Gerhard Gößler, Mag.rer.nat., BSc

Modelling the Lifespans of Lithium Ion Cells by using D-optimal Designs

MASTER'S THESIS

to achieve the university degree of a Diplom-Ingenieur

Master's degree programme: Statistics and Operations Research

Graz University of Technology

Tutor:
Univ.-Prof. Dipl.-Ing. Dr.techn. Ernst Stadlober

Institute of Statistics

Dipl.-Ing. Dr.techn. Martin Cifrain and Dipl.-Ing. Matthias Scharrer

VIRTUAL VEHICLE Research Center (ViF)

Graz, November 2015

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am ....
(Unterschrift)

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

....
date (signature)

Zusammenfassung

Im ersten Teil wird eine kurze Einführung in das Gebiet der Lithium-Ionen-Batterien gegeben. Aufbau und Funktionsweise dieses Batterietyps, sowie dessen Verwendung im Automobilbereich werden thematisiert. Weiters findet sich eine Beschreibung des Batterie-Alterungs-Experiments ALICe, welches während seiner Laufzeit statistisch begleitet und dessen Ergebnisse nach Beendigung des Experiments einer eingehenden statistischen Analyse unterzogen wurden.

Der zweite Teil befasst sich mit den Grundlagen der multiplen linearen Regression und der Versuchsplanung. Er bietet eine ausführliche Beschreibung des theoretischen Hintergrundes der verwendeten statistischen Methoden und Kenngrößen.

Der dritte Teil beinhaltet die Analyse des eingesetzten Versuchsdesigns, das vom ursprünglich geplanten Design bezüglich der Designpunkte und des parametrisierten Modells abweicht. Die Auswirkungen dieser Änderungen auf die Effizienz werden eingehend studiert und evaluiert. Die im Experiment gewonnenen Daten werden darüber hinaus genutzt, um das eigentliche Ziel des Experiments, nämlich die Auswahl und die Parametrisierung eines geeigneten multiplen linearen Regressionsmodells (MLR) für die Beschreibung des Zusammenhanges zwischen Batterienutzung und Batteriealterung durchzuführen. Neben einer ausführlichen Darstellung der Resultate der statistischen Analysen bietet dieses Kapitel auch eine kritische Diskussion der erzielten Ergebnisse.

Diese Arbeit wurde in Zusammenarbeit mit dem 'Kompetenzzentrum - Das virtuelle Fahrzeug, Forschungsgesellschaft mbH' durchgeführt.

Abstract

We start with a short introduction into the topic of lithium batteries and discuss composition, mode of operation and usage in the automotive context of this battery type. Furthermore, this chapter also provides an overview of the battery aging experiment ALICe, which was supervised from the viewpoint of statistics and experimental design during its term and whose results were thoroughly evaluated after its termination.

The second part is concerned with the basics of both multiple linear regression and experimental design and gives a rather detailed discussion of the theoretical background and the statistics used.

The third part deals with the analysis of the actual experimental setting used within the project. The main focus is on the impact of deviations of the experimental design and of the model to be parametrized on the efficiency of the initial setting. Additionally, the data gained in the experiment are used to accomplish the intended goal of the experiment, namely to select and parametrize a suitable multiple linear regression model (MLR) to describe the relationship between battery usage and battery aging. In addition to a comprehensive presentation of the results of the statistical analysis, this part also comprises a critical discussion of the obtained results.

This thesis was realized in cooperation with the VIRTUAL VEHICLE Research Center (ViF).

Danksagung

An dieser Stelle möchte ich allen Personen herzlich danken, die mich bei der Erstellung dieser Arbeit direkt oder indirekt unterstützt haben.

Mein besonderer Dank gilt meinen Betreuern - und zwar Univ.-Prof. Dipl.-Ing. Dr.techn. Ernst Stadlober vom Institut für Statistik der Technischen Universität Graz für die ausgezeichnete und freundliche Betreuung sowie Dipl.-Ing. Dr.techn. Martin Cifrain und Dipl.-Ing. Matthias Scharrer vom VIRTUAL VEHICLE Research Center (ViF) für ihre fortwährende Unterstützung, ihre Anregungen und nicht zuletzt für die wohlwollende Aufnahme in ihre Arbeitsgruppe - in diesem Zusammenhang möchte ich mich ebenfalls beim Gruppenleiter Dipl.-Ing. Dr.mont. Alexander Thaler bedanken. Außerdem möchte ich allen oben genannten dafür danken, dass sie jederzeit ein offenes Ohr für meine Fragen hatten.

Weiters möchte ich allen danken, die mir während der Jahre des Studiums zur Seite gestanden sind. Vor allem meinen Eltern, die mich im Laufe meiner gesamten Ausbildung mit großer Geduld begleitet und mich stets nach besten Kräften unterstützt haben.

Großer Dank auch an Theodora, die immer mit großem Verständnis und aufmunternden Worten sehr zur moralischen Unterstützung beigetragen hat und speziell an Florian und Stephan die mir jederzeit (nicht nur) in mathematischen Fragen stets mit Rat und Tat zur Seite gestanden sind.

Contents

1 Introduction **1**
- 1.1 Motivation . . . 1
- 1.2 Li-ion cells . . . 2
  - 1.2.1 Technological aspects . . . 2
  - 1.2.2 Aging causes . . . 4
- 1.3 The Experiment . . . 8

2 Statistical concepts and methods **13**
- 2.1 The need for statistical modelling . . . 13
- 2.2 Multiple Linear Regression . . . 14
  - 2.2.1 Least Squares Method . . . 16
  - 2.2.2 Hypothesis Testing . . . 20
- 2.3 Experimental Design and Model Selection . . . 25
  - 2.3.1 Unbalanced Datasets . . . 25
  - 2.3.2 Model Selection . . . 28
  - 2.3.3 Model Diagnostics . . . 34
  - 2.3.4 Experimental Design . . . 38
  - 2.3.5 Optimal Designs . . . 40

3 Statistical analysis **48**
- 3.1 Evaluation of the design in use . . . 48
  - 3.1.1 Comparing the initial design and the actual state of the experiment 49
  - 3.1.2 Remedies and Suggestions . . . 50
- 3.2 The cells . . . 59
  - 3.2.1 Different cell types . . . 59
- 3.3 Extrapolation of the Cells End of Life . . . 66
- 3.4 Modeling the Relationship between the Cells' Lifespans and the seven different influential Factors using multiple linear Regression (MLR) . . . 73
  - 3.4.1 Balancing the Data Set . . . 73
  - 3.4.2 Statistical Analyis and Model Selection . . . 80

4 Summary **102**

5 Outlook **105**

Bibliography **106**

1 Introduction

1.1 Motivation

Due to the finiteness of fossil fuels and the heavy (negative) impact of the emissions (exhaust fumes and noise) of combustion engines on our environment and therefore on ourselves, the call for environmentally friendly power trains respectively energy sources is becoming louder and louder.

At the moment the electric power trains constitute the most promising alternative. One possibility for storing the energy needed in this case is the use of batteries. Lithium ion batteries seem to meet the demands of electric mobility better than other battery types, because of a high energy density, hardly any memory effect, and only a slow loss of charge when not in use (Wikipedia [38]).

In addition to the above mentioned aspects, the question of longevity is, for economical reasons, crucial for the successful application in the automotive industry.

This work is concerned with the evaluation of the experimental design (constructed by W. Prochaska and G. Pregartner) and with the statistical analysis of the data produced by a preliminarily conducted experiment, which aims to investigate the topic of battery aging.

The lifespan of a battery depends on a large number of influential factors and there exists no comprehensive mechanistic understanding of the relationship between these factors and the process of cell aging. Due to the complex nature of the process under investigation, there is an obvious need for the use of statistical methods. The design as well as the analysis of the experiment can therefore only be conducted on a sound statistical basis.

This work is embedded in the K2 project E3T3 (ALiCe), which is currently carried out at the VIRTUAL VEHICLE Research Center in Graz.

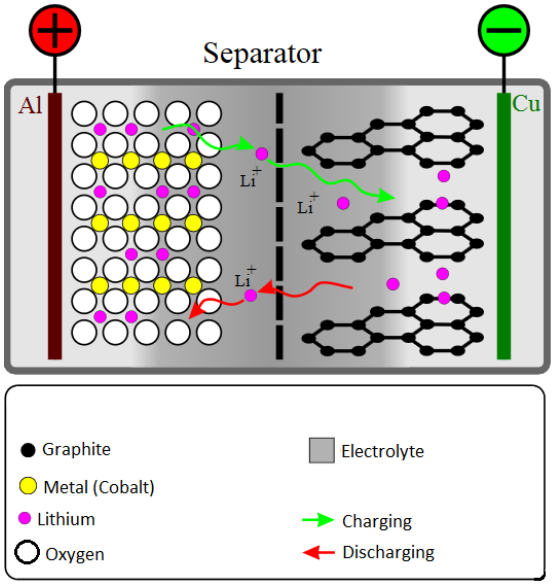


Figure 1.1: Schematic composition of a Li-ion cell (Wikipedia [39])

1.2 Li-ion cells

1.2.1 Technological aspects

The electrodes, the electrolyte and the separator constitute the main parts of a cell. The separator prevents short-circuit through physical contact of the electrodes, but allows the migration of Li^+ ions.

The electrodes have the capability to reversibly uptake the Li^+ ions. The direction of uptake and release depends on whether the battery is charged or discharged. During discharging, the Li^+ ions move out of the graphite anode and move into the cathode material. During charging, the process is reversed. Because of this seesaw of ions, the technology is also referred to as rocking-chair technology.

The anode consists of a graphite layer (often denoted as C_6), laminated on a copper current collector.

The cathode consists in most cases of an aluminum foil, which is laminated with one metal oxide or a blend of metal oxides or metal phosphates. Examples are:

- LCO = Lithium Cobalt Oxide ($LiCoO_2$), the first commercially used material invented by SONY in the early nineties. Although widely used, the material is thermally not stable and therefore rather unsafe.
- LMO = Lithium Manganese Oxide ($LiMn_2O_4$). A high-energy material, today often used in notebooks.

- NCM = Lithium Nickel Cobalt Manganese Oxide ($Li[Ni_aCo_bMn_c]O_2$ with $a+b+c = 1$). A high energy material, today often used in automotive applications.
- LFP = Lithium Iron Phosphate ($LiFePO_4$). A high-power but low-energy material, used in hybrid electric vehicles where power is more important than range. The safest cathode material at the moment.

The liquid electrolyte in lithium-ion batteries serves as a conductive medium for the Li^+ ions and allows them to move freely between the electrodes. The electrolyte consists of a mixture of organic carbonate solvents and contains a lithium salt (e.g. $LiPF_6$) to increase the lithium conductivity.

A reason for the usage of these non-aqueous electrolytes is, that the cell voltage is large enough to electrolyze aqueous solutions.

Furthermore, lithium reacts heavily with entering water, causing fire - and even explosion hazard. Therefore the interior of a lithium cell is totally free of water ($H_2O < 10$ ppm). In order to prevent water entry, the hull is sealed and impermeable.

In the charged state, the potential difference between anode and cathode (cell voltage) is in the range from 3 to 4 Volt, depending on the cathode material. In the case of a cell based on $LiCoO_2$ it is about 3.7 V.

The lithium-ion flow in the cell is necessary to compensate for the external current flow which occurs when the cell is charged or discharged (to close the circuit). During discharge, the intercalated lithium atoms emit electrons thus becoming positively, singly charged ions Li^+. The emitted electrons flow through the external circuit from the anode to the cathode and simultaneously an equal amount of lithium ions moves in the same direction through the electrolyte. At the cathode, the lithium ions do not reabsorb the electrons, this is done by compounds which are part of the positive electrode. That means, that in the discharged state, the lithium is still present at the cathode in form of ions.

The following example gives a look at what is going on in a LMO cell:

Anode (discharge): $Li_xC_6 \rightarrow C_6 + x\,Li^+ + x\,e^-$

Cathode (discharge): $Li_yMn_2O_4 + x\,Li^+ + x\,e^- \rightarrow Li_{y+x}Mn_2O_4$

Redox-equation: $Li_yMn_2O_4 + Li_xC_6 \rightarrow Li_{y+x}Mn_2O_4 + C_6$

Metallic lithium is not involved in any of these reactions.

During this (redox-) process, the electrodes alter chemically triggering a potential change, which results in a changing cell voltage (see figure 1.2).

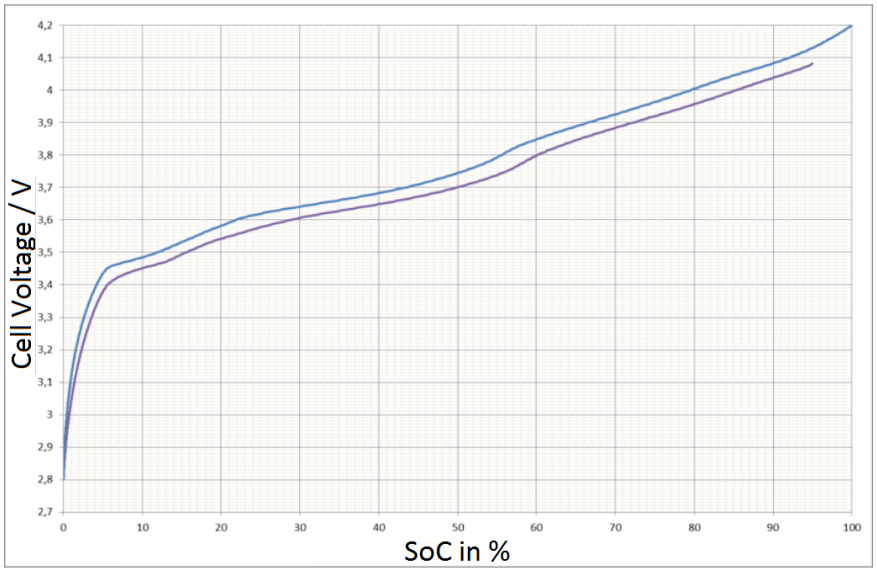


Figure 1.2: Open-circuit cell voltage (electrode potential difference) of the cell used in the ALiCe project at different states of charge (SoC). SoC = 0% ... cell discharged; SoC = 100% ... cell charged. Upper curve: charge direction; lower curve: discharge direction (ViF [34]).

1.2.2 Aging causes

As already mentioned, the design was set up prior to the beginning of this Master's thesis. We will briefly discuss the question which factors for what reason have been selected and how the design window (i.e. range of the factors) has been chosen.

The chemical expertise, which flowed in the process of designing the experiment, stems from three major sources:

1. Pre-existing knowledge at VIRTUAL VEHICLE.
2. Extensive literature search.
3. A preliminary experiment (for further details, see the Master's thesis of G. Pregartner [24]).

Initial assumptions:

These assumptions constitute a kind of 'qualitative mechanistic modeling' of the processes running inside the cell. Factors under consideration are:

(a) Temperature. As already found by S. Arrhenius in the 19-th century, every 10°C increase in temperature causes doubling of the reaction rate of many chemical reactions (Wikipedia [37]). This affects both, desirable as well as undesirable chemical reactions.

Low temperatures. Diffusion-processes become slow. This doesn't cause damages during discharging (the negative effect in this case is a reversible reduction in cell capacity - see figure 1.3), but during charging, due to the current flow which is forced into the battery by the externally applied voltage. Harmful alternative reactions instead of the desired ones can be triggered, especially at the anode. If the process of chemical conversion of the Li^+ ions on the surface of the graphitic electrode during charging happens faster than diffusion into the graphitic particles, metallic lithium is accumulating on the surface of the anode. This process is called **plating**.

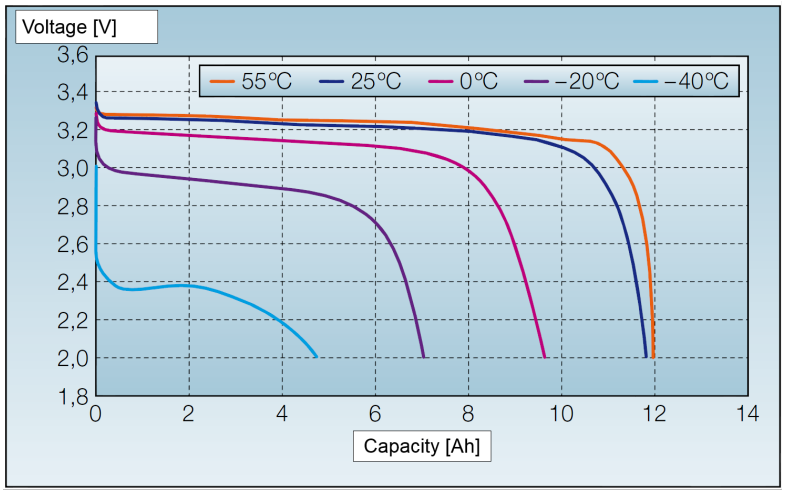


Figure 1.3: Discharge curves, showing the impact of temperature on cell capacity (Vezzini [31]).

Plating causes the growth of Li-dendrites, which can, in the extreme case, cause a short circuit. In less drastic cases, plating contributes to aging, because lithium ions get lost, e.g. due to the formation of unsolvable lithium compounds at the interface to the electrolyte. This loss of Li^+ ions results in a loss of capacity. From this it follows that the lower the temperature, the lower the charge current (CC) has to be in order to prevent aging. Hence, the influence of the charge current on aging is heavily temperature dependent. This is an example for a two-factor-interaction which we should observe in the data. If the charge current is too high at all, the temperature inside the battery will increase, causing local problems.

In the case of discharging, plating cannot happen, but the discharge current (DC) can cause local temperature increase. Especially DC-peaks are problematic. To take this aspect into account, the DC is pulsed to cover the dynamics of the automotive use case.

(b) Dynamics. From pulsing the DC, three factors arise: the peak discharge current (PDC), the average discharge current (ADC) and the frequency of pulses (F). It's not much known about the effects of dynamics, but:

- energy ($E = I^2Rt$, where I denotes the PDC, R the internal resistance and t the

pulse duration) which is brought in by the DC, especially by DC peaks, causes local temperature increase

- high voltage peaks, caused by high PDC values, can cause detrimental side reactions
- in addition, the lithium uptake into the graphite particles causes particle expansion - the faster the uptake is, the faster the particles have to expand. That means, the higher the PDC, the higher the risk of particle cracking (see figure 1.4).

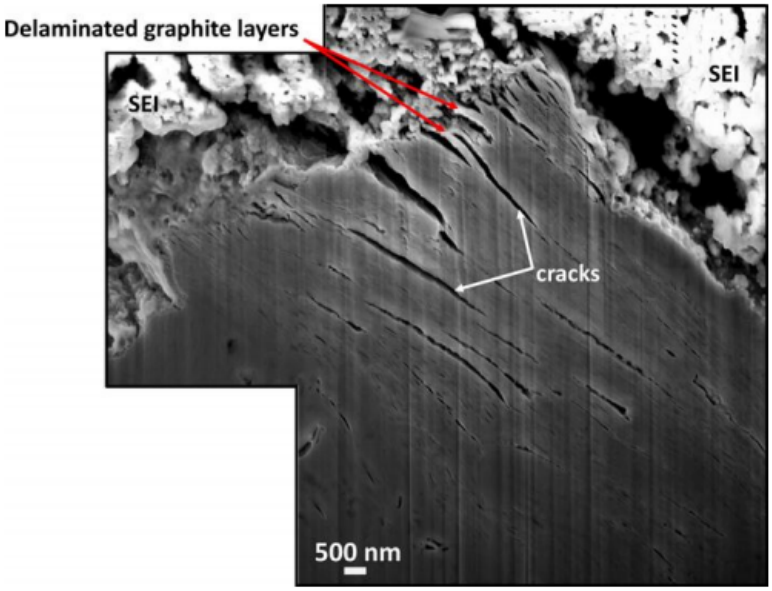


Figure 1.4: Cross-sectional SEM image of the subsurface region of the graphite electrode cycled between 0.0 and 3.0 V at a linear scan rate of 5 mV/s, showing the presence of cracks resulting in partial delamination of graphite layers at the SEI-graphite interface (ViF [34])

(c) Degree of utilization.

Two factors are investigated in this context (see figures 1.5 & 1.6):

1. The delta state of charge **(dSoC)**, which specifies the difference between the maximum and the minimum SoC levels of charge and discharge.
2. The average state of charge **(SoC)**, which specifies the arithmetic mean of the maximum and minimum levels of charge and discharge defined.

The range between the maximum charge level (max_{SoC}) and the minimum charge level (min_{SoC}), is the SoC range of the individual load point (see figure 1.6).

In the context of the experiment, the factor SoC represents the center around which the batteries actual state of charge is oscillating: $\text{SoC} = \frac{max_{SoC}+min_{SoC}}{2}$. It has an

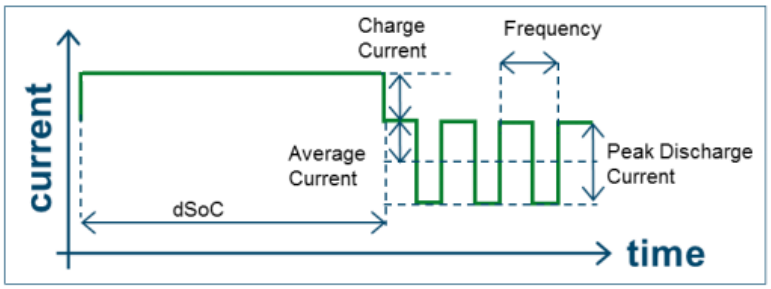


Figure 1.5: One cycle of charging and discharging in the case of a typical load point of the given experiment (ViF [34]).

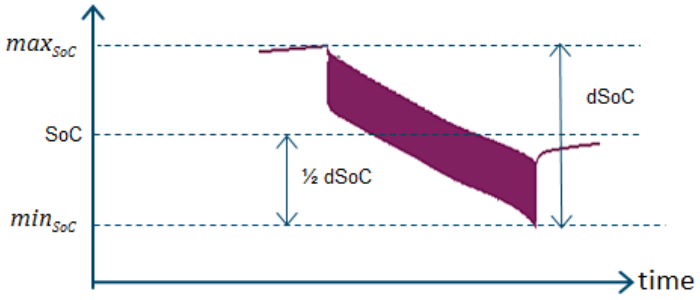


Figure 1.6: SoC change during discharging (ViF [35])

influence on particle structure and hence affects the sensitivity of the graphitic particles to changes of other factors.

The dSoC, or SoC swing, can cause mechanical damage to the graphitic particles. The larger the difference between maximum and minimum state of charge, the larger are expansion and contraction of the particles - cracks can happen to occur (see figure 1.4).

Additionally, the larger the dSoC, the larger is the voltage fluctuation at the electrodes and hence the risk of side reactions (limited electrochemical stability window of the electrolyte).

Another source of aging is the so-called solid electrolyte interphase (SEI) (see figure 1.4), a layer which is formed during initial charging by decomposing solvent on the graphitic electrode. The SEI slows down further electrolyte decomposition dramatically but is nevertheless getting thicker and thicker with time and thus increasing the internal cell resistance.

(d) Other sources of aging.

There are, of course, many more influencing factors. Under real world conditions, some of them are controllable while others are not (see figure 1.7). The factors used in the

experiment are chosen only amongst the controllable factors, all the others are held constant if possible.

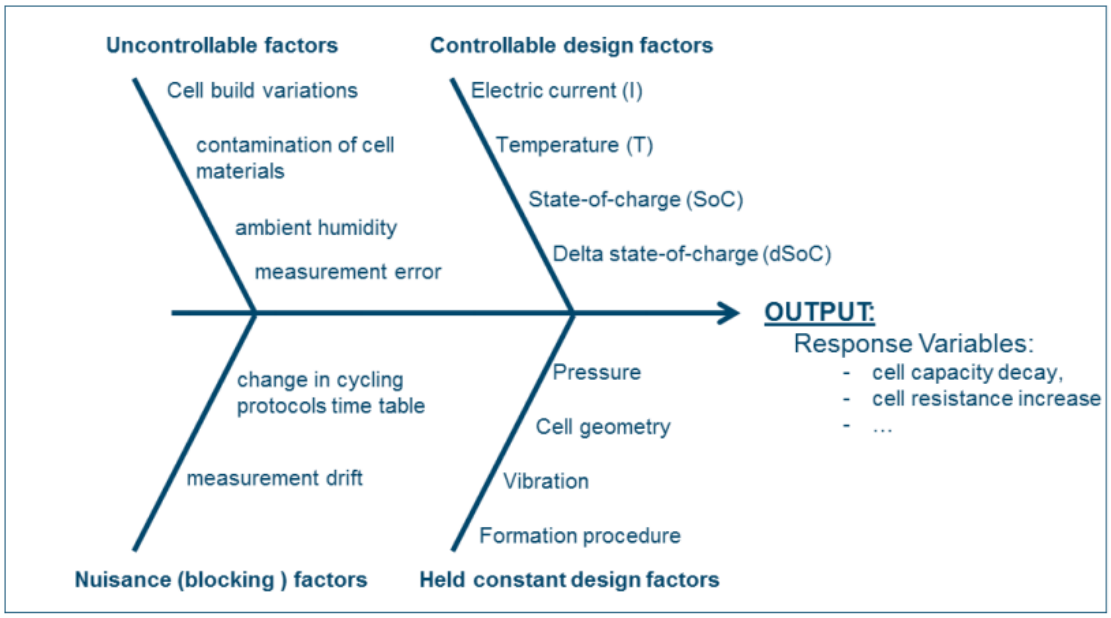


Figure 1.7: Fishbone diagram of influencing factors (ViF [33])

1.3 The Experiment

After discussing several influences, it had to be decided which one to choose and how to test for their impact on cell aging. The critical question of proper experimental design is a challenging science and will be discussed more deeply in chapter 2. At this point, we give a brief outline of the experiment that is actually run at VIRTUAL VEHICLE.

At the beginning, one has to choose the range of interest for each factor. The temperature T, for example, ranges between -10 and $+40$ degrees Celsius in the experiment. Then it has to be decided, which levels of the factors within their chosen ranges should be considered.

The experimental settings have to be chosen such that the statistical significance of the results is as high as possible. Therefore the combinations of the factor levels have to be selected (in the following often called **L**oad **P**oints or LP) on a sound statistical basis by employing proper design construction methods (e.g. usage of orthogonal designs or methods for creating optimal designs). There exists a vast literature on the topic of designing and analyzing experiments - see for example the books by Montgomery [21] and Pukelsheim [25].

The factors and factor levels which have been chosen are listed in table 1.1.

Factor	min	max	Factor Levels				
Temperature (T, Tc)	-10	40 [°C]	**-10**	**5**	**20**	**40**	
	-1	1	-1	-0,4	0,2	1	
Charge Current (CC,CCc)	0,2	2,4 [Cr]	**0,2**	**1,08**	**2,4**		
	-1	1	-1	-0,2	1		
Average Discharge Current	0	8 [Cr]	**0**	**0,2**	**1**	**4**	**8**
(ADC, ADCc)	-1	1	-1	-0,95	-0,75	0	1
Peak Current (PDC, PDCc)	0,2	14 [-]	**0,2**	**3**	**8**	**10**	**14**
	-1	1	-1	-0,5942	0,130435	0,42029	1
Frequency (F, Fc)	0,000333	0,5 [Hz]	**0,000333**	**0,03**	**0,1**	**0,5**	
	-1	1	-1	0,230596	0,559855	1	
SoC (SoC, SoCc)	15	95 [%]	**15**	**25**	**55**	**80**	**95**
	-1	1	-1	-0,75	0	0,625	1
DSoC (DSoC, DSoCc)	0,01	80 [%]	**0,01**	**2,5**	**15**	**50**	**80**
	-1	1	-1	-0,93774	-0,6252	0,249906	1

Tabel 1.1: Used factors and their levels

In addition, it has to be taken into account, that there exist constraints which prohibit the application of certain factor level combinations. On one hand, this has inherent physical reasons, which means that some factor level combinations are simply not possible (e.g. PDC lower than ADC) and on the other hand, one wants to avoid load points which would cause wrong side reactions.

One such example is the interaction of average discharge current, frequency and delta-state-of-charge (see figure 1.8). Due to physical reasons, the admissible test space covers not the whole cuboid. Instead, it is the area, which is enclosed by the red lines and the colored surface.

In total fourteen different constraints have been applied (see table 1.2), heavily reducing the number of eligible factor level combinations.

Moreover, due to limited resources, only a fraction of feasible factor level combinations can be realized.

Therefore, a classical orthogonal design can't be applied. Due to the complexity of the proper load point selection, a computer generated experimental design, a so called optimal design, is used.

After figuring out, how many and which of the feasible load points to use, the actual experiment can be put into effect.

The basic idea is to charge and discharge the batteries continuously (see figure 1.9) according to the chosen settings (factor level combinations) until a specific criterion indicates that the batteries end of life (EoL) is reached. One such sequence of charging and discharging is called a cycle. Therefore, the repeated application of cycles is called

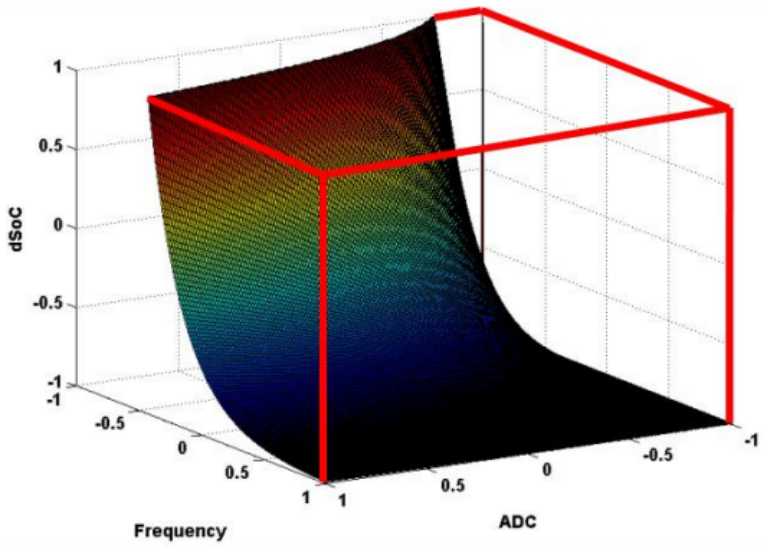


Figure 1.8: Admissible test space (ViF [33]).

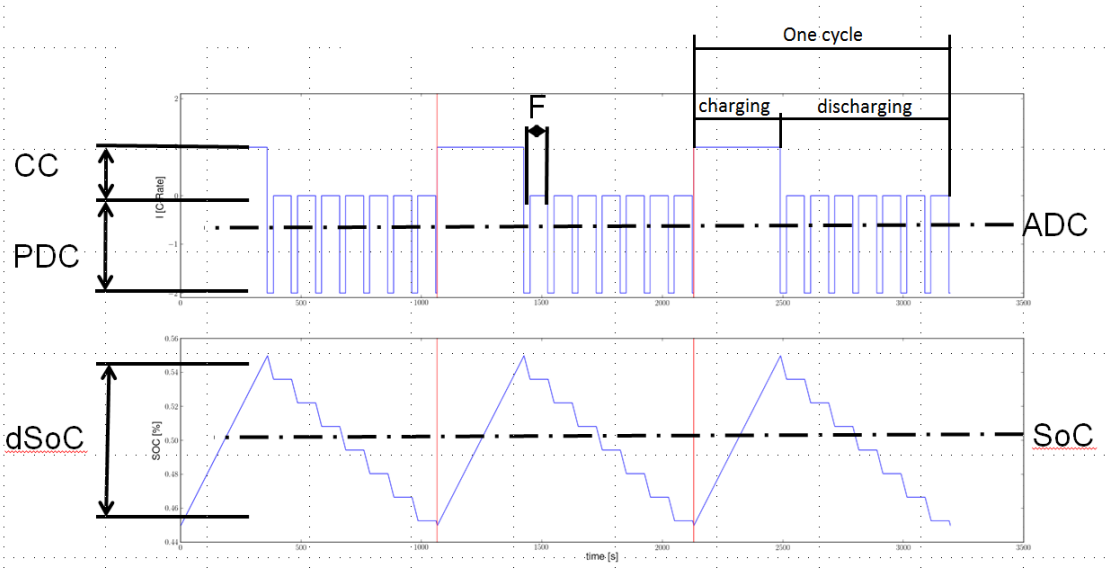


Figure 1.9: Repeated cycles of charging and discharging (ViF [32])

cycling. In the context of the actual experiment, a battery is said to have reached its end of life, if

1. the capacity of the cell has fallen below 70% of its initial capacity and/or
2. the internal resistance of the cell has reached a level above 300% of its initial internal resistance.

To establish the relation between several factors and the lifespan of the batteries, a linear regression approach has been chosen. That means, we assume a linear model describing the relationship between the chosen factors and the batteries end of life in an adequate way. More on the topic of linear modeling can be found in the next chapter.

At intervals of several weeks, the cycling has to be interrupted to assess capacity loss and resistance gain in order to stop the experiment when the EoL is reached and furthermore, to get data on the characteristics of capacity decay and resistance increase. This is done by applying the so called **R**eference **T**est **P**rocedure (RTP) - see figure 1.10. If the cells EoL is not reached yet, cycling is commenced again. This sequence of cycling and testing is repeated until at least one of the EoL criteria is reached.

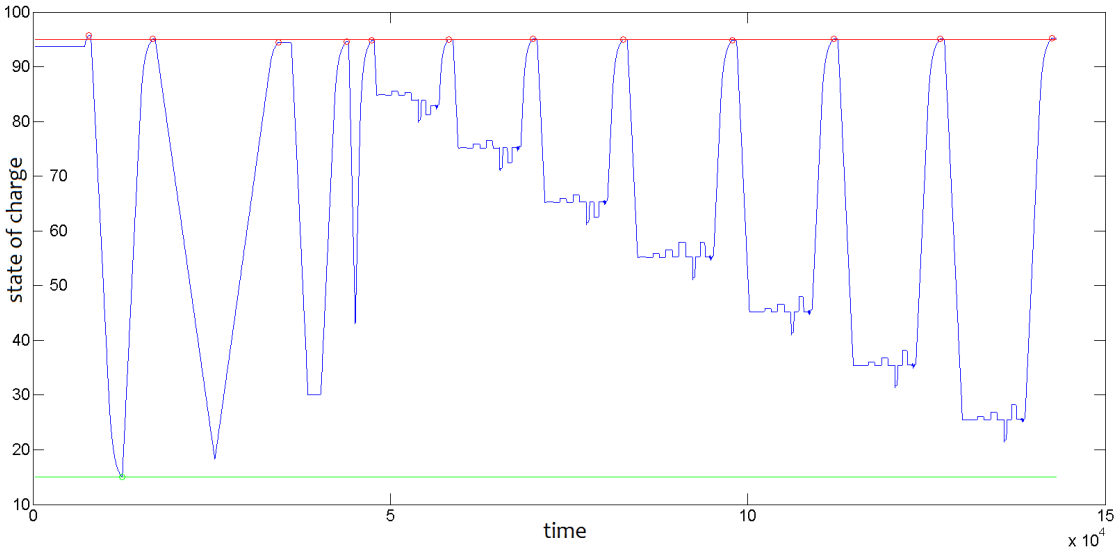


Figure 1.10: Reference Test Procedure

The data, which is gained this way, will hopefully not only lead to the building of a mathematical model, abstractly describing the relationship between several factors and the lifespan of batteries, but also, based on the mathematical model, allows for physical interpretations of factors effects. In the end, the goal is, to find a model, which empowers an industrial user to optimize his operational strategy giving a reasonable battery life.

Constraint	#
dSoCc <= 2*SoCc+1	I
dSoCc <= -2*SoCc+1	II
CCc <= ((1-g1(Tc))/1.55984)*Fc+((1+0.55984*g1(Tc))/1.55984)	III
((((0.01) +((1+dSoCc)*((80)-(0.01))/2)) >= (1/exp(log(0.000333) +((1+Fc)*(log(0.5)-log(0.000333))/2)))*(((1+ADCc)*(8)/2)) /36) | (ADCc==-1 & dSoCc == -1 & Fc == -1))	IV
PDCc<= ((0.80357-g2(Tc))/1.230596)*Fc+ ((0.80357+0.230596*g1(Tc))/1.230596)	V
((((0.01) +((1+dSoCc)*((80)-(0.01))/2)) >= (1/exp(log(0.000333) +((1+Fc)*(log(0.5)-log(0.000333))/2)))*((0.2 +((1+CCc)*(2.4-0.2)/2))) /36) | (CCc==-1 & dSoCc == -1 & Fc == -1))	VI
PDCc <= 0.2552995*Fc+0.7447005	VII
PDCc >= 0.5797077*ADCc-0.4492777	VIII
PDCc <= 8*SoCc+7	IX
PDCc <= 40*ADCc+39	X
CCc <= 40*ADCc+39	XI
dSoCc <= 40*ADCc+39	XII
Fc <= 40*ADCc+39	XIII
ADCc >= (0.875 * (PDCc + 1) * exp(log(0.000333) + ((Fc + 1) * (log(0.5) - log(0.000333)))/2)) - 1 & ADCc <= g2(Tc))	XIV

```
g1:

function (y1)
{
    -0.325221 * y1^3 + 0.00389 * y1^2 + 1.45362 * y1 + 0.148575
}

--------------------------------------------------------------------------------

g2:

function (y2)
{
    -0.082025 * y2^3 + 0.020021 * y2^2 + 0.351638 * y2 - 0.557037
}
```

Table 1.2: Constraints applied in constructing the set of candidate points.

