

DATA-DRIVEN PROGNOSTICS FOR CRITICAL
ELECTRONIC ASSEMBLIES AND
ELECTROMECHANICAL COMPONENTS

by

Lucas P. Kirschbaum



Submitted for the degree of
Doctor of Philosophy

INSTITUTE FOR SENSORS, SIGNALS AND SYSTEMS
SCHOOL OF ENGINEERING AND PHYSICAL SCIENCES
HERIOT-WATT UNIVERSITY

August 2022

The copyright in this thesis is owned by the author. Any quotation from the report or use of any of the information contained in it must acknowledge this report as the source of the quotation or information.

Abstract

The industrial digitalisation enables the adoption of robust, data-driven maintenance strategies that increase safety and reliability of critical assets such as electronics. And yet, an implementation of data-driven methods which primarily address the industrialisation of diagnostic and prognostic strategies is opposed by various, application specific challenges. This thesis collates such restricting factors encountered within the oil and gas industry, in particular for the critical electrical systems and components in upstream deep drilling tools. A fleet-level, tuned machine learning approach is presented that classifies the operational state (no-failure/ failure) of downhole tool printed circuit board assemblies. It supports maintenance decision making under varying levels of failure costs and fleet reliability scenarios. Applied within a maintenance scheme it has the potential to minimise non-productive time while increasing operational reliability. Likewise, a tailored and efficient deep learning data pipeline is proposed for a component-level forecast of the end of life of electromagnetic relays. It is evaluated using high resolution life-cycle data which has been collected as a part of this thesis. In combination with a failure analysis, the proposed method improves the prognostics capabilities compared to traditional methods which have been proposed so far in order to assess the operational health of electromagnetic relays. Two case studies underpin the need for tailored prognostic methods in order to provide viable solutions that can de-risk deep drilling operations. In consequence, the proposed approaches alleviate the pressure on current maintenance strategies which can no longer meet the stringent reliability requirements of upstream assets.

To my grandmother, Hildegard Cancik-Lindemaier.

Acknowledgements

I would like to thank my PhD supervisors Prof. Dr. David Flynn, Dr. Valentin Robu, and Prof. Dr. Jonathan Swingler for the opportunity to write this thesis and for their comprehensive supervision. I also thank my external supervisors Dr. Jens Bruns from Baker Hughes as well as Dr. Gulshan Singh for their expert advice on the topic of deep drilling. In particular, I want to thank Mark Leonard for his hands-on support in the development of the experimental platform and Donna Palmer for her guidance on all administrative aspects of this PhD. I thank both my examiners Dr. Stoyan Stoyanov and Dr. Wolf-Gerrit Fruh for taking the time to review my work.

I am grateful for the many enriching and stimulating discussions with my friends and colleagues Darius Roman, Ross Dickie, and Civan Kahraman. I would like to thank my dear friends Nimra Rasheed, Kimberly Lyon and Julián Tachella which always listened to my concerns. Finally, I would like to express my deepest appreciation to Lisa Kleinschmidt for her unconditional, relentless support and sympathy which I have always been able to rely on.

Of the many people who have supported me on this journey, I would like to express my gratitude to my family always reassuring and encouraging me throughout this endeavour.

Research Thesis Submission

Please note this form should be bound into the submitted thesis.

Name:	Lucas P. Kirschbaum		
School:	School of Engineering and Physical Sciences		
Version: <i>(i.e. First, Resubmission, Final)</i>	Final	Degree Sought:	Doctor of Philosophy

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

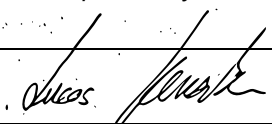
1. The thesis embodies the results of my own work and has been composed by myself
2. Where appropriate, I have made acknowledgement of the work of others
3. The thesis is the correct version for submission and is the same version as any electronic versions submitted*.
4. My thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
5. I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.
6. I confirm that the thesis has been verified against plagiarism via an approved plagiarism detection application e.g. Turnitin.

ONLY for submissions including published works

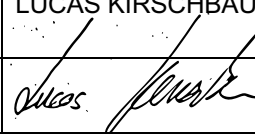
Please note you are only required to complete the Inclusion of Published Works Form (page 2) if your thesis contains published works)

7. Where the thesis contains published outputs under Regulation 6 (9.1.2) or Regulation 43 (9) these are accompanied by a critical review which accurately describes my contribution to the research and, for multi-author outputs, a signed declaration indicating the contribution of each author (complete)
8. Inclusion of published outputs under Regulation 6 (9.1.2) or Regulation 43 (9) shall not constitute plagiarism.

* Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.

