability density function using the Monte Carlo approach applied in previous studies is claimed to be computationally inefficient.

As for the Unscented Kalman Filtering (UKF) method proposed in [20], comparisons were made across several approaches that aim to combine advantages from using Relevance Vector Regression (RVR) and Kalman Filtering based methods. Results show that a combination of the RVR-UKF approach performed best against standalone UKF, RVR and PF-RVR methods. The aim is to recursively update the model parameters from using the state updates from UKF and historical information that is used in RVR for better prediction inference.

The Gaussian Process Mixture (GPM) approach employed in [21] compares its performance with Gaussian Process Regression (GPR) and SVM where it shows that the GPM method produced better prediction accuracies. The prediction accuracy is calculated by how difference between the predicted cycles and the actual cycles. GPM achieved less than 1 cycle of error while GPR and SVM obtained more than three cycles of error. However the prediction accuracy of GPM also depends on the amount of training data available to improve its performance.

The Spherical Cubature Particle Filter (SCPF) approach proposed in [22] deals with problems in using particle based approaches namely the particle degeneracy and impoverishment issues as mentioned in [26]. It compares against standard particle based methods where it is able to improve predictions by applying an importance function that gives more weights to recent capacity data.

With the widespread use of neural network approaches in recent years, [23] employed the Recurrent Neural Network (RNN) approach where results were compared against a SimRNN, PF and SVM approach. Generally the LSTM-RNN approach provides better prediction results but still suffers from longer computational time.

In summary, although the adaptive filtering technique models [20], [15], [22] improves the accuracy of the battery health prognostics but they are highly influenced by the variability in current and temperatures.

Intelligent techniques [23], [18], [16], [14], [13] uses simpler algorithms that achieves better accuracy levels compared to adaptive filter techniques but lack in analysing uncertainty in measurement results [24]. Generally these algorithms also require more training samples to improve prediction accuracy.

The main commonality among all these studies is the availability of cell data reaching end of life that is usually not the case in practice when collecting battery test data from controlled lab experiments. The number of cells used to build and test the prediction models were also small, considering only up to 4 batteries with only two studies that considered at least 12 and 26 batteries. This is important as battery ageing characteristics could differ from one battery to another, therefore using more battery cells would give better generalisation on expected battery lifespan.

A closely related work by Severson et al. used a feature based approach where linear and non-linear features were derived from the raw data. They then applied a regularised linear framework namely the Elastic net method was used. The Elastic net method is a combination of two regularisation methods; the lasso regression and ridge regression. The final model uses a linear combination of a subset of the proposed features to predict the logarithm of the cycle life [27]. The chosen model allowed for domain specific features to be selected. The choice of a linear model gives advantage with low computational cost and can be trained online and offline.

Several features were mainly calculated from the discharge voltage curve which, according to other studies, holds valuable information to capture cell degradation. From this value, summary statistics were calculated such as the minimum, mean and variance.

Using the high predictive influence of the discharge voltage curve, they experimented with three models using the feature combinations as below:

- Model 1: Variance model using only the variance of the discharge voltage curve between cycles 100 and cycle 10

- Model 2: Discharge model using features extracted drawn during discharge

- Model 3: Full model using other features such as temperature and internal resistance

The model was trained by using only the first 100 cycles of data. The training dataset consisted 41 cells, 43 cells were used as the primary test dataset and remaining 40 cells were used as secondary test dataset which was generated after model development. The train and primary test dataset was selected based on splitting two cells with the same charging conditions into train and test datasets, however, this mechanism was not applied to all the cells where in some cases the charging condition is only applied to one cell and is assumed to be tested on another cell which has almost the same end of life characteristics.

The root mean squared error (RMSE) and average percentage error were used to evaluate the models based on unit cycles.

Using the first model with only the log variance discharge voltage curve, they show a test average percentage error of around 15% on the primary test dataset and around 11% on the secondary test average percentage error. For the second model, they select six out of 13 features which resulted in a slightly improved test average percentage error of 13% and 8.6% for the primary and secondary test dataset. Adding more features, this time selecting 9 out of 20 features again reduced the average percentage error to 7.5% and 10.7% for the primary and secondary test dataset.

Their experiments involved using five naive models with various feature combinations. Out of these five benchmarked models, they show that the best performing model used only the discharge slope for cycles 91 to 100 as the feature for the model. Other multivariate models resulted in higher testing error compared to training error suggesting an overfitting problem. The study concludes that using features generated solely from the capacity fade curve does not result in accurate predictions especially for cells with longer cycles. The study also noted that the correlation between the discharge curve and the cycles increased between cycles 100 to 300 but also became non-linear as the cell ageing towards end of life usually depicts non-linearity.

The study also included other interesting findings in their supplementary materials as below:

- Considers capacity fade metric as the criteria for prediction (e.g. 5% capacity fade) instead of using cycle index. This showed a linear ageing trend emerging but this trend appears much later in the battery cell's life, therefore, predictions can also be only made at a later stage.
- Although the correlation coefficient between using the capacity fade metric against cycle life is similar to that of using the log variance of the discharge voltage curve, but using the capacity fade metric will require waiting until the capacity fades by a certain percentage. This delays prediction possibility to a later time especially for battery cells with longer lifetime.

A search on survival analysis for battery lifetime prediction resulted in a few research papers where one of which used heavy fleet data and for a different type of cell chemistry [28]. In another study, an approach to combine survival analysis with neural network is proposed [29]. However, at the time of researching for this thesis, we have not found any studies that employs both survival analysis and transfer learning as methods to predict battery end of life.

4

Transfer Learning

In this chapter, we discuss the theoretical aspects of transfer learning. We investigate the different strategies and approaches that can be considered and conclude on the most appropriate method that can be applied for our use case.

4.1 Transfer Learning Fundamentals

Transfer learning imitates the capacity for humans to learn a new skill based on the knowledge one has already developed on a related skill. For example, if a person is able to play the guitar, he or she may learn how to play the ukulele quicker than someone with no experience in playing string instruments as the foundation of the skill required to learn is similar.

The area of transfer learning is not new. According to [30], researchers were already exploring the idea more than 100 years ago. However, in the machine learning area, it is only in recent times that the approach is being actively researched and implemented. We will not get into the differences in terminologies here as there are many ways in which transfer learning is being described and implemented in literature, therefore, the focus in this thesis is on the general definition and approaches.

"..informally the definition of transfer learning in the field of machine learning is the ability of a system to recognize and apply knowledge and skills in previous domains or tasks to new or novel domains, which share some commonality."

The important part of this definition is the requirement for the domains to share some commonalities which requires preliminary analysis to determine if transfer learning can be implemented or not for this use case.

In Goodfellow et al.'s book on deep learning [31], a simpler definition is given:

"Situation where what has been learned in one setting is exploited to improve the generalisation in another setting."

So far, transfer learning is popular in the domain of image classification and computer vision with the availability of large image datasets. There are only a few examples of implementing transfer models as discussed in Chapter 3 for regression problems even more so for survival models. Therefore, in this thesis, we aim to attempt this approach with the selected survival model.

Transfer learning is first described by having a *domain D* with a feature space X and a marginal probability distribution $P(x)$, where $x \in X$. For this domain, we have a task T that includes the response variable Y and a prediction function $f(\cdot)$. This prediction function is then used to predict new unseen data denoted as x^*. To differentiate between the source and target data, we denote the source domain as D_S and the target domain as D_T. A formal definition is as follows:

Definition 1 *Given a source domain D_S and learning task T_S, a target domain D_t and a learning task T_T, transfer learning aims to help improve the learning of the target predictive function $f_T(\cdot)$ in D_T using the knowledge in D_S and T_S where $D_S \neq D_T$, or $T_S \neq T_T$.*

Based on the definition above, transfer learning only occurs when $D_S \neq D_T$ implying the features $X_S \neq X_T$ or the marginal distribution $P(x_S) \neq P(x_T)$. A task is defined as a the pair of $T = Y, P(y|x)$ where $T_S \neq T_T$ means that either the response variable values $Y_S \neq Y_T$ or the marginal probability distributions $P(y_S|x_S) \neq P(y_T|x_T)$. This differs from conventional machine learning where the domains are the same and the learning tasks are the same. Prior to implementing a transfer model it is also important to address three questions:

- *What to transfer?* addresses which part of the knowledge can be transferred from the source domain to the target domain. Here we identify the commonalities between the source and target domain.
- *When to transfer?* refers to the situation where we need to ensure that the transfer does not make the predictions for the target domain worse which is also known as *negative transfer*.
- *How to transfer?* refers to how to facilitate the transfer process. There are many techniques that can be considered depending on the answers to the *What to transfer?* question.

4.2 Transfer Learning Strategies

According Pan et al.'s, survey on transfer learning [32], there are three main settings in which transfer learning can be implemented as shown in Table 4.1 adapted from the survey as below:

Learning Type	Source and Target Domains	Source and Target Tasks
Machine Learning	the same	the same
Inductive Transfer Learning	the same	different but related
Unsupervised Transfer Learning	the same	different but related
Transductive Transfer Learning	different but related	the same

Table 4.1: Transfer Learning Strategies [32]

In a recent research work [30], a further breakdown was developed to differentiate homogeneous transfer learning and heterogeneous transfer learning. The homogeneous transfer learning concepts takes on from the formal definition given in Definition 1 but in the case of heterogeneous transfer, it is assumed that both the feature spaces and the labels spaces are different and unrelated. Here, our focus will be on homogeneous transfer learning approaches that can consist of instance based, feature based and model parameter based learning as the main ones.

4.3 Homogeneous Transfer Learning Strategies

4.3.1 Instance based approach

The main assumption used for this approach is that the source and target domains have many overlapping features. This allows for some of the labels from the source domain to be reused upon re-weighting and re-sampling for the target domain. The approach can be applied for cases where labelled data is both available or not available in the target domain.

4.3.2 Feature based approach

The main idea behind feature based approach is to learn or identify good feature representations for both the source and target dataset so that the labels from the source domain can be reused for the target domain. It aims to learn to map between the source and target domain dataset in a way that the differences between the feature spaces can be reduced. To date, two ways are discussed where one uses domain knowledge to specify the mapping mechanism while the other does not consider encoding domain knowledge. Further description on the different kinds of feature based approaches can be found in [30].

4.3.3 Model-Parameter based approach

This approach moves away from the data level approaches discussed in method one and two where we assume that the source and target domain shares some prior distributions of the hyper-parameters or parameters of the models. Using this method is based on the situation where we have a model that is considered to have a good generalisation or has been able to learn the underlying expected relationships between the features and response variable that can be transferred to learn a target model based on these known parameters from the source model.

A search on transfer learning for regression problems turned out to be a challenging task as there were not many examples of how transfer learning can be applied for a regression problem.

However, one study was found to be relevant to this use case where the main focus was to discuss covariate shift which is applied in the trans-

ductive learning case [33]. Here, the study investigates ways to adapt an existing regression model to predict for a new situation. One of the examples shown in this study was to use the prediction model trained to predict house prices in one state in the U.S. to predict house prices in other states. Their study explains that in the context of covariate shift, they apply the assumption of the probability density for the new dataset to be different from the learning dataset or source dataset and the regression model does not change. However in real applications, this implementation approach often results in poorer performance of the target dataset with a different probability distribution.

Some works makes an assumption on the probability distribution of the target domain dataset but in Bouveyron's et al.'s work, this constraint is removed by not assuming the distribution of the target domain dataset for the reason that the target domain dataset is often smaller and is not reasonable for distribution estimation.

The paper then discusses the possible transformation models that are narrowed down to seven models. In the more complex models, they introduced a scalar value that is estimated using ordinary least squares. This parameter forms a differentiation factor for the intercept and the regression parameters between the source dataset and the target dataset.

As the study uses a linear regression approach, the parameters are estimated by using the ordinary least square estimation approach. From this study, we chose to first implement one of the simpler models where the model assumes that both the source dataset and target dataset share a similar behaviour, therefore the parameters $\beta_T^0 = \beta_S^0$. The models were evaluated using Prediction Sum of Squares (PRESS) and Bayesian Information Criterion (BIC). The PRESS statistic is a leave one out re-fitting and prediction method which returns a R-squared and P-squared value. The BIC estimates several fitted models where a log-likelihood can be obtained. The BIC evaluation criteria usually favours models with low complexity. Based on these model evaluations, the results for the house prediction model favoured Model 2, 3 and 5 that was also flexible to the size of the dataset.

4.3.4 Conclusion

From the discussions above, several transfer learning strategies were investigated. The fundamental questions of what and how to transfer were discussed. The question on *when to transfer?* can be addressed upon evaluating the prediction performance of the target model. If the predictions for the target dataset becomes worse by using a transfer model compared to using the model trained on the target dataset, then we can conclude that using a transfer learning approach may not be the best way as it can lead to a situation referred to as negative transfer.

To answer the question on *what to transfer?*, initial analysis on the datasets reveals that the ALICe1 and ALICe2 battery cells were tested for the same set of features. The values for each of these features were either the same or different as shown in Table 2.2 and Table 2.3. In this case, since in reality the target domain dataset is usually much smaller than the source domain, we avoid from having any assumptions on the distribution of the target dataset. With this assumption, the model parameter based approach was selected that concentrates on the parameter levels of the datasets instead of the data level. Dealing with the data level approaches will also require in depth domain knowledge to understand how the knowledge transfer can be facilitated. Since access to this information was limited, using the model parameter based approach was considered as the most promising method at this point.

5
Selected Methods

This chapter focuses on discussing the methods chosen to achieve the two main goals for this thesis. The first section discusses the selected approach to predict battery end of life by using survival analysis while the second section discusses the transfer learning strategy selected to reuse the survival model built for the ALICe1 dataset to predict the end of life for ALICe2 battery cells.

5.1 Predicting Battery End of Life using Survival Analysis

In this chapter, we explain the use of survival analysis as a method to predict battery end of life. Survival analysis is a statistical method with the main goal of estimating the time to an event happening. In this case, the time to when the battery cell reaches it's end of life at 80% remaining discharge capacity.

Historically, survival analysis has been widely used in the medical field and biostatistics mainly to determine the effects of treatment on patients through analysing survival rates of patients with critical diseases such as cancer and heart diseases that are known to reduce life expectancy of individuals diagnosed with these ailments. However, the use of survival analysis has spread to other domains since then and includes estimating when a mechanical component will fail.