2 Statistical concepts and methods

2.1 The need for statistical modelling

Mathematical modelling is a generic tool for the description of systems, which stem from a great variety of scientific fields as well as engineering disciplines. There exists a vast literature on this topic - see for example the introductory book by Velten [30]. The goal of modelling is to explain a system in terms of its parts (**white box model**) or at least its behaviour (**black box model**). White box models describe for example mechanical systems, i.e. systems for which all relevant components and their behaviour is known. Such a model can be used to study the effects of different parts of the system and to use this knowledge to make predictions about the systems behaviour. Black box models constitute the other end of the model spectrum. In this case, the systems are too complex to be understood in a mechanistic way, for example the response of plants to different nutrition regimes. Only the relationship between input and output is modelled - this knowledge is used for predicting the output (response variable(s)) given a certain input (independent variable(s)).

To be able to cope with the great variability of systems under investigation, many different forms of models exist. Below, some model categories are shown (this categories can of course overlap):

The model can be

- linear - all operators in the model are of a linear type,
- discrete or continuous,
- deterministic or stochastic. The latter means, that randomness is involved.
- static or dynamic. The latter means, that there is time dependency involved.

In the case of the Li-ion battery, we will consider the relationship between its end of life (EoL) and seven different input variables, namely: t_1 = temperature (T), t_2 = charge current (CC), t_3 = average discharge current (ADC), t_4 = peak discharge current (PDC), t_5 = frequency (F), t_6 = state of charge (SoC) and t_7 = delta state of charge ($dSoC$), which has been discussed in more detail in chapter 1. $\boldsymbol{t} = (t_1, ..., t_7) \in \mathcal{T}$ represents the *experimental conditions*, which can freely be chosen by the experimenter from the *experimental domain* $\mathcal{T}$. In this case, obviously, $\mathcal{T} \subseteq \mathbb{R}^7$. Additionally, the experimental domain does not only depend on the number of independent variables, but

also on their ranges and on further constraints prohibiting certain treatment combinations.

The battery is, at least in the case of the end of life, a stochastic system. There is some randomness due to factors which are unknown or at least cannot be controlled properly. A deterministic system will always produce the same output for a given input, there is no uncertainty involved (e.g. the relationship among voltage, current and resistance: $V = R \cdot I$).

A stochastic system doesn't behave like this - a repeated certain input will produce different outputs - this is modelled by a noise term. The noise term makes the difference between statistical and deterministic modelling, and, due to the stochastic nature of the system under investigation, the goal is now to predict the expected outcome of an experiment and not to exactly predict the outcome of a single experiment.

In this modelling context, the battery is considered as a black box, and one only tries to model the input dependency of its lifespan, without considering its inner life (no mechanistic interrelationship flows into the process of modelling). To create a black box model, one has to estimate not only the functional relationship between in- and output variables but also the parameters in the chosen function.

Due to the afore mentioned fact, that the true functional relationship $EoL = f(t_1, ..., t_7)$ is unknown, the experimenter has to choose an (hopefully) appropriate function to approximate f. Low order polynomial models have proven to be the proper choice in most cases as approximating functions. Great care has to be taken, to avoid under- and especially overfitting, when choosing this function (the topic of model evaluation will be discussed in detail in section 2.3.3).

2.2 Multiple Linear Regression

This section is intended to comprise a short overview of the mathematical background of multiple linear regression methodology, especially in the light of unbalanced data and is, as not explicitly stated otherwise, based on the lecture notes of Stadlober [27], Stadlober [28] and the books by Montgomery [21], Montgomery et al. [22], Fahrmeir et al. [6] and Jørgensen [14].

At the beginning of the project, it was decided to explain the influence of several different factors (the independent variables t_i) on several properties of the batteries (the response or dependent variables), by means of linear regression.

In the context of this work, the response is always the batteries end of life (EoL), which is derived from the measurements of battery characteristics like cell capacity or internal resistance and the seven independent variables are $t_1, ..., t_7$ already mentioned above.

For linear regression one tries to fit a straight line (simple linear regression or SLR) or a (hyper-) plane (multiple linear regression or MLR) to the given data as best possible (least-squares-method). Thus, a model of the following form is assumed (classical linear model with normally distributed errors)

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \epsilon_i \quad \text{for} \quad i = 1, \ldots, n \tag{2.1}$$

where k indicates the number of regressors, i.e. $k = 1$ means simple linear regression and $k > 1$ means multiple linear regression. The subscript i denotes the $i-th$ observation of the experiment, whose output consists of n observations in total. The number of observations, which is necessary to be able to estimate all model parameters properly, must exceed k, so $n \geqslant k + 1$ has to be valid. Furthermore, ϵ_i is an unobserved random variable, that stands for the noise term in the $i-th$ observation. All ϵ_i are independent of each other and are identically and normally distributed: $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$.

The notion 'linear' indicates that the model is linear in the parameters β_j, which are called regression coefficients in this context, but not necessarily in the independent (or regressor respectively predictor) variables x_{ij} for which any functional form is possible. The parameter β_0 is called *intercept* and it is the expected value of the response when all regressors equal zero.

Now, lets have a look at the regression model **lm.ex**, which was stated at the beginning of the ongoing experiment based on considerations which are discussed to some extent in chapter 1:

$$\begin{aligned} y = &\beta_0 + \beta_1 T + \beta_2 CC + \beta_3 ADC + \beta_4 PDC + \beta_5 F + \beta_6 SoC + \beta_7 dSoC + \\ &\beta_8 T^2 + \beta_9 ADC^2 + \beta_{10} SoC^2 + \beta_{11} dSoC^2 + \beta_{12} T^3 + \beta_{13} T \cdot C + \\ &\beta_{14} T \cdot PDC + \beta_{15} CC \cdot PDC + \beta_{16} CC \cdot dSoC + \beta_{17} PDC \cdot F + \\ &\beta_{18} PDC \cdot SoC + \beta_{19} PDC \cdot dSoC + \beta_{20} F \cdot dSoC + \beta_{21} SoC \cdot dSoC + \\ &\beta_{22} T \cdot F + \beta_{23} T \cdot SoC + \beta_{24} F \cdot SoC + \epsilon. \end{aligned} \tag{2.2}$$

This is a model in 7 variables, but we have 24 regressors - this regressors are functions of the 7 variables. If we let $x_1 = T$, $x_2 = CC$, ... , $x_7 = dSoC$, $x_8 = T^2$, ... , $x_{24} = F \cdot SoC$ then the model can again be written as

$$y = \beta_0 + \sum_{j=1}^{24} \beta_j x_j + \epsilon.$$

This is the familiar form (2.1) of a linear regression model, which we already encountered at the beginning of this chapter.

Using some additional notation, (2.1) can also be written as:

$$y = g(\boldsymbol{t})^T \boldsymbol{\beta} + \epsilon$$

with $\boldsymbol{x} = g(\boldsymbol{t}) = (1, x_1, \ldots, x_k)$ and $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_k)$. g is called the *regression function* and $\boldsymbol{x}$ the *regression vector*. In addition to the term experimental domain $\mathcal{T}$, the term *regression range* $\mathcal{X} = \{g(t) \mid t \in \mathcal{T}\} \subseteq \mathbb{R}^{k+1}$ is introduced at this point, which will come up again in the subsection about optimal designs.

From (2.1) it follows, that the expected value or mean of each observation equals

$$E(y_i) = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}.$$

Due to the fact, that all errors are assumed to be normally and identically distributed with variance σ^2, the distributions of the observations are therefore given by

$$y_i \overset{ind}{\sim} N(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}, \sigma^2).$$

It simplifies calculus, to write model (2.1) in matrix notation:

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad , \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 I) \tag{2.3}$$

where

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} \quad \text{and} \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

Next, the already mentioned method of least squares, which is typically used for estimating the unknown model parameters β_j will be introduced.

2.2.1 Least Squares Method

Provided that $n > k$ observations on the response y are available, the model parameters can be estimated as follows:

The idea is, to choose the β's in model (2.1) such that the sum of squared errors

$$S = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - (\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}))^2 \tag{2.4}$$

is minimized.

In matrix notation (2.3), equation (2.4) can be written as follows:

$$S = \sum_{i=1}^{n} \epsilon_i^2 = \boldsymbol{\epsilon}^T \boldsymbol{\epsilon} = (\boldsymbol{y} - X\boldsymbol{\beta})^T (\boldsymbol{y} - X\boldsymbol{\beta}) = \boldsymbol{y}^T \boldsymbol{y} - 2\boldsymbol{\beta}^T X^T \boldsymbol{y} + \boldsymbol{\beta}^T X^T X \boldsymbol{\beta}.$$

Now we have to find the vector of estimates

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^{k+1}}{argmin}\ S = \underset{\boldsymbol{\beta} \in \mathbb{R}^{k+1}}{argmin}\ (\boldsymbol{y}^T \boldsymbol{y} - 2\boldsymbol{\beta}^T X^T \boldsymbol{y} + \boldsymbol{\beta}^T X^T X \boldsymbol{\beta}).$$

In the case of $\hat{\boldsymbol{\beta}}$, the following equation must hold:

$$\left.\frac{\partial S}{\partial \boldsymbol{\beta}}\right|_{\hat{\boldsymbol{\beta}}} = -2X^T \boldsymbol{y} + 2X^T X \hat{\boldsymbol{\beta}} \stackrel{!}{=} 0. \tag{2.5}$$

Rewriting equation (2.5) yields $X^T X \hat{\boldsymbol{\beta}} = X^T \boldsymbol{y}$. That leads to

$$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \boldsymbol{y} \tag{2.6}$$

which is the **least squares estimator (LSE)**.

Definition 2.1. An estimator $W : \mathbb{R}^n \to \mathbb{R}^p$ is called a *linear estimator*, if a matrix $M \in \mathbb{R}^{p \times n}$ exists, such that $W(\boldsymbol{y}) = M\boldsymbol{y}$ for all $\boldsymbol{y} \in \mathbb{R}^n$.

The LSE (2.6) is, according to definition 2.1, obviously a linear estimator.

The process of parameter estimation is called **model fitting**. The vector $\hat{\boldsymbol{\beta}}$ contains the estimated parameter values, which are used to **parametrize** the model.

The resulting, fitted linear regression model can be written as follows:

$$\hat{\boldsymbol{y}} = X\hat{\boldsymbol{\beta}} = X(X^T X)^{-1} X^T \boldsymbol{y} = H\boldsymbol{y}. \tag{2.7}$$

The $n \times n$ matrix H is called the **hat-** or **predictor-matrix**.

$\hat{\boldsymbol{y}}$ constitutes the vector of the estimated response values, which is therefore also called vector of **fitted values**.

The so called **residuals** e_i are the differences between the fitted values $\hat{y}_i$ and the corresponding observed values y_i. In vector notation this can be written as

$$\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}}. \tag{2.8}$$

The residual serves as an estimator for the unobservable error ϵ. The residuals can be used for estimating the error variance σ^2 by means of the residual (or error) sum of squares:

$$SS_E = \sum_{i=1}^{n} e_i^2.$$

If the model has (including intercept) $k+1$ parameters, than the SS_E has $n-(k+1)$ degrees of freedom. The expected value of SS_E, $E(SS_E)$ can be shown to be equal to $\sigma^2(n-(k+1))$. Therefore

$$\hat{\sigma}^2 = \frac{SS_E}{n-(k+1)} \tag{2.9}$$

is an unbiased estimator of σ^2.

The term on the right-hand side of equation (2.9) is called the mean squared error or $\boldsymbol{MS_E}$.

Definition 2.2. Let Θ denote the parameter space, that contains all possible values of a parameter vector $\boldsymbol{\theta}$. An estimator $\hat{\boldsymbol{\theta}}$ is called an *unbiased estimator* for a parameter vector $\boldsymbol{\theta} \in \Theta$ if and only if $E_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta}$ for all $\boldsymbol{\theta} \in \Theta$.

Definition 2.3. An estimator $\hat{\theta}$ is called a *best unbiased estimator* for a parameter $\theta \in \Theta$ if and only if $\hat{\theta}$ is unbiased and $Var_\theta(\hat{\theta}) \leqslant Var_\theta(\tilde{\theta})$ for all other unbiased estimators for θ, $\tilde{\theta}$, holds.

The following theorem, which is taken from Fahrmeir et al.[5] sums up some of the most important results on linear regression:

Theorem 2.1. ***Gauß-Markov-Theorem***

Let $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with $E(\boldsymbol{y}) = X\boldsymbol{\beta}$ and $Cov(\boldsymbol{y}) = \sigma^2 I$, then the following holds for the LSE $\hat{\boldsymbol{\beta}}$:

1. *$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$. Therefore, $\hat{\boldsymbol{\beta}}$ is an **unbiased** estimator of $\boldsymbol{\beta}$.*
2. *The **covariance matrix** of $\hat{\boldsymbol{\beta}}$ is given by $Cov(\hat{\boldsymbol{\beta}}) = E((\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}}))(\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}}))^T) = \sigma^2(X^TX)^{-1}$. The i-th element of the main diagonal of this symmetric matrix, is the variance of the i-th estimated parameter $\hat{\beta}_i$ and its (ij)-th element is the covariance between $\hat{\beta}_i$ and $\hat{\beta}_j$.*
3. *The LSE $\hat{\boldsymbol{\beta}}$ is the best linear unbiased estimator **(BLUE)**. That means, that the following holds for all linear estimators $\tilde{\boldsymbol{\beta}}$ with $E(\tilde{\boldsymbol{\beta}}) = \boldsymbol{\beta}$:*
   - *In general, $Var(\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}}) \leqslant Var(\boldsymbol{\lambda}^T\tilde{\boldsymbol{\beta}})$ for each linear combination with $\boldsymbol{\lambda} \in \mathbb{R}^{k+1}$.*

- *From this it follows, that especially* $Var(\hat{\beta}_i) \leqslant Var(\tilde{\beta}_i)$, $i = 0, \ldots, k$.

In terms of the above-mentioned theorem, the best possible linear prediction of the mean response to a given vector of regressors $\boldsymbol{x}$, is $\boldsymbol{x}^T\hat{\boldsymbol{\beta}}$.

Looking at the Gauß-Markov-Theorem, we can see, that the assumption of normally distributed errors is not a necessary precondition for the validity of the theorem. In addition, we assume the errors to be normally distributed (model (2.3)). Therefore, the estimator $\hat{\boldsymbol{\beta}}$ for the parameters, which constitutes a linear combination of the normally distributed observations, is itself normally distributed. Under this assumption, we can deduce the following **distributional properties of the estimators**:

1. $\hat{\boldsymbol{\beta}} \sim N_{k+1}(\boldsymbol{\beta}, \sigma^2(X^TX)^{-1})$
2. $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are independent
3. $(n-(k+1))\dfrac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-(k+1)}$

Statistical properties of the residuals:

According to (2.7) and (2.8), the vector of residuals $\boldsymbol{e}$ equals $(I-H)\boldsymbol{y}$.

From this it follows, that:

- $E(\boldsymbol{e}) = \boldsymbol{0}$
- $cov(\boldsymbol{e}) = \sigma^2(I-H) \Rightarrow$
  - $Var(e_i) = \sigma^2(1-h_{ii})$ for $i = 1, \ldots, n$ with h_{ii} denoting the i-th diagonal element of H
  - the residuals are **not uncorrelated**
  - the residuals have **heteroscedastic** variances

Due to the fact, that the residuals are correlated and have heteroscedastic variances, they can not simply be used as a mean for evaluating the model assumptions. To make better use of the residuals, we have to **standardize** or **studentize** them. Given that the model assumptions hold, we achieve homoscedasticity this way, but still have correlation. If the model is properly specified, correlation is usually weak and thus can be neglected. More on the topic of model evaluation can be found in section 2.3.3.

If the model assumptions are valid, the **standardized residuals**

$$r_i = \frac{e_i}{\hat{\sigma}\sqrt{1-h_{ii}}}$$

have homoscedastic variances and can therefore be used to test for homoscedasticity of the error terms. Since e_i shows up in the calculation of $\hat{\sigma}$, e_i and $\hat{\sigma}$ are not independent and thus, despite the fact that e_i is normally distributed and $(n-(k+1))\hat{\sigma}^2/\sigma^2 \sim \chi^2_{n-(k+1)}$, the standardized residuals are not t-distributed. To get rid of the problem of dependency, estimates are used, which are based upon the data set without the respective observation.

Studentized Residuals

$X_{(i)}$ and $\boldsymbol{y}_{(i)}$ denote the design matrix, lacking the $i-$th row of regressors and the vector of observations, with the $i-$th observation left out. Based on $X_{(i)}$ and $\boldsymbol{y}_{(i)}$, the parameters of the linear model are estimated as shown above for the case of the complete data. The parameters and residuals estimated this way are contained in the parameter vector $\hat{\boldsymbol{\beta}}_{(i)}$ and in the vector of residuals $\boldsymbol{e}_{(i)}$, respectively. They are used to calculate the predictor $\hat{y}_{(i)} = \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}_{(i)}$, the residual $e_{(i)} = y_i - \hat{y}_{(i)}$ and the estimated variance $\hat{\sigma}^2_{(i)} = \dfrac{\boldsymbol{e}_{(i)}^T \boldsymbol{e}_{(i)}}{n-k-2}$.

It can easily be shown, that

$$e_{(i)} \sim N(0, \sigma^2(1 + \boldsymbol{x}_i^T (X_{(i)}^T X_{(i)})^{-1} \boldsymbol{x}_i)).$$

It can further be shown, that $e_{(i)}$ and $\hat{\sigma}^2_{(i)}$ are independent.

In addition, we already know that $(n-k-2)\dfrac{\hat{\sigma}^2_{(i)}}{\sigma^2} \sim \chi^2_{n-k-2}$.

Putting all this together, we get the studentized residual r_i^* and its distribution:

$$r_i^* = \frac{e_{(i)}}{\hat{\sigma}_{(i)}(1 + x_i^T (X_{(i)}^T X_{(i)})^{-1} x_i)^{1/2}} = \frac{e_i}{\hat{\sigma}_{(i)}\sqrt{1-h_{ii}}} = \frac{\hat{\sigma}}{\hat{\sigma}_{(i)}} r_i \sim t_{n-k-2}$$

The studentized residuals are used for checking the model assumptions, and, in addition, knowing the distribution of the studentized residuals is helpful in identifying observations that are not in good agreement with the fitted model, given that the model is specified correctly.

2.2.2 Hypothesis Testing

In this section, a short overview on tests about model parameters is given. These procedures require the model assumptions stated in (2.3) to be true. The distributions of the model parameters, which are used for constructing the tests, depend on the validity of these assumptions.

Sum of Squares, Analysis of Variance (ANOVA) and the Significance of Regression

In the context of linear modeling, it has to be the first step in testing, to check whether there exists the assumed linear relationship between response and regressors.

The hypotheses used for testing the significance of this relationship are:

$$H_0 : \beta_1 = \beta_2 = ... = \beta_k = 0 \text{ vs. } H_1 : \beta_i \neq 0 \text{ for at least one } i = 1, \ldots, k$$

In order to do this test, we partition the total sum of squares SS_T, which reflects the total variation in the data set, into a sum of squares due to regression (or the model) and a sum of squares due to the noise (or error) in the data :

$$SS_T = SS_R + SS_E$$

where $SS_T = \sum_{i=1}^{n} (y_i - \bar{y})^2$ denotes the total sum of squares, $SS_R = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$ denotes the variation in the data which can be explained by the model and $SS_E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ denotes the variation in the data which cannot be explained by the model.

Often, the sum of squares are scaled by dividing them by their respective degrees of freedom (df). This scaled sums of squares are termed mean squares:

Source of Variation	Sum of Squares	df	Mean Squares
Total	SS_T	$n-1$	$MS_T = SS_T/(n-1)$
Regression	SS_R	k	$MS_R = SS_R/k$
Error	SS_E	$n-k-1$	$MS_E = SS_E/(n-k-1)$

Under the null hypothesis, SS_E/σ and SS_R/σ are independent and distributed as χ^2_{n-k-1} and χ^2_k respectively.

From this it follows, that

$$F = \frac{SS_R/k}{SS_E/(n-k-1)} = \frac{MS_R}{MS_E} \overset{H_0}{\sim} F_{k,n-k-1}.$$

Therefore, we reject H_0 for a given significance level α, if we have $F > F_{1-\alpha,k,n-k-1}$.

Multiple R^2 is a quantity usually considered in this context:

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T}.$$

It is a measure of the proportion of the total variability in the data, which is explained by the model. Unfortunately, R^2 increases if the number of model terms increases, even if the added terms are statistically insignificant. Therefore, a large value of R^2 does not necessarily tell that the regression model under consideration is a proper one. To overcome this problem, it is better to use a version of R^2, which is corrected for model size. This is the so called *adjusted multiple* R^2:

$$R^2_{adj} = 1 - \frac{SS_E/(n-k-1)}{SS_T/(n-1)}.$$

In many cases, in which unnecessary terms are added to the model, R^2_{adj} does not increase but does, in contrary, even decrease. If there is a great discrepancy between R^2 and R^2_{adj}, it is very likely, that there are unnecessary terms in the model. That, in turn implies, that, if R^2 and R^2_{adj} are in good agreement, it is likely that the considered model is a good choice. Several tests can be applied, to get a better impression of the significance of groups and, especially, of individual model terms.

Tests on Groups of Model Terms

To test the significance of individual model terms, we look at the following hypotheses and test statistic, respectively:

$$H_0 : \beta_i = 0 \text{ vs. } H_A : \beta_i \neq 0 \text{ which is tested using } t = \frac{\hat{\beta}_i}{\sqrt{\hat{\sigma}^2 (X^T X)^{-1}_{ii}}}$$

Due to the assumptions of model (2.3), $\hat{\beta}_i$ is normally distributed, with mean β_i, which is, under H_0, equal to zero, and variance $\sigma^2 (X^T X)^{-1}_{ii}$. Furthermore, we know that $(n-(k+1))\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-(k+1)}$ and that $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are independent. From this it follows, that t is t_{n-k-1}-distributed.

Therefore, we reject H_0 for a given significance level α if $|t| > t_{1-\alpha/2,n-k-1}$. If we can not reject H_0, there is evidence that x_i is not necessary to model the data.

A way to test, whether a group of predictor variables is significant, given that another group of predictors is already in the model, is the use of so called **extra sums of squares**. Let for example the $(k+1)$ dimensional vector of regression coefficients $\boldsymbol{\beta}$ be partitioned in the following way: $\boldsymbol{\beta} = (\boldsymbol{\beta_1}, \boldsymbol{\beta_2})$, with $\boldsymbol{\beta_i}$ denoting subvectors such that the dimension l of $\boldsymbol{\beta_2}$ fulfills $1 \leqslant l < k$ and let $SS_R(\boldsymbol{\beta_1})$ indicate the SS_R with respect to a model which just includes the parameters which are contained in $\boldsymbol{\beta_1}$. To test, whether $\boldsymbol{\beta_2}$ contributes significantly to the model, the following test can be applied:

$$F = \frac{SS_R(\boldsymbol{\beta_2}|\boldsymbol{\beta_1})/l}{MS_E} \overset{H_0}{\sim} F_{l,n-k-1}$$

with $SS_R(\boldsymbol{\beta_2}|\boldsymbol{\beta_1}) = SS_R - SS_R(\boldsymbol{\beta_1})$ denoting the extra sum of squares due to $\boldsymbol{\beta_2}$, with *df* equal to the number of parameters in this vector and provided that $\boldsymbol{\beta_1}$ is already part of the model. Under the null hypothesis, $SS_R(\boldsymbol{\beta_2}|\boldsymbol{\beta_1})/\sigma$ and SS_E/σ are independent and distributed as χ^2_l and χ^2_{n-k-1} respectively. We reject the null hypothesis $\boldsymbol{\beta_2} = \mathbf{0}$ for a given confidence level α, if we have $F > F_{1-\alpha,k,n-k-1}$ which implicates, that there is evidence in the data, that at least one of the parameters in $\boldsymbol{\beta_2}$ is unequal to zero.

Test for Lack of Fit

As last step, the topic of testing for lack of fit is discussed. This is a test, whether the relationship is linear, i.e. one wants to test whether

$$H_0 : E(Y_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{p-1} x_{i(p-1)}$$

(the assumed model in this case) holds. Replication of at least one single point (e.g. the center point in the case of the central composite design - see page 40 for more details) allows to calculate what is called sum of squares due to *pure error* (SS_{PE}). This is another type of estimator for σ^2 which does no longer depend on the assumed model, i.e. this type of error estimation does now depend solely on the true (but unknown) model. In this case, it is possible to partition the SS_E, calculated with respect to the model assumed, into two quantities:

$$SS_E = SS_{PE} + SS_{LOF}. \tag{2.10}$$

The latter quantity on the right hand side of the equation (2.10) is called sum of squares due to *lack of fit*.

Let m be the number of pairwise different design points and $n > m$ the number of independent observations. Then $df(SS_{PE}) = n - m$ and $df(SS_{LOF}) = m - p$. The test statistic of the lack of fit test then has the following form and is, given H_0 holds, distributed as follows:

$$F_{LOF} = \frac{MS_{LOF}}{MS_{PE}} \sim F_{m-p,n-m}.$$

Confidence Intervals

An essential prerequisite for constructing such intervals is that the assumptions of the error distribution, stated in (2.3) hold. That means, the observations from the experiment can be described by model (2.1).

Therefore, the estimator $\hat{\boldsymbol{\beta}}$ for the parameters is normally distributed with mean $\boldsymbol{\beta}$ and covariance matrix $\sigma^2(X^TX)^{-1}$.

Utilizing this knowledge, concerning the distribution of $\hat{\boldsymbol{\beta}}$, we cannot only construct confidence intervals for the parameters β_i, but additionally also for the response and its mean. If we let $\boldsymbol{x}_{pre} = (1, x_1, \ldots, x_k)$ denote a particular regressor, we can give confidence intervals for the mean response $\mu_{pre} = \boldsymbol{\beta}^T\boldsymbol{x}_{pre}$ as well as for the response $y_{pre} = \boldsymbol{\beta}^T\boldsymbol{x}_{pre}+\epsilon$ using the distribution of the estimated mean response $\hat{\mu}_{pre} = \hat{\boldsymbol{\beta}}^T\boldsymbol{x}_{pre}$.

At this point, we already know, that

$$t = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}^2(X^TX)^{-1}_{ii}}} \sim t_{n-k-1} \quad \text{for} \quad i = 0, \ldots, k \tag{2.11}$$

and, because $\hat{\mu}_{pre}$ is also normally distributed with $E(\hat{\mu}_{pre}) = \boldsymbol{\beta}^T\boldsymbol{x}_{pre}$ and $V(\hat{\mu}_{pre}) = \sigma^2\boldsymbol{x}_{pre}^T(X^TX)^{-1}\boldsymbol{x}_{pre}$, we have

$$t = \frac{\hat{\mu}_{pre} - \mu_{pre}}{\sqrt{\hat{\sigma}^2\boldsymbol{x}_{pre}^T(X^TX)^{-1}\boldsymbol{x}_{pre}}} \sim t_{n-k-1}. \tag{2.12}$$

To get a prediction interval for a future observation $y_{pre} = \boldsymbol{\beta}^T\boldsymbol{x}_{pre} + \epsilon$, we have to look at the error of prediction, $\hat{\epsilon}_{pre} = y_{pre} - \hat{\boldsymbol{\beta}}^T\boldsymbol{x}_{pre}$ for which

$$\hat{\epsilon}_{pre} \sim N(0, \sigma^2 + \sigma^2\boldsymbol{x}_{pre}^T(X^TX)^{-1}\boldsymbol{x}_{pre})$$

proves to be valid. From this, it follows, that

$$t = \frac{y_{pre} - \hat{\boldsymbol{\beta}}^T\boldsymbol{x}_{pre}}{\sqrt{\hat{\sigma}^2 + \hat{\sigma}^2\boldsymbol{x}_{pre}^T(X^TX)^{-1}\boldsymbol{x}_{pre}}} \sim t_{n-k-1}. \tag{2.13}$$

Due to the duality between two sided test statistics and confidence intervals, such intervals can be constructed by **inverting** the corresponding tests:

- Confidence intervals for Model Parameters:
  From inverting (2.11), we get the following $100(1-\alpha)$ percent confidence interval for the regression coefficients $\beta_i, i = 0, \ldots, k$:

$$\hat{\beta}_i - t_{\alpha/2,n-k-1}\sqrt{\hat{\sigma}^2(X^TX)^{-1}_{ii}} \leqslant \beta_i \leqslant \hat{\beta}_i + t_{\alpha/2,n-k-1}\sqrt{\hat{\sigma}^2(X^TX)^{-1}_{ii}}$$

- Confidence interval for the mean response:
  From inverting (2.12), we get the following $100(1-\alpha)$ percent confidence interval for the mean response:

$$\mu_{pre} \in [\hat{\mu}_{pre} \pm t_{\alpha/2,n-k-1}\sqrt{\hat{\sigma}^2 \boldsymbol{x}^T_{pre}(X^TX)^{-1}\boldsymbol{x}_{pre}}]$$

- Prediction interval for a future observation:
  From inverting (2.13), we get the following $100(1-\alpha)$ percent prediction interval for a future observation:

$$y_{pre} \in [\hat{\mu}_{pre} \pm t_{\alpha/2,n-k-1}\sqrt{\hat{\sigma}^2(1 + \boldsymbol{x}^T_{pre}(X^TX)^{-1}\boldsymbol{x}_{pre})}]$$

Confidence intervals constructed this way still work properly if only small departures from the distribution assumptions occur.
In addition, it can be shown, that the test statistics and confidence intervals found, using the assumption of normally distributed error terms, asymptotically still prove to be valid, even if the error terms deviate from the normal distribution.
This means, that given a sufficiently large amount of observations, the test statistics and the confidence intervals developed, can still be used, even if the error terms are no longer normally distributed.

2.3 Experimental Design and Model Selection

2.3.1 Unbalanced Datasets

Replication of an experimental design means, that the whole design is independently repeated, e.g. if we have two replicates, each of the experimental conditions is applied to two different subjects (not to confuse with repeated measurements).

Replication of a design has two advantages:

1. It allows to estimate the experimental error without assuming a certain model. This error estimate is needed in determining, whether the chosen model fits the data well or not (lack of fit test). In addition, this error is used in assessing, whether differences in observations are statistically significant.

2. Let σ^2 denote the variance of a single observation. Due to the fact that $\sigma^2_{\bar{y}_i} = \frac{\sigma^2}{n}$, using the sample means $\bar{y}_i$ for estimating the true mean responses improves accuracy the higher the number of replicates is.

If all experimental conditions under consideration are replicated an equal number of times the design is said to be balanced.

The problem of giving a sound statistical analysis of a model is complicated by having an unbalanced design. This is often not a planned result, but a common complication which often arises in conducting an experiment due to unpredictable factors, e.g. caused by problems like cells, which do not work properly - i.e. not all of them yield usable observations.

If, in the course of analysis of such data, its unbalancedness is not taken into account, i.e. the data is analyzed as it was balanced, the results of the analysis can be severely distorted and therefore suggest misleading conclusions.

Motulsky et al. [23] point out, that one of the problems, which arises in the context of unbalanced data, is, that unequal weight is given to the several treatments. They also mention, that averaging the replicated responses in order to treat the mean as a single data point, is no proper remedy: "If you have different numbers of replicates for different values, then you definitely should not average the replicates and fit to the means. The results will be wrong." In the case of a balanced design, using the averaged replicates as well as the replicates itself yields the same parameters but different error estimates and therefore different confidence intervals. Their conclusion is to use the replicates and not to average them.

But how can this be done? What does theory tell us and what types of remedies exist?

Two different approaches can be distinguished:

I.) Approximate Methods:

If the data is sufficiently close to the balanced case it might be possible to apply approximate procedures to transform the unbalanced set of data into a balanced one. In this case, one has to keep in mind, that this leads to an approximate analysis too. Thus, we have to look at the trade-off between the great ease of analysis in the case of balanced data and the error which is additionally introduced by making the unbalanced data balanced.

- **Trimming down the dataset.** If almost all experimental conditions are replicated equally often and only relatively few conditions are applied more frequently, setting aside observations which belong to the latter case, is indeed a reasonable approach. One strategy is to randomly choose the observations which should be excluded from analysis, and, instead of discarding these observations, we use them for interchange with the observations used in the analysis and the conduction of

repeated analyses to get an impression of whether there exists a significant influence of single observations, i.e. if conflicting results arise. This problem is not very likely in cases, were only few observations are excluded and the variability amongst observations with respect to the same experimental conditions is relatively small. If we observe great impact of an observation, it should be treated as an outlier and set aside anyway. Another strategy to reduce the number of observations is, to keep the smallest and the largest value and in addition, if more than two observations are needed, the observations which are most evenly distributed within the interval in between the most extreme (smallest/largest) observations. In the case of three observations needed, this means to keep in addition the observation which lies closest to the average of the most extreme observations.

- **Estimation of missing observations.** If almost all experimental conditions are replicated equally often and only relatively few conditions are applied less frequently, estimating the missing observations which belong to the latter case, is now obviously the less influential manipulation. How can this estimation be done? In the context of factorial experiments, Yates [41] suggested to use estimates such that the resulting error sum of squares for the design is minimized. To get these estimates, the error sum of squares is written as a function of the (unknown) missing observations. Differentiating with respect to the missing observations, setting the differentials to zero and solving the resulting equations with respect to the unknowns, yields the estimates of the missing observations. Analyzing the design which is balanced by using these estimates, is done in the same way as with any other balanced dataset, with the exception, that the error degrees of freedom have to be reduced by the number of estimated observations Giri [3].
- Further techniques are the **method of unweighted means** and the **weighted squares of means method**. These methods are based on the sums of squares of **cell means**.

II.) Exact Method. Approximate methods cannot be applied in all situations, e.g. situations in which the differences in the numbers of observations per experimental condition vary considerably and/or empty cells occur. In the case of conducting ANOVA, based on unbalanced data sets, there exist different types of sums of squares, which help in adjusting for unbalancedness. There are in total four different types of sums of squares, which are called Type I, Type II, Type III and Type IV (Goodnight [10]). More on this topic can be found for example in Littell et al. [17] or in the supplemental materials of the book by Montgomery [21] which can be found on the corresponding textbook website.

From a generic point of view, the problem of unbalanced designs can be seen as a generalization of the balanced case. Therefore, the balanced case has to be viewed as a special case, which is embedded in the theory of unbalanced designs. See for example the book of Searle [26] which deals with this approach.

Since only approximate methods are applied in the analyses which are part of this work, we will not go deeper into theory and application of the exact methods.

2.3.2 Model Selection

Prior to setting up the experimental design, regressors have been defined, which might be suitable to describe the relationship between in- and output. This process, which is outlined in more detail in chapter 1 and in section 2.1, leads to the formulation of a model that is intended to serve as a starting point for further analysis.

The data, describing this relationship, which is gathered in the course of the experiment (applied input and measured response) is then used to fit the model to the actual relationship, by removing regressors which do not prove to have any significant influence on the response and by calculating the proper parameter estimates for those, who prove to be necessary. That means, we are trying to trim down the model to an extent, so that we end up with a model that includes all meaningful regressors and leaves out all, which do not prove to have an effect.

One issue which should always be kept in mind when using regression analysis is what can be expected to be achieved by utilizing this method. Especially in the case of a black box model, no mechanistic explanation of the relationship between in- and output can be found by using regression. C. Geyer concisely pointed out this fact in his lecture notes (Geyer [9]): "All of this is related to our slogan about 'regression is for prediction, not explanation. If you can't decide which model is 'best' and are honest enough to admit that lots of other models are equally good, then how can you claim to have found the predictors which 'explain' the response? Of course, if you really understand 'correlation is not causation, and regression isn't either,' then you know that such 'explanations' are bogus anyway. Thus it is silly to get excited about exactly which model is chosen as the 'best' by some model selection procedure (any procedure)! When many models are equally good, the specific features of any one of them can't be very important."

There exist several different **selection criteria**, which are used as a measure for describing how well a model suits the goal of the modelling procedure. The goal is, not surprisingly, to find a model, which approximates the underlying relationship to a degree, which allows to make reasonable predictions about the systems' behaviour under conditions that lie within the feasible operating range (see Chapter 1).

In order to avoid under- and especially overfitting, a trade-off between model complexity and model fit has to be found. Especially in the case, where a large number of potential regressors are available, this is a task of high complexity, and therefore it is not easy to find a definite, unambiguous result. In order to reduce the uncertainty about the outcome of the selection procedure to a reasonably small extent, many different criteria are considered simultaneously.
Overfitting is likely to occur, when many potential regressors are available. Therefore,

special care has to be taken to avoid this problem. On Wikipedia [40], the following description can be found:

- Overfitting occurs, when a statistical model describes random error or noise instead of the underlying relationship.
- The reason for this is that the model is excessively complex.
- That means the model has too many parameters relative to the number of observations.
- Such a model has poor predictive performance, as it can exaggerate minor fluctuations in the data.

There exist several different selection criteria. In this paper we use the following:

Adjusted multiple $\mathbf{R}^2$. This is a measure of the proportion of the total variability in the data, which is explained by the model. This measure should therefore be as high as possible. Despite the fact, that it is corrected for model size, it still favors models which are unnecessarily complex. This criterion is discussed in more detail on page 22.