Signature of Candidate:		Date:	15/08/2022
-------------------------	---	-------	------------

Submission

Submitted By <i>(name in capitals)</i> :	LUCAS KIRSCHBAUM
Signature of Individual Submitting:	
Date Submitted:	15/08/2022

For Completion in the Student Service Centre (SSC)

Limited Access	Requested	Yes	No	Approved	Yes	No
<i>E-thesis Submitted (mandatory for final theses)</i>						
Received in the SSC by <i>(name in capitals)</i> :				Date:		

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Research question	3
1.2.1	BHA-PCBA fleet-level	4
1.2.2	EMR component-level	5
1.3	Research contributions	6
1.4	Thesis structure	8
	References	9
2	Prognostics and health management	13
	Introduction	13
2.1	Reliability and maintenance strategies	13
2.1.1	Reactive maintenance	14
2.1.2	Preventive maintenance	15
2.2	Overview	15
2.2.1	Key concepts	16
2.2.2	Frameworks	17
2.2.3	Methods	18
2.2.4	Condition-based, predictive, and prescriptive maintenance	19
2.2.5	Challenges	20
2.3	Data-driven techniques	21
2.3.1	Machine learning and deep learning	23
2.4	Tailoring data-driven diagnostics and prognostics	24
2.4.1	Data	25
2.4.2	Feature extraction	26
2.4.3	Depth of implementation	28
2.4.4	Embedding	30
2.4.5	Explainability	30
2.4.6	Uncertainty	30
2.5	Survey of the digital transformation in the oil and gas industry	31
	Summary	34

References	34
3 Considered electronic assembly and electromechanical component	41
Introduction	41
3.1 BHA-PCBA and deep drilling	41
3.1.1 Synopsis of the deep drilling process	41
3.1.2 BHA and downhole data transfer	42
3.1.3 Failure modes of downhole tool electronics	44
3.1.4 Data-driven downhole tool maintenance strategies	46
3.2 EMR and electrical contacts	47
3.2.1 EMR operation principle	47
3.2.2 Failure modes	48
3.2.2.1 Contacts	49
3.2.2.2 Coil and mechanical parts	50
3.2.3 Contact resistance	51
3.2.3.1 Constriction resistance	51
3.2.3.2 Film resistance	53
3.2.4 Failure mechanisms	53
3.2.4.1 DC arc erosion	53
3.2.4.2 Contact welding	55
3.2.4.3 Contact contamination	56
3.2.4.4 Fretting	57
3.2.5 Contact operation	57
3.2.5.1 Contact making	57
3.2.5.2 Contact breaking	59
Summary	61
References	62
4 Data-Driven approaches to reliability and maintenance	65
Introduction	65
4.1 Strategies in the upstream oil and gas industry	65
4.1.1 Review of topical literature	65
4.1.2 Comparative assessment	69
4.2 Strategies for EMR operation and maintenance	72
4.2.1 Contact resistance	72
4.2.2 Derived reference DIs	72
4.2.2.1 Non-intrusive DIs	73
4.2.2.2 Intrusive DIs	74
4.2.3 Review of topical literature	75
4.2.3.1 General DC EMR	75
4.2.3.2 Aviation and Aerospace DC EMR	77

4.2.3.3	High-Voltage DC EMR	79
4.2.3.4	Railway EMR	79
4.2.3.5	Automotive EMR	80
4.2.3.6	Nuclear EMR	81
4.2.3.7	Storage of EMR	81
4.2.4	Comparative assessment	83
	Summary	85
	References	85
5	Developed data-driven methods	92
	Introduction	92
5.1	BHA-PCBA maintenance support framework	92
5.1.1	Failure classification using machine learning	93
5.1.2	Data sources	93
5.1.3	Data augmentation	96
5.1.4	Ensemble algorithms	97
5.1.5	Performance evaluation	99
5.2	EMR prognostics pipeline	101
5.2.1	CNN-RNN based methods for multivariate time series data . .	102
5.2.1.1	Neural networks	102
5.2.1.2	Convolutional neural networks	103
5.2.1.3	Convolutional Auto Encoder	109
5.2.1.4	Recurrent neural networks	109
5.2.2	Time series encoding enabling CNN-RNN topologies	111
5.2.3	Review of topical literature	113
5.2.4	Principles of TCN	116
5.2.5	Applications of TCN in a prognostics context	120
5.2.6	EMRUA pipeline	122
5.2.6.1	Data extraction	122
5.2.6.2	Feature extraction	124
5.2.6.3	Sequence selection	124
5.2.6.4	RUA estimation and uncertainty	125
5.2.7	Model configuration and scoring	129
	Summary	130
	References	130
6	ALT-EMR design	137
	Introduction	137
6.1	Review of testing methodologies	137
6.2	Selected test component	140
6.3	Life-cycle platform	140