Battery cell testing often occurs within a limited time period of between six months to two years. Within this duration, many battery cells do not reach the end of life threshold at the end of the study or are removed from the study if a defect occurs while testing. This leads to not being able to

capture the true end of life time point and requires implementing extrapolation methods that are difficult to be validated.

Survival analysis deals with the problem mentioned above by considering data points that have not manifested the intended event as censored data. It is based on the assumption that at some point in the study (for drop outs) or after the study is completed, there is still the possibility that the intended event will occur for these cells but the time of event is unknown. Therefore a binary labeling method is introduced to mark if a cell has reached its end of life during the study or not. This type of censoring is called right censoring.

Two main functions exist when evaluating survival data, the survival function and the hazard function. Let T be the survival time. That is, T is the elapsed time from the beginning point, such as the point when a battery cell reaches 80% available capacity. The values of T can be formulated to follow a probability distribution.

The cumulative distribution function (c.d.f) of the random variable T is given by $F(t)$ as follows:

$$F(t) = Pr(t > T) = \int_0^t f(u)du \quad \text{for} \quad T \geqslant 0 \tag{5.1}$$

The c.d.f gives the probability that the failure event has occurred by time, t. Theoretically, the survival function is depicted as a curve from probability of 1 to 0 as at the end, no subjects are expected to survive. However, using actual datasets, this curve is often denoted in steps as observation studies are usually done within a defined timeframe and some subjects observed may not experience the intended event by the end of the study or drops out. This could also be influenced by other factors that needs to be accounted for when plotting the survival curve. The formula for estimating the survival rate, $S(t)$ is given as below:

$$S(t) = P(T \geqslant t) = 1 - F(t) \tag{5.2}$$

In contrast to the survival function, the hazard function is the probability that a subject experiences death within a small time interval between t

and δt given that the subject has survived up to time t. By definition, the p.d.f can be described as below:

$$f_T(t) = \lim_{\delta t \to 0} \frac{F_T(t+\delta t) - F_T(t)}{\delta t} \tag{5.3}$$

The hazard function can be derived as follows:

$$\begin{aligned} h_T(t) &= \lim_{\delta t \to 0} \frac{F_T(t+\delta t) - F_T(t)}{\delta t \cdot S_T(t)} \\ &= \lim_{\delta t \to 0} \frac{Pr(t \leq T < t+\delta t)}{\delta t \cdot S_T(t)} \\ &= \lim_{\delta t \to 0} \frac{Pr(t \leq T < t+\delta t \mid T \geqslant t)}{\delta t} \end{aligned} \tag{5.4}$$

The equation above can then be rewritten as below:

$$h_T(t) = \frac{f_T(t)}{S_T(t)} \tag{5.5}$$

The denominator in Eq. 5.3 is the width of the interval. Dividing one by the other we obtain a rate of event occurrence per unit of time. Taking the limit as the width of the interval goes down to zero, we obtain an instantaneous rate of occurrence.

The cumulative hazard function $H(t)$ gives the sum of the individual hazard rates from time zero to time t. The formula for the cumulative hazard function is:

$$H(t) = \int_0^t h(u)du \tag{5.6}$$

Therefore, the hazard function is the derivative, or slope, of the cumulative hazard function. The cumulative hazard function is related to the cumulative survival function by the expression:

$$S(t) = \exp^{-H(t)} \tag{5.7}$$

Or, the equation above can also be written as:

$$H(t) = -\log(S(t)) \tag{5.8}$$

Here we show that the survival function, hazard function and distribution function are mathematically related. For a graphical understanding of how these functions relate to each other, refer to Fig. 5.1 as shown below:

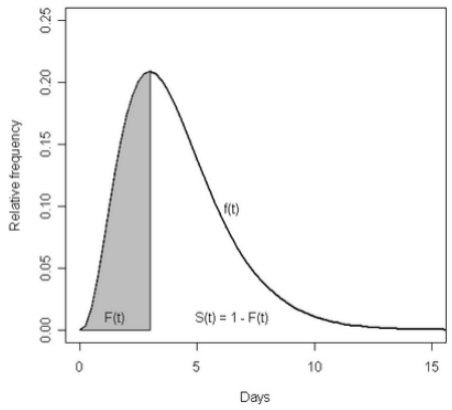


Figure 5.1: Line plot of $f(t)$ as a function of time [34]

We can describe the count of battery cells that reached end of life as a function of time. The curve fitted to the distribution function reflects the end of life density function referred as $f(t)$. According to [34], if the area under the curve of the end of life density function is set to 1, then for any given time t, the area under the curve to the left of t represents the proportion of cells that have experienced end of life. The proportion of battery cells that have reached end of life as a function of t is known as the cumulative end of life distribution function, $F(t)$ as depicted in Eq. 5.1.

As for the survival function, $S(t)$, the area under the curve on the right side of Fig. 5.1 reflects the proportion of battery cells that survived to time t. The survival function, $S(t)$, can also be plotted as a function of time that is shown by a Kaplan-Meier curve that is further explained later in this chapter. When $t = 0$, $S(t) = 1$ as there are no deaths experienced at the first time point. When $S(t) = 0$, it reflects that all the battery cells have reached end of life. Survival curves are shown in step form since we are counting battery cells that are still alive at every time point.

The hazard rate as explained in Eq. 5.4 reflects the instantaneous rate of any randomly selected cells that is still alive at time $(t-1)$ to die at time t. In other words, it is the conditional failure rate of a cell experiencing end

of life at time t given that is it still alive at time $t-1$ as reflected in function $f(t)$.

The cumulative hazard $F(t)$ as shown in Eq. 5.6, is reflected by the area under the curve until time t where the cumulative hazard curve shows the cumulative probability that the end of life has occurred up to any point in time.

5.1.1 Censoring

The uniqueness of using survival analysis methods lies in its ability to handle censored observations. This situation often occurs when observing time to event for a limited duration where not all subjects of interest will experience the event being observed. There are several types of censoring as below:

- Right censoring: a battery cell is considered right censored if it has not experienced end of life at the end of the experiment.
- Left censoring: if a battery cell has experienced end of life prior to the start of the experiment, then the battery cell is considered left censored. As this situation is not applicable for our application, it will not be considered.
- Interval censoring: this situation occurs when an event is expected to occur between two time points and the exact point of end of life is not known.

5.1.2 Likelihood Function for Censored Events

A question that may arise is on how the end of life values for censored events are estimated. For this purpose, it is important to understand the likelihood function for survival and the maximum likelihood estimator that is used to derive the coefficients for the survival models.

The maximum likelihood function aims to find the most likely values for distribution parameters by maximising the value of the likelihood function. The likelihood function is actually derived from the probability density function as follows:

$$f(x;\theta_1,\theta_2,...,\theta_k)$$

Where, x are the data points that experienced failure and $\theta_1, \theta_2, ..., \theta_k$ are the parameters to be estimated. Therefore for a dataset with all datapoints experiencing failures or end of life, the likelihood function will be the product of the p.d.f functions as shown below:

$$L = \prod_{i=1}^{n} f(x_i; \theta_1, \theta_2, ..., \theta_k) \tag{5.9}$$

Where n is the number of datapoints that experienced failures and x_i is the i-th failure timepoint. The log-likelihood function can then be formed as below:

$$\log L = \sum_{i=1}^{n} \log f(x_i; \theta_1, \theta_2, \theta_3, ..., \theta_k) \tag{5.10}$$

The equations above considers datasets without censored failure points or complete datasets that uses the probability density function. In the event that censored observations are included, the equation is extended to include the c.d.f as in the equation below:

$$L = \prod_{i=1}^{n} f(x_i; \theta_1, \theta_2, ..., \theta_k) \cdot \prod_{j=1}^{m} [1 - F(y_j; \theta_1, \theta_2, \theta_3, ..., \theta_k)] \tag{5.11}$$

Where m is the number of censored data points, y_j is the jth censored data point and $F(y_j; 1, 2, ..., k)$ is the c.d.f.

Similarly, the log likelihood for the likelihood function above is formed as below:

$$\log L = \sum_{i=1}^{n} \log f(x_i; \theta_1, \theta_2, \theta_3, ..., \theta_k) \cdot \sum_{i=1}^{m} \log[1 - F(y_j; \theta_1, \theta_2, \theta_3, ..., \theta_k)] \tag{5.12}$$

The likelihood function that includes censored observations shows one of the advantages of using the maximum likelihood estimator (MLE) approach that considers these censored observations to estimate the parameters.

5.1.3 Survival Analysis Implementation Methods

There are several ways in which survival analysis models can be implemented. The main categories fall into three categories, namely, non-parametric, semi-parametric and parametric models. A survey on survival analysis methods [35] further differentiates survival analysis approaches between statistical methods, machine learning and related topics as shown in Fig. 5.2.

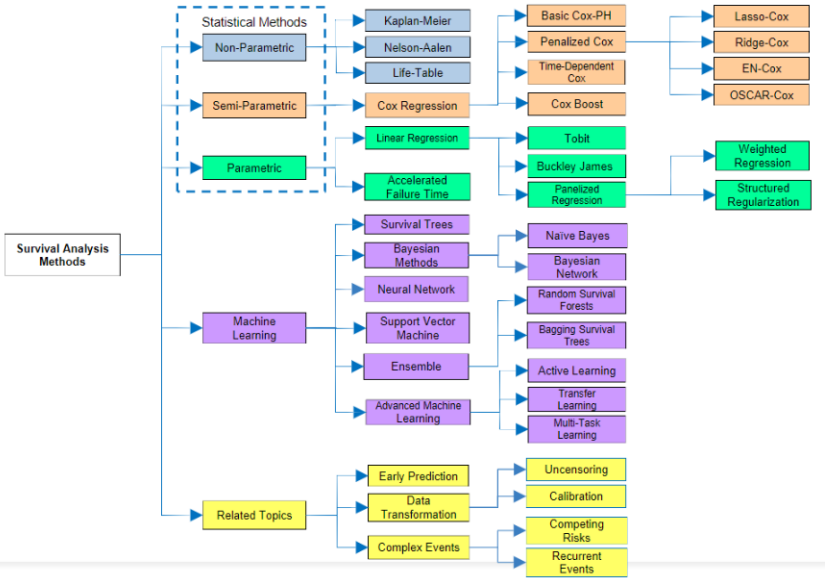


Figure 5.2: Survival Analysis Methods [35]

Based on the categories and types of survival analysis models, it is important to investigate the type of survival model that is suitable to answer the research question. Furthermore, based on the categorisation in Fig. 5.2, this thesis aims to explore statistical methods coupled with the transfer learning or multi-task learning from the machine learning methods.

Non-parametric Survival Model

There are two main non-parametric methods namely the Life Table or Actuarial Table and the Kaplan-Meier approach. The Life Table is commonly used in the insurance industry where time is defined as equally spaced intervals and the approach is usually used when dealing with large datasets with events happening within a time duration. A disadvantage of using

the Life Table approach is the dependency of the survival probabilities on the defined time interval where these values could change according to the change in the time intervals specified [36]. This issue is more apparent when dealing with small datasets, hence, we will not use this approach for this use case.

The Kaplan-Meier(KM) approach, on the other hand, calculates the survival probability when an event occurs and re-estimates this value whenever the intended event occurs. It also assumes that censored data is independent from the likelihood of the intended event happening and considers subjects that are recruited either in the beginning or later in the study. The KM plot provides a visual presentation of the survival probabilities for one or more groups.

The formula used to calculate the survival probability whenever the intended event occurs is as below:

$$\hat{S}(t) = \prod_{j:T_j<t} (1 - \frac{d_j}{r_j}) \tag{5.13}$$

where:
$T_i ... T_k$: a set of distinct times where the event is observed in the sample
d_j : number of events at T_j
c_j : number of events that are censored between T_j and T_{j+1}
r_j : number of cells that are at risk (still alive) before the j^{th} end of life.

The equation below gives the value of r_j

.

$$r_j = r_{j-1} - d_{j-1} - c_{j-1} \tag{5.14}$$

At each time point when a battery cell is observed to have reached end of life, the survival probability is calculated by the number of battery cells that are still alive after the observed incident has occurred over total cells that are at risk of reaching end of life. Battery cells that are censored are removed from the denominator. The total probability until the current time, $S(t)$ is then calculated by multiplying the survival probabilities prior to the current time to produce the cumulative probability value. It is a conditional probability to describe the probability of a battery cell surviving past time, t, given that it has survived until time t.

The log-rank test is another non-parametric approach that uses hypothesis testing where the null hypothesis H_o = no difference between the survival curves for battery cell chemistry A and battery cell chemistry B.

The log-rank test is used when the dataset is not normally distributed and contains censored data while the Wilcoxon Rank test would be employed when the data is assumed to be normally distributed and does not contain censored data. One of the main differences of these two approaches is that log-rank test gives equal weight to all timepoints while the Wilcoxon test gives higher weights to events observed at earlier timepoints.

The log-rank test compares the estimates of the survival function of two or more groups at each observed event time. It computes the observed and expected number of events for a group at each time when an event occurs and sums up these values to get the overall summary across all time points whenever an event is observed.

Using the log-rank test approach, here we consider two battery cell chemistry; A and B. Let $1,...,j$ be the distinct indexes of time points when the end of life of a cell is observed and let i be the different groups being compared. Let $N_{\mathrm{A,j}}$ and $N_{\mathrm{B,j}}$ be the number of battery cells at risk and $O_{\mathrm{A,j}}$ and $O_{\mathrm{B,j}}$ be the observed battery cells reaching end of life at timepoint j. $N_{\mathrm{j}} = N_{\mathrm{A,j}} + N_{\mathrm{B,j}}$ and $O_{\mathrm{j}} = O_{\mathrm{A,j}} + O_{\mathrm{B,j}}$.

The null hypothesis H_o is that the two survival functions are identical $H_o = h_1 = h_2$.

The expected value for the end of life distribution is given by:

$$E_{i,j} = O_j \frac{N_{i,j}}{N_j} \tag{5.15}$$

From computing the expected number of events for each group, we then use the formula below to sum up the values and calculate the log-rank χ^2.

$$LogRank = \frac{(\sum O_{i,j} - \sum E_{i,j})^2}{\sum E_{i,j}} \tag{5.16}$$

Semi-parametric Survival Models

The objective of the Cox-PH model is to examine the effects of covariates on the hazard rate of an event happening at a particular timepoint, in this case, the battery cell end of life.