Mallow's Cp. Let p denote the number of regressors of the full model. If a reduced model, possessing $q < p$ regressors is sufficiently complex to explain the underlying relationship, the standardized sum of squared errors due to this model, $MSE(q)$, should therefore be an unbiased estimator of the error variance. Otherwise, due to bias, the estimation should tend to values larger than the true one. In addition, the standardized sum of squared errors due to the full model, $MSE(p)$, is still an unbiased estimator of the error variance. The Cp statistic compares the two measures of error variance in the following way:

$$Cp(q) = \frac{SSE(q)}{MSE(p)} - (n - 2q).$$

In the case, that the bias of the reduced model equals zero, the following equation obviously holds:

$$E(Cp(q)|bias = 0) = \frac{(n-q)\sigma^2}{\sigma^2} - (n - 2q) = q.$$

Therefore, plotting $Cp(q)$ for varying values of q with respect to q, yields a diagram, where the best models in terms of this criterion show Cp values close to the bisectrix.

Akaike Information Criterion. This criterion is very frequently used in the context of maximum likelihood estimators. The theory behind is a 'little bit' more challenging than in the case of the two criteria discussed above. A good introduction to this topic can be found in the book by Konishi et al. [15] on which the following summary is based:

Akaike proposed, that the goodness of a model should be measured in terms of its prediction capability. Based on this reasoning, he decided to measure the distance between the predictive distribution $f(x)$ which is given by the model and the true distribution $g(x)$ and to use this as a measure of the predictive power of the model. Akaike based this measure on the **Kullback-Leibler information (divergence)** (Kullback-Leibler [16]):

Let $\boldsymbol{x} = (x_1, \ldots, x_n)$ denote a n-dimensional vector consisting of n independent observations from a population with unknown distribution function $G(x)$ (true distribution function). In addition, the model, which has been specified by the researcher is associated with the distribution function $F(x)$. The goodness of the model is measured by assessing how close $F(x)$ is to $G(x)$. Letting $f(x)$ and $g(x)$ denote the respective probability (density) functions, this can be done by using the Kullback-Leibler information:

$$I(g;f) = E_G\left[\log\frac{g(X)}{f(X)}\right] = \begin{cases} \int_{-\infty}^{\infty} \log\frac{g(x)}{f(x)} g(x)dx, \text{ for continuous distributions,} \\ \sum_{i=1}^{\infty} \log\frac{g(x_i)}{f(x_i)} g(x_i), \text{ for discrete distributions,} \end{cases}$$

with $E_G[\ ]$ denoting the expected value with respect to the true distribution.

It is obvious, that, the smaller (closer to zero) $I(g;f)$ is, the smaller is the discrepancy between model and reality. Due to the decomposition of $I(g;f) = E_G[\log(g(X))] - E_G[\log(f(X))]$, we can see, that, because $E_G[\log(g(X))]$ is a constant, which depends exclusively on the true model, it is sufficient, to just look at the latter quantity on the right hand side to compare different models. Therefore, since $I(g;f) \geqslant 0$, the larger $E_G[\log(f(X))]$, the smaller $I(g;f)$ and the better the model.

The problem we are facing now is, that $E_G[\log(f(X))]$ depends on the unknown true distribution. To overcome this problem, we have to estimate this quantity. In order to do this, we replace $G(x)$ with an empirical distribution function $\hat{G}(x)$ with support $\boldsymbol{x}$ by defining $\hat{g}(x_i) = 1/n$ for $i = 1, \ldots, n$. This yields

$$E_{\hat{G}}[\log f(X)] = \int_{-\infty}^{\infty} \log f(x)\hat{g}(x)dx = \sum_{i=1}^{n} \hat{g}(x_i)\log f(x_i) = \frac{1}{n}\sum_{i=1}^{n} \log f(x_i). \tag{2.14}$$

If the number of observations, n, approaches infinity, (2.14) converges, according to the law of large numbers, to the expectation of $\log f(X)$:

$$\frac{1}{n}\sum_{i=1}^{n} \log f(x_i) \xrightarrow[n\to\infty]{} E_G[\log f(X)].$$

Therefore, (2.14) can be used as an estimate of $E_G[\log(f(X))]$. In addition, we can see, that (2.14) is nothing else than n^{-1} times the log-likelihood of the observed data with respect to the density function $f(x)$ which is associated with the model under consideration. Hence, n^{-1} times the frequently used log-likelihood can be seen as an approximation to the Kullback-Leibler information!

In the process of model selection, we deal with the situation that multiple models exist, with f_j , $j = 1, \ldots, n_m$, denoting the respective probability (density) functions. These models arise from the preliminary selection of potential regressors, and therefore the best model under all possible combinations of this regressors should be found. Parametrization of these models gives n_m different estimated parameter vectors $\hat{\boldsymbol{\theta}}_j$ of model-specific dimension n_j.

Using the K-L information for comparing the different parametrized models, according to the argumentation above, is done by calculating the expected log-likelihood $E_G[\log(f_j(X|\hat{\boldsymbol{\theta}}_j))]$ for each of the models under consideration, which can be approximated by using $\frac{1}{n_a}\sum_{i=1}^{n_a} \log f_j(z_i|\hat{\boldsymbol{\theta}}_j)$, with z_i , $i = 1, \ldots, n_a$, denoting n_a additional observations, which are independent from the first n observations x_i gathered during the experiment and used for model parametrization.

The statements, given above, unfortunately do not indicate that comparing the maximum likelihoods $l(\hat{\boldsymbol{\theta}}_j)$ of different models, which are calculated using the first n observations x_i, can be a sufficient method for selecting the most appropriate model under all available ones. This problem arises due to the fact, that $l(\hat{\boldsymbol{\theta}}_j)$ is a biased estimator of $n \cdot E_G[\log(f_j(X|\hat{\boldsymbol{\theta}}_j))]$ with a bias that depends on the number of model terms n_j. This bias stems from the fact, that the observations used for parametrization of the models are again used in comparing these models by using this data for calculating the maximum log-likelihoods $l(\hat{\boldsymbol{\theta}}_j)$ of the respective models.

From this it follows, that we need a **bias correction for the log-likelihood**.

The bias is definded by:

$$b(g; f) = E_{G(\boldsymbol{x})}\left[\sum_{i=1}^{n} \log f(X_i|\hat{\boldsymbol{\theta}}(\boldsymbol{X})) - nE_{G(z)}\left[\log f(Z|\hat{\boldsymbol{\theta}}(\boldsymbol{X}))\right]\right],$$

with $G(\boldsymbol{x})$ denoting the true joint distribution of the sample $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ and $G(z)$ denoting the true distribution G, with respect to an additional univariate random variable Z.

Based on the considerations, discussed so far, Akaike defined his information criterion as follows:

$$AIC = -2[log{-}likelihood(statistical\ model) - bias].$$

The factor -2 which shows up in the formula for the AIC comes into play due to information theoretical reasons (connection to Kullback-Leibler Information). If one wants to go deeper into this matter, one could for example read the original paper by Akaike [1] or the commented version by deLeeuw [4].

Akaike derived, given the validity of several regularity conditions, a formula for the bias. This formula gives the approximate bias, which is calculated with respect to the true (but unknown) distribution $G(x)$. To overcome the problem of not knowing the distribution $G(x)$, he additionally derived an asymptotic bias estimator and showed, that, under the assumption that the true model is contained within the set of models to choose from, the bias can be approximated by the number of parameters in the model. This yields the familiar formula for the AIC of a model under consideration:

$$AIC = -2[log{-}likelihood(statistical\ model) - p],$$

with p denoting the number of parameters in the model.

The formula above can be interpreted in a simple, heuristic way: The AIC rewards goodness of fit (the higher the likelihood, the better the fit) and penalizes complexity (the penalty term rises with the number of model parameters) to avoid overfitting. Therefore, the AIC is in favor of the model, that shows the lowest AIC value.

Unfortunately, the penalty term seems not to be large enough to serve as a reliable means against overfitting. Therefore, other information criteria, like the BIC, which will be discussed next, penalize complexity more strongly.

It is important, to point out, that the AIC does not provide information about the overall quality of a model. If all models under consideration do not fit well, the AIC will not indicate this problem and just relying on the AIC would therefore lead to choosing an improper model.

Bayesian Information Criterion. The BIC was developed by Gideon E. Schwarz in 1978. He was born in Austria in 1933 and emigrated to Israel, where he became a professor of statistics at the Hebrew University. The BIC, as the name already suggests, is based on a Bayesian argument - it should be noted, that AIC and BIC are justified by using totally different arguments. The BIC is given by:

$$BIC = -2 \cdot log{-}likelihood(statistical\ model) + p \cdot log(n),$$

with p denoting the number of parameters in the model and n denoting the number of observations made during the experiment.

Formally, the AIC and the BIC look quite similar. The BIC also rewards goodness of fit in the same way, the AIC does (the higher the likelihood, the better the fit) and penalizes complexity (the penalty term also rises with the number of model parameters) to avoid overfitting. Therefore, the BIC is, like the AIC in favor of the model, that

shows the lowest value of the criterion. But there exists an important difference - the BIC penalty term penalizes model complexity much stronger, than the corresponding AIC term. Now, the number of model terms is multiplied by $\log(n)$ instead of 2, as it is the case for the AIC. Therefore, the BIC is in favour of less complex models than the AIC. Due to this property, we decided to use the Bayesian information criterion.

The four criteria mentioned above (*adjusted multiple* R^2, *Mallow's* Cp, AIC and BIC), can, amongst others, now be used in the process of model selection to choose the most suitable model amongst the more or less large set of potential models. There exist several different procedures, some of them will be described in the following.

Forward Selection. This algorithm starts with the model that shows the lowest possible level of complexity. This is the model which consists of the intercept only. Then, regressors are added one at a time, by comparing all regressors which are not already in the model. This is done in a way, that the incorporation of the regressor which is finally chosen leads to the highest improvement of the goodness of the model in terms of the criterion used. This procedure stops, when no further improvement of the model is achievable.

Backward Selection. This algorithm starts with the model which shows the highest level of complexity possible. This is the model which consists of all regressors under consideration. Then, step by step, one regressor after the other is excluded in a way, that leads to the highest stepwise improvement of the goodness of the model in terms of the criterion used. This procedure stops, when no further improvement of the model is achievable.

Bidirectional Selection. A combination of the two afore mentioned procedures, i.e. it considers at each step whether a single regressor should be dropped or added in order to improve the goodness of the model.

All Subset Selection. If the number p of potential regressors is not too large (i.e. less than 40), an **exhaustive search**, which yields the best model of each size (from 1 to p) in terms of the criteria mentioned above can be performed. To keep the needed amount of computation reasonably low, a break and bound algorithm is used (Furnival & Wilson [8]), which is, for example implemented in the R-routine `leaps`. The algorithm uses a certain feature of the residual sum of squares: a bigger model has always a smaller sum of squares than a smaller submodel - this can be used to avoid evaluation of many potential models and thus speed up the search tremendously. Due to the fact that, in the case of linear regression and given a certain model size, the Cp-statistic, the *adjusted* R^2, the AIC and the BIC are all monotone functions of the residual sum of squares, the result of the algorithm doesn't depend on a specific selection criterion.

2.3.3 Model Diagnostics

As already mentioned, several assumptions are made, when applying linear regression. Only when all of them are fulfilled, the conclusions drawn from the model prove to be valid. If one or more of the assumptions are violated, the conclusions drawn can be biased or even misleading. Therefore, checking the validity of the underlying assumptions is a crucial step in statistical analysis.

The assumptions we have to check, are:

- Linear relationship between regressor- and response variables.
- The errors are normally distributed.
- The errors are independent of each other.
- The distributions of the errors are homoscedastic (homogeneity of variances).

In addition, **outliers** have to be detected and treated, and, depending on the experimental design, one might have to check for **multicollinearity**.

What actions can be taken in the case of detected violations to obtain data which suits linear regression? The data can be transformed. There exist many different transformations that can be applied. Very common is the usage of the so called Box-Cox-(power)-transformations (Box & Cox [2]). In the case of a response $y > 0$, this type of transformation has the following generic representation:

$$y^*(\lambda) = \begin{cases} \dfrac{y^\lambda - 1}{\lambda}, \text{ if } \lambda \neq 0, \\ \log(y), \text{ if } \lambda = 0, \end{cases}$$

with λ denoting the transformation parameter. The underlying assumption of this transformation is, that given a certain set of observed data, $(y_i, \boldsymbol{x_i})$, $i = 1, ..., n$, there exists a parameter λ, which can be estimated by using a certain technique, such that

$$y_i^*(\lambda) \overset{ind}{\sim} N(\boldsymbol{x_i}^T\boldsymbol{\beta}, \sigma^2).$$

Therefore, the Box-Cox-transformation can be used to transform for normality and linearity. But transforming the data doesn't always yield the result to be expected. In such cases, other types of regression should be considered, the generalized linear model for example or nonlinear regression.

There exist several means for checking the assumptions, the most common are **graphical methods** which are utilizing the residuals.

The residuals $r_i = y_i - \hat{y}_i$ are the estimates for the unobservable error terms ϵ_i and, in order to make good use of these quantities, they have to be standardized or studentized. For more details on these topics, see the corresponding sections on pages 19 to 20.

Linearity assumption. To check, whether there are unappreciated nonlinearities in the relationship between regressors and response, one can draw a scatter plot of standardized (or studentized) residuals versus fitted values. If the model assumptions are fulfilled, the residuals should be randomly distributed around zero, no particular pattern should be recognizable. If this plot shows something unexpected, e.g. a curved structure, the model should by all means be over-worked, because it would likely yield severe, misleading interpretations.

The assumption of normally distributed errors. To check this assumption, a so called **QQ-Plot** can be applied. In this plot, the quantiles of the empirical distribution of the residuals r_i are plotted versus the quantiles of the corresponding normal distribution. Let F denote the cumulative distribution function of the true underlying distribution and $X_1, \ldots, X_n$ a random sample of size n. It can be shown, that

$$X_{(i)} \approx F^{-1}\left(\frac{i}{n+1}\right), \text{ with } i = 1, \ldots, n \text{ and } X_{(i)} \text{ denoting the } i\text{-}th \text{ order statistic.}$$

Therefore, the $(i/(n+1))$-*th* quantiles of the normal distribution with the estimated parameters $\hat{\mu} = \bar{r}$ and $\hat{\sigma} = s$ are plotted versus the *i-th* order statistic of the residuals $r_{(i)}$.

If the resulting graph shows no major deviation from a straight line, then there is no evidence for a departure from the normality assumption.

In addition, there exist many different tests of normality amongst which the **Shapiro-Wilk test** is one of the most powerful.

Independence of errors. If violated, the errors are called **autocorrelated**. Autocorrelated errors yield autocorrelated residuals allowing for detection of this type of deviation. This is a problem predominantly associated with time series, i.e. in the context of regression analysis this means that the measurements used for the analysis are taken over a certain time period at several distinct points in time.

Autocorrelation is often caused by using a misspecified model, e.g. important regressors are not included in the model.

In regression analysis, autocorrelation has the consequence that the Gauß-Markov-Theorem is no longer valid. This means, that the estimators are no longer BLUE. This doesn't effect the coefficient estimates itself, but distorts the estimation of their standard errors and therefore falsifies the statistical analysis concerning their significance, the estimation of confidence intervals and, in the end, the whole conclusion drawn from the model.

Several tools are available to test for autocorrelation. There exist graphical methods as well as statistical tests. The most common test for first-order autocorrelation is the **Durbin-Watson** statistic. The **Breusch-Godfrey** test (Urban & Mayerl [29]) constitutes a more elaborated type of test, which allows in addition to test for the occurrence of autocorrelation of higher orders and can also cope with the presence of lags of the dependent variables. Graphical methods are, for example, plotting the residuals in the chronological order of their appearance. Such a plot shows distinct patterns in case of positive or negative autocorrelation.

Homogeneity of variances. Heteroscedasticity can be detected by means of graphical methods as well as statistical tests. One such graphical method is plotting the studentized residuals versus the fitted values. The existence of variance inhomogeneity is likely, if a distinct pattern is shown. For example, if a funnel shaped pattern is visible, one can conclude that, depending on the orientation of the "funnel", the variances of the errors grow or shrink with increasing magnitude of the fitted values. In addition, there exist several different statistical tests for heteroscedasticity, e.g. the Levene-test, the Fligner-test, the Breusch-Pagan-test, etc. A remedy, often applied to fix this problem is the above mentioned transformation of the data.

The issue of multicollinearity. The more the regressors are correlated, the larger the variance of the parameter estimates. Thus, it is desirable to have independent variables, which show as little correlation as possible in order to get reliable estimates. If the regressors are totally uncorrelated, the variances are as small as possible. This case is referred to as *orthogonal design*.
On the other hand, if we have highly correlated regressors, the estimates become very unreliable. A number, which quantifies the amount of linear association between a given regressor x_j and the rest of the regressors, is the so called variance inflation factor: $VIF_j = (1 - R_j^2)^2$, with R_j^2 denoting the coefficient of determination of a regression of x_j on all the other regressors. A $VIF_j > 10$ tells us, that there exists a serious multicollinearity problem.
To minimize this type of problem in the ongoing battery life experiment, the whole experiment is based on an *optimal experimental design*. The term "optimal" refers to the effort of keeping the contribution of the design to the variance of the parameter estimates as small as possible by using a computer-aided method of design construction. More on the topic of optimal designs can be found in subsection 2.3.5.

Outliers: detection and treatment

There doesn't exist a generally accepted definition of this notion, but it is commonly considered as an observation which shows an eye-catching distance to what one would expect in the context of the assumed model. Outliers can occur due to several reasons:

1. by chance
2. measurement errors
3. model assumptions do not hold

Outliers constitute a problem, which could cause serious damage to the validity of a regression analysis. They at least inflate the estimated value of the error variance, which will, concerning the estimated parameters, result in biased statistical tests and inflated confidence intervals. In addition, depending on their "position" (stronger influence in the border area of the range of the values than in its centre), they can exert an unduly large influence on the parameter estimation itself and therefore lead to a poorly parametrized model.

How can outliers be detected? One mean is *checking the studentized residuals.* If its value is unexpectedly large for observation i, then this observation might be an outlier. Due to the fact that, given the assumed model is correct, the distribution of the studentized residuals is known, statistical tests can be applied. This procedure is, due to several reasons, not working very well. A graphical representation of the studentized residuals constitutes a better means of outlier detection.

Because outliers often show a disproportionate influence on the outcome of a regression analysis, the estimation of this kind of *influence of a data point* is another widely used tool for assessing whether an observation should be considered as an outlier (but: not all outliers are highly influential points and vice versa).
Which types of measures are used to quantify this influence? Two very common characteristics are given as follows:

The *leverage score* of the influence of observation i is defined as the $i-$th diagonal element h_{ii} of the hat-matrix H. This value falls between $1/n$ and 1. The closer to 1, the more influential the observation can be. If $h_{ii} > 2p/n$, one should have a closer look at the corresponding observation.
The *Cook-distance* D_i is defined as:

$$D_i = \frac{(\hat{\boldsymbol{y}}_{(i)} - \hat{\boldsymbol{y}})^T(\hat{\boldsymbol{y}}_{(i)} - \hat{\boldsymbol{y}})}{p \cdot \hat{\sigma}^2} \quad \text{for} \quad i = 1, \ldots, n.$$

The subscript (i) denotes, that the respective quantity is determined using all but the $i-$th observation, whereas in the case of $\hat{\boldsymbol{y}}$ all observations are used. D_i quantifies the effect of deleting the observation by measuring the standardized distance between these two quantities, which is in case of $D_i > 1$ considered critical (thresholds different to 1 have also been suggested). There exist also other common measures of influence, namely *DFFITS* and *DFBETA* - more on this topic can be found, for example in Montgomery et al. [21].

The *treatment* of outliers embraces several different strategies. One is, if the outlier is likely due to an erroneous measurement or a faulty experiment, to exclude the observation from regression analysis. This strategy should be chosen with caution - all excluded observations should at least be mentioned in the documentation of the experiment and the reason for their exclusion should be pointed out comprehensively. Another way of treatment is the method of robust regression, e.g. quantile regression. This methods are much less sensitive to outliers, but show other disadvantages like reduced efficiency.

2.3.4 Experimental Design

In the battery aging experiment, as well as very frequently in scientific work in general, the goal is to investigate and mathematically model some cause and effect relationship between quantities of specific interest. Due to, at least in most cases, limited resources, the experimenter has to find a trade-off between validity of the results and costs of the experiments needed. Therefore, if planning of the experiments is possible, the topic of design of experiments (**DOE**) constitutes a big issue.

The first, and most crucial step is, to precisely formulate the questions, which the experimenter wants to investigate. In the case of planned experiments, statistics provides sophisticated methods for setting up the experiments such that not only the gathered data is suitable to work on answers to the questions under investigation, but also to allocate the given resources best possible.

That means, the design should depend on the question(s) stated and fulfill all, or at least the first three of the following requirements:

- The data produced has to be suitable to do the statistical tests needed to investigate the hypotheses which have been stated previously or come up in the course of the statistical analysis of the data.
- The number of experiments should be the minimum number needed in order to investigate the stated problem accurately.
- The precision of the results of the statistical analysis should not fall below a predefined level.
- The data produced has to be suitable to investigate the type of relationship between in- and output or parametrize an already given model.

As already mentioned, due to the stochastical nature of the data generating process, which underlies many experimental settings, the data generated in those cases show considerable variation whose magnitude depends on the error associated with the underlying process. This variation influences the process of parameter estimation in a way, that the estimators also show a distribution, which now not only depends on the inherent error of the underlying process but also on the design of the experiment. According to Theorem 2.1, the parameter estimators are distributed as follows: $\hat{\boldsymbol{\beta}} \sim N_{k+1}(\boldsymbol{\beta}, \sigma^2(X^TX)^{-1})$. Obviously, the errors of the estimators' distributions depend not only on σ^2 but also on $(X^TX)^{-1}$ and thus on the design matrix X (see also the subsection on multicollinearity). One of the goals of DOE is to keep the variability of the estimators as low as possible which in turn maximizes the reliability of the statistical analysis.

Special desirable design properties are:

- **Orthogonality.** In the case of an orthogonal design, $X^T X$ is a diagonal matrix. In such a design, the regressor variables are linearly independent of each other and the variance of the estimators is minimized. The opposite constitute designs with highly correlated regressor variables (multicollinearity). Correlation of the regressors can inflate the estimators variance to an extent, that reliable statistical analysis can become impossible. In this case, regressors have to be removed from analysis in order to sufficiently reduce the correlation of the regressors.
- **Rotatability.** If a design has this property, the variance of a predicted response $\hat{y}(x)$ to a given point x (in the design space) does only depend on the points distance from the center of the design and not on its direction. Therefore, the prediction variance is constant on spheres. This is a desirable property for response surface designs for quadratic models.

Due to the fact, that very often only little is known about the influential factors and their impact on the output, a frequently used approach is to start with a screening phase to determine the important influences by means of a (fractional) factorial design. After that screening procedure has been employed, one sets up an experimental design, which is capable to yield data suitable to parametrize a more sophisticated model. Which designs are considered suitable, depends on two factors - the experimental region, which has to be explored and the model which should be parametrized. There exist a lot of classical designs, which, in many cases are the method of choice. Examples are:

- **Full Factorial Designs.** This family of designs possibly constitutes the most important one. The term 'full' indicates that those designs consist of all possible factor level combinations. These designs are orthogonal and allow to identify main effects and interactions. A drawback is, that these designs quickly become outsized and therefore impractical due to growing numbers of factors and/or number of factor levels. This is why the most commonly used designs of this type are the 2^k and the 3^k designs, with k denoting the number of different factors, 2 or 3 indicating the number of levels considered for each factor and 2^k or 3^k denoting the needed number of experimental runs respectively. In the case of a 2^k design all k main effects and all possible interactions can be estimated, but it offers no possibility to fit models in which the factors show up with an order greater than one. Quadratic models can be fit, using a 3^k design and in general, $l + 1$ factor levels of a factor A are necessary in order to estimate the parameter for the model term A^l.
- **Fractional Factorial Designs.** To overcome the problem of the exploding number of runs, fractions of full factorial designs can be used. The drawback in this case is, that these designs no longer allow to estimate all effects separately, i.e. the effects are, to some extend confounded (aliased). The amount of aliasing is determined by how much the full factorial design is reduced. This design type is very frequently used for screening experiments.

- **Central Composite Design.** Extending the 2^k design by including center points, allows to test for curvature and constitutes an economical alternative to 3^k designs. A CCD consists of a full or fractional 2^k and in addition some center runs (all factors are set to their middle (or center) levels) and $2k$ axial runs for each of the factors (all but one of the factors are set to their middle level and the remaining one is set to some especially chosen levels, called α and $-\alpha$). These design parts can, as well, be applied sequentially. The factorial can be used to estimate linear and interaction effects, the center points to test for curvature (which in addition allow for an independent error estimation) and the axial points to estimate quadratic effects.

Coded design variables. When constructing designs, classical as well as optimal, usually the range of the factors used, is transformed to a certain interval, in most cases to the interval $[-1, 1]$, i.e. the coded factors are therefore dimensionless. In this context one speaks of using coded design variables. The factors, of course, exist originally in their natural coding, i.e. they are given with respect to their proper unit of measurement. Let the interval $[a, b]$ be the range of a factor F in natural coding and x be an arbitrary number in $[a, b]$. Now, the transformation is done as follows:

$$x_c = \frac{x - (a+b)/2}{(b-a)/2}.$$

Factors are transformed to get orthogonal designs, which allow to estimate the effects independently of each other and to make the effects of different factors comparable, i.e. the estimated effects do not longer depend on the magnitudes of the factors in natural units which allows to determine the relative size (or importance) of factor effects.

The *classical designs* have many advantageous features, but there exist also some *drawbacks*. They cannot be applied on irregular shaped experimental regions, i.e. each such design needs a specifically shaped design space (e.g. a hypercube in the case of the factorial designs). Secondly, there do not exist classical designs for any model one intends to parametrize. In such cases, the use of optimal designs, which can be constructed by means of computer algorithms constitutes the remedy of choice.

2.3.5 Optimal Designs

As already mentioned, classical designs are not always appropriate. The reasons for applying an optimal design instead are manifold, some of the main motivations are:

- There exists no classical design to parametrize the chosen model.
- The *experimental region* $\mathcal{T}$ (also known as *design region* or *design space* - see page 13 for more details) is not suitable for classical designs, i.e. it is irregular shaped due to constraints concerning forbidden factor combinations.

- The number of experiments, which can be performed is, due to economical restrictions lower than what is needed for conducting the corresponding classical design.
- One wants to incorporate experiments, which have been carried out preliminarily.
- One wants to adapt the design during experimentation, to incorporate growing knowledge about the system under investigation.

At this point a short review concerning experimental designs is given to understand the idea of the optimality of a design.

In creating optimal designs, one strives for finding minimum variance unbiased linear estimators (the 'optimal' estimators). In the context of the discussion of the classical linear model given so far (see (2.1)), the emphasis lies on the variance minimality of the parameter estimators. That means, that the question of how to choose the regression vectors $\boldsymbol{x}_i$, $i = 1, \ldots, n$, and the design matrix $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^T$ respectively, will now be in the focus of our interest, since the covariance matrix of $\hat{\boldsymbol{\beta}}$ is given by $\sigma^2(X^TX)^{-1}$.

Now, the definition of an experimental design in the context of design optimization will be given (Pukelsheim [25]):

Definition 2.4. An **experimental design for sample size n**, ξ_n, is given by a finite number of $l \leqslant n$ distinct regression vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_l$ in the regression range $\mathcal{X}$, and non-zero integers $n_1, \ldots, n_l$ such that $\sum_{i \leqslant l} n_i = n$. The set, which consists of all possible designs of size n is called Ξ_n.

The values n_i give the number of times the according vectors $\boldsymbol{x}_i$ show up in the full design with design matrix $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$, i.e. the design ξ_n tells the experimenter to conduct n_i single experiments, applying conditions according to the regression vector $\boldsymbol{x}_i$.

Definition 2.5. The l distinct regression vectors $\boldsymbol{x}_i$, which show up in the design ξ_n constitute what is called the **support** of ξ_n, *supp* $\xi_n = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_l\}$.

Definition 2.6. The vectors that appear in *supp* ξ_n are used to compute the so called **moment matrix** of the design, $M(\xi_n) = \sum_{i \leqslant l} (n_i/n) \boldsymbol{x}_i \boldsymbol{x}_i^T = X^TX/n$.

As will soon become clear, $M(\xi_n)$ plays *the* central role in the process of design optimization. The problem with the optimization of experimental designs in the case of finite sample size n is, that one often ends up with discrete optimization problems, which can hardly be solved. Therefore, the theoretical concept of so called *experimental designs for infinite sample size*, ξ, emerged. As can be shown for any design ξ with an infinite support, constructed to parametrize a model with $k+1$ parameters, there exists a finite design $\tilde{\xi}$ with $l \leqslant (k+2)(k+1)/2 + 1$ distinct support points, such that $M(\tilde{\xi}) = M(\xi)$ (Melas [18]). Therefore, dealing with approximate designs with a finite support only is

sufficient.
The idea is, to consider the design ξ as a distribution on $\mathcal{X}$ assigning its whole mass to just l points. In the case of a finite design, the l rational weights $w_i = n_i/n$, $i = 1, \ldots, l$, are placed on the $\boldsymbol{x}_i \in supp\ \xi_n$ and in the case of an infinite design, the l weights assigned to the elements of the support are allowed to vary freely in the closed interval $[0, 1]$ under the condition, that $\sum_{i \leqslant l} \xi(\boldsymbol{x}_i) = \sum_{i \leqslant l} w_i = 1$.

Such a generalized design ξ is therefore determined by $l \geqslant 1$ vectors $\boldsymbol{x}_i$ carrying nonzero weights w_i:

$$\xi = \begin{pmatrix} \boldsymbol{x}_1 & \ldots & \boldsymbol{x}_l \\ w_1 & \ldots & w_l \end{pmatrix}.$$

This l vectors are called the support points of ξ, and thus again constituting the support of ξ, $supp\ \xi$. The set of all such designs is denoted by Ξ.

The moment matrix of such a general design $\xi \in \Xi$ is therefore consequently defined to be the $(k+1) \times (k+1)$ matrix given by

$$M(\xi) = \sum_{\boldsymbol{x} \in supp\ \xi} \xi(\boldsymbol{x}) \boldsymbol{x}\boldsymbol{x}^T = \int_{\mathcal{X}} \boldsymbol{x}\boldsymbol{x}^T d\xi.$$

Finally, to make a design for infinite sample size realizable, it has to be approximated by a design for finite sample size. Such apportionment methods, can for example be found in Pukelsheim [25].

How are optimal designs constructed? As already mentioned, the task is to minimize the variability of the parameter estimators. It can be shown, that the joint confidence region for the parameters constitute a (hyper-) ellipsoid, i.e. the whole information about the 'uncertainty' regarding the parameter estimation is contained in this ellipsoid and can be represented by its volume, the area of its surface or the lengths of its axes. Thus simultaneous minimization of the estimators' variances is equal to minimizing this ellipsoid. It can further be shown, that shape and volume of the confidence ellipsoid are determined by $Cov(\hat{\boldsymbol{\beta}}) = \sigma^2 n^{-1} (M(\xi))^{-1}$. Therefore different optimization methods, which make use of various characteristics of the confidence ellipsoid, are present.

There exist many different optimality criteria, which asses the variance and covariance of the estimators in their own, unique way. The most common criteria are the so called alphabetic optimality criteria, amongst which the D-, A-, G- and E-optimality criterion are the most widely used.

In the course of this work, two optimality criteria (which are the most commonly used criteria) have been applied.

D-Criterion. Using this criterion, one tries to minimize the product of the squared lengths of the ellipsoids axes which is proportional to the volume of the confidence ellipsoid. The smaller this volume is, the smaller is the variance of the predictors, the correlation between them and the variance of the predicted response.

Minimizing the volume is equivalent to maximizing the determinant of the moment matrix or, equivalently, minimizing the determinant of the variance-covariance matrix $Cov(\hat{\boldsymbol{\beta}})$. This is why this criterion is named *D- or Determinant-Criterion.* As already stated, $Cov(\hat{\boldsymbol{\beta}})$ equals $\sigma^2 n^{-1}(M(\xi))^{-1}$. Therefore, if σ^2 and n are fixed, it remains to deal with the determinant of $M(\xi)$, which, in the case of a finite design is proportional to $X^T X$.

In the theory of infinite designs, the D-optimal design ξ^* is defined as follows:

$$\xi^* = \underset{\xi \in \Xi}{argmax}(|M(\xi)|).$$

A drawback of this approach is, that it can happen, that, due to minimization of the ellipsoids volume, one axis is much shorter than the others, which in turn means, that it might be the case, that just one parameter is estimated efficiently, while the others are not.

G-Criterion. Using this criterion, one tries to minimize the maximum scaled prediction variance over the whole experimental region - that's why this criterion is called G- or Global-Criterion. Thus a G-optimal design $\xi_G^* \in \Xi$ is one, that satisfies

$$\sup_{\boldsymbol{x} \in \mathcal{X}} \left(\frac{n \cdot var(\hat{y}(\xi_G^*, \boldsymbol{x}))}{\sigma^2} \right) = \inf_{\xi \in \Xi} \sup_{\boldsymbol{x} \in \mathcal{X}} \boldsymbol{x}^T M(\xi)^{-1} \boldsymbol{x}$$

with $var(\hat{y}(\xi, \boldsymbol{x}))$ denoting the variance of a predicted mean value with respect to a certain experimental design ξ and a certain regression vector $\boldsymbol{x}$.

A crucial advantage of both, the *D-* and the *G-Criterion* is, that they are independent of linear transformations of the factor ranges, i.e. that they do not depend on units of scale.

A well-known result developed by Kiefer and Wolfowitz connects the D- and G-Criterion (Melas [18]):

Theorem 2.2. ***Kiefer-Wolfowitz Equivalence Theorem***

Given that the assumptions of the classical linear model with normally distributed homoscedastic errors hold, the following conditions are equivalent in the class of approximate designs:

1. ξ^* *is a D-optimal design.*

2. *ξ^* is a G-optimal design.*
3. $\sup_{\boldsymbol{x}\in\mathcal{X}} \left(\boldsymbol{x}^T M(\xi^*)^{-1}\boldsymbol{x}/n\right) = k+1$

Therefore, a design ξ^* is D-optimal if and only if point 3 of theorem 2.2 is fulfilled - this constitutes an important necessary and sufficient criterion for D-optimality.

Two finite designs ξ_1 and ξ_2 can be compared with each other by computing a quantity named *relative-efficiency*, which compares their 'goodness' with respect to the D-criterion. This quantity is computed as

$$RE = \left(\frac{|X_1^T X_1|}{|X_2^T X_2|}\right)^{1/p},$$

with X_1 and X_2 denoting the respective design matrix and p the number of model parameters. Taking the *p-th* root gives, in the case of uncorrelated estimators, the average *D-efficiencies* of the p parameter estimators and enhances therefore interpretability.

A resulting RE value of d indicates, that ξ_1 has to be replicated about $1/d$ times to yield the same efficiency as ξ_2. But, comparing two designs by this measure is of very limited benefit. The goal is to find a design which is as efficient as possible, and therefore, only comparing the design to that which proves to be the most efficient design under all designs possible is of interest in the end. Most efficient in this context means most efficient in terms of D-optimality and therefore, this most efficient design is called *D-optimal*. It is denoted by ξ^* and the quantity computed is the so called *D-efficiency*:

$$D_{eff} = \left(\frac{|X^T X|}{|nM(\xi^*)|}\right)^{1/p},$$

with X denoting the design matrix of the design ξ, whose *D-efficiency* is computed and p the number of model parameters. Again, we can interpret $(1/D_{eff} - 1)$ as the amount of additional observations needed in order to make ξ as efficient as ξ^*.

This works fine in theory, but in practice, one of course encounters usually the problem of not knowing the optimal design.

Compared to the *D-efficiency*, the *G-efficiency* can, due to the Kiefer-Wolfowitz Equivalence Theorem, be calculated more easily. Since we know, that, in case of the G-optimal design, $\sup_{\boldsymbol{x}\in\mathcal{X}} \left(\boldsymbol{x}^T M(\xi^*)^{-1}\boldsymbol{x}/n\right) = k+1$ the *G-efficiency* of a design ξ is given as

$$G_{eff} = \frac{k+1}{\sup_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{x}^T M(\xi)^{-1}\boldsymbol{x}/n}.$$

In case of the *D-efficiency*, only lower bounds can be given, one of these bounds is a result of the *Kiefer-inequality* (see Pukelsheim [25]):

$$D_{eff} \geqslant e^{1-1/G_{eff}}.$$

Algorithms for D-optimal Design-Construction

As already explained, a design for infinite sample size can usually not be used due to the fact, that the weights can be irrational numbers. A second problem one encounters when trying to find the D-optimal design is, that in order to do this one would have to make a search over the whole experimental region, which would result in a computationally intractable task.

Therefore, in order to obtain a specific design for conducting an experiment, several algorithms exist, which yield more or less good approximations to the D-optimal design. In most cases these algorithms use a so called **point exchange method**. These algorithms need a predefined set of points which are termed candidate points, i.e. a list of points which are selected from the experimental region by constructing a grid over the experimental region, with mesh size depending on the chosen levels of the factors under consideration and avoiding all factor combinations which are not allowed due to the constraints under consideration.

Based on a randomly chosen, non-singular start-design these algorithms exchange one or more design points for the same number of candidate points in each step, in order to achieve an improvement in the chosen optimality criterion. The algorithm stops, when substantial improvement is no longer possible or a predefined number of iterations is reached. Unfortunately, there is no guarantee that the algorithm did not just find a local optimum and therefore it is highly recommended to repeat the search several times.

The main disadvantage is, that no other design than one that is covered by the candidate points can be found this way, and therefore, the result can almost never be the optimal design, but a more or less good approximation.

Examples are:

- Fedorov's Exchange Algorithm (Fedorov [7])
- k-Exchange Algorithm (Johnson and Nachtsheim [13])
- Mitchell's DETMAX algorithm (Mitchell [20])

In addition, other types of algorithms exist, which are based on different construction principles, for example:

- *Coordinate exchange methods.* They do not require a list of candidate points and modify the design points by doing a coordinate-wise optimization for each design point (Meyer and Nachtsheim [19]).
- *Genetic algorithms.* A type of iterative heuristic search based on the model of biologic evolution - for more information on this topic see for example Heredia-Langner et al. [12].

Due to the fact, that the experimental design used in the ongoing experiment has been constructed by applying the R-routine `optFederov` (Wheeler [36], who named his routine (by mistake?) Federov instead of Fedorov), which makes use of Fedorov's Exchange Algorithm, this algorithm is now explained in more detail:

Fedorov's Exchange Algorithm

It exchanges points by simultaneously adding and deleting points (a so called rank-2 exchange algorithm). At each step, it exchanges a point of the design matrix with a point of the candidate set in order to increase the determinant of the resulting information matrix. During this process, the size of the design is held constant according to the desired number of experiments (in contrast to, for example the DETMAX-algorithm, which temporarily allows for differing numbers of experiments during the construction process).

If design point $\boldsymbol{x}_i$, being part of the 'old' design matrix $\tilde{X}$ is exchanged with a candidate point $\boldsymbol{x}_j$, thus transforming $\tilde{X}$ into the 'new' design matrix $\hat{X}$, the following relations prove to be valid:

$$\hat{X}^T\hat{X} = \tilde{X}^T\tilde{X} + \boldsymbol{x}_j \cdot \boldsymbol{x}_j^T - \boldsymbol{x}_i \cdot \boldsymbol{x}_i^T$$

and

$$|\hat{X}^T\hat{X}| = |\tilde{X}^T\tilde{X}|(1 + \Delta(\boldsymbol{x}_i, \boldsymbol{x}_j))$$

with

$$\Delta(\boldsymbol{x}_i, \boldsymbol{x}_j) = d(\boldsymbol{x}_j) - d(\boldsymbol{x}_i) - d(\boldsymbol{x}_i)d(\boldsymbol{x}_j) + d(\boldsymbol{x}_i, \boldsymbol{x}_j)^2$$

and

$$d(\boldsymbol{x}_k) = \boldsymbol{x}_k^T(\tilde{X}^T\tilde{X})^{-1}\boldsymbol{x}_k \text{ for } k = i, j$$
$$d(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i^T(\tilde{X}^T\tilde{X})^{-1}\boldsymbol{x}_j$$

Basically, Fedorov's algorithm works as follows: In each step, Δ-values for all possible $(\boldsymbol{x}_i, \boldsymbol{x}_j)$-combinations are calculated, with $\boldsymbol{x}_i$ denoting the points in the design matrix and $\boldsymbol{x}_j$ denoting all points in the list of candidate points. The combination with the highest value is selected for exchange. As long as such combinations lead to positive Δ-values, the algorithm can improve the design - therefore it should not terminate until such successful combinations do not appear any more. What seems to be a good approach in theory, is often not feasible in practice, because reaching this point might be too expensive according computation time. Therefore, a threshold is chosen, and the algorithm stops when all Δ-values become smaller than the value of the threshold.

Dynamic Design of Experiment (DDoE)

The algorithm sets up a D-optimal experimental design given the number of design points, n, and the structure of the model. This means, that **the constructed design is suitable only for the given design space, a specific model and a certain number of design points**.

Growing information about the relationship between battery-usage and cell ageing during the ongoing experiment makes the revision of the proposed model necessary before the end of the experiment is reached. As mentioned before, a new model needs a new design for achieving the data quality needed for proper parametrization.

The optFederov-Routine enables us to add additional design points to the original design in order to restore the desired design feature of minimized estimator variance for parametrization of the new model.

This stepwise process of

- setting up a model,
- designing a proper experiment,
- model revision in the light of new information due to the ongoing experiment
- and designing an extended experiment with respect to the new model

is called DDoE (Haselgruber [11]).

Further design revisions can be conducted unless one ends up with a reasonable and well parametrized model.

The number of distinct design points needed

In order to be able to estimate p different parameters, the number of distinct design points, l, must at least be equal to p. In many cases p is also the optimal number of distinct design points. Generally, the optimal number lies in the interval $[p\ ,\ p(p+1)/2]$. If l is outside of this interval, the efficiency of the design can drop severely.

3 Statistical analysis

3.1 Evaluation of the design in use

In the preceding chapters much has been said about the concrete situation for which a proper experimental design is needed and the theory behind design construction was discussed.

In chapter 1 the chemical and physical background of the battery-aging experiment, the question which factors to investigate and the special requirements concerning factor ranges and the problem of constraints imposed on the design region is addressed. In chapter 2 the theory of design construction is discussed in more detail. Especially the application of optimal designs in the light of economical restrictions and irregular shaped experimental regions due to forbidden factor-level-combinations is considered.

The ongoing experiment is based on the initial design (table 3.1) constructed by G. Pregartner and W. Prochaska who have put lots of effort in its creation. It's Deffbound- or Dea-value of about 0.701 (a lower bound on D-efficiency - see the corresponding sections of chapter 2 and 3 for details) made it a reasonable candidate for acting as a starting point to commence the experiment. Unfortunately, the implementation of this design was not possible due to several physical restrictions concerning the technical equipment. That was the beginning of several changes concerning the design. This fact as well as the evolving linear model under investigation in combination with the omitted adaptation of the design (see the comment on DDoE, page 47), makes the lower bound for the designs D-optimality, computed for the initial design, no longer valid.

In order to again establish such a lower bound, the experiment in its actual state is thoroughly evaluated. Finally, suggestions what actions could be taken for maintaining the quality of the experimental design in the course of its execution are given. Calculation of the D-efficiency of an existing design can be done by using the R-routine `eval.design`. In addition, in order to repair (i.e raising its D-efficiency) an existing design the augmentation feature provided by the R-routine `optFederov` can be used. Both routines are part of the R-package `AlgDesign` (see [36] for details). In order to quantify the D-efficiency of and/or repair an existing design two major ingredients are inevitable: a proper candidate set and a model which which has to be parametrized. If one of those ingredients cannot be provided no meaningful statement about the efficiency of the experiment can be given.

3.1.1 Comparing the initial design and the actual state of the experiment

In contrast to the number of factor-levels given in the corresponding internal report, Prochaska & Pregartner used different numbers of factor-levels in the actual process of design construction:

Factor	Factor-Levels									# of Factor-Levels
T	-10	5	20	40						4
CC	0.2	0.8	2.4							3
ADC	0.0	0.2	1.0	2.0	3.0	4.0	6.0	8.0		8
PDC	0.2	1.0	3.0	4.0	6.0	8.0	10.0	12.0	14.0	9
F	0.000333	0.03	0.06	0.1	0.2	0.5				6
SoC	15	25	55	80	95					5
dSoC	0.01	2.50	15	50	80					5

For this 129600 different factor-level-combinations exist. Applying the constraints mentioned in more detail in chapter 1 (see table 3.2 for an explicit list) reduces the number of possible factor-level-combinations by a factor of more than 30. In total, there are exactly 4233 feasible factor-level-combinations left (the candidate points), from which the algorithm for calculating the D-optimal design can choose the appropriate design points.

The process of design construction based on the regression model `lm.ex` (see chapter 2, page 40 ff. for more details) and the set of 4233 candidate points mentioned above, aims to yield a design with a D-efficiency that is, under the given conditions, as high as possible. One of the best designs found this way is `des_701` (table 3.1), which has a lower D-efficiency boundary of 0.701. The design reported in the respective internal report, termed there as final-design-of-experiment (in the following termed `des_WP1300` - see table 3.2), which constitutes a major outcome of WP1300, shows a few deviations from `des_701` (table 3.3). Evaluation of `des_WP1300` shows, that the constraints in the case of five design-points are no longer fulfilled.

Does it make sense to evaluate the D-efficiency of a design, which is no longer within the allowed test space with respect to that test space? Since evaluation of the efficiency of a design with respect to an enlarged test space (and therefore enlarged set of candidate points) yields a bound which is always lower or equal to the bound which is found doing the respective calculations with respect to the smaller initial set of candidate points (which is a subset of the larger set). The lower bound calculated using the initial set can be seen as an upper bound for the bound which would be calculated using the larger set, i.e. designs which do not fit into the initial design space have a lower bound for D-efficiency which is at best as large as the bound calculated using the initial set. Evaluating `des_WP1300` in this way yields a significantly lower D-efficiency boundary of

only 0.116 compared to the initial design `des_701` having a lower bound of 0.701. This tells us, that a designs D-efficiency can be severely effected by just a few modifications.

Therefore, if changes of the design are inevitable (due to technical and/or economical reasons) an accompanying statistical monitoring is highly recommended - this is the only way to keep the benefit of the experiment as high as possible. Modifying the experiment without statistical supervision can cause poor data quality and hence dissatisfactory experimental results.

The design `des_m` (see table 3.4), which is the design actually in use, shows many deviations from `des_701`. Not only several load points have changed with time, but also the factor ranges. The lowest levels of CC and PDC are now set to zero (initially the lowest possible value of this factors was 0.2) and in the case of F the lowest level dropped from 0.000333 to 0.0003. If this design is coded with respect to the initial factor ranges, coded factor values smaller than -1 emerge. This will likely raise the chance of having load points in the design which violate one or more of the constraints - in fact 21 of the load points violate the constraints. Incorporating the new factor ranges in coding the design is obviously a remedy for the problem of coded values not within the interval $[-1, 1]$, but still many of the design points (17 in total) are violating constraints.

Evaluation of this design, coded and analyzed with respect to the initial factor ranges, the initial set of candidate points and the initial model `lm.ex` yields a lower bound for D-efficiency smaller than 10^{-1}. Due to the fact, that in the course of the experiment, it soon turned out that model `lm.ex` is not the model of choice and, instead a model of much lower complexity, namely model `lm6_bal` is emerging (see page 83 for details), the same design evaluation is done with respect to model `lm6_bal`. This yields a much better, but still unsatisfactory Dea-value of 0.337.

3.1.2 Remedies and Suggestions

This rather inferior result of course raises the question of how to deal with such a situation. The only remedy is repairing the design. Unfortunately, this is not always possible due to economical and/or time restrictions - in this situation, there is no alternative to sticking to the experimental outcome, hoping that the data gained this way still will do the job.

If repairing the design is a reasonable possibility, two different ways can be chosen. First, all load points violating constraints can be removed from the design, but can subsequently be used as additional points for model validation. Now, the remaining part of the design can be augmented such that again a sufficient lower D-efficiency bound for the experimental design is reached.

A second, and maybe more economical way, would be, to revise the constraints and thus the set of candidate points such that the new factor ranges are incorporated and the existing design no longer violates the constraints. Building on that, one can repair the

design by augmenting it with, hopefully, not as many additional runs as will be needed using the first method.

The procedure of design-repair will now be illustrated for the first method suggested above. After removing all load points which are violating the constraints, 21 load points remain (see table 3.5), which can be augmented until we have again a reasonable design.

LP	T	CC	ADC	PDC	F	SoC	dSoC
L01	-1	-1	-0.5	0.71	0.23	-0.75	-0.94
L02	1	1	-0.75	0.71	0.56	-0.75	-0.94
L03	-1	-1	-0.5	0.71	0.75	-0.75	-0.94
L07	1	1	-0.75	-0.59	0.23	0	-0.94
L09	-1	-1	-0.75	0.42	0.75	0.63	-0.94
L13	-1	-1	-0.75	-0.59	0.23	-0.75	-0.63
L14	1	1	-0.5	-0.59	0.23	-0.75	-0.63
L15	-1	-1	-0.95	-0.16	0.23	0	-0.63
L16	1	1	-0.95	0.71	0.23	-0.75	-0.63
L18	1	1	-0.5	-0.59	0.23	0.63	-0.63
L19	-0.82	-0.82	-0.95	-0.16	0.23	0.38	-0.63
L20	1	1	-0.95	0.71	0.23	0.63	-0.63
L22	-1	-1	-0.95	-1	-1	0	0.25
L23	-1	-1	-0.95	-1	-1	0	0.25
L24	-1	-1	-0.5	0.71	0.75	0	0.25
L25	-1	-1	-0.95	-1	-1	0	0.25
L26	-1	-1	-0.95	-1	-1	0	1
L29	-1	-1	-0.95	0.71	0.23	0.13	0.75
L31	-1	-1	-0.75	0.71	0.56	0	1
L32	-0.09	-0.09	-0.75	0.71	0.56	0	0.25
L35	-1	-1	-0.5	-0.59	1	0	1

Table 3.5: 21 `des_m` - runs not in conflict with the constraints

In order to find out how many runs we have to add, we construct designs with sizes ranging from 24 up to 66 runs. This is done with respect to the initial set of candidate points and model `lm6_bal`. It can be observed, that the D-efficiency is increasing (despite some fluctuations) up to the size of 41 runs, where the D-efficiency has its maximum and then drops continuously (figure 3.1).

This means, that the best repaired design has the same number of runs as the initial design. The new design (see table 3.6) consists of the 21 remaining runs taken from `des_m` and 20 new runs, which have to be started time-delayed and thus prolonging the duration of the whole experiment. The lower D-efficiency boundary of the repaired design is 0.721 and thus reasonably large.

Finally, due to limited resources, none of these two possibilities has been realized. Instead the free capacities have been used for replicating some of the existing experimental

conditions and the execution of validation experiments. Therefore, the experiment is finally brought to an end based on `des_m`, which is, for the purpose of statistical analysis coded with respect to the new extended factor ranges (table 3.7).

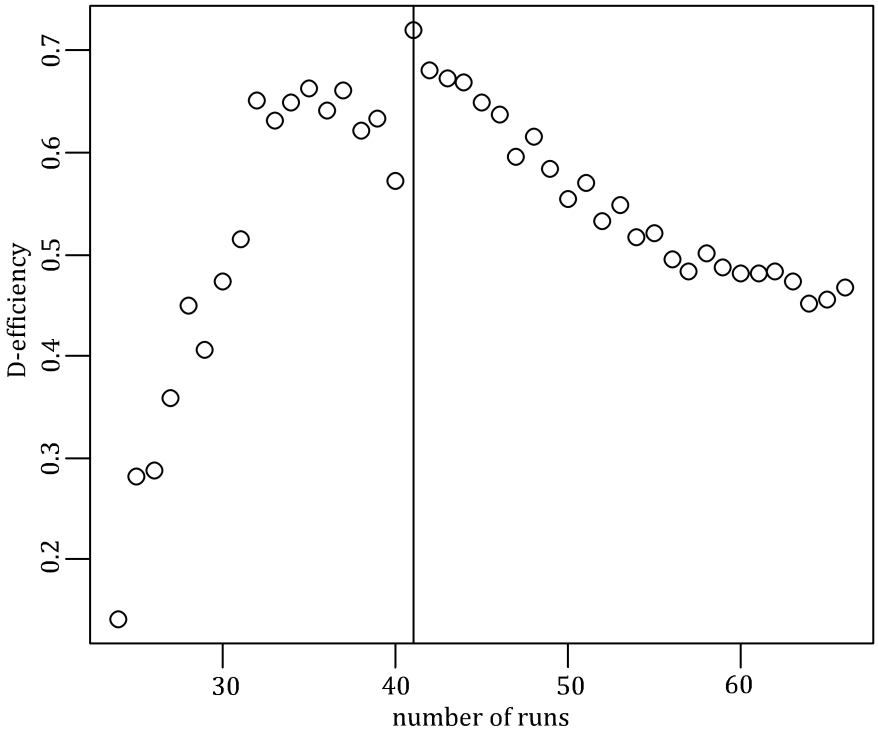


Figure 3.1: D-efficiency as a function of design size

LP	T	CC	ADC	PDC	F	SoC	dSoC
C01	-10	0.2	0	0.2	0.000333	15	0.01
C02	40	0.2	0	0.2	0.000333	15	0.01
C03	5	0.2	0	0.2	0.000333	55	0.01
C04	40	0.2	0	0.2	0.000333	55	0.01
C05	-10	0.2	0	0.2	0.000333	95	0.01
C06	40	0.2	0	0.2	0.000333	95	0.01
L01	20	0.2	2	12	0.03	25	2.5
L02	5	2.4	1	12	0.1	25	2.5
L03	40	0.2	2	12	0.2	25	2.5
L04	-10	0.2	0.2	0.2	0.5	25	2.5
L05	40	0.2	0.2	0.2	0.5	25	2.5
L06	20	2.4	0.2	0.2	0.5	25	2.5
L07	40	2.4	1	3	0.03	55	2.5
L08	40	0.2	1	12	0.03	80	2.5
L09	5	0.2	1	10	0.2	80	2.5
L10	20	2.4	2	12	0.2	80	2.5
L11	40	0.2	0.2	0.2	0.5	80	2.5
L12	-10	2.4	0.2	0.2	0.5	80	2.5
L13	20	0.2	1	3	0.03	25	15
L14	40	2.4	2	3	0.03	25	15
L15	-10	0.2	0.2	12	0.03	25	15
L16	40	2.4	0.2	12	0.03	25	15
L17	5	2.4	0.2	0.2	0.5	25	15
L18	40	2.4	2	3	0.03	80	15
L19	-10	0.8	0.2	12	0.03	80	15
L20	40	2.4	0.2	12	0.03	80	15
L21	5	0.2	0.2	0.2	0.5	80	15
L22	20	0.2	0.2	0.2	0.000333	55	50
L23	-10	2.4	0.2	0.2	0.1	55	50
L24	40	0.2	2	12	0.2	55	50
L25	-10	0.2	0.2	0.2	0.000333	55	80
L26	40	0.2	0.2	0.2	0.000333	55	80
L27	40	0.2	0.2	6	0.000333	55	80
L28	40	2.4	0.2	0.2	0.03	55	80
L29	5	0.2	0.2	12	0.03	55	80
L30	40	2.4	2	12	0.03	55	80
L31	20	0.2	1	12	0.1	55	80
L32	5	2.4	1	12	0.1	55	80
L33	-10	0.2	0.2	0.2	0.5	55	80
L34	20	2.4	0.2	0.2	0.5	55	80
L35	40	0.2	2	3	0.5	55	80

Table 3.1: Initial design (`des_701`) with 41 load points

LP	T	CC	ADC	PDC	F	SoC	dSoC
C01	-10	0.2	0	0.2	0.000333	15	0.01
C02	40	0.2	0	0.2	0.000333	15	0.01
C03	5	0.2	0	0.2	0.000333	55	0.01
C04	20	0.2	0	0.2	0.000333	55	0.01
C05	-10	0.2	0	0.2	0.000333	95	0.01
C06	40	0.2	0	0.2	0.000333	95	0.01
L01	20	0.2	2	12	0.03	25	2.5
L02	5	2.4	1	12	0.1	25	2.5
L03	40	0.2	2	12	0.2	25	2.5
L04	-10	0.2	0.2	0.2	0.5	25	2.5
L05	40	0.2	0.2	0.2	0.5	25	2.5
L06	20	2.4	0.2	0.2	0.5	25	2.5
L07	40	2.4	1	3	0.03	55	2.5
L08	-10	0.2	1	12	0.03	80	3
L09	5	0.2	1	10	0.2	80	2.5
L10	-10	2.4	2	12	0.2	80	3
L11	40	0.2	0.2	0.2	0.5	80	2.5
L12	-10	2.4	0.2	0.2	0.5	80	2.5
L13	20	0.2	1	3	0.03	25	15
L14	40	2.4	2	3	0.03	25	15
L15	-10	0.2	0.2	12	0.03	55	15
L16	40	2.4	0.2	12	0.03	25	15
L17	5	2.4	0.2	0.2	0.5	25	15
L18	40	2.4	2	3	0.03	80	15
L19	-10	0.8	0.2	12	0.03	80	15
L20	40	2.4	0.2	12	0.03	80	15
L21	5	0.2	0.2	0.2	0.5	80	15
L22	20	0.2	0.2	0.2	0.000333	55	50
L23	-10	0.2	0.2	0.2	0.1	55	50
L24	40	0.2	2	12	0.2	55	50
L25	-10	0.2	0.2	0.2	0.000333	55	80
L26	40	0.2	0.2	0.2	0.000333	55	80
L27	-10	0.2	0.2	6	0.000333	55	80
L28	40	2.4	0.2	0.2	0.03	55	80
L29	5	0.2	0.2	12	0.03	55	80
L30	40	2.4	5	12	0.03	55	80
L31	20	0.2	1	12	0.1	55	80
L32	5	2.4	1	12	0.1	55	80
L33	-10	0.2	1.3	2	0.5	55	80
L34	20	2.4	0.2	0.2	0.5	55	80
L35	40	0.2	2	3	0.5	55	80

Table 3.2: Design reported in WP1300 (`des_WP1300`, 41 load points)

LP	T	CC	ADC	PDC	F	SoC	dSoC
C01	0	0	0	0	0	0	0
C02	0	0	0	0	0	0	0
C03	-20	0	0	0	0	0	0
C04	0	0	0	0	0	0	0
C05	0	0	0	0	0	0	0
C06	0	0	0	0	0	0	0
L01	0	0	0	0	0	0	0
L02	0	0	0	0	0	0	0
L03	0	0	0	0	0	0	0
L04	0	0	0	0	0	0	0
L05	0	0	0	0	0	0	0
L06	0	0	0	0	0	0	0
L07	-50	0	0	0	0	0	0.5
L08	0	0	0	0	0	0	0
L09	-30	0	0	0	0	0	0.5
L10	0	0	0	0	0	0	0
L11	0	0	0	0	0	0	0
L12	0	0	0	0	0	0	0
L13	0	0	0	0	0	0	0
L14	0	0	0	0	0	30	0
L15	0	0	0	0	0	0	0
L16	0	0	0	0	0	0	0
L17	0	0	0	0	0	0	0
L18	0	0	0	0	0	0	0
L19	0	0	0	0	0	0	0
L20	0	0	0	0	0	0	0
L21	0	0	0	0	0	0	0
L22	0	-2.2	0	0	0	0	0
L23	0	0	0	0	0	0	0
L24	0	0	0	0	0	0	0
L25	0	0	0	0	0	0	0
L26	-50	0	0	0	0	0	0
L27	0	0	0	0	0	0	0
L28	0	0	0	0	0	0	0
L29	0	0	3	0	0	0	0
L30	0	0	0	0	0	0	0
L31	0	0	0	0	0	0	0
L32	0	0	1.1	1.8	0	0	0
L33	0	0	0	0	0	0	0
L34	0	0	0	0	0	0	0
L35	0	0	0	0	0	0	0

Table 3.3: Differences between `des_701` and `des_WP1300`

LP	T	CC	ADC	PDC	F	SoC	dSoC
C01	-10	0	0	0	3.00E-04	15	0.01
C02	40	0	0	0	3.00E-04	15	0.01
C03	5	0	0	0	3.00E-04	55	0.01
C04	20	0	0	0	3.00E-04	55	0.01
C05	-10	0	0	0	3.00E-04	95	0.01
C06	40	0	0	0	3.00E-04	95	0.01
L01	20	0.2	2.0	12.0	0.0300	25	2.5
L02	5	2.4	1.0	12.0	0.1000	25	2.5
L03	40	0.2	2.0	12.0	0.2000	25	2.5
L04	-10	0.2	0.2	0.2	3.00E-04	25	2.5
L05	40	0.2	0.2	0.2	3.00E-04	25	2.5
L06	20	2.4	0.2	0.2	3.00E-04	25	2.5
L07	40	2.4	1.0	3.0	0.0300	55	2.5
L08	-10	0.2	1.0	10.0	0.0300	80	2.5
L09	5	0.2	1.0	10.0	0.2000	80	2.5
L10	-10	1.2	2.0	10.0	0.2000	60	2.5
L11	40	0.2	0.2	0.2	3.00E-04	80	2.5
L12	-10	1.2	0.2	0.2	3.00E-04	60	2.5
L13	20	0.2	1.0	3.0	0.0300	25	15.0
L14	40	2.4	2.0	3.0	0.0300	25	15.0
L15	-10	0.2	0.2	6.0	0.0300	55	15.0
L16	40	2.4	0.2	12.0	0.0300	25	15.0
L17	5	2.4	0.2	0.2	3.00E-04	25	15.0
L18	40	2.4	2.0	3.0	0.0300	80	15.0
L19	-10	0.4	0.2	6.0	0.0300	70	15.0
L20	40	2.4	0.2	12.0	0.0300	80	15.0
L21	5	0.2	0.2	0.2	3.00E-04	80	15.0
L22	20	0.2	0.2	0.2	0.0003	55	50.0
L23	-10	0.2	0.2	0.2	3.00E-04	55	50.0
L24	40	0.2	2.0	12.0	0.2000	55	50.0
L25	-10	0.2	0.2	0.2	0.0003	55	50.0
L26	40	0.2	0.2	0.2	0.0003	55	80.0
L27	-10	0.2	0.2	3.0	0.0003	55	50.0
L28	40	2.4	0.2	0.2	3.00E-04	55	70.0
L29	5	0.2	0.2	12.0	0.0300	60	70.0
L30	40	2.4	5.0	12.0	0.0300	55	70.0
L31	20	0.2	1.0	12.0	0.1000	55	80.0
L32	5	1.2	1.0	12.0	0.1000	55	50.0
L33	-10	0.2	1.3	2.0	0.5000	55	50.0
L34	20	2.4	0.2	0.2	3.00E-04	50	60.0
L35	40	0.2	2.0	3.0	0.5000	55	80.0

Table 3.4: Design actually in use (`des_m`, 41 load points)

T	CC	ADC	PDC	F	SoC	dSoC
0.2	-1	-0.5	0.71	0.23	-0.75	-0.94
-0.4	1	-0.75	0.71	0.56	-0.75	-0.94
1	-1	-0.5	0.71	0.75	-0.75	-0.94
1	1	-0.75	-0.59	0.23	0	-0.94
-0.4	-1	-0.75	0.42	0.75	0.63	-0.94
0.2	-1	-0.75	-0.59	0.23	-0.75	-0.63
1	1	-0.5	-0.59	0.23	-0.75	-0.63
-1	-1	-0.95	-0.16	0.23	0	-0.63
1	1	-0.95	0.71	0.23	-0.75	-0.63
1	1	-0.5	-0.59	0.23	0.63	-0.63
-1	-0.82	-0.95	-0.16	0.23	0.38	-0.63
1	1	-0.95	0.71	0.23	0.63	-0.63
0.2	-1	-0.95	-1	-1	0	0.25
-1	-1	-0.95	-1	-1	0	0.25
1	-1	-0.5	0.71	0.75	0	0.25
-1	-1	-0.95	-1	-1	0	0.25
1	-1	-0.95	-1	-1	0	1
-0.4	-1	-0.95	0.71	0.23	0.13	0.75
0.2	-1	-0.75	0.71	0.56	0	1
-0.4	-0.09	-0.75	0.71	0.56	0	0.25
1	-1	-0.5	-0.59	1	0	1
-1	-1	-1	-1	-1	-1	-1
-0.4	-1	-1	-1	-1	-1	-1
0.2	-1	-1	-1	-1	-1	-1
1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	1	-1
-0.4	-1	-1	-1	-1	1	-1
0.2	-1	-1	-1	-1	1	-1
1	-1	-1	-1	-1	1	-1
-1	1	-0.95	-1	0.56	-0.75	-0.63
0.2	1	-0.5	-0.59	0.56	-0.75	-0.63
-1	1	-0.95	-1	1	-0.75	-0.63
1	-1	-0.5	0.13	1	-0.75	-0.63
1	-1	-0.5	-0.59	0.23	0.63	-0.63
-1	1	-0.95	-1	0.56	0.63	-0.63
0.2	1	-0.5	0.13	1	0.63	-0.63
1	1	-0.95	-1	0.23	0	1
1	1	-0.95	-0.16	0.42	0	1
0.2	1	-0.5	-0.59	0.56	0	1
-1	1	-0.95	-1	1	0	1
0.2	1	-0.95	-1	1	0	1

Table 3.6: The repaired design (41 load points)

LP	T	CC	ADC	PDC	F	SoC	dSoC
C01	-1	-1	-1	-1	-1	-1	-1
C02	1	-1	-1	-1	-1	-1	-1
C03	-0.4	-1	-1	-1	-1	0	-1
C04	0.2	-1	-1	-1	-1	0	-1
C05	-1	-1	-1	-1	-1	1	-1
C06	1	-1	-1	-1	-1	1	-1
L01	0.2	-0.83	-0.5	0.71	0.24	-0.75	-0.94
L02	-0.4	1	-0.75	0.71	0.57	-0.75	-0.94
L03	1	-0.83	-0.5	0.71	0.75	-0.75	-0.94
L04	-1	-0.83	-0.95	-0.97	-1	-0.75	-0.94
L05	1	-0.83	-0.95	-0.97	-1	-0.75	-0.94
L06	0.2	1	-0.95	-0.97	-1	-0.75	-0.94
L07	1	1	-0.75	-0.57	0.24	0	-0.94
L08	-1	-0.83	-0.75	0.43	0.24	0.63	-0.94
L09	-0.4	-0.83	-0.75	0.43	0.75	0.63	-0.94
L10	-1	0	-0.5	0.43	0.75	0.13	-0.94
L11	1	-0.83	-0.95	-0.97	-1	0.63	-0.94
L12	-1	0	-0.95	-0.97	-1	0.13	-0.94
L13	0.2	-0.83	-0.75	-0.57	0.24	-0.75	-0.63
L14	1	1	-0.5	-0.57	0.24	-0.75	-0.63
L15	-1	-0.83	-0.95	-0.14	0.24	0	-0.63
L16	1	1	-0.95	0.71	0.24	-0.75	-0.63
L17	-0.4	1	-0.95	-0.97	-1	-0.75	-0.63
L18	1	1	-0.5	-0.57	0.24	0.63	-0.63
L19	-1	-0.67	-0.95	-0.14	0.24	0.38	-0.63
L20	1	1	-0.95	0.71	0.24	0.63	-0.63
L21	-0.4	-0.83	-0.95	-0.97	-1	0.63	-0.63
L22	0.2	-0.83	-0.95	-0.97	-0.97	0	0.25
L23	-1	-0.83	-0.95	-0.97	-1	0	0.25
L24	1	-0.83	-0.5	0.71	0.75	0	0.25
L25	-1	-0.83	-0.95	-0.97	-0.97	0	0.25
L26	1	-0.83	-0.95	-0.97	-0.97	0	1
L27	-1	-0.83	-0.95	-0.57	-0.97	0	0.25
L28	1	1	-0.95	-0.97	-1	0	0.75
L29	-0.4	-0.83	-0.95	0.71	0.24	0.13	0.75
L30	1	1	0.25	0.71	0.24	0	0.75
L31	0.2	-0.83	-0.75	0.71	0.57	0	1
L32	-0.4	0	-0.75	0.71	0.57	0	0.25
L33	-1	-0.83	-0.68	-0.71	1	0	0.25
L34	0.2	1	-0.95	-0.97	-1	-0.13	0.5
L35	1	-0.83	-0.5	-0.57	1	0	1

Table 3.7: Design actually in use coded using the new (wider) factor-ranges

3.2 The cells

3.2.1 Different cell types

In the actual experiment two different types of cells are considered . Large ones - these are the cells used in automotive applications and whose behaviour should be modeled in the end. Due to economical restrictions a second, smaller type of batteries has especially been made to gain the data needed for parametrizing the linear model. The smaller cells are intended to constitute an economical model-system and will hopefully show, despite controllable scaling effects, the same behaviour as the large cells.

The variation in the data observed stems from different sources, one of it are differences between cells due to the manufacturing process. These differences are much more obvious in the case of the small cells. These are, in contrast to the large cells not machine-made, but hand-made which causes higher variability concerning internal resistance and cell capacity. An interesting, albeit expected observation in this context is that deviations are more likely in directions which are undesirable, i.e we can see more cells with higher resistance and/or lower cell-capacity than would be expected if these quantities are normally distributed. These facts are depicted in figure 3.2 - it shows histograms concerning resistance and cell-capacity supplemented with the empirical density function and the density function of the corresponding normal distribution.

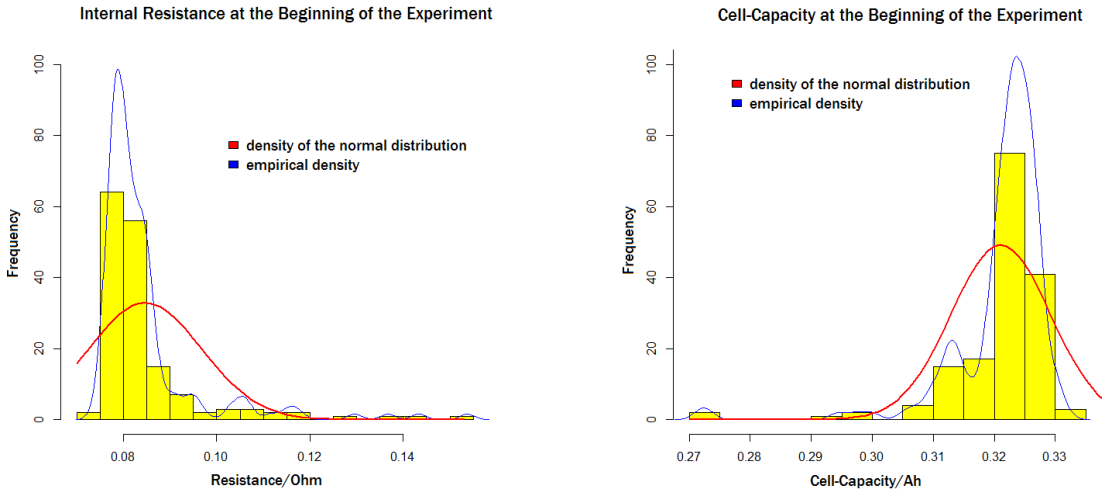


Figure 3.2: State of characteristic quantities at the beginning of the experiment for small cells

Both histograms show considerable skewed distributions. The distribution is positively skewed in case of the internal resistance and negatively skewed in case of the cell capacity. The skewness is in both cases almost similar concerning its magnitude but opposite in direction.

skewness(RTP1_Ri) = 3.185988 skewness(RTP1_Capacity) = -3.289153

Testing for normality yields the expected result - the hypothesis of normally distributed observations is in both cases rejected with very high levels of confidence (the p-values of the tests are smaller than 10^{-16}).

shapiro.test(RTP1_Ri)	shapiro.test(RTP1_Capacity)
Shapiro-Wilk normality test data: RTP1_Ri W = 0.6108, p-value < 2.2e-16	Shapiro-Wilk normality test data: RTP1_Capacity W = 0.693, p-value < 2.2e-16

To investigate whether there is a relationship between resistance and cell-capacity, a scatter-plot is drawn (figure 3.3).

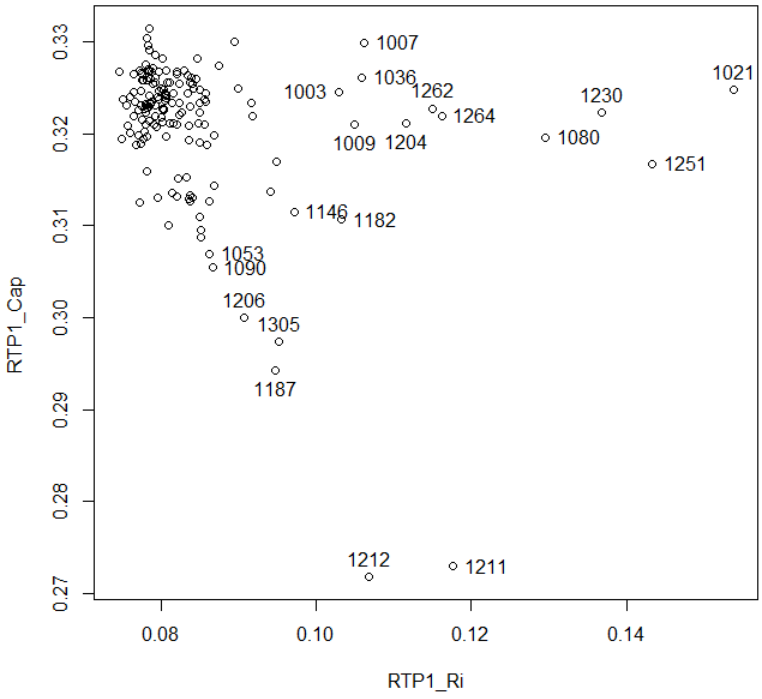


Figure 3.3: Scatter-plot comparing internal resistance and cell-capacity of the small cells at the beginning of the experiment

This plot does obviously not show any evidence for an interrelation between these quantities. To maintain a reasonable homogeneous cell-population to start the experiment with, it was decided to sort out all cells, which showed a deviation from the mean capacity of more than 4%.