The Cox-PH model uses the hazard function denoted as $h(t)$ to determine the hazard rate of subjects at time t. The two main components that makes this model semi-parametric is by having an unspecified baseline hazard where no assumptions are made on the survival distribution. As mentioned above, this function calculates the hazard rate of a battery cell dying at time t and is estimated using the formula below:

$$h(t) = h_0(t)\exp(b_1x_1 + b_2x_2 + b_3x_3...b_nx_n) \tag{5.17}$$

where:
$h(t)$: is the hazard function defined by a set of n covariates $(x_1, x_2, ..., x_n)$
$h_o(t)$: baseline hazard or the intercept which corresponds to the value of the hazard. (e.g.: if all x_i=0 then $exp(0)$ = 1. The t in $h(t)$ indicates that this hazard may vary over time
$b_1...b_n$: measurement of the impact of the covariates.

The Cox-PH model makes several assumptions that needs to be tested in order to determine if the selected model is appropriate for the dataset used. The proportional hazards component comes from having the linear predictors linked through the log transformation to the hazard ratio. If the hazard ratio obeys the proportional hazards assumption, and thus does not depend on time, the formula can be re-written as follows:

$$\log[HazardRatio(x)] = \log\frac{h(t|x)}{h_0(t)} = \beta_1x_1 + ... + \beta_px_p \tag{5.18}$$

Here, $h(t|x)$ is the hazard at time t for an observation with covariate value x, and $h_0(t)$ is the baseline hazard function, defined as the hazard at time t for observations with all predictors equal to zero. Therefore the value of the baseline hazard may change with time.

Within the process of investigating the appropriateness of using the Cox-PH model for this use case, we discover the Cox-PH approach models the

hazard rate yielding prediction results in the form of hazard rate values. As the goal of the thesis is to predict time to event, this approach was deemed inappropriate for the use case intended.

However, using the Cox-PH approach provided further understanding on the influence of each covariate on the survival rate and hazard rate as depicted in Fig. 6.2 in Chapter 6.

Given that the goal of this thesis is to predict when a battery cell is expected to reach end of life, other survival model approach needs to be considered. This requirement points us towards adopting parametric models specifically accelerated failure time models that are more commonly used in analysing reliability data [37].

Parametric Survival Models

As an alternative to the limitations found in the Cox-PH regression model, we investigate the use of Accelerated Failure Time (AFT) models that can model the effects of covariates on survival time as opposed to the hazard rate in the Cox-PH model. It assumes that the effect of the covariate can either accelerate or decelerate survival time and is useful when modeling end of life of mechanical components that results from some underlying mechanical process that affects the performance over time.

Parametric methods are used when some assumptions are made on the baseline hazard or survival distribution where a predictable pattern can be observed. An advantage of using parametric methods is its ability to predict time to event for periods after an event has occurred that is often used when we need to predict failure time. It is often applied in reliability analysis and engineering.

In reliability or lifetime data analysis, the common parametric distributions that are applied are the exponential, Weibull, log-normal and log-logistic distributions [37]. As we are interested in predicting time to event, the AFT model defines the relationship of the survival function for every time $t \in T$, $S(t|X)$ and the covariates as shown in the formula below:

$$S(t|X) = S_o[t \exp(\beta^t X)] \tag{5.19}$$

Here, T is the random variable of survival times and X is a column vector of covariates $X_1, X_2, ..., X_p$. S_o is the baseline survival function and $\beta^t = (\beta_1, \beta_2, ..., \beta_p)$ is a vector of regression coefficients and $n \in N$.

The factor $\exp(\beta^t X)$ is known as the accelerator factor which accelerates the survival function with covariate $X = 0$. The AFT model assumes that the effects of the covariates are fixed and is multiplied by the accelerated factor on the time scale, t.

The relationship between the covariates and the survival time can also be expressed as a linear relation between the logarithm of survival time and the covariate X as follows:

$$Y = \log(T) = \mu + \theta^t X + \sigma W \tag{5.20}$$

Where μ is the slope, $\sigma > 0$ is an unknown scale parameter, $\theta^t = (\theta^1, \theta^2, ..., \theta^p)$ is a vector of regression coefficients, $\theta = -\beta$, σ is a scale parameter and W is a distribution error that is a random variable that follows a certain parametric distribution. For every distribution of W, there is a related parametric for T. Some of the commonly used distributions are described in the table below:

Distribution	$f(t)$	$h(t)$	$H(t)$	$S(t)$
Exponential	$\lambda \exp[-\lambda t]$	λ	λt	$\exp[-\lambda t]$
Weibull	$\lambda p t^{p-1} \exp[-\lambda^p]$	$\lambda p t^{p-1}$	λt^p	$\exp[-(\lambda t)^p]$
Log-logistic	$abt^{b-1}/(1+at^b)^2$	$abt^{b-1}/1+at^b$	$log(1+at^b)$	$(1+at^b)^{-1}$
Log-normal	$\frac{\exp^{-((\log((x-\theta)/m))^2/(2\sigma^2))}}{(x-\theta)\sigma\sqrt{2\pi}}$	$\frac{(\frac{1}{x\sigma})\phi(\frac{\log x}{\sigma})}{\Phi(\frac{-\log x}{\sigma})}$	$-\log(1-\Phi(\frac{\log(x)}{\sigma}))$	$1-\Phi(\frac{\log(t)-\mu}{\sigma})$

Table 5.1: Parametric Survival Distributions [34]

$f(t)$ = probability density

$h(t)$ = instantaneous hazard

$H(t)$ = cumulative hazard

$S(t)$ = survival rate

Below gives the description of the shape for each kind of distribution based on the explanations given in [38]:

1. Exponential Distribution

The mean λ describes the exponential function and using this function, the instantaneous hazard is assumed to remain constant with time. To assess if the survival distributions follow an exponential distribution, we can plot the instantaneous hazard, $h(t)$ as a function of time that yields a straight line. The log of the cumulative hazard plotted against the log of time yields a 45 degree slope.

2. Weibull Distribution

   Two parameters describe the Weibull distribution which are the scale, λ parameter and the shape parameter,p. There are three shapes in which the Weibull function can be defined. If shape, $p <$ 1, this results in a monotonically decreasing instantaneous hazard, $h(t)$ with time, $p = 1$ yields a constant hazard with time (exponential) and $p > 1$ reflects an instantaneous hazard that increases with time. To check if the data follows the Weibull distribution, we plot the log cumulative hazard as a function of log time.

3. Log-normal Distribution

   The log-normal model is derived from taking the logarithm of random variable X where the lifetime of X is assumed to have a normal distribution. It reflects the assumption made when using a normal distribution taking the sum of independent variables that are identically distributed as well. This distribution works effectively when the data does not have censored observations but the computation increases when censored observations are involved [38]. It has also been observed that the hazard form of the distribution decreases at the tail end which is not suitable for situations where hazard rate is expected to increase with time.

4. Log-logistic Distribution

   The log-logistic model is another flexible parametric model where the hazard rate can be decreasing, increasing or takes on a hump-shaped form where the hazard rate first increased than decreases towards the end. Another feature of the log-logistic model is its similarity with the log-normal distribution which can be further observed when comparing prediction results.

5.1.4 Distribution Fit Selection

One way to assess distribution fit is by using non-parametric visualisation methods where we observe how well the parametric distribution fits the survival function, $S(t)$. The limitation with this method is that we do not consider the influence of the covariates which may change the shape of the function. However, as a first step to obtain a rough estimation and assurance on the assumed distribution, we show the different distribution pattern against the non-parametric survival rate as shown in Fig. 6.6 - Fig. 6.9.

The second method is by using domain knowledge where the life dataset is assumed to follow a certain distribution. Using this approach, the Weibull distribution is often selected. However, care is needed as when theoretical assumptions are used, the shape of the time dependency is difficult to be determined when covariates are added into the model.

In Chapter 6, we show the implementation of the selection method to decide which distribution shape to use for the parametric model method.

5.2 Transferring Knowledge From ALICe1 to ALICe2

Referring back to Fig. 5.2, we now consider implementing a machine learning approach that can reuse the prediction model build on the ALICe1 dataset to predict the end of life of battery cells in the ALICe2 dataset of a different cell chemistry type. In this category, we focus on implementing a transfer learning approach where we consider the different battery chemistries to be of different domains. Therefore, the ALICe1 battery cells is considered as the source domain while the ALICe2 battery cells is considered as the target domain.

Several transfer learning approaches were investigated as described in Chapter 4. As shown in Table 4.1, we consider the problem we are dealing with here to be in the transductive transfer learning category as we are dealing with related domains of different battery chemistry types but for the same prediction task to predict the end of life for the battery cells.

Within the transductive learning approach, the main idea is to achieve some level of generalisation in a model so that it will be able to perform

well when predicting unseen observations. Here, our aim is to build some level of generalisation in terms of the ageing patterns observed so that the knowledge captured from predicting the end of life for the ALICe1 battery cells can be used to improve the prediction of the battery cells end of life in ALICe2.

According to [32], in the transductive learning strategy, the target domain would be expected to have no labelled data available. This needs to be further clarified where there could be two approaches within the transductive learning approach that can be considered as below:

- Feature spaces between source and target domain is different
- Feature spaces between source and target are the same $X_S = X_T$ but the marginal probability distributions are different $P(X_S)$!= $P(X_T)$

Therefore, here we need to emphasise that we are considering the second case because we are assuming that the feature spaces are the same between the two domains but marginal probability distributions are different. This will allow for also using some data collected in ALICe2 to augment the model as shown in the models considered.

To facilitate the transfer of knowledge from ALICe1 to ALICe2 ageing prediction, we employ the model parameter based approach. This approach was selected because it allows for no assumptions to be made on the ageing distribution of the target dataset that is usually small and unknown. Furthermore, time limitations did not permit to consider other possible methods such as the instance based and feature based approaches which would have also required deep domain knowledge at the data level.

Formally the mapping of the parameters can be written as follows:

$$f(x) = \langle \theta, x \rangle = \theta^T x = \sum_{i=1}^{m} \theta_i x_i \tag{5.21}$$

From this mapping, several models can be derived from making assumptions on the similarities shared between the datasets either with the coefficients or the intercept.

Inspired by the work done in [33] for transfer learning in regression cases, we consider that we may not know the distribution of the target dataset

since the idea is to also reduce the requirement for labelled target data. Therefore, several models can be derived and experimented as follows:

- Model 1: Assume that the source and target dataset shares the same coefficients and intercepts stays the same
- Model 2: Assume that the source and target dataset shares the same intercepts but not the model parameters
- Model 3: Assume that the source and target dataset shares the same intercept and model parameters
- Model 4: Combine the weighted sum coefficient and weighted sum covariance to form the new coefficient and covariance for the ALICe2 prediction model. Here we are using the knowledge gained from both datasets to improve the model performance.

In model 4, we aim to test the possibility of leveraging the ageing patterns observed from fewer observations collected from the target dataset to augment the prediction model for better performance. Eventhough we may not be expecting enough target data available, but the idea is mainly to show whether combining knowledge gained from different types of battery cell chemistries could actually further improve the generalisability of the model.

6

End of Life Prediction Model

In this chapter, we test on several survival analysis approaches that may be applied to model and predict the end of life for ALICe1 battery cells. The search for the best approach here was done purely through independent investigation of approaches within the survival analysis family of models.

6.1 Survival Analysis

Although survival analysis is the approach selected to build the battery end of life prediction model, we still needed to consider which type of survival model would be most suitable for the intended use case. There were no specific direction on the most appropriate survival model approach to be considered, therefore we researched and tested the different algorithms to reach a favourable conclusion.

6.1.1 Non-Parametric Survival Data Analysis

A non-parametric approach simply means that we are not considering any parameters or variables in the model.

Using the non-parametric approach, we can visualise and understand the survival patterns between different populations. Most of the time, the Kaplan-Meier approach is applied as explained in Chapter 5.

A tabular example of how the survival probability is calculated in R using the survfit function is as shown in Listing 6.1 for the ALICe1 dataset and Listing 6.2 for the ALICe2 dataset.

Listing 6.1: Kaplan-Meier Table for ALICe1 Fit

```
> summary(fit_KM)
Call: survfit(formula = Surv(Days, Status) ~ 1, data = a1)

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
  104    122       1    0.992 0.00816        0.976        1.000
  105    121       1    0.984 0.01150        0.961        1.000
  155    120       2    0.967 0.01612        0.936        0.999
  212    113       1    0.959 0.01811        0.924        0.995
  235    112       1    0.950 0.01987        0.912        0.990
  259    101       2    0.931 0.02351        0.886        0.979
  285     99       1    0.922 0.02508        0.874        0.972
  288     98       1    0.912 0.02653        0.862        0.966
  333     95       1    0.903 0.02794        0.850        0.959
  339     94       1    0.893 0.02925        0.838        0.952
  363     93       7    0.826 0.03645        0.758        0.901
  387     85       2    0.807 0.03810        0.735        0.885
  422     83       7    0.739 0.04269        0.659        0.827
  446     54       1    0.725 0.04403        0.644        0.817
  470     53       3    0.684 0.04749        0.597        0.784
  520     50       1    0.670 0.04847        0.582        0.772
  658     46       1    0.656 0.04955        0.565        0.760
  680     45       1    0.641 0.05055        0.549        0.748
  714     44       1    0.626 0.05146        0.533        0.736
  740     43       1    0.612 0.05228        0.518        0.723
> summary(fit_KM)$table
   records      n.max    n.start     events     *rmean *se(rmean)
median    0.95LCL    0.95UCL
 122.00000  122.00000  122.00000   37.00000  672.60139   23.10079
NA          NA          NA
```

Listing 6.2: Kaplan-Meier Table for ALICe2 Fit

```
> summary(fit_KM)
Call: survfit(formula = Surv(Days, Status) ~ 1, data = a2)

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   57     97       2    0.979  0.0144        0.952        1.000
   84     95       2    0.959  0.0202        0.920        0.999
  109     93       2    0.938  0.0245        0.891        0.987
  134     90       4    0.896  0.0310        0.838        0.959
  161     86       2    0.876  0.0336        0.812        0.944
  183     84       3    0.844  0.0369        0.775        0.920
  210     81       3    0.813  0.0397        0.739        0.895
  301     78       1    0.803  0.0406        0.727        0.886
  350     77       2    0.782  0.0421        0.703        0.869
  379     75       2    0.761  0.0435        0.680        0.851
  406     73       3    0.730  0.0453        0.646        0.824
  431     69       3    0.698  0.0469        0.612        0.796
  455     66       1    0.687  0.0474        0.601        0.787
  477     65       2    0.666  0.0482        0.578        0.768
  504     63       2    0.645  0.0489        0.556        0.748
  525     58       1    0.634  0.0493        0.544        0.738
> summary(fit_KM)$table
   records      n.max    n.start     events     *rmean *se(rmean)
median    0.95LCL    0.95UCL
  97.00000   97.00000   97.00000   35.00000  435.31360   15.24688
NA          NA          NA
```