To determine which of the two characteristic quantities considered for each cell, internal resistance or cell-capacity, should be used as an indicator of cell aging, the evolution of these (standardized) quantities with respect to the time elapsed is graphed:

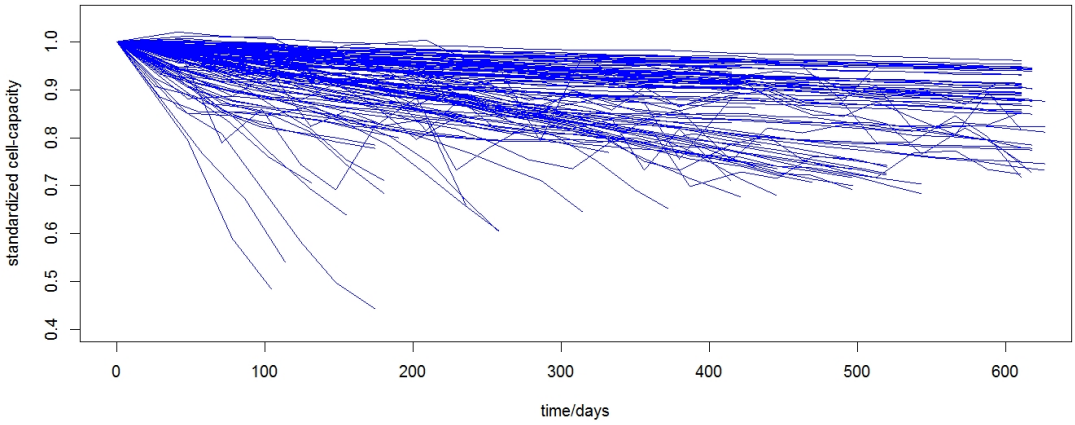


Figure 3.4: Evolution of standardized cell-capacity

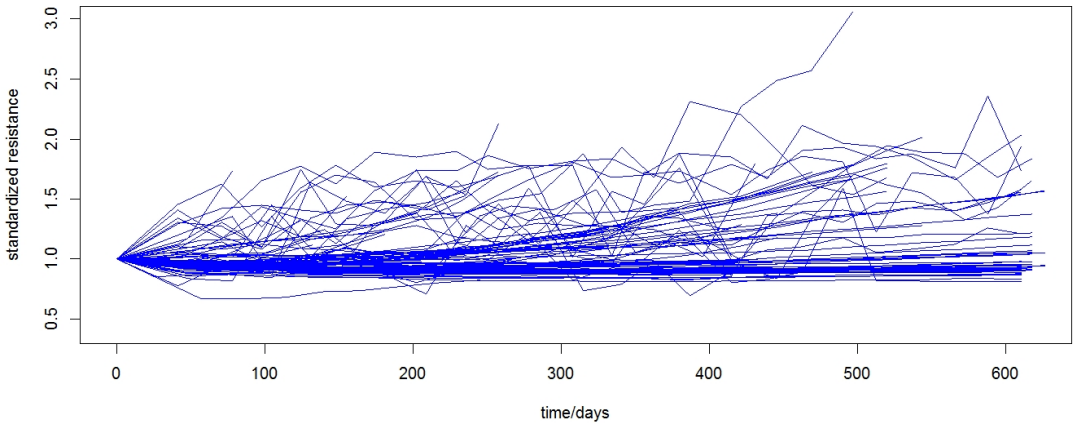


Figure 3.5: Evolution of standardized resistance

Figure 3.5 shows, that the evolution of the internal resistance does not take place in the expected way. Instead of increasing, the resistance seems to decrease in many cases and in addition, looking at single cells, the resistance evolves rather erratically in some cases, i.e. shows irregular behaviour like alternating de- and increasing. The reason for this undesirable behaviour might be measurement errors producing noisy data and adding a

further source of variability this way. Therefore, the internal resistance is not useful for modeling the process of aging.

In opposition to the behaviour of the internal resistance, the cell capacity shows an evolution, which is much more as expected (3.4). The cells capacities are, in most cases, decreasing as anticipated, but still there are some cells, whose capacities show an irregular behaviour. These irregularities are caused by some inherent problems of the method of capacity determination. Data of cells whose measurements are effected by this problem is not used in the analysis of the experiment. In summary, after sorting out problematic cells, the characteristic of capacity evolution makes this quantity the ideal candidate for assessing a cells status of aging.

During the experiment, each experimental condition is imposed on several cells (three or more) and it can be seen that, even after standardizing the respective quantity (internal resistance, cell-capacity) with respect to the initial value, different cells show considerable variability in their response to the same experimental condition. It has been discussed in detail in chapter 2 how this variability is incorporated in modeling by means of a noise term. The model chosen demands, that this design-point specific variability is the same across all experimental settings under consideration - whether this proves to be true has to be verified by running the proper model diagnostic tools.

If the precondition of homoscedasticity is not valid, the varying error variances, which are connected with different experimental conditions then constitute an additional source of variability. In this case it is necessary to at least transform the data in a proper way (see page 34) enabling the application of linear regression analysis, or, if this does not yield the desired result, switch to a modeling concept of greater flexibility like GLM.

Looking at the raw data, plotting the standardized cell-response versus the according load point, could give an idea whether heteroscedasticity can in fact constitute an issue in analyzing this data. Unfortunately, this is problematic in our case because it is not meaningful to compare the current capacity- and resistance-values concerning different load points due to the fact that, at least in most cases, even for one and the same load point the cell specific data is gathered with respect to different RTPs. The reason for this is twofold. First, in the case of most load points, some of the cells have been taken out without having reached their end of life for chemical characterization while others are still running. The second reason is, that not all cells are part of the experiment from the very beginning - some of them have been introduced later in order to use the experimental equipment to full capacity when other cells in the course of the experiment were taken out due to several reasons (e.g. chemical characterization, reached EoL, abnormal behaviour). Because the extrapolated lifespans change with the duration of the experiment (see figures 3.9 to 3.11), it is not reasonable to compare extrapolations based on cells which have experienced different times of operation. By using the nascent capacity of the equipment it was aimed to get better data by replicating critical load points, i.e. load points where the cells do not work properly or show conflicting behaviour.

Figure 3.6 illustrates this situation. For each observation, the highest RTP number experienced by the cell so far is indicated by a certain symbol - same symbol indicates same RTP.

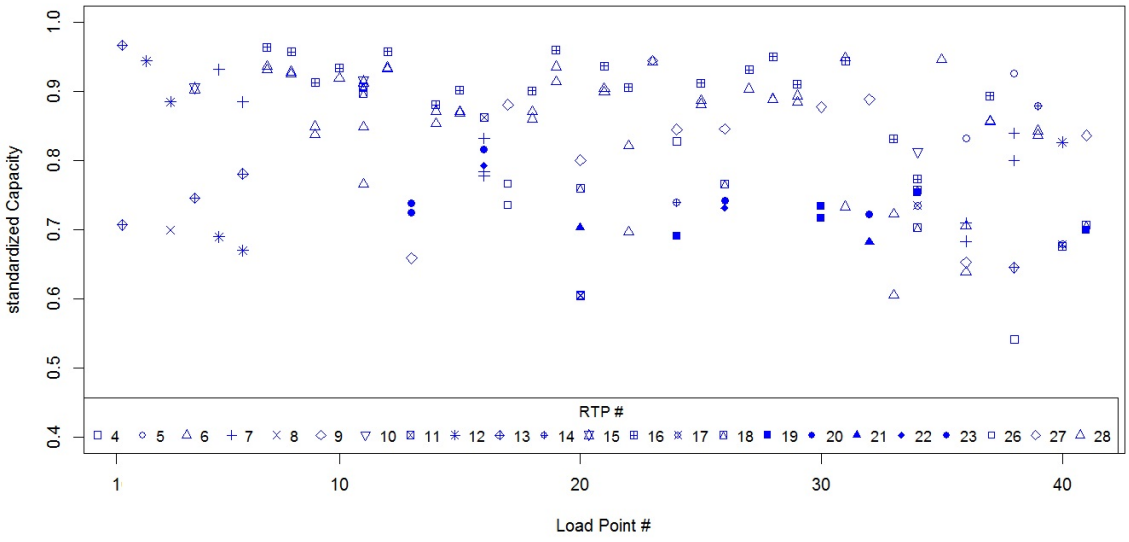


Figure 3.6: Standardized cell-capacity versus load point

If not taking into account that the data belongs to different RTP's, figure 3.6 could possibly be misleading by wrongly attributing the whole differences concerning location as well as variability to the effect of different load points only.

One remedy is, to choose a certain RTP for each load point, which is available for all cells running under its regime and plotting this data versus the load point (figure 3.7).

This is still not a very satisfying solution since a large amount of data available is not used in this examination. However, it can be seen that, despite the fact that the variability within data concerning single load points is obviously smaller now (compare figures 3.6 & 3.7), the variability varies considerably between different load points. Hence, the conclusion is, that heteroscedasticity might become an issue, but in fact, to be able to see the true differences between the effects of the load points in use, we have to compare their lifetimes and the corresponding variabilities. Due to the fact, that many of the cells haven't reached their end of life yet, in these cases we will estimate the cells lifetimes (see the section about extrapolating the cells end of life on page 66 ff.).

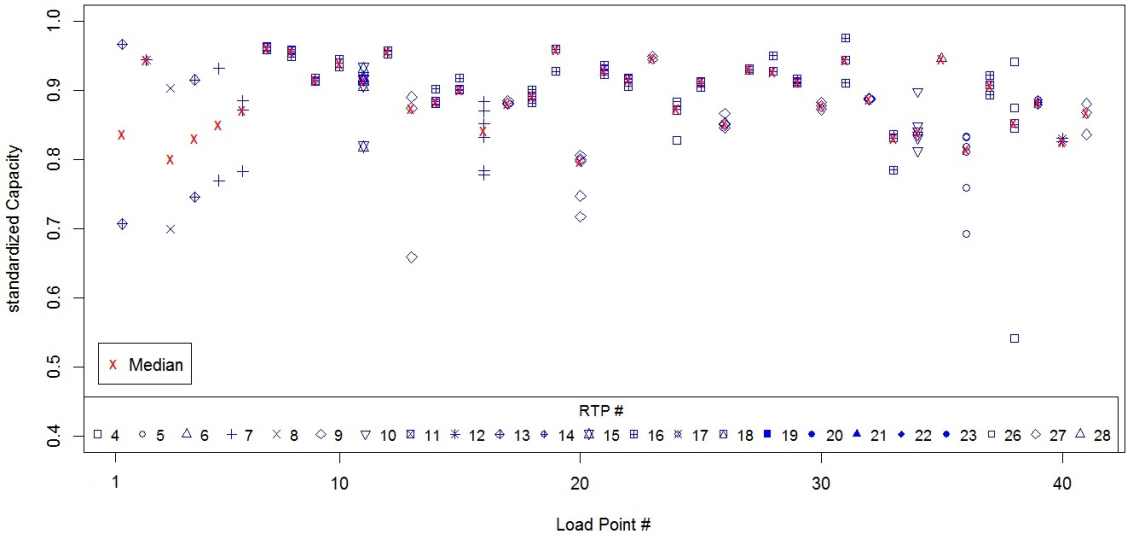


Figure 3.7: Standardized cell-capacity versus load point - showing data which for each load point belongs to the same RTP, but not necessarily to the same RTP for all load points

Interaction-Plots

These are plots intended to discover first order interactions between the factors used in the experiment.

Due to the rather scarce data, proper interaction-plots can hardly be drawn, but nonetheless evidence for interactions can be seen. Therefore, considering interaction terms in describing the relationship between the cells life time and the factors under investigation constitutes an essential part of the modeling process.

To quantify the interaction effects for this large number of factors may require much more data than can be produced due to economical restrictions.

Additionally, at least some qualitative statements about the influence of single factors can be given. On the diagonal, the interaction plot shows the effect according to the single factors, totally ignoring the others. It can be seen, that extreme levels of T, especially in the case of the highest level, have a strong negative effect on the cells lifespans. Additionally, it seems, that the influence of T is rather non-linear (quadratic and/or cubic). A similar behaviour can be observed in the case of ADC and dSoC. In the case of PDC and SoC, the effect seems to be the other way around - the life expectancy seems to be higher for the lowest and the highest level, respectively. A rather low effect can be found for CC and almost none for F.

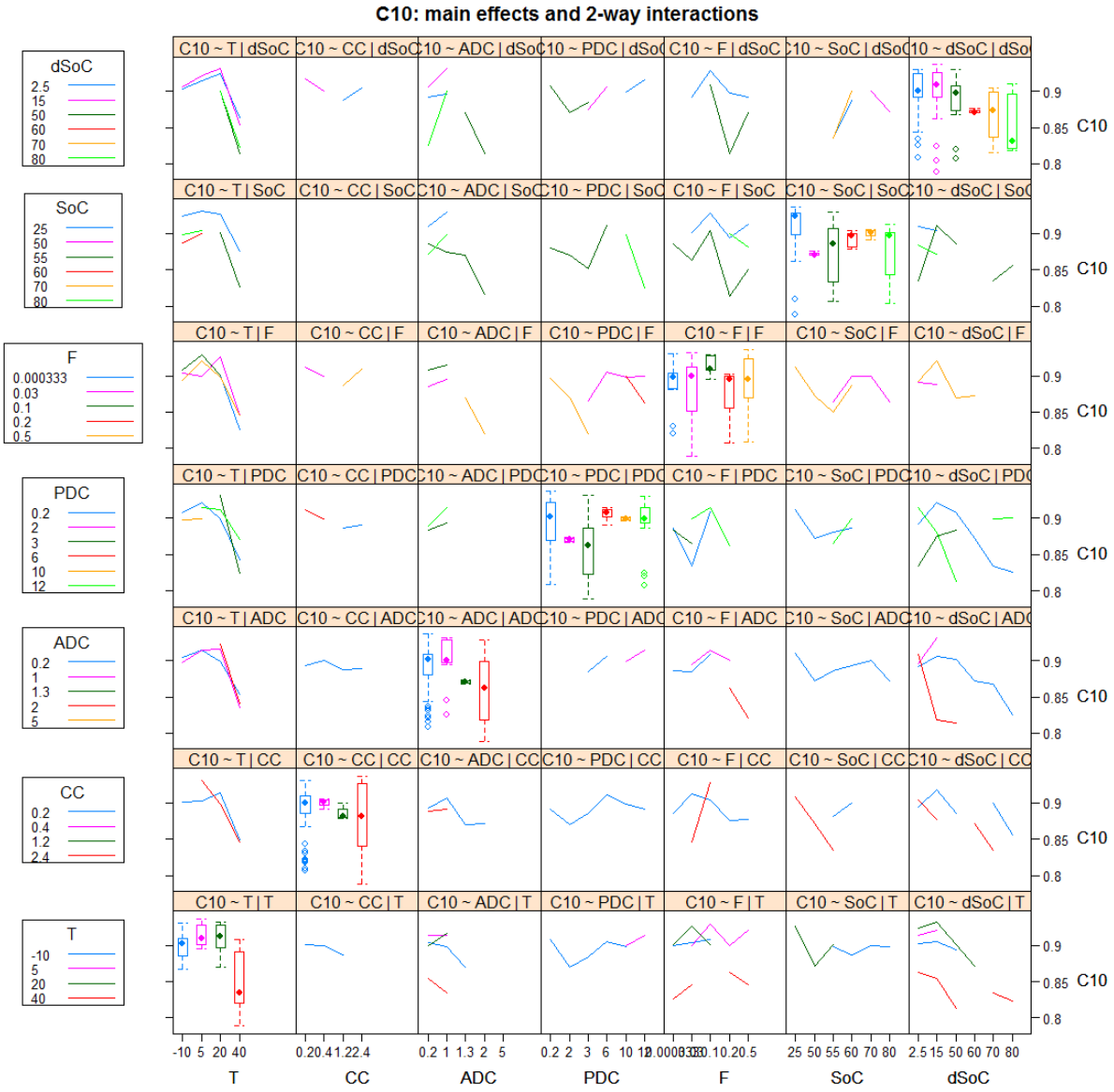


Figure 3.8: Interaction-plot, using the the cell-capacity as response in the case of the RTP 10 data

3.3 Extrapolation of the Cells End of Life

Due to the fact, that the duration of the experiment is to short for reaching the end of life (EoL) of many of the batteries, the lifespan of these cells has to be extrapolated.

This is done as follows. Using the available capacity data, but without the data belonging to RTP1 in order to avoid initial effects, the following simple model has been applied:

$$y(t) = \beta_1 e^{-\beta_2 t}$$

with $y(t)$ denoting the remaining capacity after t days of operation and $\beta_1, \beta_2 \in \mathbb{R}_{>0}$.

Modeling the capacity decay in this way reflects the tendency of flattening the decay by the monotonous growth of the exponential functions slope. This type of modeling doesn't take into account the factors applied, it just establishes a load point specific relationship between time of operation and capacity decay. The question whether this type of modeling is adequate or not, cannot be conclusively decided until (almost) all cells have finished their lives, but all data available at the moment seem to justify this approach, not implying that no other or even better ways exist.

The figures 3.9 to 3.11 show the outcome of this extrapolation process. Each of the diagrams represents one single load point. The measurements are drawn in blue, the extrapolated development of the capacity decay is drawn in black. The horizontal line indicates the 70% level of the initial cell capacity and the vertical lines mark the extrapolated lifetime of the individual cells.

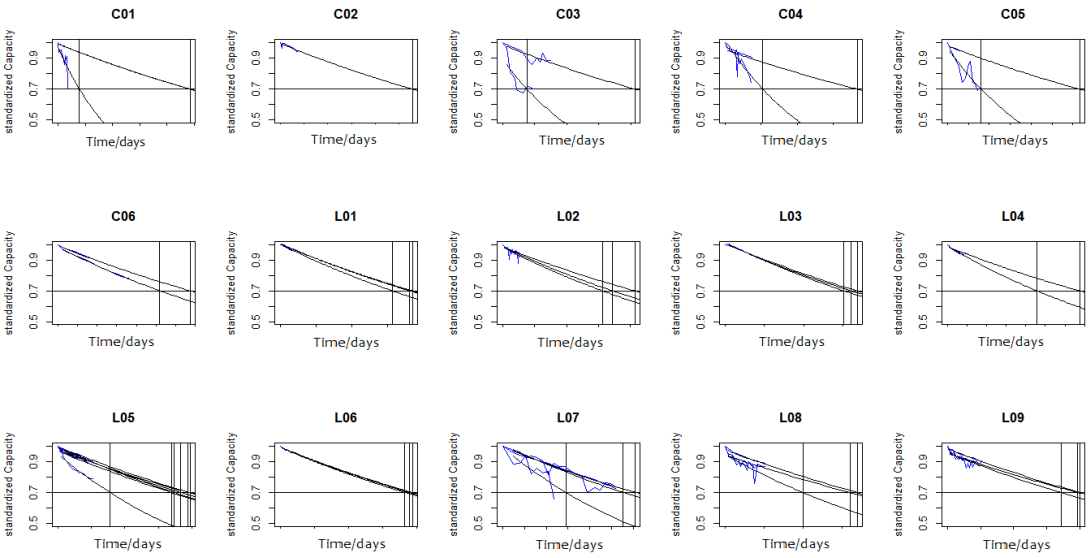


Figure 3.9: EoL - extrapolation C01 - L09

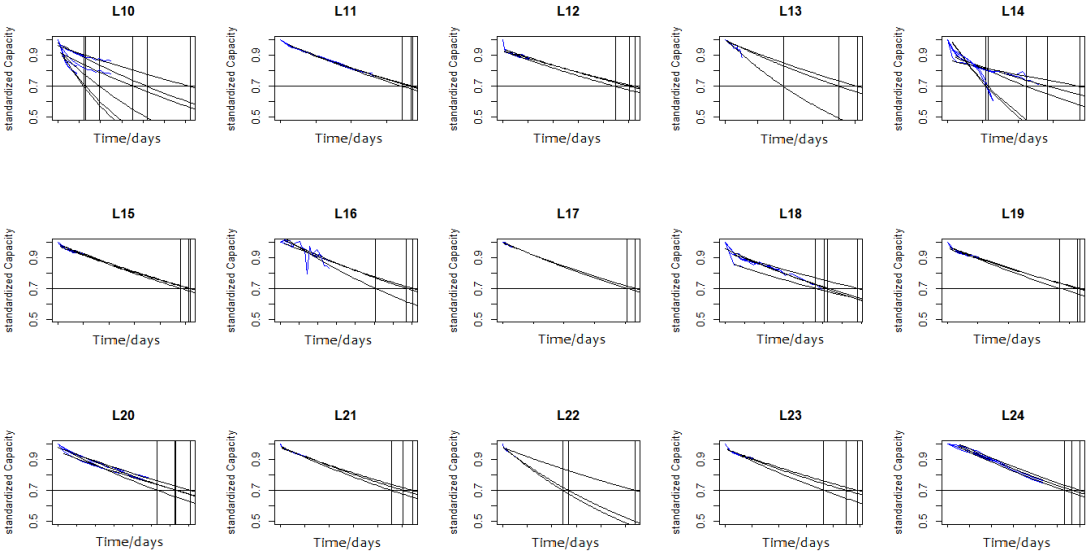


Figure 3.10: EoL - extrapolation L10 - L24

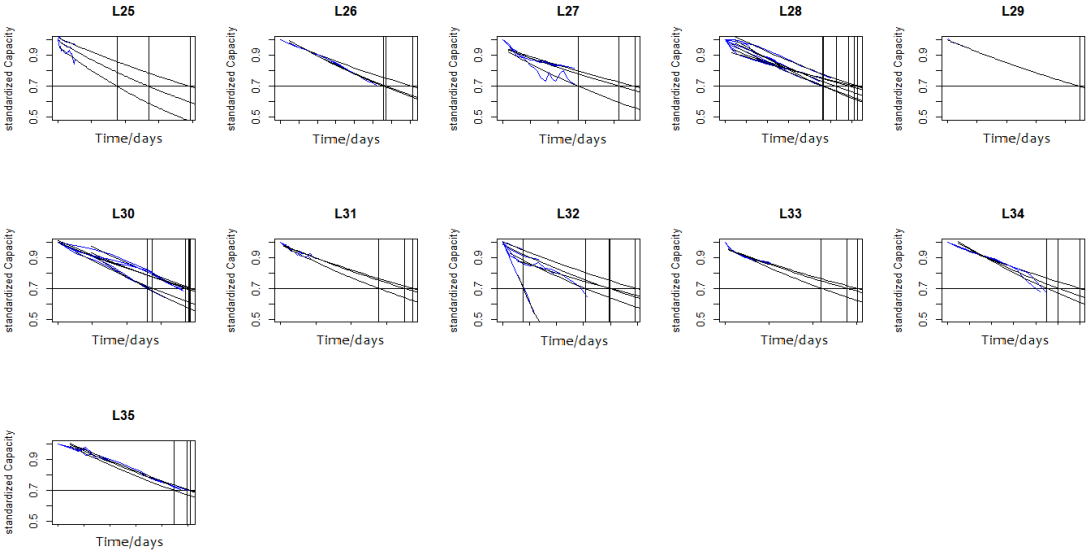


Figure 3.11: EoL - extrapolation L25 - L35

Some of the load points in use have already caused the EoL of the cells which are operated under their regimes whilst others have not. This shows, that at least some of the factors have a massive impact on the life expectancy of the cells. Figure 3.12 exhibits a histogram depicting the distribution of the extrapolated lifespans.

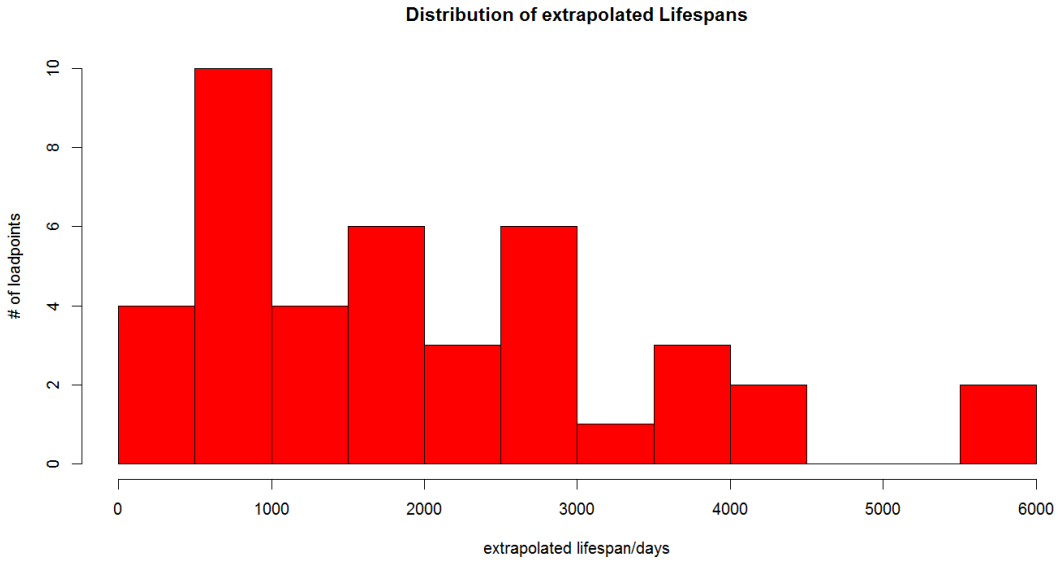


Figure 3.12: Distribution of extrapolated lifespans according to the data gained during RTP20

Another effect, which is obvious, is that in many cases the speed of the capacity-decay is slowing down with increasing duration of the experiment.

This behaviour rises the question about the stability of the extrapolated EoLs, i.e. do the extrapolations change significantly in the course of the experiment and, even more essential, is the character of the load point, i.e. is the load point causing slow or fast aging, significantly affected.

Figures 3.13 to 3.17 show the evolution of the extrapolated lifespans with respect to the data available. One can observe different types of trends, even for the same load point. In some cases the values are stabilizing, in others they are in- or declining.

To get an impression, whether at least the basic characteristic of the load points is not changing over the course of the experiment, the extrapolated lifespans due to data gained during several load points are compared for RTP8, RTP12, RTP16, RTP20 and RTP24 to those calculated with respect to RTP28 (figure 3.18). As expected, the extrapolations are more and more approaching those carried out with respect to the RTP28 data and, despite the fact that, with increasing RTP number, the estimates of the lifespans are by trend becoming longer, we can see that the character of the load points, at least in most cases, is not changing, since in all diagrams there is a strong positive correlation between the extrapolations made with respect to different RTP's. This means that all

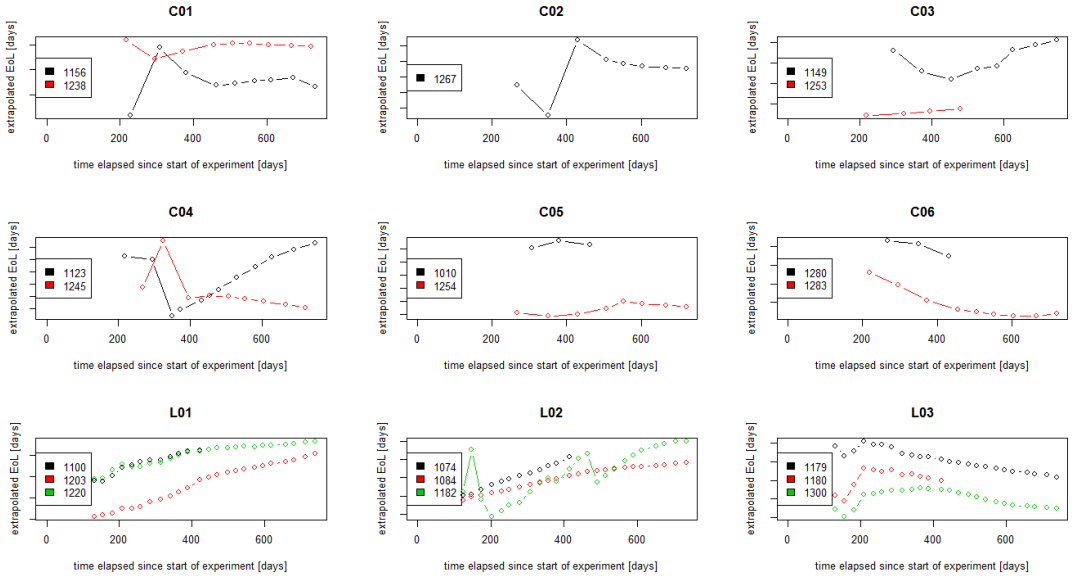


Figure 3.13: EoL - extrapolation vs. duration of experiment, C01 - L03

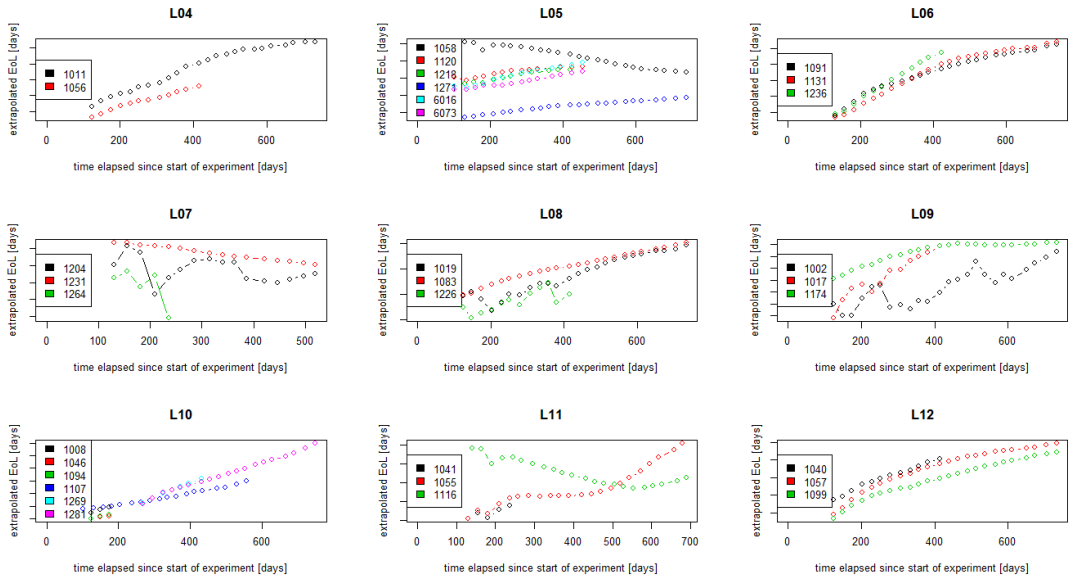


Figure 3.14: EoL - extrapolation vs. duration of experiment, L04 - L12

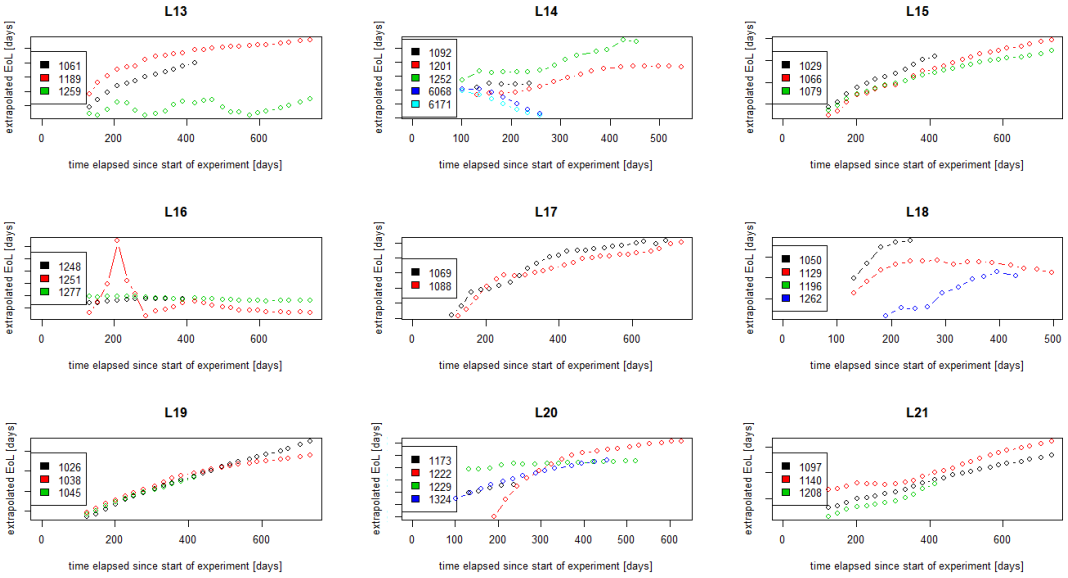


Figure 3.15: EoL - extrapolation vs. duration of experiment, L13 - L21

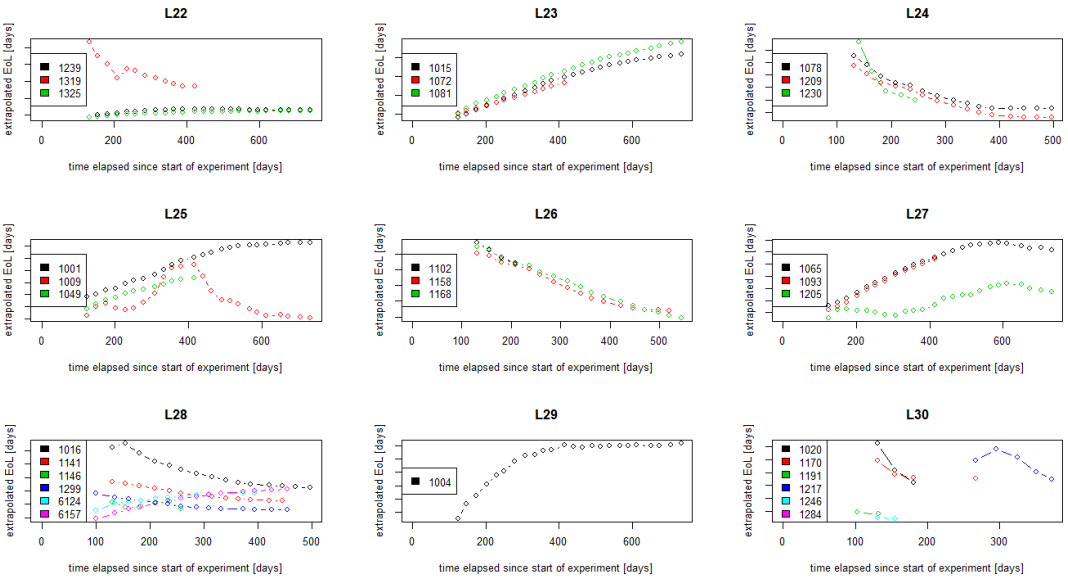


Figure 3.16: EoL - extrapolation vs. duration of experiment, L22 - L30

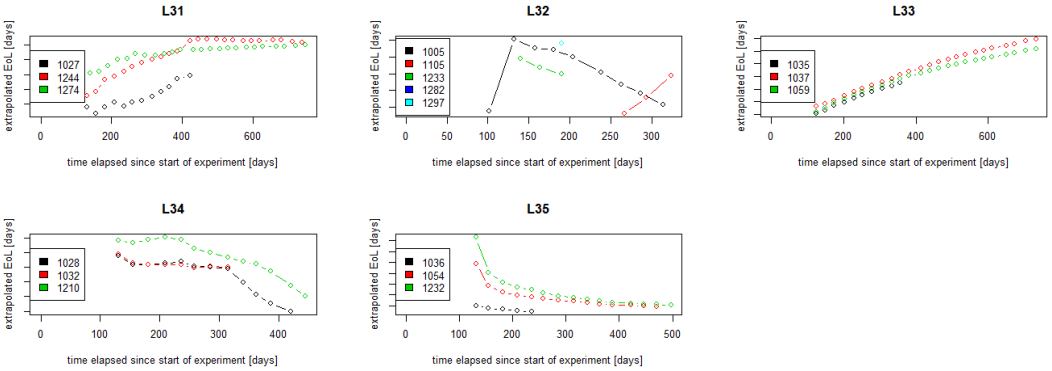


Figure 3.17: EoL - extrapolation vs. duration of experiment, L31 - L35

sub-plots show only relatively small deviations of the earlier extrapolations from that made using the RTP28 data (no 'starry sky').

If successive extrapolations (calculated with respect to RTP8 and RTP12, RTP12 and RTP16,...) are compared, we can in addition see, that the differences between successive extrapolations become smaller and smaller (see figure 3.19), which indicates, that the extrapolations are stabilizing.

Because the scatter-plots show in all cases clearly the same trend, namely that the character of the load points stays the same throughout the whole experiment and is not changing from RTPx to RTPy, it is reasonable to use the extrapolated lifespans for modeling the relationship between factors and capacity decay or equivalently factors and life expectancy of the cells.

But still, the question whether all cells performed in a way, that the data can be used for model selection without hesitation is not fully investigated at this point.