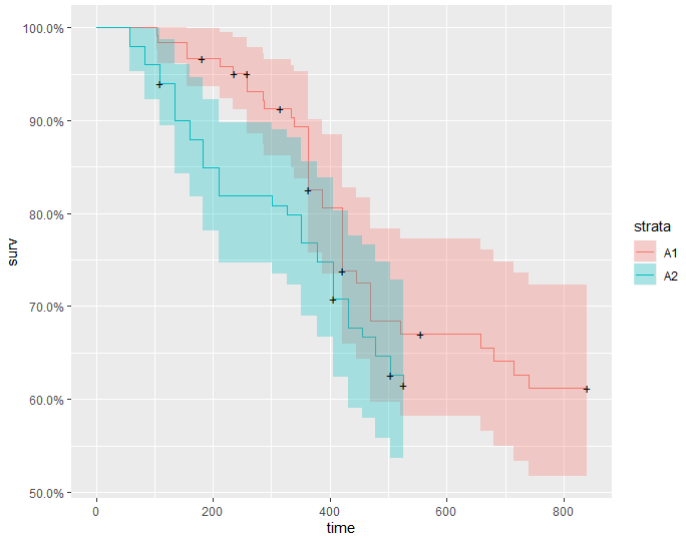


Figure 6.1: Kaplan-Meier Curve for ALICe1 and ALICe2 Dataset

As shown in Fig. 6.1, the drop in the step signifies when battery cells are experiencing end of life and the plus sign points indicate when battery cells are censored from the dataset. As the ALICe2 experiment period was shorter than ALICe1 we can only compare the distributions upto about 500 days.

Based on this non-parametric curve alone, the survival rate for ALICe1 seem to begin to rapidly drop from the point of about 400 days into testing and reaches a plateau again between about 550 and before 650 days into testing.

The plateaued trend observed for ALICe1 after 500 days of testing needs to be further investigated but this is due to no battery cells reaching the end of life threshold during this time window.

As for ALICe2, the survival probability showed a decrease much earlier than ALICe1 and rapidly drops until it reached a plateau between 200 to 300 days into the experiment.

Upon verification, the trend observed is as expected where three phases of capacity fading occurs. The first phase involves strong ageing in the beginning followed by a second phase of plateau ageing before a third phase of accelerated ageing taking place in the end.

To compare the survival rates between the two types of cell chemistries, we run the log-rank test using the survdiff function in R. The null hypothesis of the log-rank test assumes that the survival distributions of ALICe1 and ALICe2 are statistically the same.

Listing 6.3: Log-Rank Results

```
Call:
survdiff(formula = Surv(data$Days, data$Status) ~ data$ChemType,
    data = data)

                    N Observed Expected (O-E)^2/E (O-E)^2/V
data$ChemType=A1  122       37     40.4     0.289     0.707
data$ChemType=A2   97       35     31.6     0.370     0.707

 Chisq= 0.7  on 1 degrees of freedom, p= 0.4
```

The results in Listing 6.3 show that the test statistic is at 0.7. Using the Chi-Square value table, the value given at 1 degree of freedom for a p-value of 0.05 is 3.8414. The value of 0.7 is indeed smaller than 3.8414 that suggest that the difference between the observed results and the expected results is significantly small.

Based on these results, we cannot reject the null hypothesis that the survival rates for these two battery cell chemistries are statistically the same. However looking back at the Kaplan-Meier curve, it is now difficult to assume that the survival probabilities of these chemistries are indeed similar as the log-rank results may be influenced by the survival similarities observed towards the tail-end of the dataset.

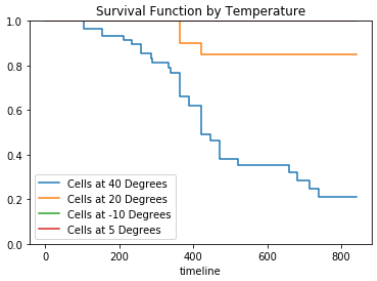
As the log-rank test is a non-parametric method, we will need to use approaches that allows the use of explanatory variables such as a semi-parametric or parametric approach. However, this non-parametric approach is helpful to at least determine if some similarities in the ageing trends can be observed between the different types of battery cell chemistry.

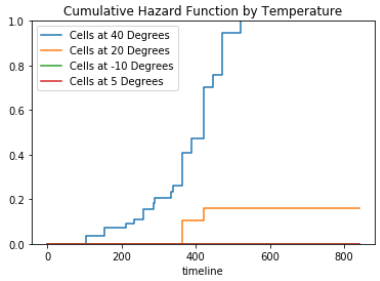
Furthermore, since these are univariate models, it does not take into account the effects of the covariates with the ageing of the cells which are required for the purpose of predicting battery end of life. Nevertheless, it is a useful approach as an initial step to understand the survival trend in different datasets.

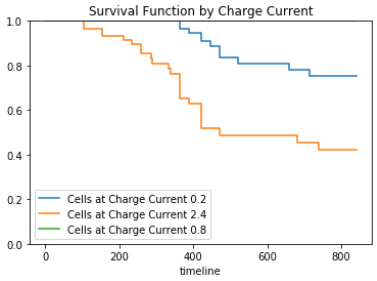
6.1.2 Semi-Parametric Survival Model

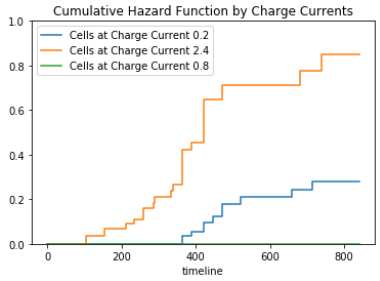
The second approach in survival analysis is called the semi-parametric survival model. Here, the main assumption is that the Cox-PH model can be used to model the ALICe1 dataset.

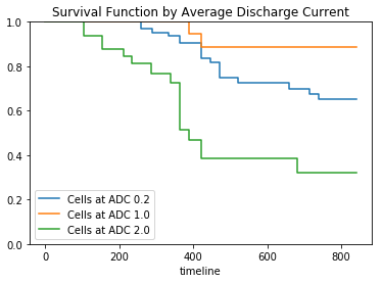
As discussed in Chapter 5, we first employ visual inspection by plotting the survival function and cumulative hazard rates for each covariate. It can also be used to determine if the values of the covariate influences survival time.

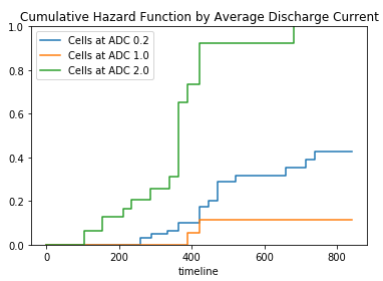












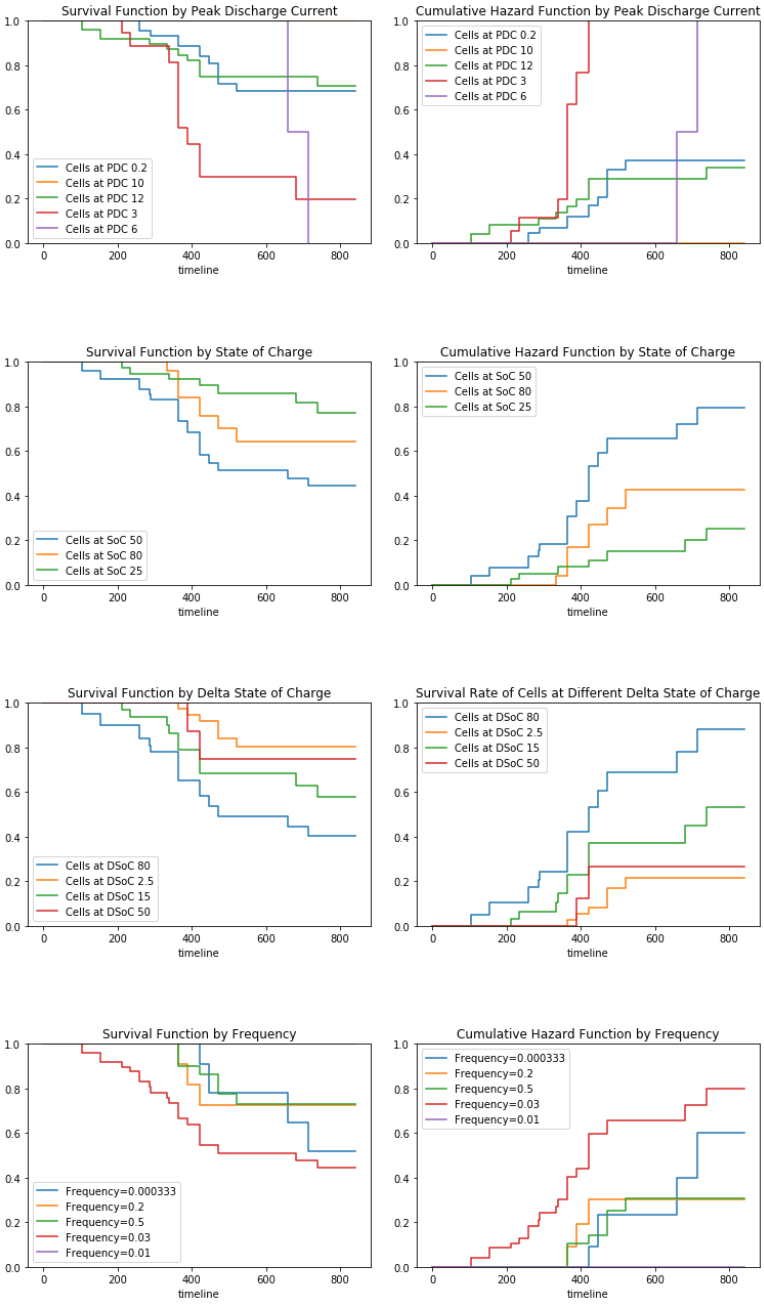


Figure 6.2: Survival curves and cumulative hazards for ALICe1 covariates

The plots above clearly showed that the survival probabilities are influenced by the different load point factors. For example, if we look at the influence of temperature, battery cells that are exposed to higher temperatures degraded quicker than those exposed to medium temperatures.

However, this is a weak conclusion as there was not enough survival data to show the influence of battery cells exposed to other temperature values as most of these battery cells did not experience end of life during the observation period. Generally there were clear differences observed when battery cells were exposed to different discharge currents and the swing in state of charge that can help to confirm how the different influencing factors contributes to the rapidity of the battery cell ageing process.

Cox-Proportional Hazard (PH) Model

The objective of the Cox-PH model is to examine the effects of covariates on the hazard rate of an event happening at a particular timepoint, in this case, the cell end of life.

The Cox-PH model uses the hazard function denoted as $h(t)$ to determine the hazard rate of subjects at time t as shown in Eq. 6.1. The two main components that makes this model semi-parametric is by not having the baseline hazard specified and no assumptions are made on the survival distribution. As mentioned above, this function calculates the hazard rate of a battery cell dying at time t and is estimated using the formula below:

$$h(t) = h_0(t)\exp(b1x1 + b2x2 + b3x3...bnxn) \tag{6.1}$$

where:
$h(t)$: is the hazard function defined by a set of n covariates $(x_1, x_2, ..., x_n)$
$h_0(t)$: baseline hazard or the intercept which corresponds to the value of the hazard. (e.g.: if all x_i=0 then $exp(0) = 1$. The t in $h(t)$ indicates that this hazard may vary over time
$b_1...b_n$: measurement of the impact of the covariates.

The Cox-PH model makes several assumptions that needs to be tested in order to determine if the selected model is appropriate for the dataset used. The proportional hazards component comes from having the linear predictors linked through the log transformation to the hazard ratio. If the hazard ratio obeys the proportional hazards assumption, and thus does not depend on time, the formula can be re-written as in Eq. 6.2.

$$log[HR(x)] = log\frac{h(t|x)}{h_0(t)} = \beta_1 x_1 + ... + \beta_p x_p \tag{6.2}$$

$h(t|x)$ is the hazard at time t for an observation with covariate value x, and $h_0(t)$ is the baseline hazard function, defined as the hazard at time t for observations with all predictors equal to zero. Therefore the value of the baseline hazard may change with time.

Model Fitting

Initial results by just using the seven covariates without further feature engineering and trained on all datapoints results in the following for the ALICe1 dataset as shown in Listing 6.4 :

Listing 6.4: Log-Rank Results

```
Call:
coxph(formula = Surv(Days, Status) ~ Temp + CC + ADC + PDC +
    Freq + SoC + DSoC, data = a1)

          coef exp(coef)  se(coef)      z        p
Temp  0.171810  1.187452  0.036421  4.717 2.39e-06
CC    1.233561  3.433435  0.250035  4.934 8.07e-07
ADC   0.821447  2.273789  0.227741  3.607  0.00031
PDC  -0.001387  0.998614  0.041903 -0.033  0.97359
Freq  3.698333 40.379948  1.458322  2.536  0.01121
SoC   0.031827  1.032339  0.010089  3.155  0.00161
DSoC  0.030267  1.030730  0.005755  5.259 1.45e-07

Likelihood ratio test=125.6  on 7 df, p=< 2.2e-16
n= 122, number of events= 37
```

The concordance index (CI) is often referred as a global value that shows the predictive ability of the model. The computation is based on the probability of agreement between two randomly chosen observations. It is also defined as the chance of selecting and correctly predicting an event with higher risk of failure. A value closer to one indicates a better model performance. However, it has to be noted that the CI index should not be the only performance measurement to rely on.

The likelihood ratio shows significant p-values that indicates the covariates do have an influence on the model against the null hypothesis of the survival function being merely a function of time.