When extrapolating the lifespans, there is one additional problem which we already mentioned but not discussed in detail so far, namely the problem of cell data belonging to the same load point but to different RTPs. For illustration purposes, let us have a look at load point L04. By examining the evolution of the extrapolated life spans for the cells under its regime, we can see (figure 3.14), that both cells show a very similar behaviour, but one of the cells (1056) is running shorter than the other (1011). In this case, it is not reasonable to use the extrapolations calculated with respect to the different RTPs available - this would be equivalent to assume, that in the case of cell 1056 the evolution of the lifespan would not show any change (i.e. proceed horizontally) from the time of the last measurement until the actual experimental duration of cell 1011 is reached. Instead of that, we assume the evolution to be similar to that shown by cell 1011 and therefore we finally extrapolate the lifespan of 1056 by subtracting the difference between the extrapolated lifespans of the two cells according to the latest RTP experienced by cell 1056 from the extrapolated lifespan of cell 1011 according to

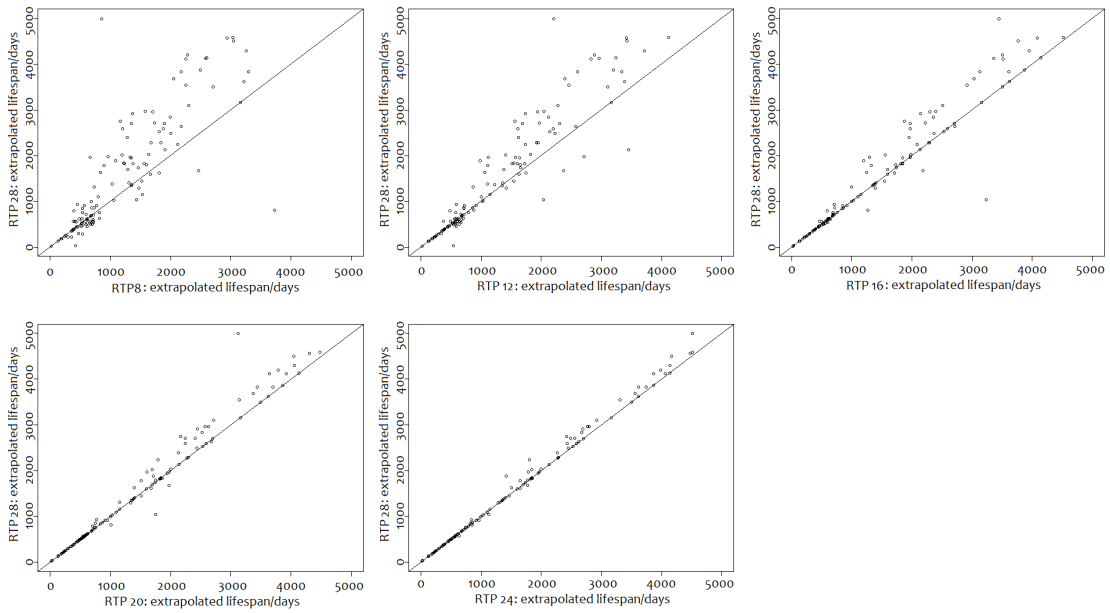


Figure 3.18: Extrapolated lifespans according to RTP: 8, 12, 16, 20 and 24 vs. extrapolated lifespans according to RTP28

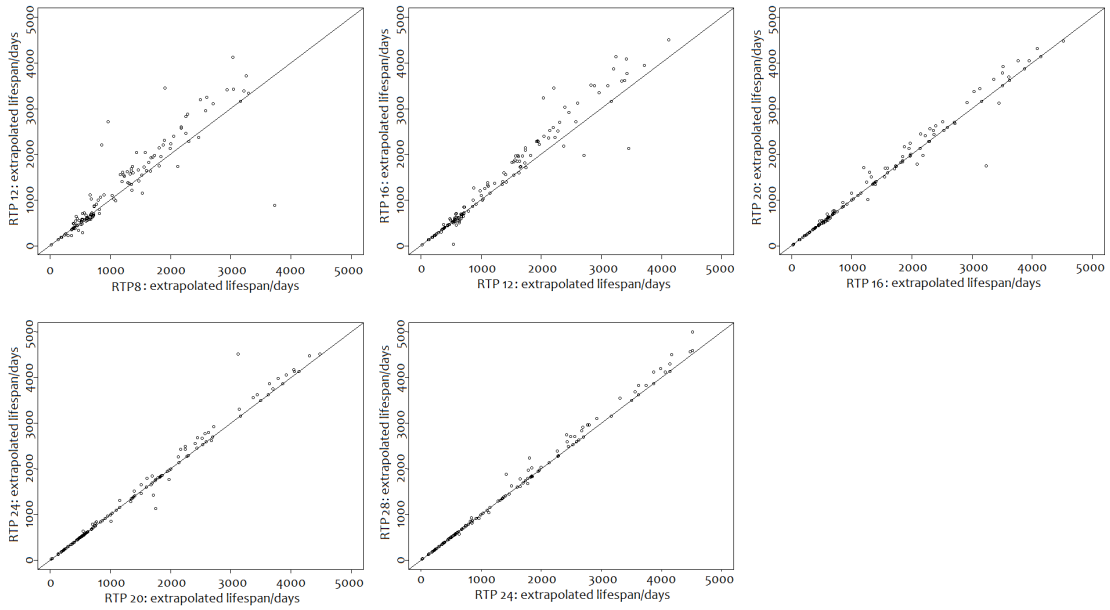


Figure 3.19: Extrapolated lifespans according to RTPx vs. extrapolated lifespans according to RTP(x+4), with x = 8, 12, 16, 20 and 24

the latest RTP experienced by it. This procedure, although rather rough, gives surely a more reasonable result than we get by mixing up data due to different RTPs.

In other cases, some of the cells should be considered as outliers and left aside in analysis. For example cell 1319 (load point 22 - see figure 3.16). The extrapolations regarding cell 1319 have not stabilized yet, while the others are extraordinary stable. In addition, the extrapolations for cell 1319 are tending obviously towards those belonging to the other two cells under the regime of L22.

In the case of C01, we have conflicting information about the lifespans which can be reached by cells operated under its regime. The data belonging to one cell suggest a lifespan which is around 6 times longer than in the second case. By examining the available data a bit closer, the involved chemist came to the conclusion, that a reasonable lifespan in the case of C01 is very likely something in between these two differing lifespans - so to solve this conflict, instead of the two extrapolations available, only one estimated lifespan in between the two observed values is used for further analysis.

Also in the case of C02, the involved chemist concluded due to similar reasons as in the case of C01, that the extrapolated life span should be reduced by around 1000 days.

3.4 Modeling the Relationship between the Cells' Lifespans and the seven different influential Factors using multiple linear Regression (MLR)

3.4.1 Balancing the Data Set

The experimental design consists of 41 different experimental settings or load points (LP) which are initially applied to three cells each. This means, that for most experimental conditions, we have three independent observations. Due to unexpected problems with some of the cells, there exist load points which yield less than three usable results and in the case of four load points, namely C01, C02, C03 and LP29, there exists only one cell each considered working reliably. In other cases, there exist more than three observations, due to subsequent replications of the experimental conditions (see page 62). This means, that we are now in the rather unfavourable situation of dealing with an unbalanced dataset. Since we want to benefit from the pros of a balanced design when doing the statistical analysis of the data, we artificially balanced the data by

- trimming down the number of observations if more than 3 are available or
- adding artificial observations if less than 3 observations have been made

by following rules.

Actually, we set up a procedure in R to balance the data automatically:

- More than three observations: keep the lowest and the highest observation and in addition the observation, which is closest to their average.
- Two observations: add the average of the two available observations. In statistical analysis, introducing an artificial observation of this type has to be accounted for by reducing the degree of freedom by one.
- One observation: we take its value as the mean of a normal distribution with variance empirically determined using the data produced by the experiment. In a next step we add two randomly generated values by using this distribution. To avoid producing outliers this way, we limit the distance between the randomly generated observation and the mean value of the distribution used by the maximum observed distance between the observations made in the experiment and the mean values of the observations made under the regime of the respective load points.

Then we analyze the dataset balanced this way (i.e. fit the model `lm.ex`), repeat the whole process of balancing and analyzing ten times in total, compare the different results (i.e. compare the parameters and p-values generated by parametrizing the model `lm.ex`) and, as a last step, if no significant differences occur, we choose one of the balanced data sets randomly for use in the subsequent analysis.

Additionally, the same analysis is done with the unbalanced dataset, to see whether the process of balancing is of significance for the subsequent model selection process.

When looking at the 'diagnostic'-plots, which are the sub-figures of figures 3.20 to 3.23, we can see, that balancing the dataset as specified above does not have a huge impact on the significance of most regressors when parametrizing the model `lm.ex` by using 10 different automatically generated balanced datasets and additionally the unbalanced dataset. The larger deviations can be found between the values of the quantities calculated in the balanced cases and the unbalanced case, but these values are still of the same magnitude and do not yield a significantly different picture when compared with the results of the balanced cases. The diagnostic-plots show the p-values and the parameters in the case of the intercept and the 23 regressors of `lm.ex` which are generated when parametrizing this model. The subplots showing the p-values are ordered such that the first ten subplots show the p-values of the significant regressors, the next three subplots those of regressors with borderline significance and the remaining 11 subplots show the p-values of the regressors which are not significant. The numbers on the left hand side of the sub-figures indicate the position of the regressors in the model `lm.ex`.

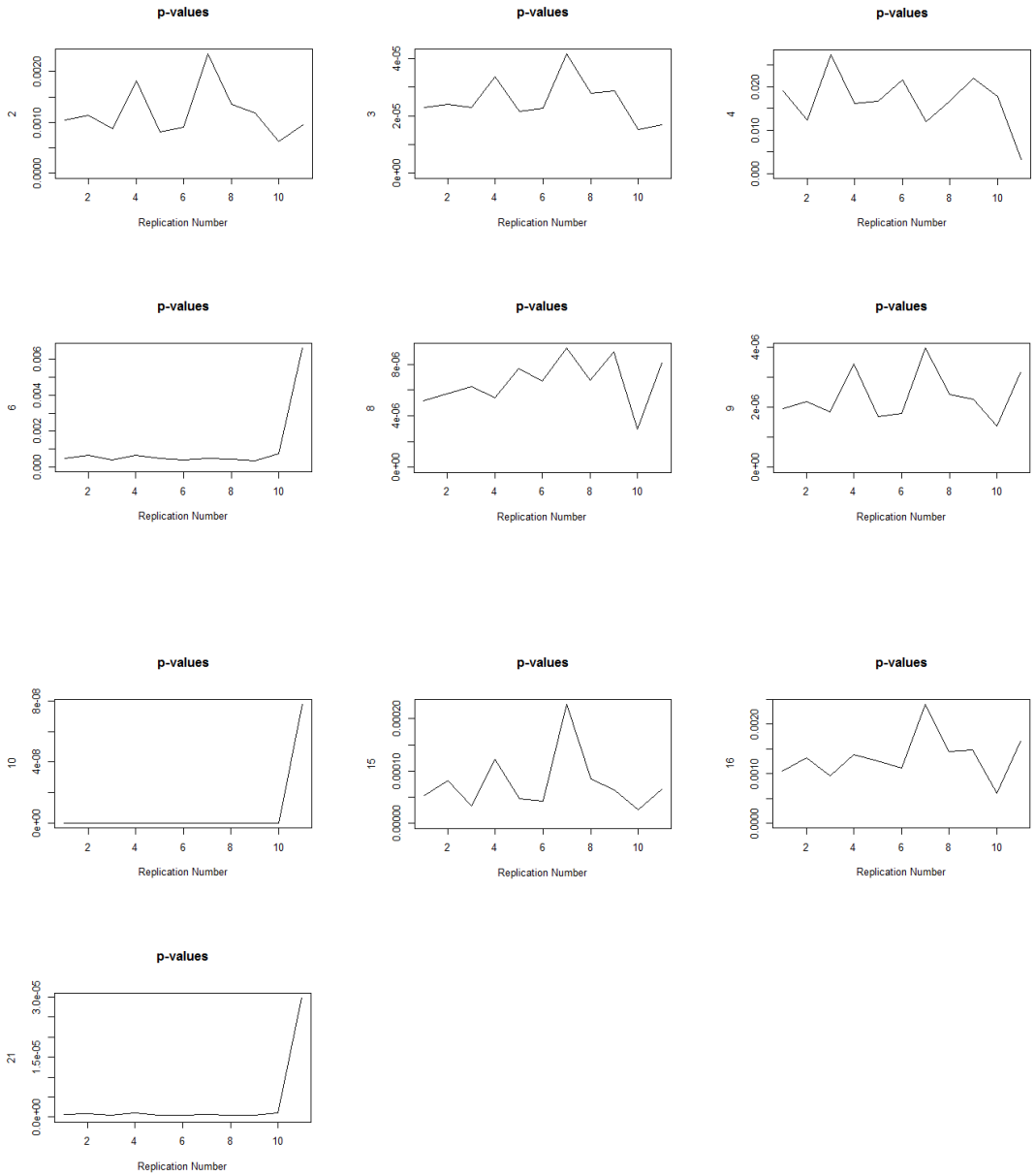


Figure 3.20: p-values (on y-axis) of the 10 regressors of `lm.ex` turning out to be highly significant when 10 different balanced datasets and the unbalanced dataset (1 to 11 on x-axis) are used for parametrization of `lm.ex` (10 sub-figures above). The numbers on the left hand side of the sub-figures indicate the position of the regressors in the model `lm.ex`

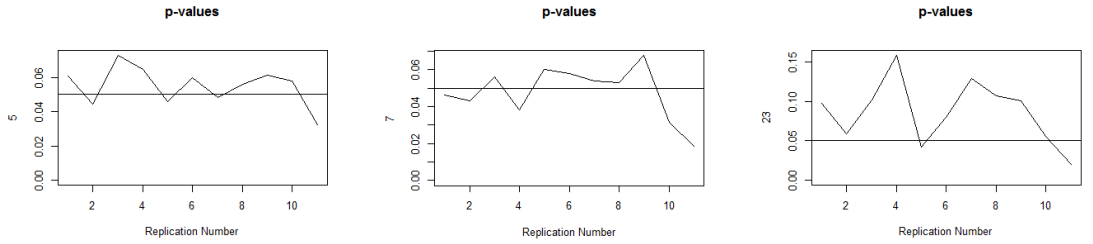
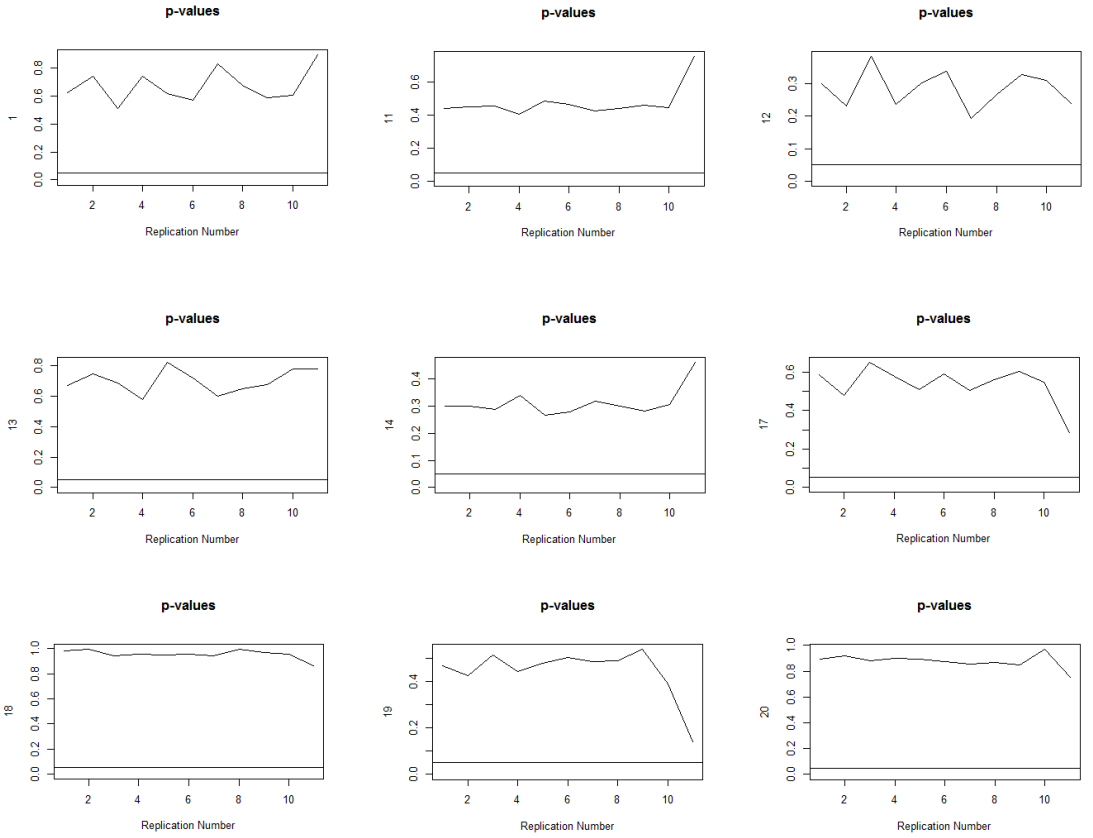


Figure 3.21: p-values (on y-axis) of the 3 regressors of `lm.ex` showing borderline significance when 10 different balanced datasets and the unbalanced dataset (1 to 11 on x-axis) are used for parametrization of `lm.ex` (3 sub-figures above). The numbers on the left hand side of the sub-figures indicate the position of the regressors in the model `lm.ex`



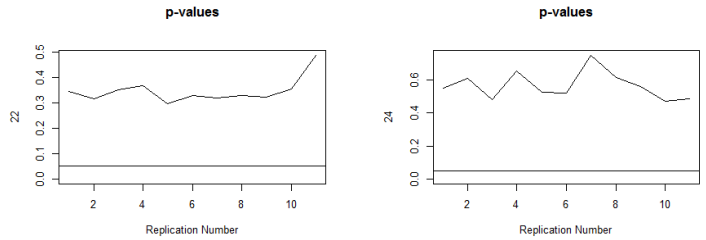
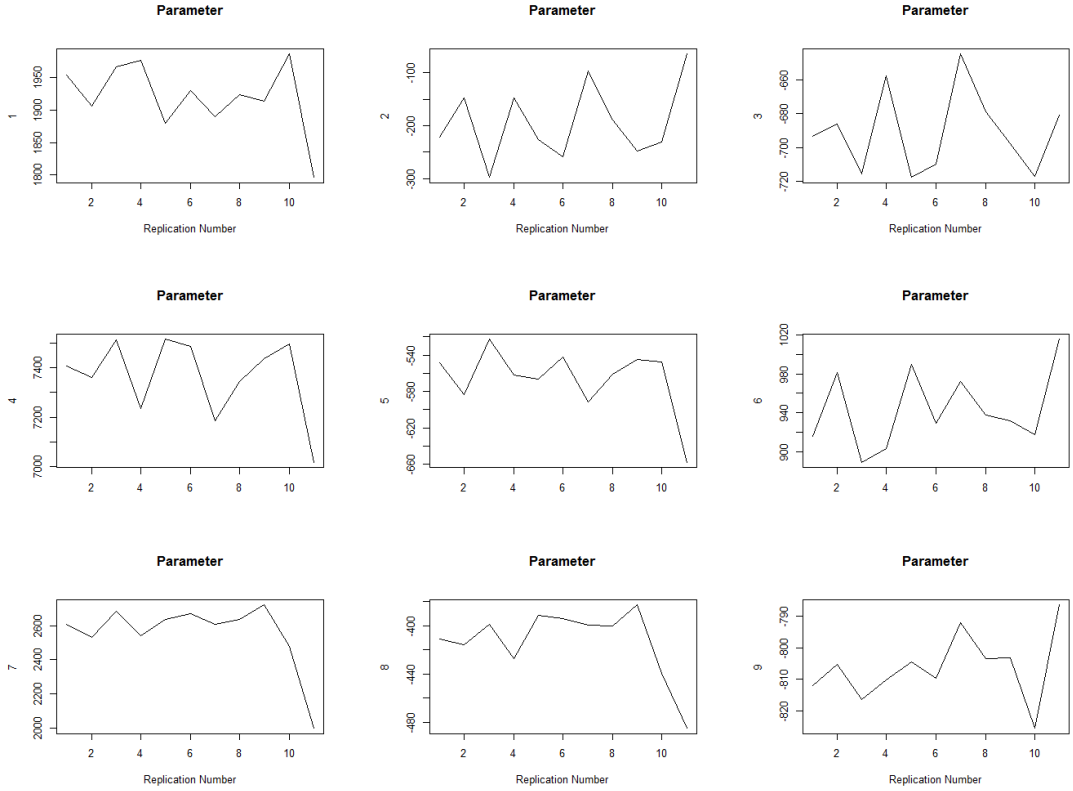


Figure 3.22: p-values (on y-axis) of the intercept and 10 regressors of `lm.ex` turning out to be not significant when 10 different balanced datasets and the unbalanced dataset (1 to 11 on x-axis) are used for parametrization of `lm.ex` (11 sub-figures above). The numbers on the left hand side of the sub-figures indicate the position of the regressors in the model `lm.ex`



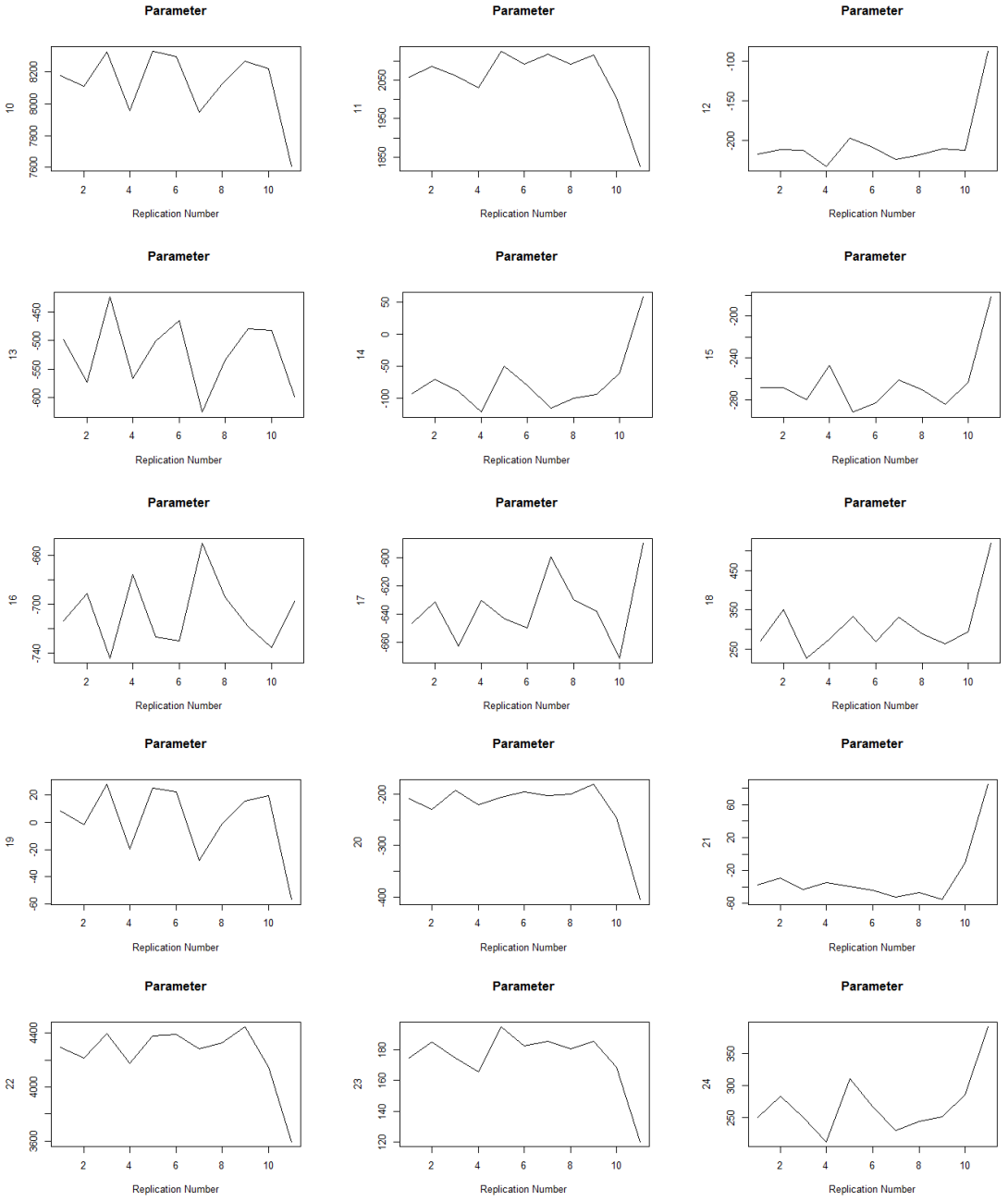


Figure 3.23: Parameter-values (on y-axis) in the case of the intercept and the 23 parameters of `lm.ex` arising when 10 different balanced datasets and the unbalanced dataset (1 to 11 on x-axis) are used for parametrization of `lm.ex` (24 sub-figures above). The numbers on the left hand side of the sub-figures indicate the position of the regressors in the model `lm.ex`

In addition, we also compared the values of the residual standard error which are calculated when parametrizing `lm.ex` using the 10 different balanced datasets and the unbalanced data as well (figure 3.24). There are again only minor differences between the balanced datasets. As anticipated, the larger deviations can be found between the values of the quantities calculated in the balanced cases and the unbalanced case.

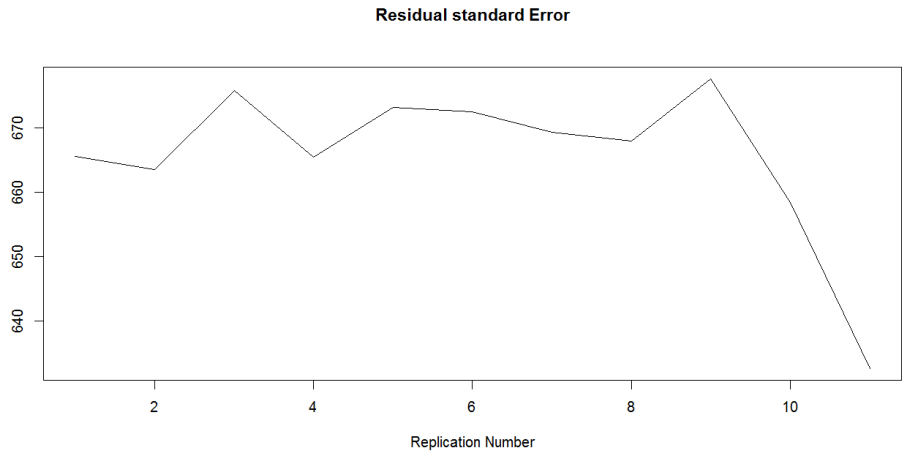


Figure 3.24: Residual standard error obtained when parametrizing `lm.ex` by using 10 different balanced datasets (replication number 1 - 10) and the unbalanced data (replication number 11).

All together, it appears to be reasonable to balance the data set as described above. The analysis, i.e. the 'diagnostic' - plots shown do not provide evidence, that balancing will significantly distort the results of the subsequent analysis.

Degrees of Freedom

In the case of the unbalanced data set there are 8 load points in total where we have only two (reliable) extrapolations at hand and therefore had to add the mean values of these two extrapolations in order to balance it. This also means, that we have to reduce the available degrees of freedom in the subsequent statistical analysis by 8. Eight observations out of 123 - this is a fraction of around 6.5 % which is of course not a totally marginal part of the observations, but we think still a fraction which will, due to the fact that the regressors chosen in the model selection process (using the unreduced degrees of freedom) are highly significant, have no significant effect on the outcome of the analysis. We therefore rely on the 'approximate result' of the analysis using the uncorrected degrees of freedom.

3.4.2 Statistical Analyis and Model Selection

All analyses which will be discussed from now on, will always be based on the balanced dataset. Therefore, it will not be mentioned each time separately.

When parametrising the model `lm.ex`, we see (figure 3.25), that there are many parameters considered insignificant (i.e. have p-values greater than 0.05). This tells us, that there is a considerable amount of redundancy in the model, i.e. not all parameters are needed to describe the relationship between in- and output.

In order to get rid of non-informative parameters, we use the R-routine `leaps` to perform an all-subsets regression (see page 33 for more details). By using several criteria (adjR^2, BIC and Mallows' Cp), we try to select the model which is the most suitable under all models at hand, i.e. this procedure of course does in general not yield the best possible model for describing the data, it just yields the most suitable under all linear models which can be constructed by using the regressors available in the model `lm.ex` (for more information on this topic see page 29 ff.).

When plotting the best model (in terms of the criteria mentioned above) of each size (i.e. number of regressors in the model) ordered by its corresponding BIC value (figure 3.26), then we get a good impression of the order of magnitude concerning the size of suitable models. In terms of the BIC, the 7 best models are ranging in between 6 - 12 regressors, showing BIC-values between -130 and -140, i.e. are pretty close together concerning this criterion. In addition we cannot only see which regressors are comprised by each of these models, but also get an impression of which of them show up in almost all of the models, thus being likely the most informative of the 24 regressors in total. The regressors showing up most often are the temperature (T), the state of charge (SoC), the squared average discharge current (ADC^2), the squared state of charge (SoC^2), the charge current (CC), the average discharge current (ADC) and the interactions of charge current with peak discharge current (CC:PDC), of charge current with state of charge (CC:dSoC) and of state of charge with delta state of charge (SoC:dSoC).

The interaction SoC:dSoC is literally ubiquitous - it is present in all models. The temperature is included in all models with more than one regressor in form of at least one of the regressors T, T^2 and T^3 and can often be found also in one or more interactions present in the model. The same is also true for SoC, with the exception of not having SoC^3 as a potential regressor. Therefore, the selected model will include T^2 or T^3, T, SoC, SoC^2 and last but not least the interaction SoC:dSoC, i.e. it will be a model comprising at least 5 regressors.

To refine our quest for the most adequate model, additionally we consider further criteria, namely Mallows' Cp and the adjusted R^2. To get a comprehensive picture concerning the criteria used, we plot the number of parameters (from 1 to 24) versus the values of the respective criterion associated with the best model of each size (figure 3.27). Two criteria, namely adjR^2 and the BIC suggest models with 6 and 10 regressors and Mallows' Cp indicates a model with approximately 10 regressors. In the following analysis we will

investigate the best models comprising 6 and 10 regressors, subsequently named `lm6_bal` and `lm10_bal`.

```
> summary(lm.ex)

Call:
lm(formula = EoL ~ T + CC + ADC + PDC + F + SoC + dSoC + I(T^2) +
    I((ADC)^2) + I(SoC^2) + I(dSoC^2) + I(T^3) + T:CC + T:PDC +
    CC:PDC + CC:dSoC + PDC:F + PDC:SoC + PDC:dSoC + F:dSoC +
    SoC:dSoC + T:F + T:SoC + F:SoC)

Residuals:
     Min       1Q   Median       3Q      Max
-1447.36  -382.52    14.57   369.39  1879.94

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1917.566    395.576   4.848 4.69e-06 ***
T           -143.484    444.466  -0.323 0.747516
CC          -690.093    203.651  -3.389 0.001013 **
ADC         7375.934   1652.470   4.464 2.16e-05 ***
PDC         -585.218    227.971  -2.567 0.011768 *
F            981.516    479.903   2.045 0.043511 *
SoC         2484.768    715.090   3.475 0.000763 ***
dSoC        -425.840    202.033  -2.108 0.037602 *
I(T^2)      -809.211    167.125  -4.842 4.80e-06 ***
I((ADC)^2)  8106.520   1604.501   5.052 2.02e-06 ***
I(SoC^2)    2067.414    271.875   7.604 1.75e-11 ***
I(dSoC^2)   -211.893    277.889  -0.763 0.447586
I(T^3)      -574.891    473.150  -1.215 0.227274
T:CC         -62.975    217.012  -0.290 0.772284
T:PDC       -265.421    256.698  -1.034 0.303689
CC:PDC      -694.979    167.581  -4.147 7.17e-05 ***
CC:dSoC     -637.380    190.099  -3.353 0.001138 **
PDC:F        358.832    492.446   0.729 0.467940
PDC:SoC       -0.738    368.464  -0.002 0.998406
PDC:dSoC    -240.697    284.026  -0.847 0.398811
F:dSoC       -21.246    281.464  -0.075 0.939983
SoC:dSoC    4160.783    798.870   5.208 1.05e-06 ***
T:F          182.375    181.774   1.003 0.318185
T:SoC        291.665    147.807   1.973 0.051280 .
F:SoC       -223.990    414.816  -0.540 0.590439
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 660.7 on 98 degrees of freedom
Multiple R-squared:  0.8226,    Adjusted R-squared:  0.7791
F-statistic: 18.93 on 24 and 98 DF,  p-value: < 2.2e-16
```

Figure 3.25: Printout of the R routine `lm`, when parametrizing model `lm.ex` using the balanced data.

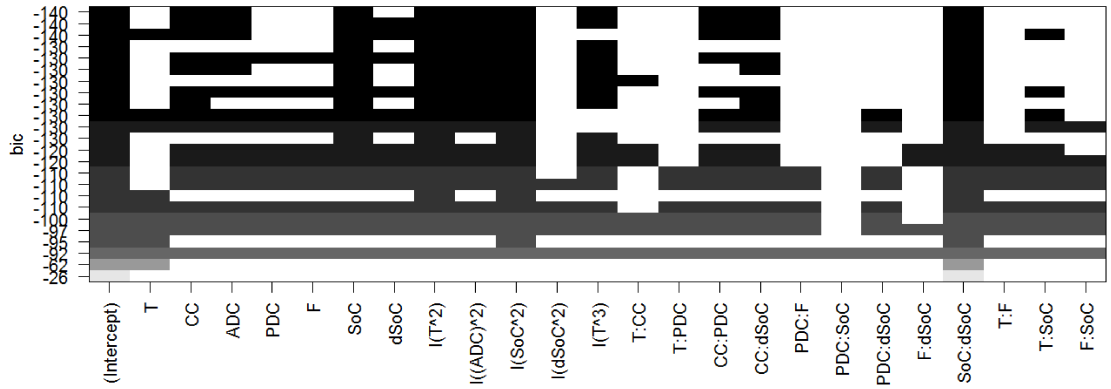


Figure 3.26: Best models with 0,1,...,24 regresors orderd by their respective BIC-value.

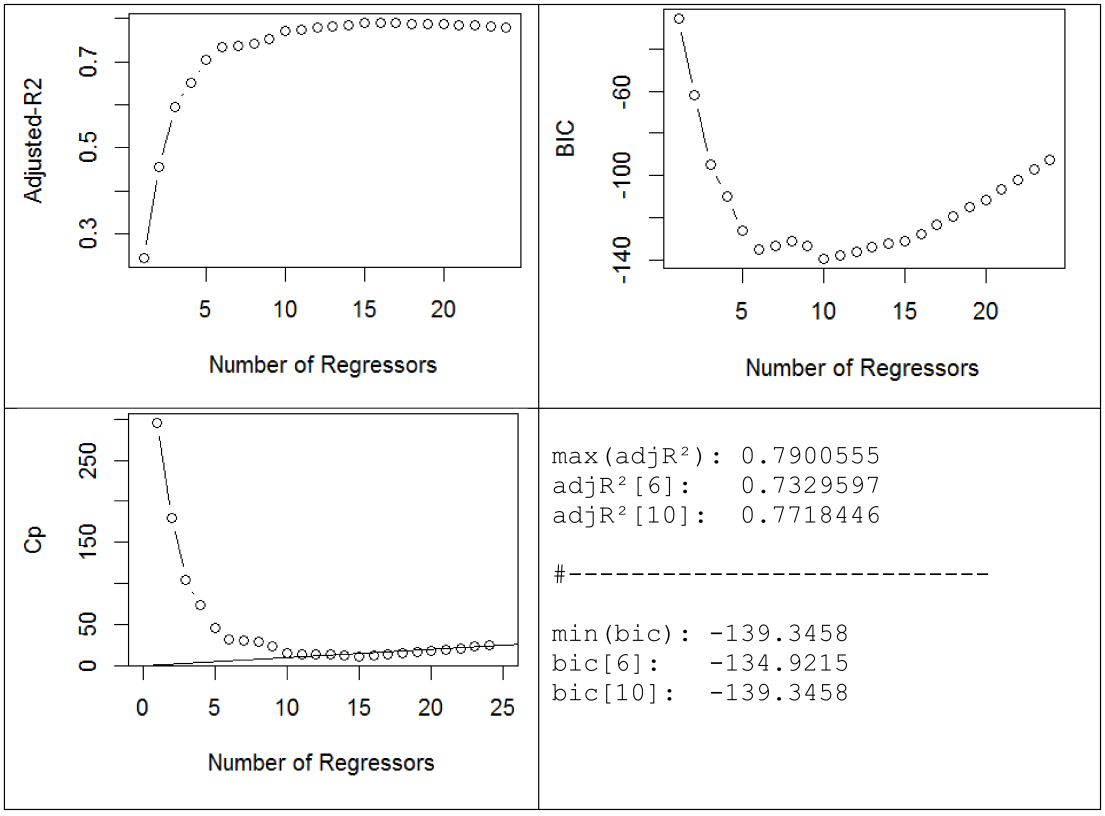


Figure 3.27: Selection criteria vs. model-size.