Model Diagnostics

As the Cox-PH model uses the assumption that the effect of the covariate of the hazard is proportional with time, it is necessary to check if this assumption holds for the fitted model. As in a linear regression model, we observe the fit by comparing the predicted value against the observed to assess linearity but in the case of survival models, the residuals are handled differently and uses several methods as discussed below:

- **Schoenfeld residuals**: The Schoenfeld residuals test is used to check the proportional hazards assumption. The Schoenfeld residuals describes the difference between the observed covariate and the expected risk at time t. If the proportional assumption holds, the line is expected to be flat around the zero axis. Essentially the covariate effect should be time independent and does not show any pattern. In a simpler example, if the effect of temperature on battery cell ageing is 40% then this effect is expected to stay constant regardless of the timepoints until the battery cell dies. To interpret the inferred results, if the p-value is less than 0.05, then the model is said to violate the null hypothesis that the proportional hazards assumption holds. The formula for the Schoenfeld residual is given in Eq. 6.3.

$$residual = x_{ik} - \sum_{i=1}^{j \in R(t_i)} x_{jk} p_j \tag{6.3}$$

  As an example, if the covariate being examined is age, if the person who experienced death was 60 years old, based on the fitted model, how likely is the person who died was 60 and older.

  The cox.zph diagnostic test in R is used to test if the proportional hazards assumptions is violated for every covariate and at the global level.

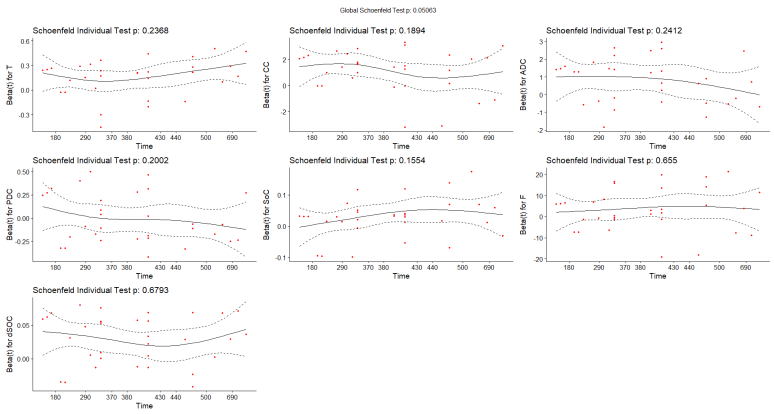


Figure 6.3: Schoenfeld Residuals for ALICe1

The p-value recorded for each variable (>0.05) suggest that there is no correlation between the covariate with time. Graphically, the PH assumption is supported by a random pattern with time. A global p-value of 0.0506 gives further indication that the proportional hazards assumption holds although it is rather at the borderline level. This means that the proportional hazards assumption holds suggesting that the Cox-PH model would be appropriate for the given dataset.

- **Martingale residuals**: As discussed in [39], the Martingale residuals test is used to assess the goodness of fit for the model. In some settings, it is also used to assess the functional form of the model by plotting the residuals when excluding variables from the model. In other words, it can be used to select features as well.

  Another use of this measure is to check if the model fulfills the linearity assumption that is required when fitting a Cox-PH model. To know whether the model fulfills this assumption, we expect a line that does not deviate far from the 0 axis. Usually martingale residuals will not cause an issue if the data is categorical. In our case, although the variable value is numeric, it can be presented in categorical form as well which would explain why we observe the linearity assumption being valid in this case.

  The formula to calculate the martingale residuals is as follows:

$$r_{Mi} = \delta_i - \hat{\lambda}_i(y_i|x_i) \tag{6.4}$$

Note that $\hat{\lambda}_i(y_i|x_i)$ is from hazard rate that is estimated using the Cox-PH model equation and δ_i refers to the count for each subject whether an event occurred or not.

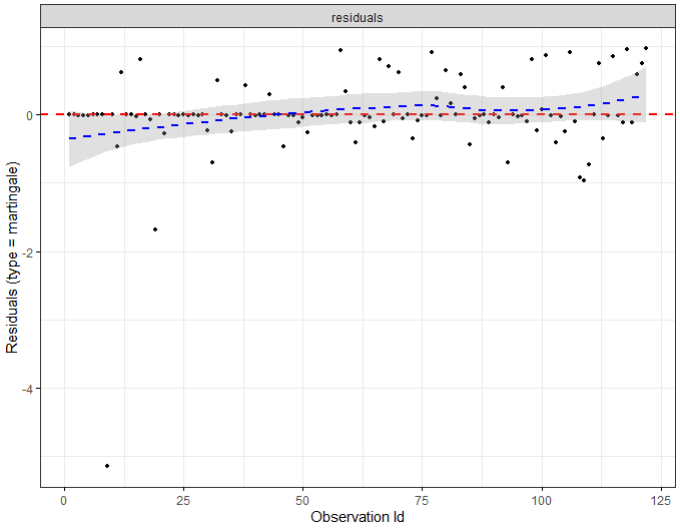


Figure 6.4: Martingale Residuals for ALICe1

From the figure above, we observe that the residuals are close to the red line, however there are certain observations on the left side that seem to suggest outliers in the data.

- **Deviance residuals**:Deviance residual (symmetric transformation of the Martinguale residuals) is used to observe any influential datapoints and is also used to detect outliers if they exist in the dataset. The formula for deviance residuals is given by the equation below:

$$r_{Di} = sign(r_{Mi})[-2r_{Mi} + \delta_i log(\delta_i - r_{Mi})]^{1/2} \tag{6.5}$$

We observe that r_{Di} has the same sign as r_{Mi} and the value in [] is positive where the square root value can be obtained and the sign is used to ensure that the deviance residual has the same representation as the martingale residual.

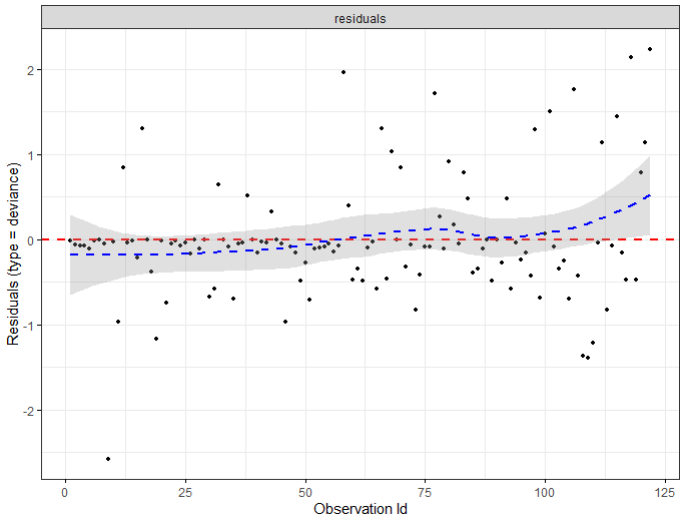


Figure 6.5: Deviance Residuals for ALICe1

Model Limitations

One of the main limitations observed when implementing the Cox-PH model is that the Cox-PH models the hazard rate where prediction values would provide a hazard rate value instead of time to event.

From the various testings conducted, it was difficult to ascertain if the Cox-PH model would be a suitable choice. This was observed from the results of the statistical tests that were mostly at borderline numbers. Although the Cox-PH assumption was satisfied, we found that the model did not explain the data well enough and with borderline statistical results, there is not enough confidence to decide on proceeding with a semi-parametric model for this use case.

Given that the goal of this thesis is to predict when a battery cell is expected to reach end of life, other survival model approach needs to be considered. This requirement points us towards adopting parametric models specifically accelerated failure time models that are found to be commonly used in analysing reliability data.

6.1.3 Parametric Survival Modeling

Since we are now making a rough estimation on the distribution shape we evaluate the non-parametric visualisations along with using the generalised gamma distribution that nests several distributions namely the Weibull, exponential, log-normal and standard gamma distribution.

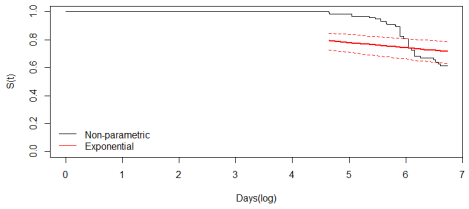


Figure 6.6: ALICe1 Exponential Distribution

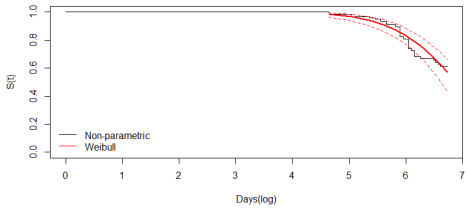


Figure 6.7: ALICe1 Weibull Distribution

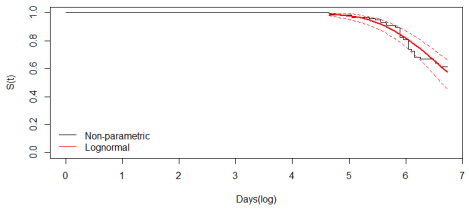


Figure 6.8: ALICe1 Log-normal Distribution

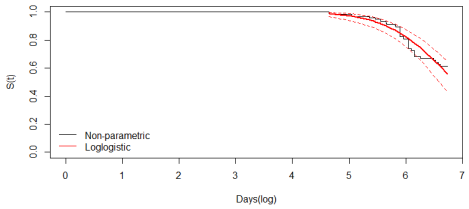


Figure 6.9: ALICe1 Log-logistic Distribution

From Fig. 6.6 to Fig. 6.9, we observe that the Weibull, Log-normal and Log-logistic gives a close approximate fit to the survival rate in general. Another approach to visualise the distribution fits is by using empirical plots from the fitdistplus package in R that produces the empirical density, cumulative density, Q-Q and P-P plots as shown in Fig. 6.10 to Fig. 6.13. A limitation with using this function is it only works with uncensored data. Therefore, the visualisation below only considers cells that have reached end of life. The "Data" notation along the x-axis refers to the days.

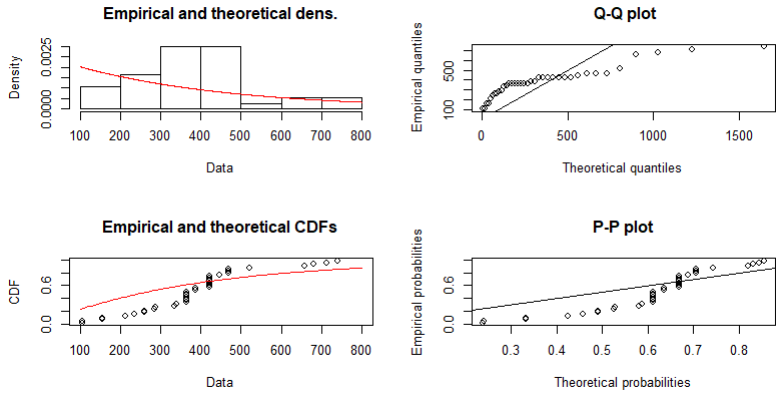


Figure 6.10: ALICe1 Empirical Exponential Distribution

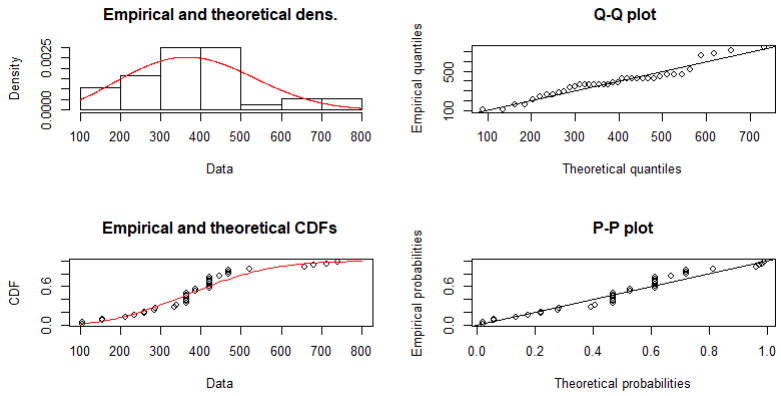


Figure 6.11: ALICe1 Empirical Weibull Distribution

The Q-Q plot, or quantile-quantile plot, is used as a way to visualise and assess if the dataset came from a theoretical distribution such as a Weibull, exponential, log-normal or log-logistic distribution. In other words it is

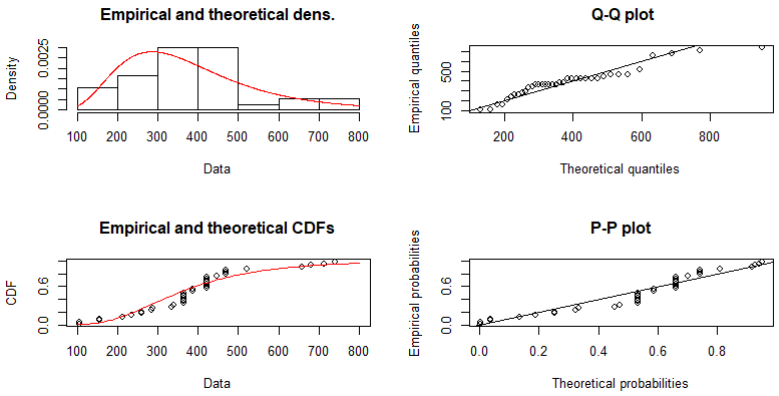


Figure 6.12: ALICe1 Empirical Log-normal Distribution

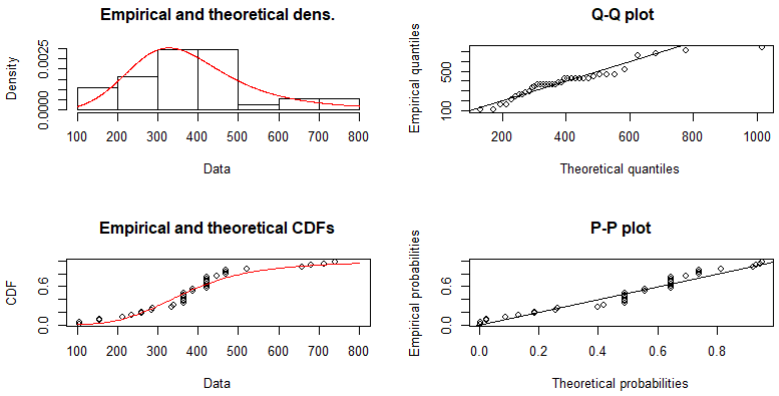


Figure 6.13: ALICe1 Empirical Log-logistic Distribution

used to check the assumption of the dataset to belong to a pre-specified distribution.

In this case, it plots the quantiles generated from the data points itself against the assumed distribution. If both sets of quantiles came from the same distribution, the points would align along the assumed distribution.

The probability-probability (P-P) plot is a graph of the empirical c.d.f values plotted against the theoretical c.d.f values. It is used to determine how well a specific distribution fits to the observed data. This plot will be approximately linear if the specified theoretical distribution is the correct model.

Based on the observations above, the fit between Weibull, log-normal and log-logistic shows close approximation to the dataset and therefore this

results in needing to build the prediction models considering each distribution and to then compare the prediction error results for the best fit.

6.2 Model Selection

Model selection is an important step to ensure that original and transformed features are considered for model building. As mentioned in Chapter 2, 24 regressors were derived to develop the initial linear regression model [11]. While the selection criteria above can be used to select a model, care is still needed in selecting the parameters which is usually guided by domain experts. The usual saying "all models are useless, but some are useful" is when other selection considerations is required.

There exist several selection criteria that are used to determine the number of suitable covariates and which combination of features are suitable.

- Akaike Information Criterion (AIC) (Akaike,1974)

  The AIC selection criteria is a technique to estimate the likelihood of a model to predict future values. A good model is defined by having the smallest AIC value compared to other models. Suppose that we have a statistical model of some data. Let k be the number of estimated parameters in the model. Let $\hat{L}$ be the maximum value of the likelihood function for the model. Then the AIC value of the model is the following.

  $$AIC = -2 \cdot \log(L) + 2 \cdot k \tag{6.6}$$

  L = likelihood value
  k = number of estimated parameters

  The AIC function favours more complex models that is able to fully explain the training data. This selection approach leads to models that includes as many parameters as possible because increasing the number of parameters in the model almost always improves the goodness of the fit.

- Adjusted R-squared

The adjusted R-squared is a modified version of the R-squared metric to adjust to the number of predictors in the model. The adjusted R-squared increases only if adding a new predictor improves the model performance more than what would be expected by chance. A model with a larger R-squared value means that the independent variables explains a larger percentage of variation in the dependent variable.

The adjusted R-squared can be calculated based on the value of R-squared, number of independent predictors and total sample size as given in the formula below:

$$R^2_{\text{adjusted}} = 1 - \frac{(1-R^2)(N-1)}{N-p-1} \tag{6.7}$$

R^2 = sample R-squared
p = number of predictors
N = total sample size

- Mallow's C_p

  The Mallow's C_p statistic is estimated based on a sample of data estimates where the mean squared prediction error (MSPE) is estimated. The idea is to select features when this value reaches the minimum value. The formula for C_p is given by the formula below:

$$C_p = \frac{1}{n}(\text{RSS} + 2d\hat{\sigma}^2) \tag{6.8}$$

  RSS = residual sum of squares
  d = number of predictors
  $\hat{\sigma}^2$ = estimate of variance associated with each response in the linear model

  It should be noted that the Mallow's C_p is usually valid for large datasets hence care is needed when deciding to use this selection criteria.

- Bayesian Information Criterion (BIC)(Stone, 1979)

  The BIC selection criteria measures the trade-off between model fit and complexity of the model. Similar to the AIC, a lower BIC value

indicates a better model. However unlike the AIC measure, the BIC gives a bigger penalty on complex models and gives preferences for simpler models.

The formula for BIC is as follows:

$$BIC = -2 \cdot \log(L) + 2 \cdot \log(N) \cdot k \tag{6.9}$$

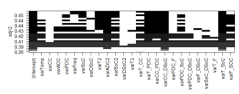
R^2 = sample R-squared
p = number of predictors
N = total sample size
$\hat{L}$ = the maximized value of the likelihood function of the model

6.2.1 Feature Selection

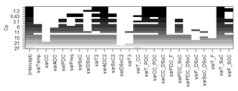
Some of the common feature selection approaches that can be considered are as follows:

- Backward elimination: the procedure starts with the full set of attributes. At each step, it removes the worst attribute remaining in the set.
- Forward selection: the procedure starts with an empty set of features (reduced set). The best of the original features is determined and added to the reduced set. At each subsequent iteration, the best of the remaining original attributes is added to the set.
- Exhaustive search: like breadth first search (BFS), the exhaustive search enumerates all possible feature combinations. These methods are rarely used in practice since the time complexity would be $O(2^n)$.

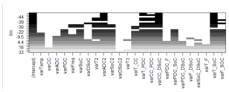
The quest to find an appropriate model requires several considerations among which is the suitability of including certain covariates in the model. In this approach we employ several approaches to at least yield a few models that can be considered. The first approach is to consider probabalistic model selection criteria namely the Mallow's C_p, Adjusted R-squared and BIC using backward elimination, forward selection and exhaustive approach which resulted in several model suggestion as shown in Fig. 6.14(a-i).



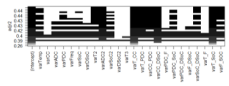
(a) Exhaustive Search- AdjR2



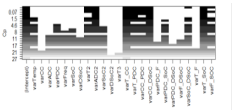
(b) Exhaustive Search- Cp



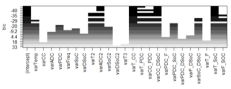
(c) Exhaustive Search- BIC



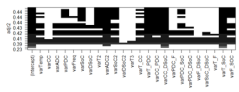
(d) Forward Select- AdjR2



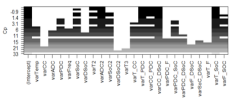
(e) Forward Select-Cp



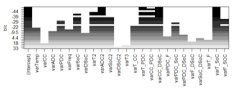
(f) Forward Select-BIC



(g) Backward Elimination- AdjR2



(h) Backward Elimination- Cp



(i) Backward Elimination- BIC

Figure 6.14: Model Selection Process

This process resulted in seven models as below:

- model1 = "Freq + SoC + Temp2 + ADC2 + SoC2 + Temp:CC + Temp:PDC + CC:PDC + CC:DSoC + Temp:SoC"
- model2 = "Temp2 + ADC2 + Temp:PDC + CC:PDC + CC:DSoC + Temp:SoC"
- model3 = "ADC2 + Temp:PDC + CC:DSoC + Temp:SoC"
- model4 = "Temp2 + ADC2 + Temp:CC + Temp:PDC + CC:PDC + CC:DSoC + SoC:DSoC + Temp:SoC + Freq:SOC"
- model5 = "ADC2 + Temp:CC + Temp:PDC + CC:DSoC + Temp:SoC"
- model6 = "SoC + Temp2 + ADC2 + SoC2 + Temp:PDC + CC:PDC + CC:DSoC + Temp:SoC + Freq:SOC"
- model7 = "Temp:SoC + Temp:PDC + CC:DSoC + ADC2 + CC:PDC + Temp2"

We also consider the initial models used in the linear regression approach [11]:

- lm6 = "Temp + SoC + Temp2 + ADC2 + SoC2 + SoC:DSoC"

- lm7 = "Temp + SoC + DSoC + Temp2 + ADC2 + SoC2 + SoC:DSoC"
- lm10 = "Temp + CC + ADC + SoC + DSoC + Temp2 + ADC2 + SoC2 + CC:DSoC + SoC:DSoC"
- lm11 = "Temp + CC + ADC + SoC + DSoC + Temp2 + ADC2 + SoC2 + CC:PDC + CC:DSoC + SoC:DSoC"

In order to select which model to use for the remaining experiments, we employ a resampling approach we train and test with a split ratio of 70:30 on each of these models and on each distribution type with a seed loop of 500. The root mean squared error (RMSE) metric based on battery cells that have reached end of life is used as the evaluation criteria to select the combination of features that gives the lowest prediction error. The reason for not using censored data here is because the estimated values will be beyond the survived days therefore will not give an accurate result.

6.3 Results

In Table 6.1 below, we show the top three out of sample prediction results for ALICe1.

Model	Log-logistic	Log-normal	Weibull
model5 = ADC2 + Temp:CC + Temp:PDC + CC:DSoC + Temp:SoC	219.03	226.10	313.72
model3 = ADC2 + Temp:PDC + CC:DSoC + Temp:SoC	243.59	255.72	353.45
model6 = SoC + Temp2 + ADC2 + SoC2 + Temp:PDC + CC:PDC + CC:DSoC + Temp:SoC + Freq:SOC	1453.80	340.12	1.35198E+18

Table 6.1: Top 3 Out of Sample Prediction Results

Based on the results above, unanimously all distribution types favoured Model 5 with five regressors but we are now faced with the question of whether to choose the distribution type with the lowest RMSE or to choose the distribution type based on domain application.

The Weibull distribution was first introduced by Dr. Waloddi Weibull. Here, he suggested a distribution that can represent a wide range of fail-

ure trends by easily changing two parameters or constants of the failure distribution. Given this reason and based on the non-parametric distribution fits shown in Fig. 6.7- Fig. 6.13 that showed very close results, we select the Weibull distribution fit for the remaining experiments. However, knowing that the dataset could follow other distribution shapes will need to be further validated by domain experts in the field.

In the next section, we look at applying the AFT survival model using a Weibull distribution trained on ALICe1 dataset to investigate if a transfer learning method can be employed to improve the prediction function for ALICe2 dataset.

7

Transfer of Battery Cell Ageing Prediction Models

In this chapter, we discuss the second goal of the thesis which is to investigate the transferability of the battery ageing or end of life prediction model to predict the end of life for a different type of battery cell chemistry.

7.1 Aim and Approach

The second aim of this thesis is to introduce a way in which a prediction model that was built to predict the end of life for one type of battery chemistry (source domain) can be reused to improve the end of life prediction for another type of battery chemistry (target domain) without needing a huge amount of labelled dataset in the target domain. This process is referred to as transfer learning. Here, we assume that the prediction model trained on the source dataset is able to capture the general patterns of battery ageing and that these patterns are shared in the ageing process of other battery cell chemistries as well.

In this section, we show how the model parameter based transfer strategy can be implemented by taking the intercept and the coefficient values estimated for the source dataset to be used for the target data prediction model as well. We also introduce another method in which we leverage the little knowledge gained from the target dataset by introducing a weighted coefficient and covariance mechanism. In this way we may consider situations in which a reduced amount of battery life dataset from another type of battery chemistry can also be used to train the prediction model.

7.2 Evaluation Criteria

The evaluation criteria that is used to compare the performance of the prediction models is based on calculating the root mean squared error (RMSE). The RMSE shows the standard deviation of the prediction errors between the actual predicted end of life in and the predicted end of life in days. While the prediction errors depicts the difference between the prediction value and the actuals, the RMSE shows the spread of these prediction errors or what is called as residuals.

The RMSE formula is as shown in Eq. 7.1 below:

$$RMSE = \sqrt{(\frac{1}{n}) \sum_{i=1}^{n} (\hat{y}_i - y_i)^2} \tag{7.1}$$

Where,
n = total number of battery cells tested
i = battery cell i
$\hat{y}_i$ = predicted end of life value in days battery cell i
y_i = actual end of life value in days for battery cell i

How do we know if the transfer method is actually improving the predictions for the target dataset? To answer this question, we begin with building a prediction model using the target dataset alone. The model is trained and tested using the ALICe2 dataset. This model acts as the baseline for which the subsequent experiments using the model parameter based transfer approaches is compared against. If we observe a reduction in the RMSE, this shows that leveraging the knowledge captured from modeling the ageing pattern in ALICe1 can be used to predict the end of life of a different battery chemistry used in the ALICe2 dataset.

As there are several ways in which the knowledge transfer can be facilitated, we decided to design five model scenarios including the baseline model as explained further in the next section.

7.3 Experiment Settings and Results

The general idea of transfer learning is as depicted in Fig. 7.1. Based on the discussion on the researched transfer learning strategies in Chapter 5,

we are leaning towards using the model parameter based approach and tackling this use case as a covariate shift problem.

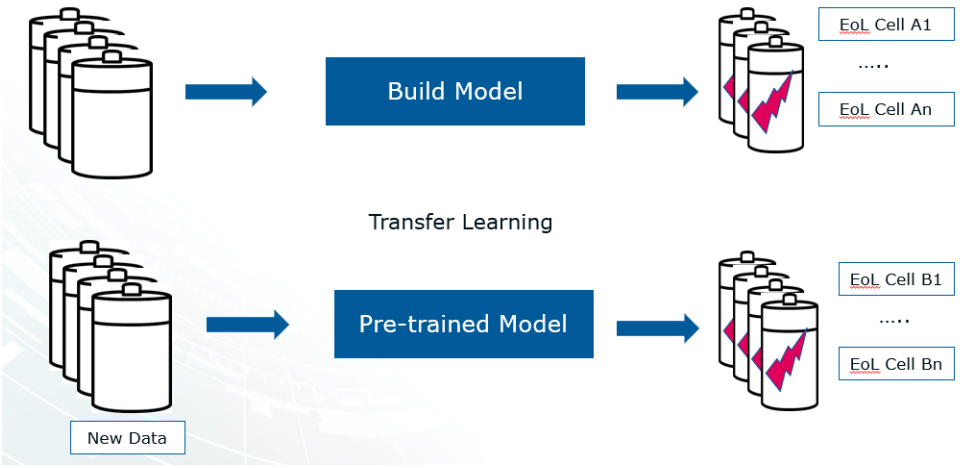


Figure 7.1: Main Idea For Battery Cell Ageing Prediction Transferability

When applying transfer learning, one of the key assumption that is tested is to know whether the two datasets are coming from different types of distribution. But as mentioned in [33], this is difficult to identify in practice especially when the target domain dataset is small or is not random enough to make any conclusion. We see that one of the issues with the ALICe2 dataset is the duration of the experiment which was much shorter than ALICe1 and this led to more than half of the observations to be right censored. To tackle this situation, we will have to ensure that the dataset is randomised through looping the train and validation splits across a number of seeds. Eventhough, a test split was also required to ensure that we can test the prediction model against unseen data, but we were grossly limited by the size of the dataset that did not allow for enough data to be set aside for testing.