The printout of the R routine `lm`, when parametrizing model `lm6_bal` (figure 3.28) shows, that all regressors in the model are highly significant. It contains the regressors stated above (T^2, T^3, SoC, SoC^2, SoC:dSoC) and additionally ADC^2.

```
> lm6_bal<-lm(EoL~ SoC + I(T^2) + I((ADC)^2) + I(SoC^2) + I(T^3) + SoC:dSoC)
> summary(lm6_bal)

Call:
lm(formula = EoL ~ SoC + I(T^2) + I((ADC)^2) + I(SoC^2) + I(T^3) +
    SoC:dSoC)

Residuals:
     Min       1Q   Median       3Q      Max
-1773.97  -365.41   -12.99   390.30  1931.51

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1343.30     221.33   6.069 1.66e-08 ***
SoC          3010.92     614.68   4.898 3.16e-06 ***
I(T^2)       -879.09     157.60  -5.578 1.61e-07 ***
I((ADC)^2)    905.09     246.76   3.668 0.000371 ***
I(SoC^2)     1718.08     213.80   8.036 8.67e-13 ***
I(T^3)       -712.55      85.36  -8.348 1.68e-13 ***
SoC:dSoC     4574.14     691.78   6.612 1.21e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 726.5 on 116 degrees of freedom
Multiple R-squared:  0.7461,    Adjusted R-squared:  0.733
F-statistic: 56.81 on 6 and 116 DF,  p-value: < 2.2e-16
```

Figure 3.28: Printout of the R routine `lm`, when parametrizing model `lm6_bal` using the balanced data.

By looking at the diagnostic-plots (figure 3.29), which are automatically generated by applying the plot function on an lm-object (which is a result of the lm-function), we can see that the QQ-plot indicates a deviation from the assumption of normally distributed errors and that the scale-location plot does also allow for doubting the assumption of homoscedasticity.

Though the result is not at all clear, we apply some statistical tests (table 3.8).

`shapiro.test(rstudent(lm6_bal))`	`ncvTest(lm6_bal)`

`Shapiro-Wilk normality test`
`data:  rstudent(lm6_bal)`
`W = 0.9856, p-value = 0.2199`
	`Non-constant Variance Score Test`
	`Variance formula: ~ fitted.values`
	`Chisquare = 4.445739    Df = 1`
	`p = 0.03498844`

Table 3.8: Results of Shapiro-Wilk- and ncv-test in the case of `lm6_bal`.

The Shapiro-Wilk normality test does not reject the assumption of normally distributed errors but the ncv-test rejects the assumption of homoscedasticity (even though the p-value of 0.035 is only slightly smaller than the chosen significance level of 0.05).

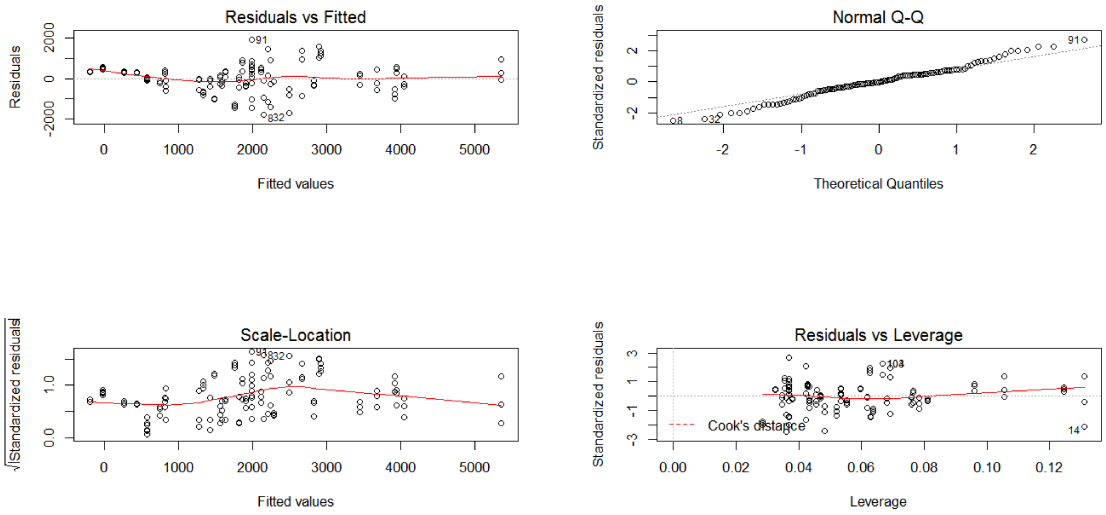


Figure 3.29: Diagnostic-plots in the case of `lm6_bal`.

The Box-Cox-routine (figure 3.30) suggests a transformation with $\lambda = 0.412$.

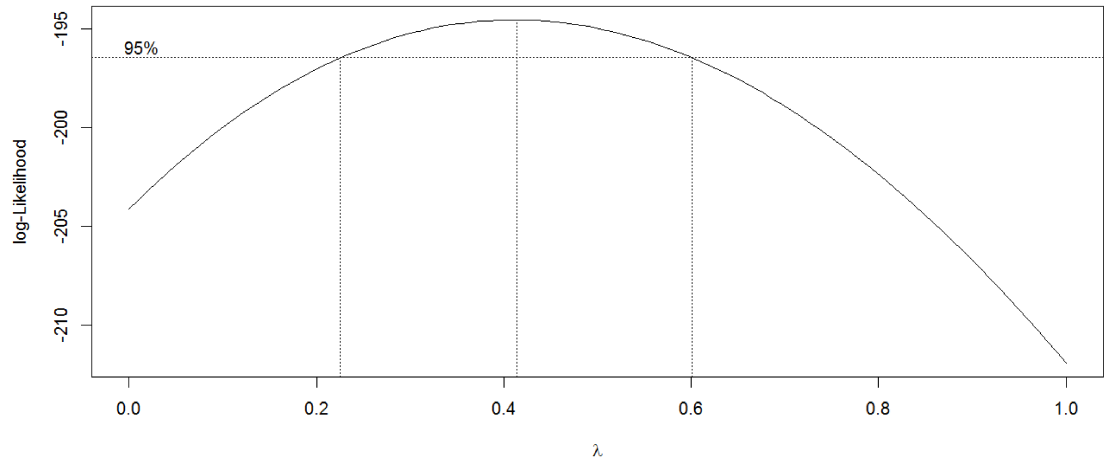
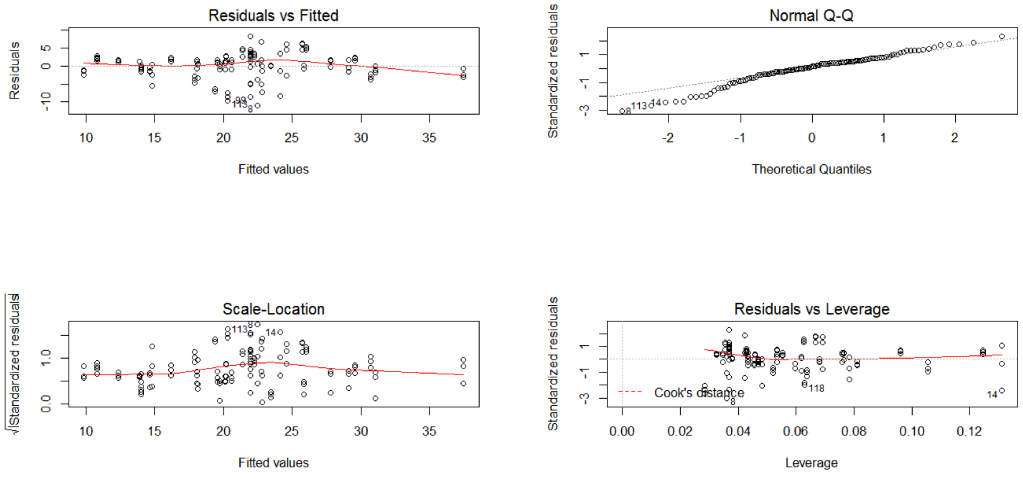


Figure 3.30: Box-Cox-analysis of `lm6_bal`.

But transforming the data as suggested by the Box-Cox-routine does not yield the desired result. Tables and sub-figures of figure 3.31 show, that the data has now approximately constant variance, but the error distribution is now consistently rated not being normal by both, the QQ-plot and the Shapiro-Wilk test. Therefore, we suggest to stick to the untransformed data in this case: the assumption of normally distributed errors seems to be fulfilled and the deviation from the assumption of homoscedasticity seems to be rather moderate.

```
> ld3 <- EoL^0.412
> lm6_trans<-lm(ld3~SoC + I(T^2) + I((ADC)^2) + I(SoC^2) + I(T^3) + SoC:dSoC)

> summary(lm6_trans)

Call:
lm(formula = ld3 ~ SoC + I(T^2) + I((ADC)^2) + I(SoC^2) + I(T^3) +
    SoC:dSoC)

Residuals:
     Min       1Q   Median       3Q      Max
-10.9306  -1.4641   0.5125   2.0839   8.3140

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  17.9766     1.1198  16.054  < 2e-16 ***
SoC          13.6249     3.1099   4.381 2.60e-05 ***
I(T^2)       -4.4793     0.7974  -5.618 1.35e-07 ***
I((ADC)^2)    4.9676     1.2485   3.979 0.000121 ***
I(SoC^2)      8.3692     1.0817   7.737 4.11e-12 ***
I(T^3)       -3.9514     0.4319  -9.150 2.33e-15 ***
SoC:dSoC     20.3094     3.4999   5.803 5.77e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.675 on 116 degrees of freedom
Multiple R-squared:  0.7359,    Adjusted R-squared:  0.7223
F-statistic: 53.88 on 6 and 116 DF,  p-value: < 2.2e-16
```



Shapiro-Wilk normality test	Non-constant Variance Score Test
W = 0.9583, p-value = 0.0007732	Chisquare = 1.563276 Df = 1 p = 0.2111862

Figure 3.31: Results of statistical analysis concerning `lm6_trans`.

The printout of the R routine `lm`, parametrizing model `lm10_bal` shows (figure 3.32), that all of the regressors in the model are highly significant. It consists of the regressors stated above (T^2, T^3, SoC, SoC^2, SoC:dSoC) and additionally of CC, ADC, ADC^2, CC:PDC and CC:dSoC.

```
> lm10_bal<-lm(EoL~ CC + ADC + SoC +I(T^2)+ I((ADC)^2)+ I(SoC^2)+ I(T^3) + CC:PDC +
                   CC:dSoC+ SoC:dSoC)
> summary(lm10_bal)

Call:
lm(formula = EoL ~ CC + ADC + SoC + I(T^2) + I((ADC)^2) + I(SoC^2) +
    I(T^3) + CC:PDC + CC:dSoC + SoC:dSoC)

Residuals:
     Min       1Q   Median       3Q      Max
-1699.57  -394.96    58.07   389.94  1856.53

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1766.09     280.34   6.300 6.03e-09 ***
CC           -469.27     113.22  -4.145 6.64e-05 ***
ADC          3216.82     845.73   3.804 0.000233 ***
SoC          2553.04     586.04   4.356 2.94e-05 ***
I(T^2)       -809.43     151.74  -5.334 5.05e-07 ***
I((ADC)^2)   3708.67     740.12   5.011 2.04e-06 ***
I(SoC^2)     2152.40     224.44   9.590 2.92e-16 ***
I(T^3)       -699.41      83.76  -8.350 2.03e-13 ***
CC:PDC       -403.73     124.59  -3.241 0.001571 **
CC:dSoC      -470.26     125.10  -3.759 0.000273 ***
SoC:dSoC     4021.03     658.26   6.109 1.49e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 671.5 on 112 degrees of freedom
Multiple R-squared:  0.7905,    Adjusted R-squared:  0.7718
F-statistic: 42.27 on 10 and 112 DF,  p-value: < 2.2e-16
```

Figure 3.32: Printout of the R routine `lm`, when parametrizing model `lm10_bal` using the balanced data.

By looking at the diagnostic-plots (figure 3.33), we can again observe that the QQ-plot indicates a slight deviation from the assumption of normally distributed errors and that the scale-location plot does also allow for doubting the assumption of homoscedasticity. Again, in order to enhance diagnostics, we apply some statistical tests (table 3.9) and get essentially the same results as before.

Shapiro-Wilk normality test	Non-constant Variance Score Test
W = 0.9813, p-value = 0.08573	Chisquare = 4.951782 Df = 1 p = 0.02606379

Table 3.9: Results of Shapiro-Wilk- and ncv-test in the case of `lm10_bal`.

The Shapiro-Wilk normality test does again not reject the assumption of normally distributed errors, but showing a smaller p-value of 0.086 now compared to the p-value of

0.22 in the case of `lm6_bal`. The ncv-test again rejects the assumption of homoscedasticity (also the p-value of 0.026 is smaller compared to the p-value of 0.035 in the case of `lm6_bal`). Again, the Box-Cox-routine suggests a transformation with λ equal to 0.4.

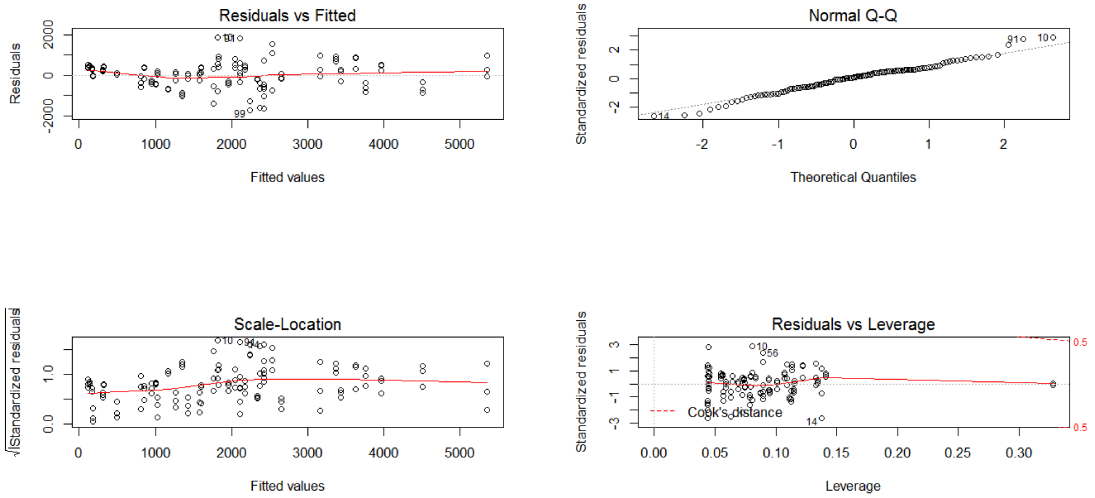


Figure 3.33: Diagnostic-plots in the case of `lm10_bal`.

But, as before, transforming the data as suggested by the Box-Cox-routine does not yield the desired result. Again, the homoscedasticity of the data is improved, but the error distribution is again consistently rated not being normal by the QQ-plot as well as the Shapiro-Wilk test. Therefore, we again recommend to stick to the untransformed data.

In addition we did a leaps-based model-selection using transformed data, to investigate the topic of data-transformation by exponentiation of the response with λ as suggested by the Box-Cox-routine.

Applying the Box-Cox-routine to `lm.ex` yields a λ of 0.265 for transforming the response in advance. Using the same strategy as described above (see figures 3.34 & 3.35), the outcome of this model-selection process is model `lm11_trans` consisting of 11 regressors which are the same as in the case of `lm10_bal` plus dSoC (see figure 3.36).

Looking at the diagnostic-plots and statistics respectively (figure 3.36), it is obvious, that we encounter in this case the same problems with the transformation suggested by the Box-Cox-routine, as we had in the case of `lm6_bal` and `lm10_bal`. Therefore, we conclude, that transforming the data in this case is not an appropriate remedy for solving the problem of slight deviations from homoscedasticity in the statistical analysis of models `lm6_bal` and `lm10_bal`. Hence, it is reasonable to stick to the models found in the context of untransformed data.

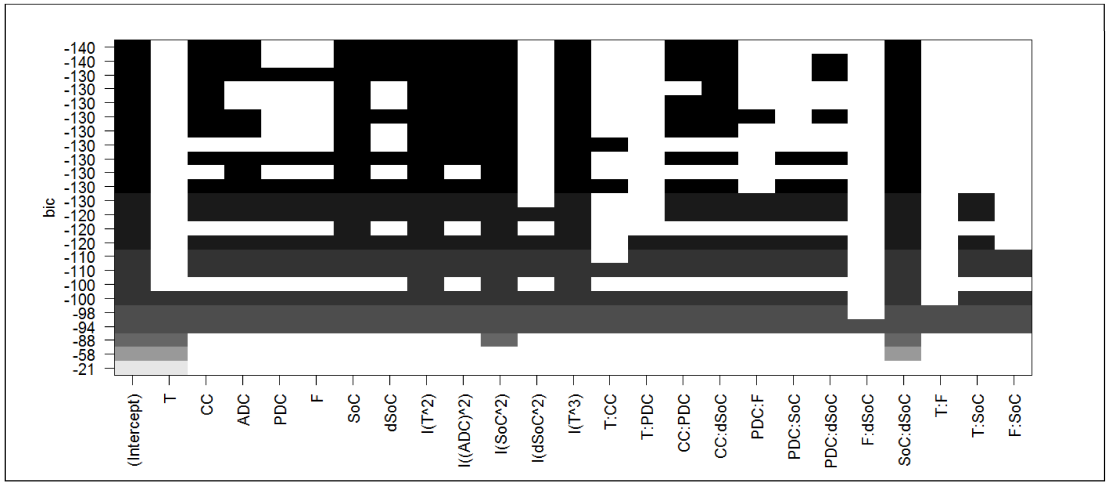


Figure 3.34: Output of the R routine leaps, when doing an all subsets regression using the balanced and transformed data ($\lambda = 0.265$).

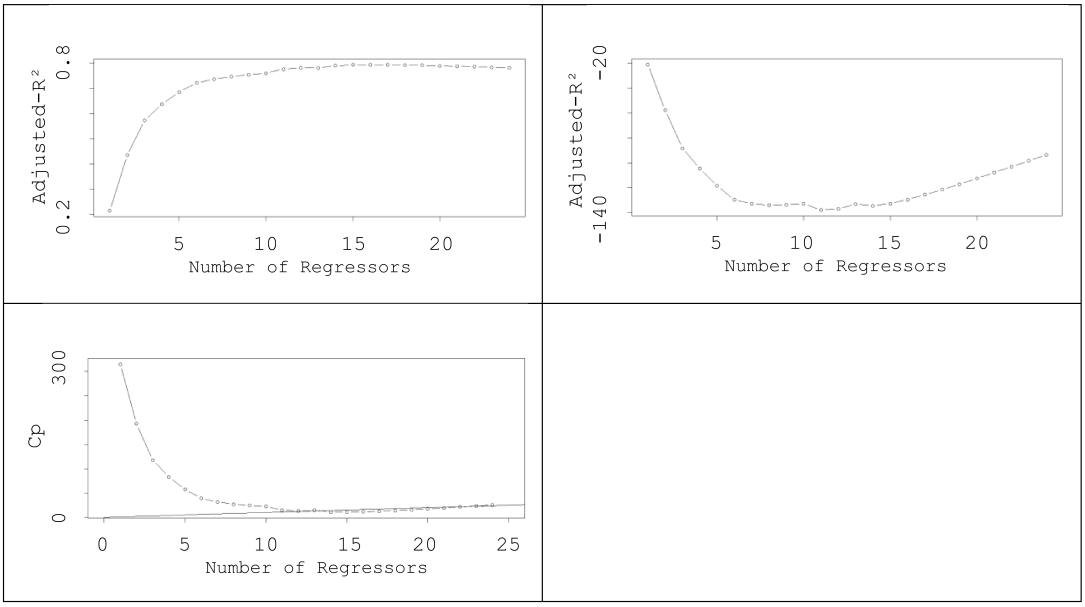
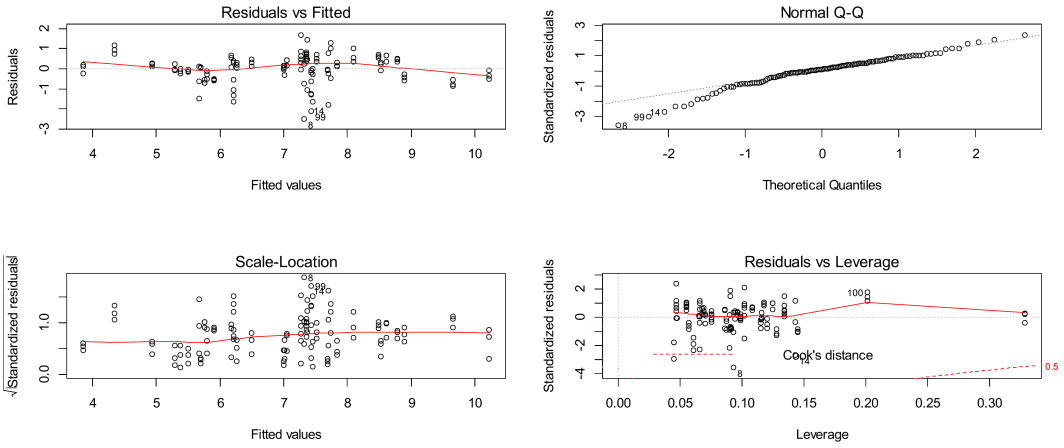


Figure 3.35: Selection criteria vs. model-size.

```
> ld3 <- EoL^0.265
>
> lm11_trans<-lm(ld3~CC + ADC + SoC + dSoC + I(T^2) + I((ADC)^2) + I(SoC^2) +
                      I(T^3) + CC:PDC + CC:dSoC + SoC:dSoC)
>
> summary(lm11_trans)

Call:
lm(formula = ld3 ~ CC + ADC + SoC + dSoC + I(T^2) + I((ADC)^2) + I(SoC^2) + I(T^3)
                      + CC:PDC + CC:dSoC + SoC:dSoC)

Residuals:
    Min      1Q  Median      3Q     Max
-2.4934 -0.3020  0.0794  0.4471  1.6891

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.49227    0.31476  20.626  < 2e-16 ***
CC          -0.77949    0.13827  -5.637 1.33e-07 ***
ADC          2.90993    1.03800   2.803  0.00597 **
SoC          1.93054    0.64689   2.984  0.00350 **
dSoC        -0.49521    0.15182  -3.262  0.00147 **
I(T^2)      -0.84643    0.16505  -5.128 1.25e-06 ***
I((ADC)^2)   3.75884    0.90163   4.169 6.09e-05 ***
I(SoC^2)     1.68104    0.28606   5.877 4.47e-08 ***
I(T^3)      -0.71974    0.09462  -7.607 9.81e-12 ***
CC:PDC      -0.49013    0.14038  -3.492  0.00069 ***
CC:dSoC     -0.71853    0.15627  -4.598 1.14e-05 ***
SoC:dSoC     3.19715    0.72619   4.403 2.47e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.73 on 111 degrees of freedom
Multiple R-squared:  0.7936,    Adjusted R-squared:  0.7732
F-statistic:  38.8 on 11 and 111 DF,  p-value: < 2.2e-16
```



Shapiro-Wilk normality test	Non-constant Variance Score Test
data: rstudent(lm11_trans) W = 0.9503, p-value = 0.0001883	Variance formula: ~ fitted.values(lm11_trans) Chisquare = 0.8134279 Df = 1 p = 0.3671087

Figure 3.36: Results of statistical analysis concerning `lm11_trans` when using the balanced and transformed data ($\lambda = 0.265$) for parametrization.

Additional regressors

In order to investigate whether additional regressors, i.e interactions not included in `lm.ex` might be of importance we expanded the list of potential regressors by allowing many more potential interactions, especially those of the type A:B^2, with A and B denoting any of the 7 influential factors except frequency. We have already seen in the foregoing model-selection process based on the regressors occurring in `lm.ex`, that F does not seem to play any role in the modeling of the aging process. In addition, in case of temperature T, also interactions of the type A:T^3 are considered. This strategy yields 64 potential regressors in total (table 3.10).

```
T , CC , ADC , PDC , SoC , dSoC , I(T^2) , I(CC^2) , I((ADC)^2) , I((PDC)^2) ,
I(SoC^2) , I(dSoC^2) , I(T^3) , T:CC , T:PDC , T:ADC, , T:SoC, T:dSoC, T:I((ADC)^2)
, T:I(SoC^2) , T:I(dSoC^2) , T:I(CC^2) , T:I((PDC)^2) , CC:PDC , CC:dSoC , CC:ADC ,
CC:SoC , CC:I(T^2) , CC:I((ADC)^2) , CC:I(SoC^2) , CC:I(dSoC^2) , CC:I(T^3) ,
CC:I((PDC)^2), ADC:PDC ,ADC:SoC ,ADC:dSoC ,ADC:I(T^2) ,ADC:I(SoC^2), ADC:I(dSoC^2)
, ADC:I(T^3) , ADC:I(CC^2) , ADC:I((PDC)^2), PDC:SoC , PDC:dSoC , PDC: I(T^2) ,
PDC:I((ADC)^2), PDC:I(SoC^2) , PDC:I(dSoC^2) , PDC:I(T^3) , PDC:I(CC^2) , SoC:dSoC
, SoC:I(T^2) , SoC:I((ADC)^2), SoC:I(dSoC^2) , SoC:I(T^3) , SoC:I(CC^2) ,
SoC:I((PDC)^2) , dSoC: I(T^2) , dSoC:I((ADC)^2) , dSoC:I(SoC^2) , dSoC:I(T^3) ,
dSoC:I(CC^2) , dSoC:I((PDC)^2)
```

Table 3.10: Enlarged set of regressors.

Due to the large number of potential regressors, new challenges emerge. Now it is not possible to do an all-subsets regression using the R-package `leaps`. `Leaps` is, as already mentioned using a branch-and-bound algorithm, which is capable of dealing with at most around 40 regressors. Beyond 40 regressors, the computational effort becomes too large to be handled. Therefore, we decided to use a heuristic strategy:

- We first perform a for- and a backward regression based on the set of 64 regressors.
- The regressors, which emerge this way, are then added to the initial set of 24 regressors comprised by `lm.ex`. From this new set of regressors, all regressors with the frequency term are removed.
- On this new set of regressors, which does fortunately contain fewer than 40 regressors, again an all-subsets regression is performed.

But, it is almost impossible to find the 'best' model this way. Due to the fact that this model selection is not based on an exhaustive procedure, it is very likely that many more well performing models exist which perform even better in terms of the criteria used for model selection.

The result of the exhaustive search (figure 3.37) indicates that a model with around 16 regressors will be of sufficient complexity. This model is called `lm16_heu` (figure 3.38) and a brief discussion of its properties is given in the following.

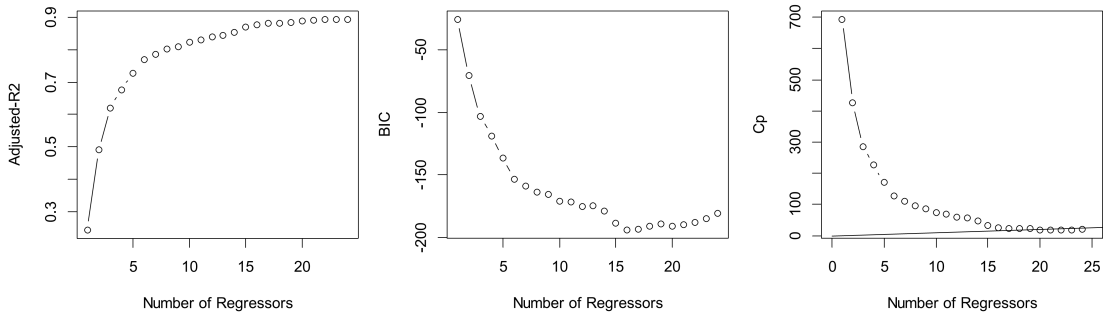


Figure 3.37: Selection criteria concerning the all-subsets-regression based on the new set of regressors.

```
lm16_heu<-lm(EoL~SoC+I(T^2)+I(SoC^2)+I(T^3)+I(CC^2)+T:CC+PDC:SoC+SoC:dSoC+
                 CC:ADC+CC:SoC+PDC:I((ADC)^2)+PDC:I(T^3)+SoC:I(dSoC^2)+
                 SoC:I(CC^2)+dSoC:I((ADC)^2)+dSoC:I((PDC)^2))

summary(lm16_heu)

Call:
lm(formula = EoL ~ SoC + I(T^2) + I(SoC^2) + I(T^3) + I(CC^2) +
    T:CC + PDC:SoC + SoC:dSoC + CC:ADC + CC:SoC + PDC:I((ADC)^2) +
    PDC:I(T^3) + SoC:I(dSoC^2) + SoC:I(CC^2) + dSoC:I((ADC)^2) +
    dSoC:I((PDC)^2))

Residuals:
    Min      1Q  Median      3Q     Max
-1689.6  -176.8    36.1   210.3  1800.8

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)      1043.90     205.73   5.074 1.67e-06 ***
SoC             12485.88     913.33  13.671  < 2e-16 ***
I(T^2)          -1006.13     128.92  -7.804 4.47e-12 ***
I(SoC^2)         2111.13     223.06   9.464 9.01e-16 ***
I(T^3)          -1026.04     109.84  -9.341 1.71e-15 ***
I(CC^2)          1166.64     284.44   4.102 8.07e-05 ***
T:CC             -710.11     113.80  -6.240 9.16e-09 ***
SoC:PDC          -459.86     153.38  -2.998 0.003384 **
SoC:dSoC        22895.42    1751.29  13.073  < 2e-16 ***
CC:ADC           -496.70     106.59  -4.660 9.24e-06 ***
SoC:CC           1146.93     204.14   5.618 1.57e-07 ***
PDC:I((ADC)^2)   -887.04     128.38  -6.909 3.76e-10 ***
I(T^3):PDC        360.25      99.25   3.630 0.000438 ***
SoC:I(dSoC^2)   13987.36    1273.02  10.988  < 2e-16 ***
SoC:I(CC^2)     -4612.05     771.95  -5.975 3.13e-08 ***
dSoC:I((ADC)^2) -1872.89     359.94  -5.203 9.63e-07 ***
dSoC:I((PDC)^2)  2247.59     397.91   5.648 1.37e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 491.6 on 106 degrees of freedom
Multiple R-squared:  0.8937,    Adjusted R-squared:  0.8777
F-statistic: 55.72 on 16 and 106 DF,  p-value: < 2.2e-16
```

Figure 3.38: Printout of the R-routine `lm` in the case of `lm16_heu`.

The printout of the R-routine `lm` in the case of `lm16_heu` (figure 3.38) tells us, that all of the regressors comprised by the model are highly significant and that the rather high adjR^2 value of almost 0.88 does not differ much from the multiple R^2 value of 0.894. This means that most of the variability in the data is explained by `lm16_heu` and that there is almost no redundancy in the model. The diagnostics show (3.39 & table 3.11), that the problem of not meeting the basic assumptions of regression analysis is even more severe now compared to the situation in the case of `lm6_bal` and `lm10_bal`. The Box-Cox-routine suggests transforming the data by exponentiation of the response with $\lambda = 0.65$ (figure 3.40).

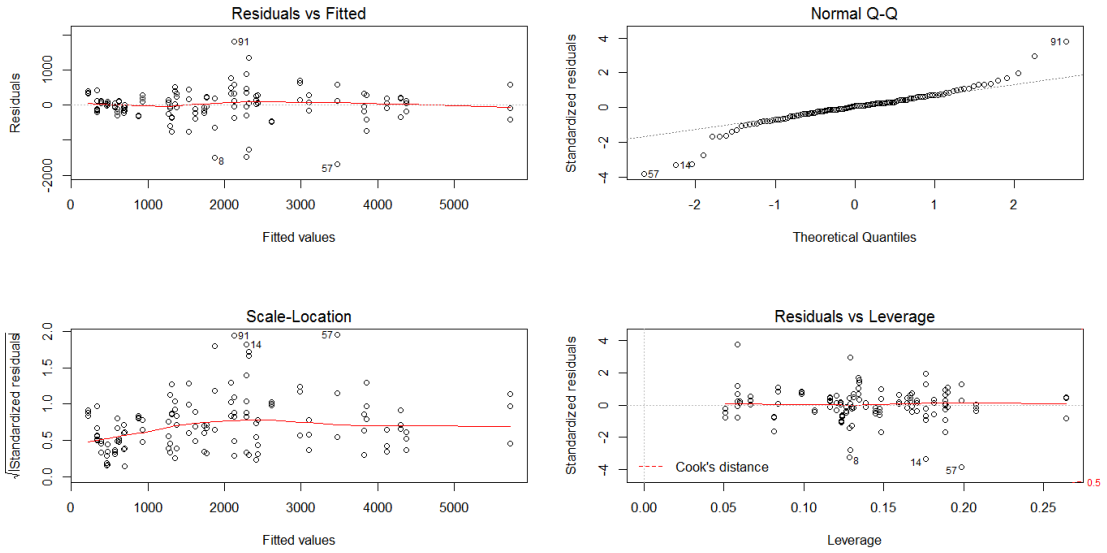


Figure 3.39: Diagnostic plots in the case of `lm16_heu`.

```
shapiro.test(rstudent(lm16_heu))

        Shapiro-Wilk normality test

data:  rstudent(lm16_heu)
W = 0.8963, p-value = 9.561e-08
```

```
ncvTest(lm16_heu)

Non-constant Variance Score Test

Variance formula: ~ fitted.values
Chisquare = 6.505041    Df = 1     p = 0.01075691
```

Table 3.11: Testing for normally and homoscedastic distributed errors in the case of `lm16_heu`.

But again, applying this transformation to the data constitutes no remedy (figure 3.41 & table 3.12) and therefore, even more than before, the question remains whether this type of modeling should be used at all.

The result of this strategy is a model of higher complexity than encountered in the context of the regressors comprised by `lm.ex`. It consists of 16 instead of at most 10 regressors. This means, that the model fit is better, but also raises the question of

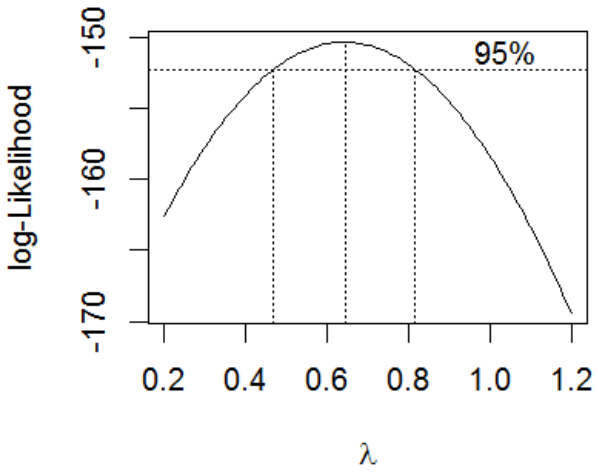


Figure 3.40: Box-Cox-analysis of `lm16_heu`.

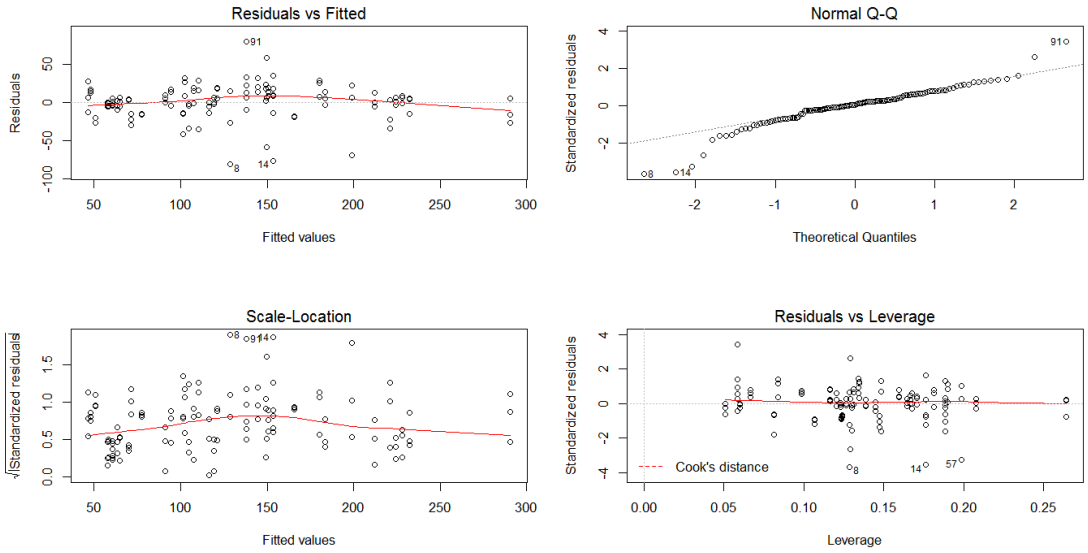


Figure 3.41: Diagnostic plots in the case of `lm16_heu` parametrized using the transformed data as suggested by the Box-Cox-routine.

shapiro.test(rstudent(lm16_heu))	ncvTest(lm16_heu)
Shapiro-Wilk normality test	Non-constant Variance Score Test
data: rstudent(lm16_heu) W = 0.9144, p-value = 8.852e-07	Variance formula: ~ fitted.values Chisquare = 2.824197 Df = 1 p = 0.09285333

Table 3.12: Testing for normally and homoscedastic distributed errors in the case of `lm16_heu` parametrized using the transformed data as suggested by the Box-Cox-routine.

overfitting. A behaviour, which can indeed be observed, when applying this model to the validation points (see next subsection for details).

Validation of the chosen models by examining the validation points

Validation points are additional experimental conditions applied to at least three different cells each. The data gathered this way is not used for parametrizing the models, but instead used for testing the models quality by comparing the models prediction for these experimental conditions with the extrapolated life spans calculated for these load-points.

The results of extrapolation for the validation-points is depicted in figure 3.42. Since the validation-points V10, V11 and V12 have been established recently, i.e. the respective cells are not part of the experiment from the beginning, it remains questionable, whether it makes sense to already use this data because there is some kind of stabilization-phase at the beginning of most of the runs, as can be seen when cells are tested for longer periods. For cells under the regimes of V02, V03 and V04 we can find longer times of operation, i.e. there are cells which are running since hundreds of days, therefore giving a better impression of the evolution of the extrapolations which are, as already discussed (see page 68 ff.) changing with time. By looking at the respective subfigures of 3.42, we can see, that there is at least one cell per load-point, which is not as long part of the experiment as the other cells belonging to this load-point. This problem is handled as described before (see page71 ff.).

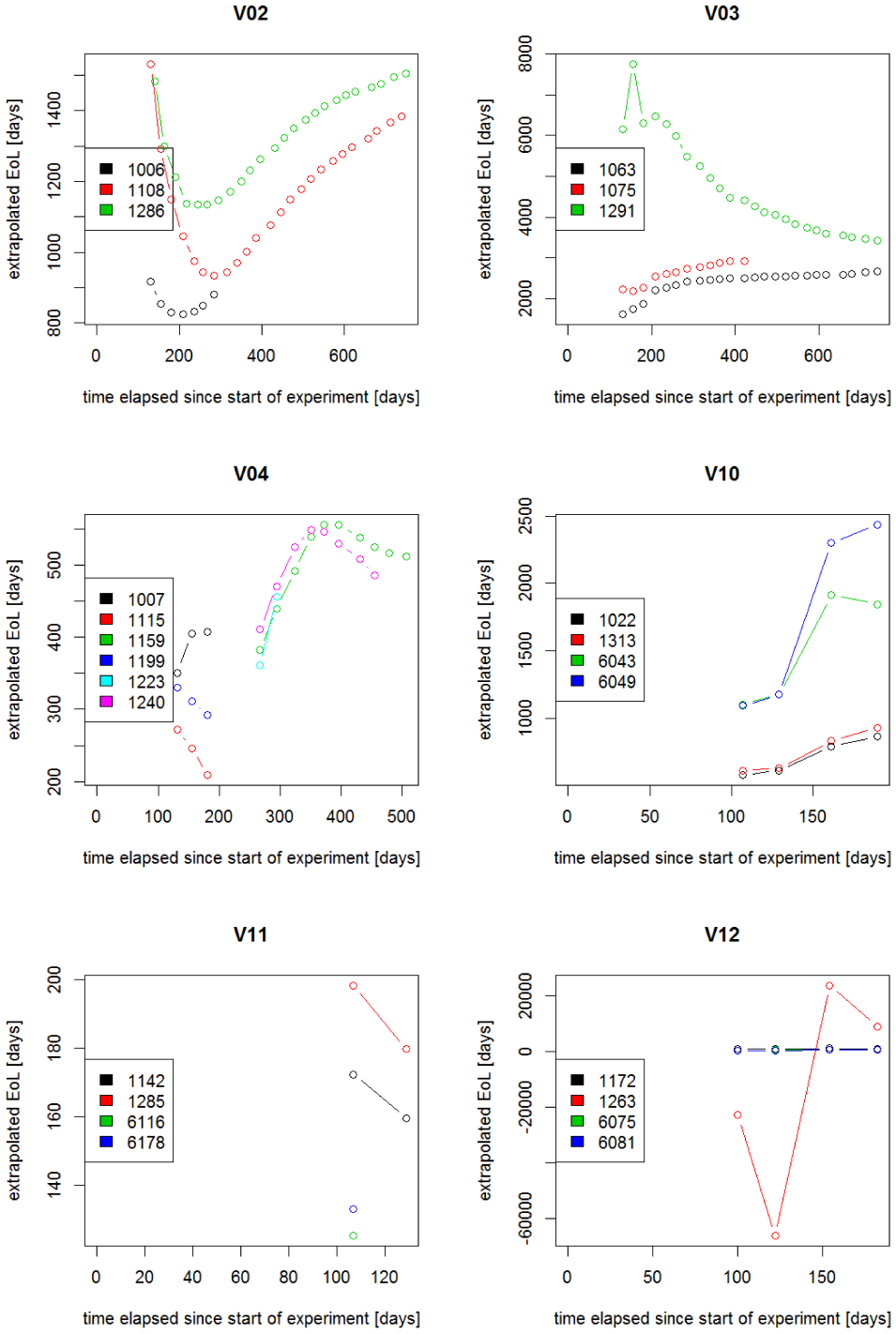


Figure 3.42: Evolution of extrapolation in the case of the validation points.

Table 3.13 shows the mean values of the extrapolated lifespans and the lifespans predicted by the three models of interest, `lm6_bal`, `lm10_bal` and `lm16_heu`.

First, let us consider `lm6_bal` and `lm10_bal`, which show a different behaviour compared to `lm16_heu`.
In both cases, the models give reasonable results for load-points resulting in a moderate or rather slow aging (extrapolated lifespans between 1233 and 3141 days), but giving rather poor results in the case of the harsher load-points (extrapolated lifespans between 149 and 695 days). In all cases, except in the case of V12, both models are underestimating the extrapolations, showing even negative values in the case of the harshest load-point V11. Additionally, `lm10_bal` also gives a negative prediction in the case of the load-point causing the second shortest lifespan, V04, while `lm6_bal` gives a rather reasonable result in this case. In the case of V12, both models are overestimating the extrapolation by a factor of three, but due to the short time of operation of the cells tested under the regime of this load-point, it might be, that the extrapolation is not very reliable. Despite the fact, that negative predicted life-spans are void of meaning, both models seem to be able to discriminate between harsh and moderate load-points. It is for sure better, to use a model giving negative predictions in the case of harsh load-points and is apart from that behaving well, than to use a model which doesn't identify harsh conditions as such.

In the case of `lm16_heu`, the situation is different. This model is not able to distinguish between harsh and moderate load-points. Summing up the models' behaviour it can be said, that literally all predictions in the case of the validation points seem to be absolutely useless. This result, at least in our opinion, indicates overfitting. A result, which is somewhat surprising in the light of the criteria used for model selection (figure 3.37) and the printout of the R-routine `lm` (figure 3.38), because all criteria unambiguously suggest a model of such a size and the printout shows that all of the regressors chosen are highly significant. This is a good example for the complexity of modeling such a relationship and it tells, that just relying on a few criteria might yield a poor result. Therefore, it is of great importance to incorporate as many ways of model evaluation as possible (another way of evaluating the models is applying them to real-world driving cycles as done by the involved chemist).

load-point / lifespan/days	V02	V03	V04	V10	V11	V12
extrapolated	1233	3141	394	1521	149	695
`lm6_bal`	929	2543	238	1320	-202	1969
`lm10_bal`	857	2728	-90	1405	-1348	1725
`lm16_heu`	-259	-132	-1497	-2025	1513	3139

Table 3.13: Extrapolated and predicted lifespans.

Accuracy of the models

To get an impression concerning the accuracy of the models, the plots 3.43 and 3.44 compare the magnitude of the predicted values and the associated confidence intervals.

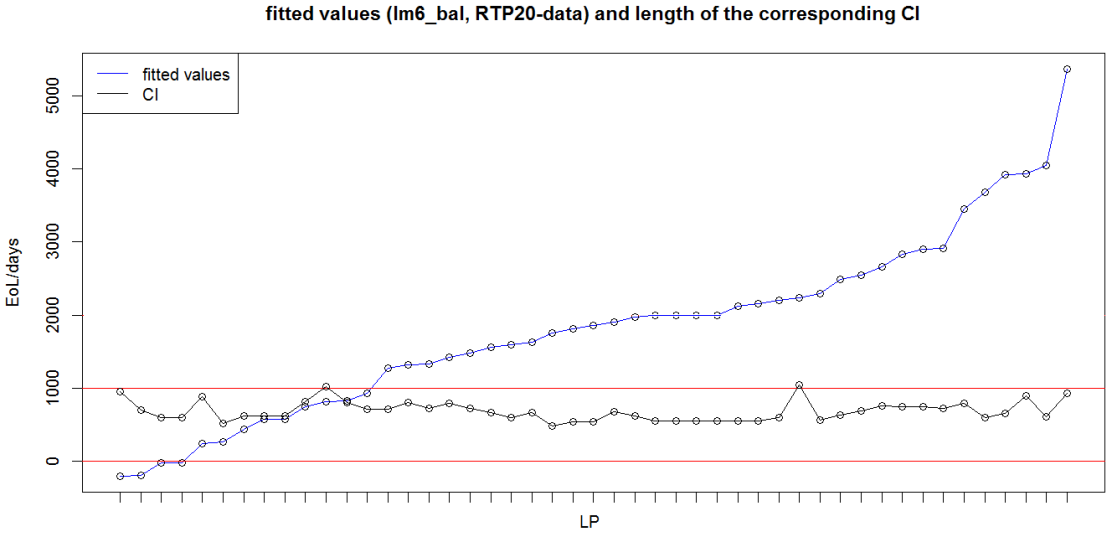


Figure 3.43: Predicted values and length of CI in the case of `lm6_bal`.

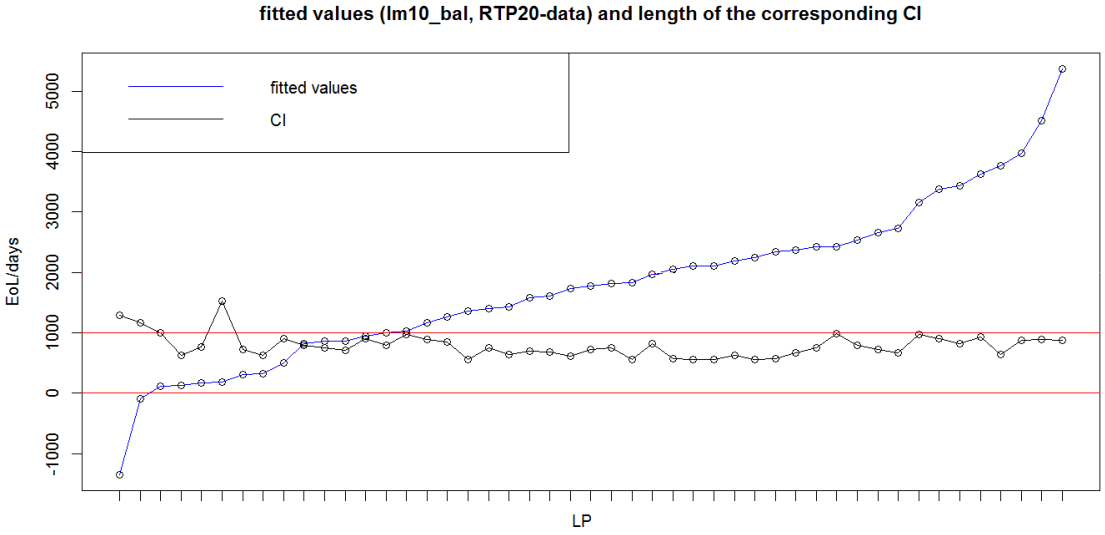


Figure 3.44: Predicted values and length of CI in the case of `lm10_bal`.

We can see that the confidence intervals are rather wide. This might, at least partly, be caused by the rather low D-efficiency of the design in use (see page 50 for further details).

Despite this fact, there are a lot of means of the observations not covered by the confidence intervals of the predictions (figures 3.45 & 3.46). Still in most cases, both models catch the trend rightly, but the accuracy of model `lm6_bal` is not as high as that of `lm10_bal`. It is obvious, that `lm6_bal` is over- and also underestimating the extrapolations somewhat more often severely than `lm10_bal`.

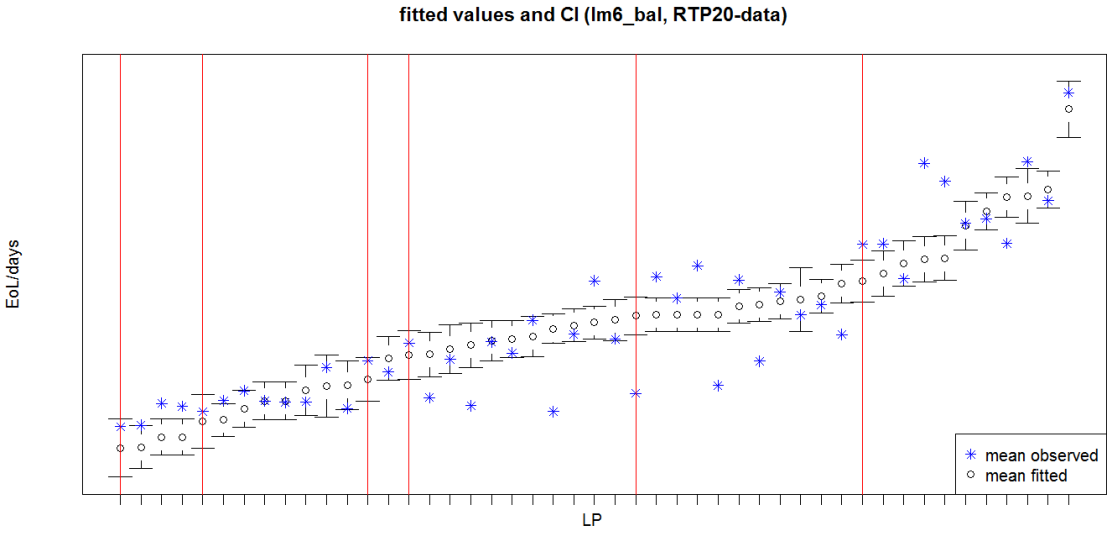


Figure 3.45: Coverage of the observed means by the confidence intervals of the predictions in the case of `lm_6` (vertical lines indicate validation points).

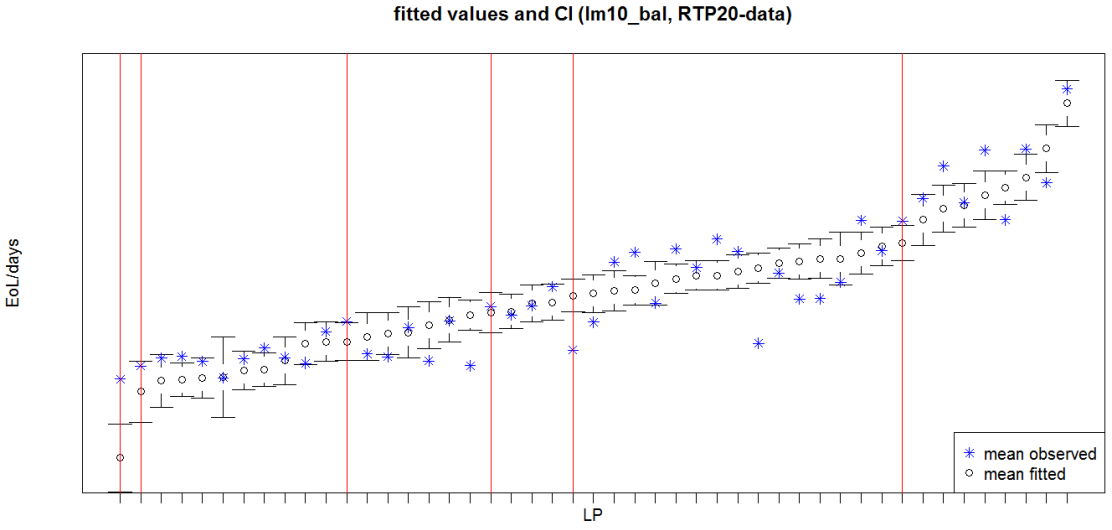


Figure 3.46: Coverage of the observed means by the confidence intervals of the predictions in the case of `lm10_bal` (vertical lines indicate validation points).

Looking at the residuals of the different models (figures 3.47 & 3.48), the overall picture is that `lm6_bal` and `lm10_bal` do not differ very much in most cases.

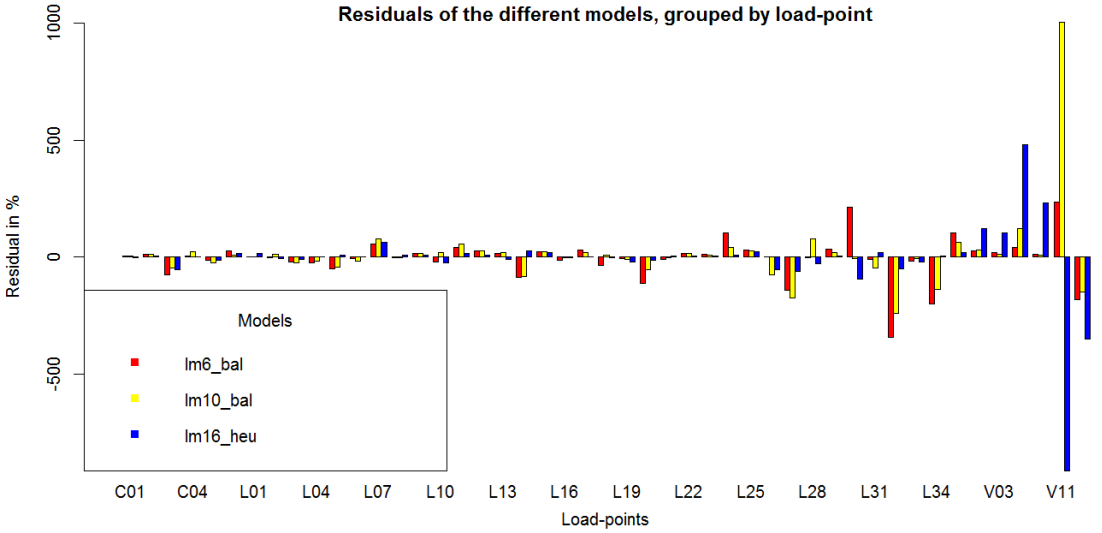


Figure 3.47: Residuals of different models given in percent of the according extrapolation.

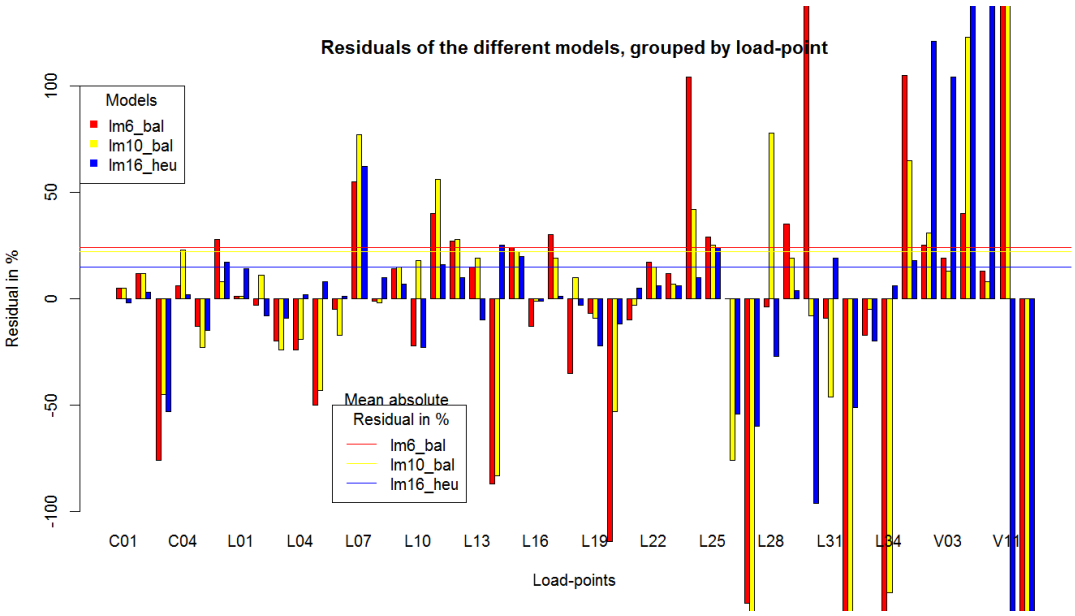


Figure 3.48: Residuals of different models given in percent of the according extrapolation and truncated at $\pm$ 150%.

But, leaving the validation points aside and taking a closer look, it can also be seen that the residuals of `lm6_bal` are especially in the case of great deviations by trend larger

than those of `lm10_bal`. Interestingly, the means of the absolute values of the residuals do not differ much - the more complex model `lm10_bal` shows only a slightly smaller value than the less complex model `lm6_bal`. In the case of the most complex model, `lm16_heu`, we have much smaller residuals in the case of load points C01 to L35, which are the ones used for model parametrization. Also the mean of the absolute values of the residuals is significantly smaller than in the case of the less complex models, but `lm16_heu` performs worse in the case of the validation points compared to the other models.

The next figure shows a comparison of the mean values of the lifespan-extrapolations (ordered by size) with the predictions according to the models `lm6_bal`, `lm10_bal` and `lm16_heu`. We can see again, that `lm16_heu` is not able to cope with the validation points (indicated by vertical lines) whilst showing a good fit in the case of the load points used for model parametrization. For the other two models, there is a somewhat contrary situation - these models perform much more reasonable concerning the validation points but show more severe deviations in the case of the load points used for parametrization.

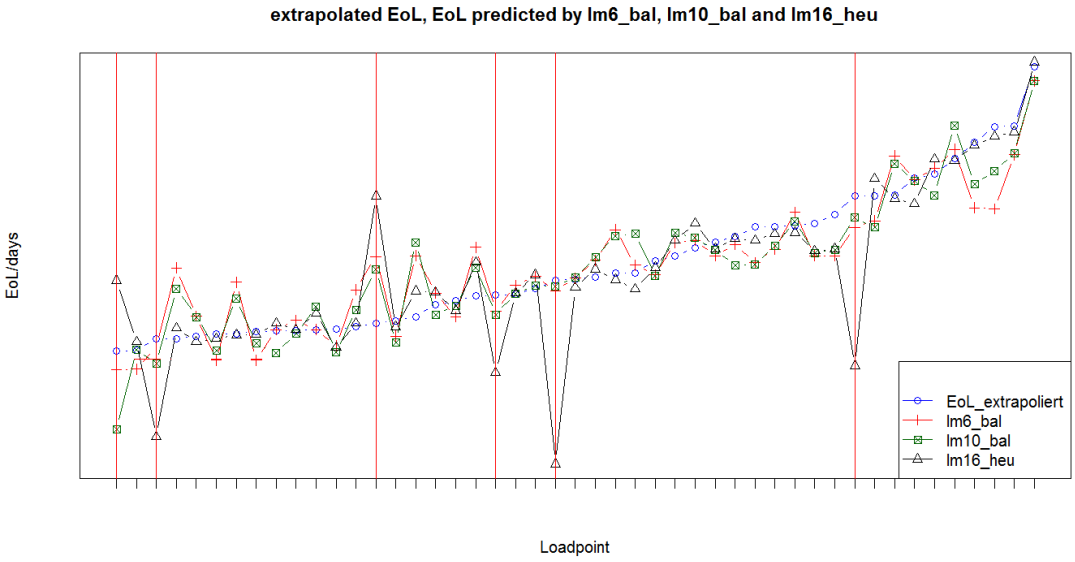


Figure 3.49: Comparison of extrapolated EoL and EoL predicted by `lm6_bal`, `lm10_bal` and `lm16_heu`.

All in all, due to the fact, that `lm16_heu` shows a devastating performance in case of the validation points and is in this context totally outperformed by both models of lower complexity, we would not suggest to consider `lm16_heu` as a suitable candidate for modeling the aging of the Li-ion cells. Since `lm6_bal` and `lm10_bal` have, in many cases, equal strengths and weaknesses but `lm6_bal` is, as discussed above, in some load points inferior to `lm10_bal` (it seems, that `lm6_bal` might be of too low complexity to be able to cope with as many different operating conditions as appearing in our setting),

we would suggest to consider `lm10_bal` to be the most promising, albeit not perfect candidate for modeling cell-aging in the context of multiple linear regression.

4 Summary

Design of Experiment

One of the goals of DOE is to choose the design matrix X such that the variability of the estimators is as low as possible given a certain error associated with the underlying stochastic process (i.e. to optimize the efficiency of the design). The errors of the estimators' distributions depend not only on the error of the stochastic process but also on the design matrix X.

The constructed design is suitable only for a given design space, a specific model and a certain number of design points. Therefore, one has to be careful when altering the design or the model. As shown in section 3.1, only a few changes concerning the design can have a devastating impact on the designs D-efficiency - its lower bound can even drop to almost zero. In addition, growing information about the relationship between in- and output gathered during the ongoing experiment, makes a revision of the proposed model necessary before the end of the experiment is reached. As mentioned before, a new model needs a new design for achieving the data quality needed for proper parametrization.

Therefore, if changes of the design are inevitable (due to technical and/or economical reasons) and/or modifications of the model are inevitable, an accompanying statistical monitoring is highly recommended. This is the only way to keep the benefit of the experiment as high as possible. Modifying the experiment without statistical supervision can cause poor data quality and hence dissatisfactory experimental results. How such a design revision works and what remedies exist in order to restore a design possessing a satisfactory degree of efficiency, is also illustrated in 3.1 in form of two different strategies. Both ways have in common, that additional experiments are needed. Obviously, one will also in this case have to find a balance between constraints concerning the available resources and the increase of the designs D-efficiency.

Statistical Modeling of the Relationship between Cell Usage and Cell Aging

At the beginning, some preliminary questions have to be answered concerning the experimental data, in order to establish a reasonable model describing this relationship.

First, the question which cell characteristics constitute a proper indicator of cell aging has to be answered. It turns out, that the capacity decay seems to be the ideal candidate for assessing a cell's status of aging (see section 3.2). Since the intended response is the cells' end of life, it was decided to model the relationship between cell usage (represented by 7 different factors, namely: temperature, charge current, average discharge current,

peak discharge current, frequency, state of charge and delta state of charge - see chapter 1 for more details) and the time needed for the cells to reach 70% of their initial capacity. In addition it is necessary to establish a method for extrapolating the time needed for reaching this state, because a significant number of cells does not reach their end of life according to the afore-mentioned criterion until the experiment is terminated. We decided to use a simple exponential model which is described in detail on page 66. By investigating the behavior of these extrapolations in the light of increasing information due to the progressive gain of data in the course of the experiment, we show that using this strategy yields reasonable, i.e. stable extrapolations with respect to the characteristic of the experimental settings (i.e. harsh conditions stay harsh conditions, intermediate conditions stay intermediate conditions and soft conditions stay soft conditions - see page 68 ff. for further details).

Second, it has to be decided which type of data to use for statistical analysis. The raw data of the experiment is unbalanced, because some of the cells did not work as expected. We decided to balance the data by using the strategy described on page 73 ff. in order to benefit from the ease of analyzing balanced data-sets. We also investigated, whether the impact of balancing on statistical analysis proves to be significant. This is done by applying the balancing strategy several times and comparing the results of the statistical analyses of the resulting data-sets balanced using different estimates. No such evidence was found and therefore we decided to use a randomly balanced data-set for further analysis, i.e. model selection and model parametrization.

Modeling Results

Using the set of 23 regressors, which have initially been selected (see page 4 ff.) yields two models of rather low complexity showing a reasonable performance: `lm6_bal` and `lm10_bal` (pages 83 & 86), comprising 6 and 10 regressors respectively. One significant finding of this modeling process is, that one of the factors considered influential, namely the frequency, does not prove to have any effect and thus can be neglected. Additionally, in order to investigate whether supplemental regressors, i.e regressors not included in the initial list of 23 regressors might be of importance, we expanded the list of potential regressors by allowing many more potential interactions. The result of these approach is a model of higher complexity comprising 16 regressors and denoted `lm16_heu` (page 91).
By assessing the goodness of these models, we tried to select the most reasonable candidate under the three models found. This is done by examining the validation points and assessing the accuracy of the models. The validation points are experimental conditions not used for modeling but instead used for testing the models' quality by comparing their predictions for these experimental conditions with the extrapolated life spans calculated for these load-points (table 3.13 on page 96). Both models, `lm6_bal` and `lm10_bal`, seem to be able to discriminate between harsh and moderate validation points, albeit the fact that they show negative predictions concerning some harsh load-points. In the case of `lm16_heu`, we can see a completely different picture. This model is not able to

distinguish between harsh and moderate validation points. Summing up this models' behaviour it can be said, that literally all predictions in the case of the validation points seem to be absolutely useless. This result, at least in our opinion, indicates overfitting (page 28 f.). It is for sure better, to use a model giving negative predictions in the case of harsh conditions and is apart from that behaving well, than to use a model which doesn't identify such conditions reliably.
The accuracy of the models is assessed by looking at the confidence intervals of the predictions and the magnitudes of the according residuals (page 97 ff.). It can be seen, that confidence intervals are rather wide. This might, at least partly, be caused by the rather low D-efficiency of the design in use (see page 50 for further details). Despite this fact, there are a lot of means of the observations not covered by the confidence intervals of the predictions. Still in most cases, both models catch the trend rightly, but the accuracy of model `lm6_bal` is not as high as that of `lm10_bal`. The residuals of `lm6_bal` are especially in the case of great deviations by trend larger than those of `lm10_bal`. The most complex model, `lm16_heu`, shows much smaller residuals concerning the load points used for model parametrization, but `lm16_heu` performs worse in the case of the validation points compared to the other models. All in all, due to the fact, that `lm16_heu` shows a devastating performance in case of the validation points and is in this context totally outperformed by both models of lower complexity, we would not suggest to consider `lm16_heu` as a suitable candidate for modeling the aging of the Li-ion cells. Since `lm6_bal` and `lm10_bal` have, in many cases, equal strengths and weaknesses but `lm6_bal` is, as discussed above, in some load points inferior to `lm10_bal` (it seems, that `lm6_bal` might be of too low complexity to be able to cope with as many different operating conditions as appearing in our setting), we would suggest to consider `lm10_bal` to be the most promising, albeit not perfect candidate for modeling cell-aging in the context of multiple linear regression.

5 Outlook

For conducting further experiments, the following issues should be kept in mind:

- It is highly recommended to revise the whole topic of test space construction since the problem of violated constraints became an issue in the course of the experiment.
- Due to the fact that only minor changes of the experimental setting (concerning the DoE and/or the model which should be parametrized) could lead to a drastic reduction of the efficiency of the experimental design two important points should be considered:
  1. The experiment should be statistically supervised throughout its whole duration, i.e. each change should be evaluated immediately in terms of the designs efficiency and all alterations negatively effecting the experimental setting should under all circumstances be avoided or at least be repaired.
  2. A dynamic experimental design (DDoE) is recommended in order to maintain a reasonable data quality level since it is likely that the model of interest is changing over the course of the experiment.
- One of the influential factors, namely the frequency (F), does not prove to play a significant role in explaining the relationship between battery usage and battery aging and can thus be neglected. This will likely improve the study of the effects of the remaining factors.
- We recommend using the most promising model `lm10_bal` as a starting point for design construction, since using transformed data and/or models of higher complexity do not prove to be a reasonable choice.

Bibliography

[1] H. Akaike. Information theory and an extension of the maximum likelihood principle. 2^{nd} *International Symposium on Information Theory (Petrov, B. N. and Csaki, F., eds.), Akademiai Kiado, Budapest*, pages 267–281, 1973.

[2] G. E. P. Box and D. R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2):211–252, 1964.

[3] M. N. Das and N. C. Giri. *Design and Analysis of Experiments.* Wiley Eastern Ltd., New Delhi, 1^{st} edition, 1979.

[4] J. de Leeuw. Introduction to akaike (1973) information theory and an extension of the maximum likelihood principle. *Breakthroughs in statistics, Springer-Verlag, London*, 1:599–609, 1992.

[5] L. Fahrmeir and A. Hamerle. *Multivariate statistische Verfahren.* Walter De Gruyter, Berlin, New York, 1^{st} edition, 1984.

[6] L. Fahrmeir, T. Kneib, and S. Lang. *Regression: Modelle, Methoden und Anwendungen.* Springer, Heidelberg, Dordrecht, London, New York, 2^{nd} edition, 2009.

[7] V. V. Fedorov. *Theory of Optimal Experiments.* Academic Press, New York, 1^{st} edition, 1972.

[8] G. M. Furnival and R. W. Wilson. Regressions by leaps and bounds. *Technometrics*, 16(4):499–511, 1974.

[9] C. J. Geyer. *Model Selection in R; Lecture Notes.* 2003.

[10] J. H. Goodnight. Tests of hypotheses in fixed effects linear models, 1978. SAS Technical Report R-101.

[11] N. Haselgruber. Sampling and Design of Large-Scale Life Time Experiments. PhD thesis, Graz University of Technology, 2007.

[12] A. Heredia-Langner, W. M. Carlyle, D. C. Montgomery, C. M. Borror, and G. C. Runger. Genetic algorithms for the construction of D-optimal designs. *Journal of Quality Technology*, 35:28–46, 2003.

[13] M. E. Johnson and C. J. Nachtsheim. Some guidelines for constructing exact D-optimal designs on convex design spaces. *Technometrics*, 25:271–277, 1983.

[14] B. Jørgensen. *The Theory of Linear Models.* Chapman & Hall, New York, London, 1^{st} edition, 1993.

[15] S. Konishi and G. Kitagawa. *Information criteria and statistical modeling.* Springer, New York, 1^{st} edition, 2008.

[16] S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86, 1951.

[17] R. C. Littell, J. F. Freund, and P. C. Spector. *SAS System for Linear Models.* SAS Institute Inc., Cary, 3^{rd} edition, 1991.

[18] V. B. Melas. *Functional approach to optimal experimental design.* Springer, New York, 1^{st} edition, 2006.

[19] R. K. Meyer and C. J. Nachtsheim. The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics*, 37:60–69, 1995.

[20] T. J. Mitchell. An algorithm for the construction of D-optimal experimental designs. *Technometrics*, 16:203–210, 1974.

[21] D. C. Montgomery. *Design and Analysis of Experiments: International Student Version.* Wiley, Singapore, 8^{th} edition, 2012.

[22] D. C. Montgomery, E. A. Peck, and G. G. Vining. *Introduction to Linear Regression Analysis.* Wiley, Hoboken, 5^{th} edition, 2012.

[23] H. Motulsky and A. Christopoulos. *Fitting Models to Biological Data Using Linear and Nonlinear Regression.* GraphPad Software Inc., San Diego, 1^{st} edition, 2003.

[24] G. Pregartner. Design-of-Experiment and Statistical Modeling of Large Scale Lithium Ion Cells. Master thesis, Graz University of Technology, 2012.

[25] F. Pukelsheim. *Optimal Design of Experiments.* Wiley & Sons, New York, 1^{st} edition, 1993.

[26] S. R. Searle. *Linear Models.* John Wiley & Sons, New York, 1^{st} edition, 1971.

[27] E. Stadlober. *Angewandte Statistik; Lecture Notes.* Institut für Statistik, Technische Universität Graz, 2012.

[28] E. Stadlober. *Versuchsplanung; Lecture Notes.* Institut für Statistik, Technische Universität Graz, 2014.

[29] D. Urban and J. Mayerl. *Regressionsanalyse: Theorie, Technik und Anwendung.* Springer Fachmedien Wiesbaden GmbH, Wiesbaden, 4^{th} edition, 2011.

[30] K. Velten. *Mathematical Modeling and Simulation: Introduction for Scientists and Engineers.* Wiley-VCH, Weinheim, 2008.

[31] A. Vezzini. Lithiumionen-Batterien als Speicher für Elektrofahrzeuge. *Bulletin SEV/AES*, (3), 2009. https://www.ti.bfh.ch/fileadmin/_migrated/content_uploads/2009-03-09_SEV_Bulletin_Lithiumionen-Batterien_als_Speicher_fuer_Eektrofahrzeuge__ganz.pdf.

[32] Virtual Vehicle. M18Presentation_v4.

[33] Virtual Vehicle. M12 semiannual report, 08/2013. Internal Report.

[34] Virtual Vehicle. E3T3_ZB-023-2013, 2013. Internal Report.

[35] Virtual Vehicle. Presentation E3T3, September 25^{th}, 2013.

[36] R. E. Wheeler. efficient.rounding.AlgDesign, 2004. URL http://www.r-project.org/. Last visited 2015-11-18.

[37] WIKIPEDIA. Arrhenius equation. URL https://en.wikipedia.org/wiki/Arrhenius_equation. Last visited 2015-11-18.

[38] WIKIPEDIA. Lithium-ion battery. URL https://en.wikipedia.org/wiki/Lithium-ion_battery. Last visited 2015-11-18.

[39] WIKIPEDIA. Lithium-Ionen-Akkumulator. URL https://de.wikipedia.org/wiki/Lithium-Ionen-Akkumulator. Last visited 2015-11-18.

[40] WIKIPEDIA. Overfitting. URL https://en.wikipedia.org/wiki/Overfitting. Last visited 2015-11-18.

[41] F. Yates. The design and analysis of factorial experiments. *Commonwealth Bureau of Soil Science, Technical Communication*, (35):1–95, 1937.