Experiment settings were designed for five scenarios to evaluate the transferability of the prediction model from ALICe1 to ALICe2. Every experiment was iterated through 500 loops to gain the average RMSE results at the end as described in the following points:

1. Experiment 1

- Build prediction model by only using ALICe2 dataset (target dataset). This model acts as the baseline model to evaluate model performance improvement.
- Split training and validation data using 70:30 ratio. This split ratio is selected based on the dataset size and to ensure there is sufficient data for training and validation.
- Iterate split over 500 loops to randomise dataset.

2. Experiment 2

   - Train ALICe1 (source dataset) and train 70% of ALICe2 as target dataset.
   - Transfer coefficients (model parameters) only from ALICe1 source model to ALICe2 target model.
   - Validate on 30% of ALICe2 target model.

3. Experiment 3

   - Train ALICe1 (source dataset) and train 70% of ALICe2 as target dataset.
   - Transfer intercept only from ALICe1 source model to ALICe2 target model.
   - Validate on 30% of ALICe2 target model.

4. Experiment 4

   - Train ALICe1 (source dataset) and train 70% of ALICe2 as target dataset.
   - Transfer both model parameters and intercept from ALICe1 source model to ALICe2 target model.
   - Validate on 30% of ALICe2 target model.

5. Experiment 5

   - Using a weighting approach, weights are determined by getting the weighted mean for ALICe1 and ALICe2 datasets.

$$w_{ALICe1} = \frac{n_{ALICe1}}{n_{ALICe1} + n_{ALICe2}} \tag{7.2}$$

$$w_{ALICe2} = \frac{n_{ALICe2}}{n_{ALICe1} + n_{ALICe2}} \tag{7.3}$$

- Multiply weights with the covariance of ALICe1 and ALICe2 models and sum the ALICe1 and ALICe2 covariance. Assign these new covariance values to the target prediction function for ALICe2.

$$X_{ALICe2} = w_{ALICe1} X_{ALICe1} + w_{ALICe2} X_{ALICe2} \tag{7.4}$$

- Multiply weights with the coefficients of ALICe1 and ALICe2 models and sum these to become the new coefficients for the target prediction function for ALICe2.

$$\beta_{ALICe2} = w_{ALICe1} \beta_{ALICe1} + w_{ALICe2} \beta_{ALICe2} \tag{7.5}$$

- Using these new covariance and coefficient weights, we validate on 30% of ALICe2 dataset to observe if there will be further improvements to the model.

We calculate an average RMSE on end of life events from the 500 iterations and the results of the average RMSE for each experiment were as follows:

Experiment Type	Data Split 70:30
Train dataset A2 and validate on dataset A2	923.2991
Transfer model coefficients only and validate on dataset A2	846.7398
Transfer model intercept only and validate on dataset A2	708.2903
Transfer model coefficients and intercept from ALICe1 model to ALICe2 model	691.4873
Use weighted sum coefficient and weighted sum covariance	550.9872

Table 7.1: Experiment Results

Based on the results shown in Table 7.1, the first experiment is meant to show situations where only a small validation dataset can be obtained. In the case of ALICe2 dataset, since the experiment duration was short and had many censored observations, the dataset is not sufficient to make reliable predictions especially for battery cells that have reached end of life.

In the second experiment, we leveraged the knowledge gained from training on ALICe1 dataset by transferring only the model coefficients to the target model. We observe a prediction error reduction by 8.3%.

In the third experiment, we tested on retaining the target model coefficients but reuse only the intercept from the source model. We observed a further prediction error reduction by 23.3%.

In the fourth experiment, we transferred both the source model coefficients and intercept into the target model. Here we are making a bold assumption that the two battery chemistries share the same behaviour. From the results obtained, the target prediction function showed reduction in RMSE by about 25% suggesting that this assumption is valid.

In the fifth experiment, we aimed to use the best of both worlds. Here we leveraged the knowledge from ALICe1 and ALICe2 and surprisingly the results showed further reduction in the the RMSE. This suggest that this model can be considered to facilitate better predictions when using available source data along with a smaller set of available target data.

The R code used to perform these experiments is shown in Appendix A. 6

8

Conclusion and Future Work

8.1 Conclusion and Future Work

This thesis aimed to address the need to reduce battery testing requirements by investigating the possibility to reuse the knowledge gained from one type of battery cell chemistry to predict the end of life for another type of battery cell chemistry. From the conceptual framework in Fig. 1.1, the following conclusions were derived.

Collect and Process Battery Life Dataset

The datasets used were from two types of battery chemistry namely the NMC and NCA battery of type 18650. We were able to collect the capacity measurements from every reference test procedure and defined the capacity degradation to be linear in the beginning. Upon reaching 80% available capacity, a rapid decline of available capacity was observed that led to defining the battery end of life at 80% or lower available capacity. Where some abnormalities were observed, domain expert opinions were sourced to explain the trend. Preprocessing the battery life dataset was essential to ensure that the raw dataset collected is transformed into the dataset needed when implementing survival models.

Characterise Battery Ageing Trend

To characterise the battery ageing trend, we mainly used non-parametric survival models. Using this approach gave a visual understanding on how the battery cells were ageing over time and their survival patterns. The log-rank model was able to statistically evaluate whether there were any simi-

larities between the ageing trend in ALICe1 compared to the ALICe2. This resulted in initial acceptance of the chemistries showing some similarities in their survival patterns. However, relying only on these results were not sufficient as it did not consider influences of any ageing parameters.

Develop Battery End of Life Prediction Model

In order to facilitate the model transfer, a prediction model was first developed for the ALICe1 dataset. Since many cells did not reach the end of life threshold that was selected at 80%, we explored survival models that are often used in reliability engineering for its capability to handle censored observations.

As it was not clear on which type of survival model will be suitable for this use case, several approaches were considered. The Cox-PH model was first developed that showed the influences of each variable on the ageing of battery cells over time. Although this analysis was useful to determine the change in survival rates according to the different load point settings, the model was not suitable for the purpose of time to event predictions and it was found to be normally used to determine hazard ratios between different groups of factors. Using semi-parametric models does not require the knowledge of the distribution but since the main goal of the thesis is to predict battery end of life, other methods had to be considered.

With further research, end of life predictions were found to be common in reliability analysis with applications in engineering and manufacturing. Accelerated Failure Time (AFT) models were found to be mostly used for this kind of use cases where time to event prediction models can be developed. This fell into the parametric survival analysis approach. However, care was needed to determine which distribution type would best fit the data.

Interestingly, several distributions were closely fitting the data namely the log-normal, log-logistic and Weibull distribution. Although the log-normal model gave the lowest RMSE predictions but since the Weibull distribution is most commonly used for failure predictions, precedence was given to the domain application. However this selection step could have been improved with having more domain knowledge to determine whether other distributions should be considered as well. The domain

knowledge here can be acquired from having further discussions with reliability engineers who would be more familiar with the use of reliability analysis models for failure predictions.

Investigate Transferability of Battery End of Life Prediction Models

Using the prediction model built for the ALICe1 dataset using the Weibull AFT approach, the next goal was to determine what part of the model can be transferred for the ALICe2 dataset and how. Several transfer learning strategies were investigated. However as much time was taken in just determining which prediction model and distribution type was suitable to model the ALICe1 dataset, only one of the transfer learning approach was applied that is the model parameter based approach. Using the model parameter based approach allowed for not making any assumptions on the dataset distribution and to simply conduct several experiments by either transferring the whole model or partially as inputs for the target prediction model.

A new approach was also developed where a weighting mechanism was implemented that combined the weighted parameters from both ALICe1 and ALICe2 datasets. This approach seemed to have worked the best with the lowest RMSE compared to simply transferring the ALICe1 model coefficients and intercepts. At this point, these results can be regarded as a stepping stone for the possibility of requiring only a small amount of target dataset to predict battery end of life. However, future work would benefit from further experiments with different configurations of training, validation and testing dataset to further confirm these findings and to test on other battery types of battery cell chemistry.

Future Work

In summary, this thesis was mostly driven by discovery of new approaches to build a more reliable battery end of life prediction model and to determine whether transfer of knowledge gained from predicting the end of life of one type of battery cell chemistry can be used for another type of battery cell chemistry. It was encouraging to observe the use of survival models that assisted in a clearer understanding of the survival patterns between different types of battery cell chemistry. This research also provided a way

in which survival models can be reused by way of implementing a transfer learning method. However, since purely data driven methods were considered and showed promising results, we suggest the following steps to be considered for future work:

- To seek an opinion from domain experts especially reliability engineers who can assist with validating the results and methods used.
- To acquire testing datasets from other types of battery cell chemistry to improve model robustness and further confirm the tested hypotheses.
- To acquire a larger dataset so that the trained model can be tested on purely unseen data and to allow for different splits of training, validation and testing data.

A

R Code

```
##Experiment 1: split A2 into 70% for training and 30%
   for validation

rmse_data2 = data.frame()
pred_d2 = data.frame()

for (k in 1:500){
  set.seed(k)

  ind = createDataPartition(data2$Status, p=.7,list=FALSE,
     times=1)
  train_d2 = data2[ind,]
  test_d2= data2[-ind,]

  summary(as.factor(train_d2$Status))
  summary(as.factor(test_d2$Status))

  surv_func <- Surv(log(train_d2$Days),train_d2$Status)
  weibull_train_d2 <- survreg(surv_func ~ ADC2+Temp_CC+Temp
     _PDC+ CC_DSoC+Temp_SoC,data=train_d2,dist="weibull")

  #predict on validation data A2
  predInt = add_pi(fit = weibull_train_d2, tb = test_d2)
  pred = predict(object = weibull_train_d2,
  newdata = test_d2,se.fit=T, type="response")
```

```
  predictions <- as.data.frame(cbind(actual=test_d2$Days,
     status=test_d2$Status, predicted=exp(pred$fit), LB=
     exp(predInt$LPB0.025), UP=exp(predInt$UPB0.975)))

  y <- predictions$actual[predictions$status==1]
  y_hat <- predictions$predicted[predictions$status==1]
  rmse = sqrt(mean((y-y_hat)^2))

  df =cbind(predictions,rmse)
  df2= cbind.data.frame(CellID=test_d2$CellID,LP=test_d2$
     LP,df)
  pred_d2 = rbind(pred_d2,df2)

  rmse_df = data.frame(rmse)
  rmse_data2 = rbind(rmse_data2,rmse_df)
}
##Experiment2: Train A1 (source model) and train on 70%
    of A2 (target model), then transfer coefficients only
    from source model to target model and validate on 30%
    of target model

surv_func <- Surv(log(data1$Days),data1$Status)
weibull_train_d1 <- survreg(surv_func ~ ADC2+Temp_CC+Temp_
    PDC+CC_DSoC+Temp_SoC,data=data1,dist="weibull")
print(weibull_train_d1)

rmse_s2 = data.frame()
pred_d1d2 = data.frame()

for (k in 1:500){
  set.seed(k)

  #loop the train and test data to get a randomise
     combination of datasets
  ind = createDataPartition(data2$Status, p=.7,list=FALSE,
     times=1)
  train_2 = data2[ind,]
```

```
  test_2= data2[-ind,]

  summary(as.factor(train_2$Status))
  summary(as.factor(test_2$Status))
  surv_func <- Surv(log(train_2$Days),train_2$Status)

  weibull_train_d2 <- survreg(surv_func ~ ADC2+Temp_CC+Temp
     _PDC+CC_DSoC+Temp_SoC,data=train_2,dist="weibull")
  weibull_train_d2$coefficients <- weibull_train_d1$
     coefficients

  #on test data
  predInt = add_pi(fit = weibull_train_d2, tb = test_2)
  pred = predict(object = weibull_train_d2, newdata = test_
     2,se.fit=T, type="response")
  predictions <- as.data.frame(cbind(actual=test_2$Days,
     status=test_2$Status,predicted=exp(pred$fit), LB=exp(
     predInt$LPB0.025),UP=exp(predInt$UPB0.975)))

  y <- predictions$actual[predictions$status==1]
  y_hat <- predictions$predicted[predictions$status==1]
  rmse_s2 = sqrt(mean((y-y_hat)^2))

  df =cbind(predictions,rmse_s2)
  df2= cbind.data.frame(CellID=test_2$CellID,LP=test_2$LP,
     df)
  pred_d1d2 = rbind(pred_d1d2,df2)

  rmse_df = data.frame(rmse_s2)
  rmse_s2 = rbind(rmse_s2,rmse_df)
}

##Experiment 3: Transfer model intercept only and
   validate on dataset A2

surv_func <- Surv(log(data1$Days),data1$Status)
```

```
weibull_train_d1 <- survreg(surv_func ~ ADC2+Temp_CC+Temp_
   PDC+CC_DSoC+Temp_SoC,data=data1,dist="weibull")
print(weibull_train_d1)

rmse_s3 = data.frame()
pred_d1d2 = data.frame()

for (k in 1:500){
  set.seed(k)

  #loop the train and test data to get a randomise
     combination of datasets
  ind = createDataPartition(data2$Status, p=.7,list=FALSE,
     times=1)
  train_2 = data2[ind,]
  test_2= data2[-ind,]

  summary(as.factor(train_2$Status))
  summary(as.factor(test_2$Status))

  surv_func <- Surv(log(train_2$Days),train_2$Status)

  weibull_train_d2 <- survreg(surv_func ~ ADC2+Temp_CC+Temp
     _PDC+CC_DSoC+Temp_SoC,data=train_2,dist="weibull")
  weibull_train_d2$coefficients[1] <- weibull_train_d1$
     coefficients[1]

  #on test data
  predInt = add_pi(fit = weibull_train_d2, tb = test_2)
  pred = predict(object = weibull_train_d2, newdata = test_
     2,se.fit=T, type="response")
  predictions <- as.data.frame(cbind(actual=test_2$Days,
     status=test_2$Status,predicted=exp(pred$fit), LB=exp(
     predInt$LPB0.025),UP=exp(predInt$UPB0.975)))

  y <- predictions$actual[predictions$status==1]
  y_hat <- predictions$predicted[predictions$status==1]
```

```
  rmse_s3 = sqrt(mean((y-y_hat)^2))

  df =cbind(predictions,rmse_s3)
  df2= cbind.data.frame(CellID=test_2$CellID,LP=test_2$LP,
     df)
  pred_d1d2 = rbind(pred_d1d2,df2)

  rmse_df = data.frame(rmse_s3)
  rmse_s3 = rbind(rmse_s3,rmse_df)
}

##Experiment 4: Transfer model coefficients and intercept
    from A1 model to A2 model

surv_func <- Surv(log(data1$Days),data1$Status)
weibull_train_d1 <- survreg(surv_func ~ ADC2+Temp_CC+Temp_
   PDC+CC_DSoC+Temp_SoC,data=data1,dist="weibull")
print(weibull_train_d1)

rmse_s4 = data.frame()
pred_d1d2 = data.frame()

for (k in 1:500){
  set.seed(k)

  #loop the train and test data to get a randomise
     combination of datasets
  ind = createDataPartition(data2$Status, p=.7,list=FALSE,
     times=1)
  train_2 = data2[ind,]
  test_2= data2[-ind,]

  summary(as.factor(train_2$Status))
  summary(as.factor(test_2$Status))

  surv_func <- Surv(log(train_2$Days),train_2$Status)
```

```
  weibull_train_d2 <- survreg(surv_func ~ ADC2+Temp_CC+Temp
     _PDC+CC_DSoC+Temp_SoC,data=train_2,dist="weibull")
  weibull_train_d2$coefficients <- weibull_train_d1$
     coefficients


  #on test data
  predInt = add_pi(fit = weibull_train_d2, tb = test_2)
  pred = predict(object = weibull_train_d2, newdata = test_
     2,se.fit=T, type="response")
  predictions <- as.data.frame(cbind(actual=test_2$Days,
     status=test_2$Status,predicted=exp(pred$fit), LB=exp(
     predInt$LPB0.025),UP=exp(predInt$UPB0.975)))

  y <- predictions$actual[predictions$status==1]
  y_hat <- predictions$predicted[predictions$status==1]
  rmse_s4 = sqrt(mean((y-y_hat)^2))

  df =cbind(predictions,rmse_s4)
  df2= cbind.data.frame(CellID=test_2$CellID,LP=test_2$LP,
     df)
  pred_d1d2 = rbind(pred_d1d2,df2)

  rmse_df = data.frame(rmse_s4)
  rmse_s4 = rbind(rmse_s4,rmse_df)
}


##Experiment 5: Use weighted coefficients and weighted
    covariance approach
surv_func <- Surv(log(data1$Days),data1$Status)
weibull_train_d1 <- survreg(surv_func ~ ADC2+Temp_CC+Temp_
    PDC+CC_DSoC+Temp_SoC,data=data1,dist="weibull")

rmse_s5 =data.frame()
pred_weighted = data.frame()
```

```
for (j in 1:500){
  set.seed(j)
  ind = createDataPartition(data2$Status, p=.7,list=FALSE,
     times=1)

  train2 = data2[ind,]
  test2 = data2[-ind,]

  summary(as.factor(train2$Status))
  summary(as.factor(test2$Status))

  #find weights for data 1 and data2
  d1_rows = nrow(data1)
  d2_rows = nrow(train2)

  weight_d1 = d1_rows/(d1_rows+d2_rows)
  weight_d2 = d2_rows/(d1_rows+d2_rows)

  #train data2 model
  weibull_train_d2 <- survreg(Surv(log(train2$Days),train2$
     Status) ~ ADC2+Temp_CC+Temp_PDC+CC_DSoC+Temp_SoC,data
     =train2,dist="weibull")

  #weight the covariances and sum them up
  X_d1 = weight_d1 * weibull_train_d1$var
  X_d2 = weight_d2 * weibull_train_d2$var
  var = X_d1 + X_d2

  #transfer new convariace into target model
  weibull_train_d2$var = var

  #weight the coefficients and sum them up
  beta_d1 = weight_d1 * weibull_train_d1$coefficients
  beta_d2 = weight_d2 * weibull_train_d2$coefficients

  beta_weights = beta_d1 + beta_d2
  weibull_train_d2$coefficients = beta_weights
```

```
  #transfer new convariace into target model
  weibull_train_d2$coefficients = beta_weights

  #test on 30% of data2
  predInt = add_pi(fit = weibull_train_d2, tb = test2)
  pred = predict(object = weibull_train_d2, newdata = test2
     ,se.fit=T, type="response")

  predictions <- as.data.frame(cbind(actual=test2$Days,
     status=test2$Status,predicted=exp(pred$fit), LB=exp(
     predInt$LPB0.025),UP=exp(predInt$UPB0.975)))

  y <- predictions$actual[predictions$status==1]
  y_hat <- predictions$predicted[predictions$status==1]
  rmse = sqrt(mean((y-y_hat)^2))

  df =cbind(predictions,rmse)
  df2= cbind.data.frame(CellID=test2$CellID,LP=test2$LP,df)

  pred_weighted = rbind(pred_weighted,df2)

  rmse_df = data.frame(rmse)
  rmse_s5 = rbind(rmse_s5,rmse_df)
}
```

B

Abbreviations and Terminologies

Battery Unit Measurements

Acronym	Meaning
Ah	Ampere-hour: Unit of Electric Charge
W/kg	Watt per kilogram: Power to weight ratio/mass
Wh/kg	Watt hours per kilogram: Unit of specific energy (energy per unit mass of battery)
Wh	Watt hours: Power consumption of one watt for one hour
Ω	Resistance, Ohm: Unit of electric resistance

Battery Terminologies

Terminology	Meaning
Battery capacity, **Ah**	Amount of charge can be stored in a cell
Power density, **W/kg**	Amount of power per kg can be delivered
Energy density, **Wh/kg**	Amount of energy per kg can be delivered
Energy content, **Wh**	Amount of energy that can be discharged
State of Charge (SoC)	Amount of charge remaining in the battery

State of Function (SoF)	Reflects battery readiness in terms of usable energy by observing state of charge in relation to the available capacity (computes power limits that is the extend to which the battery power can be used)
State of Energy (SoE)	Reflects the residual energy of the battery as the ratio between remaining energy and total energy that is stored in a battery
Begin of Life (BoL)	Condition of battery when it leaves production or the beginning of operational life in production terms
State of Health (SoH)	Represents the ratio of capacity or internal resistance and other factors such as charge acceptance, voltage and self discharge measured versus values at BoL
End of Life (EoL)	Condition when the cell cannot fulfill its application requirements or when the battery SoH has reached 80% or 70% remaining capacity
C-Rate	Measure of the rate at which a battery is discharged relative to its maximum capacity
I, current	Rate of charge flowing
Internal resistance, Ir	Internal resistance is measured in Ohms. It is a concept used in direct current (DC) and is in opposition to steady electrical current. It determines how well electrical current flows through a material or device

Impedance	Consept used in alternating current (AC) and is a combination of resistance and reactance. While resistance is the opposition of current through a resistor, reactance is the opposition of current through an inductor or capacitor. When the systems involves using a resistor, inductor and capacitor, it is then called impedance
Charge, voltage and current	By way of an anology, charge is the water amount, voltage is the water pressure and current is the water flow. When there is higher resistance, then there will be less current flowing
Self-discharge	Occurs when batters is not in use and internal chemical reactions reduces the stored charge of the batters without any connection between electrodes
Specific Energy **Wh/kg**	Indicates the capacity a battery can hold

List of Figures

D

List of Tables

E

Bibliography

[1] I. E. Agency, *Global EV Outlook 2019*. 2019.

[2] D. N. P. M. Wallace R. Blischke, M. Rezaul Karim, *Warranty Data Collection and Analysis*. Springer-Verlag London, 2011.

[3] R. S. A. of Sciences, "Scientific background on the nobel prize in chemistry 2019 - lithium-ion battery." URL: *nobelprize.org/uploads/2019/10/advanced-chemistryprize2019.pdf*.

[4] 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, p. 1820, 2018.

[5] C. Iclodean, B. Varga, N. Burnete, D. Cimerdean, and B. Jurchiş, "Comparison of different battery types for electric vehicles," *IOP Conference Series: Materials Science and Engineering*, vol. 252, p. 012058, 2017.

[6] K. D. Stetzel, L. L. Aldrich, M. S. Trimboli, and G. L. Plett, "Electrochemical state and internal variables estimation using a reduced-order physics-based model of a lithium-ion cell and an extended Kalman filter," *Journal of Power Sources*, vol. 278, pp. 490–505, 2015.

[7] G. Plett. Coursera: Algorithms for Battery Management System. URL: *https://www.coursera.org/specializations/algorithms-for-battery-management-systems?*.

[8] I. Buchmann. Battery University: Series and Parallel Battery Configurations. URL: *https://batteryuniversity.com/learn/article/serial_and_parallel_battery_configurations*.

[9] A. Barré, B. Deguilhem, S. Grolleau, M. Gérard, F. Suard, and D. Riu, "A review on lithium-ion battery ageing mechanisms and estimations for automotive applications," *Journal of Power Sources*, vol. 241, pp. 680–689, 2013.

[10] W. Prochazka, G. Pregartner, and M. Cifrain, "Design-of-experiment and statistical modeling of a large scale aging experiment for two popular lithium ion cell chemistries," *Journal of The Electrochemical Society*, vol. 160, no. 8, pp. 1039–1051, 2013.

[11] G. Gossler, "Modelling the lifespans of lithium-ion cells by using d-optimal designs," Master's thesis, 2015.

[12] W. He, N. Williard, M. Osterman, and M. Pecht, "Prognostics of lithium-ion batteries based on Dempster-Shafer theory and the Bayesian Monte Carlo method," *Journal of Power Sources*, vol. 196, no. 23, pp. 10314–10321, 2011.

[13] Y. Chen, Q. Miao, B. Zheng, S. Wu, and M. Pecht, "Quantitative analysis of lithium-ion battery capacity prediction via adaptive bathtub-shaped function," *Energies*, vol. 6, no. 6, pp. 3082–3096, 2013.

[14] B. Long, W. Xian, L. Jiang, and Z. Liu, "An improved autoregressive model by particle swarm optimization for prognostics of lithium-ion batteries," *Microelectronics Reliability*, vol. 53, no. 6, pp. 821–831, 2013.

[15] Q. Miao, L. Xie, H. Cui, W. Liang, and M. Pecht, "Remaining useful life prediction of lithium-ion battery with unscented particle filter technique," *Microelectronics Reliability*, vol. 53, no. 6, pp. 805–810, 2013.

[16] D. Wang, Q. Miao, and M. Pecht, "Prognostics of lithium-ion batteries based on relevance vectors and a conditional three-parameter capacity degradation model," *Journal of Power Sources*, vol. 239, pp. 253–264, 2013.

[17] S. S. Ng, Y. Xing, and K. L. Tsui, "A naive bayes model for robust remaining useful life prediction of lithium-ion battery," *Applied Energy*, vol. 118, pp. 114–123, 2014.

[18] S. Wang, L. Zhao, X. Su, and P. Ma, "Prognostics of lithium-ion batteries based on battery performance analysis and flexible support vector regression," *Energies*, vol. 7, no. 10, pp. 6492–6508, 2014.

[19] S. Tang, C. Yu, X. Wang, X. Guo, and X. Si, "Remaining useful life prediction of lithium-ion batteries based on the wiener process with measurement error," *Energies*, vol. 7, no. 2, pp. 520–547, 2014.

[20] X. Zheng and H. Fang, "An integrated unscented kalman filter and relevance vector regression approach for lithium-ion battery remaining useful life and short-term capacity prediction," *Reliability Engineering and System Safety*, vol. 144, pp. 74–82, 2015.

[21] L. Li, P. Wang, K. H. Chao, Y. Zhou, and Y. Xie, "Remaining useful life prediction for lithium-ion batteries based on Gaussian processes mixture," *PLoS ONE*, 2016.

[22] D. Wang, F. Yang, K. L. Tsui, Q. Zhou, and S. J. Bae, "Remaining Useful Life Prediction of Lithium-Ion Batteries Based on Spherical Cubature Particle Filter," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 6, pp. 1282–1291, 2016.

[23] Y. Zhang, R. Xiong, H. He, and M. G. Pecht, "Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 7, pp. 5695–5705, 2018.

[24] M. S. Lipu, M. A. Hannan, A. Hussain, M. M. Hoque, P. J. Ker, M. H. Saad, and A. Ayob, "A review of state of health and remaining useful life estimation methods for lithium-ion battery in electric vehicles: Challenges and recommendations," *Journal of Cleaner Production*, vol. 205, pp. 115–133, 2018.

[25] W. He, N. Williard, M. Osterman, and M. Pecht, "Prognostics of lithium-ion batteries using extended kalman filtering," 2011.

[26] T. Li, S. Sun, T. P. Sattar, and J. M. Corchado, "Fight sample degeneracy and impoverishment in particle filters: A review of intelligent approaches," 2014.

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

[28] S. Voronov, E. Frisk, and M. Krysander, "Data-driven battery lifetime prediction and confidence estimation for heavy-duty trucks," *IEEE Transactions on Reliability*, vol. 67, no. 2, pp. 623–639, 2018.

[29] J. Zhang, S. Wang, L. Chen, G. Guo, R. Chen, and A. Vanasse, "Time-dependent survival neural network for remaining useful life prediction," in *Advances in Knowledge Discovery and Data Mining* (Q. Yang, Z.-H. Zhou, Z. Gong, M.-L. Zhang, and S.-J. Huang, eds.), (Cham), pp. 441–452, Springer International Publishing, 2019.

[30] C. C. Aggarwal, *Data Classification: Algorithms and Applications*. Chapman & Hall/CRC, 1st ed., 2014.

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

[32] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345 – 1359, 2010.

[33] C. Bouveyron and J. Jacques, "Adaptive linear models for regression: Improving prediction when population has changed," *Pattern Recognition Letters*, vol. 31, no. 14, pp. 2237–2247, 2010.

[34] M. Stevenson. Lecture Notes on Introduction to Survival Analysis, Massey University, 2007.

[35] P. Wang, Y. Li, and C. K. Reddy, "Machine learning for survival analysis: A survey," *ACM Computing Surveys*, vol. 51, no. 6, p. 110, 2019.

[36] L. Sullivan, “Survival analysis.” URL: *http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Survival/BS704_Survival_print.html*.

[37] W. Q. Meeker and L. A. Escobar, *Statistical methods for reliability data*. New York : Wiley, 1998. "A Wiley-Interscience publication.".

[38] A. Wienke, *Frailty models in survival analysis*. 2010.

[39] T. M. Therneau, P. M. Grambsch, and T. R. Fleming, “Martingale-based residuals for survival models,” *Biometrika*, vol. 77, no. 1, pp. 147–160, 1990.