6.3.1	Design and hardware selection	142
6.3.2	Control and data acquisition	145
6.3.3	Specifics of the experiments conducted	146
6.3.4	Data set DA1	149
6.3.5	Data set DA2	149
	Summary	150
	References	150
7	Results and analysis	152
	Introduction	152
7.1	BHA-PCBA maintenance support framework	152
7.1.1	Model accuracy	152
7.1.2	Maintenance threshold interpretation	155
7.1.3	Deployed API	159
7.1.4	Discussion of PdM integration and data constraints	159
7.2	EMR failure analysis of silver-plated contacts	160
7.2.1	Observed failure modes and mechanisms	161
7.2.1.1	Sealed EMR	162
7.2.1.2	Unsealed EMR	169
7.2.2	The influence of the operational environment on CR	173
7.2.3	Analysis of DA2 CR and temperature data	174
7.2.4	Implications for the usage of CR	178
7.3	EMR prognostics pipeline	179
7.3.1	Derived feature sets	179
7.3.2	Hyperparameters and sequence selection	180
7.3.3	EMRUA inference	184
7.3.4	Performance comparison with LSTM	186
7.3.5	Discussion and research prospects	188
7.3.6	Computational hardware and software	188
	Summary	189
	References	189
8	Conclusion	192
8.1	BHA-PCBA maintenance support framework	194
8.2	EMR life-cycle data and prognostics pipeline	195
8.3	General suggestions for future work	196
	Appendices	197
A	Experimental design	198
A.1	ALT-EMR test PCB	198
A.2	LabVIEW user interface	199

A.3	H-PC LabVIEW main program	200
A.4	DAQ LabVIEW EMR switching loop	202
A.5	DA1 actuations till failure	204
A.6	DA2 actuations till failure	204
B	Supplementary EMR failure analysis results	205
B.1	DA2 sealed EMR making waveform	205
B.2	DA2 sealed EMR breaking waveform	206
B.3	DA2 unsealed EMR making waveform	206
B.4	DA2 unsealed EMR breaking waveform	206
B.5	EDX analysis	207
C	Supplementary EMRUA results	208
C.1	DA1 time-based feature set	208
C.2	DA1 statistical feature set	210
C.2.1	Making actuation	210
C.2.2	Breaking actuation	211
C.3	DA1 inference performance evaluation	212
C.3.1	DA1-01U	212
C.3.1.1	LI	212
C.3.1.2	GI	213
C.3.2	DA1-02U	215
C.3.2.1	EI	215
C.3.2.2	LI	216
C.3.2.3	GI	218
C.3.3	DA1-03U	219
C.3.3.1	EI	219
C.3.3.2	LI	221
C.3.3.3	GI	222
C.3.4	DA1-04U	224
C.3.4.1	EI	224
C.3.4.2	LI	225
C.3.4.3	GI	227
C.3.5	DA1-05U	228
C.3.5.1	EI	228
C.3.5.2	LI	230
C.3.5.3	GI	231
C.3.6	DA1-06U	233
C.3.6.1	EI	233
C.3.6.2	LI	234
C.3.6.3	GI	236

C.4	Batch size effects on uncertainty estimation	237
-----	--	-----

List of Tables

2.1	Software tools for PHM.	18
2.2	Comparison of PHM methods.	19
2.3	Embedding of PHM.	29
2.4	Oil and gas survey.	32
3.1	Downhole tool data sources.	45
3.2	Contact related failure modes and mechanisms.	51
4.1	Overview of data-driven maintenance applications for downhole tools.	71
4.2	An overview of classical EMR DIs.	74
4.3	Summary of reviewed methodologies.	83
5.1	Selected NN activation functions.	103
5.2	Comparison of DL time series related methods.	113
5.3	Evaluated TCN model hyperparameters.	130
6.1	Aspects to consider when testing EMR.	138
6.2	Specifications of the selected EMR under test.	140
6.3	ALT comparison.	146
6.4	Overview of the recorded data in DA1.	149
6.5	Overview of the recorded data in DA2.	150
7.1	Classifier performance comparison.	154
7.2	Maintenance support for BHA units.	156
7.3	Calculation of selected time-based reference DIs.	166
7.4	EMRUA performance evaluation - metrics.	184
7.5	TCN-LSTM performance comparison.	187
8.1	Methodological comparison.	193

List of Figures

1.1	Thesis road-map.	9
2.1	Bathtub curve.	14
2.2	PHM concept.	16
2.3	RUL schematic.	17
2.4	Prognostic methods overview.	22
2.5	Prognostic method selection criteria.	25
2.6	PHM data sources.	26
2.7	Feature extraction through waveform windowing.	27
2.8	Data preprocessing steps.	28
3.1	Oil and gas drilling.	43
3.2	BHA-PCBA failures	46
3.3	EMR working principle.	48
3.4	Electrical arcing.	54
3.5	Contact welding.	56
3.6	Fretting.	58
3.7	EMR contact making.	58
3.8	EMR contact breaking.	61
4.1	Data-driven methods deployed throughout the oil and gas industry.	66
4.2	Data-driven methods used for different downhole tools.	67
4.3	Data-driven methods as part of novel maintenance strategies.	68
4.4	Time-based reference DIs.	73
4.5	Network graph of reviewed EMR literature.	86
5.1	BHA-PCBA maintenance support framework.	94
5.2	BHA-PCBA features overview	95
5.3	BHA-PCBA time features	96
5.4	DT schematic.	98
5.5	Confusion matrix.	100
5.6	ROC-AUC interpretation.	101
5.7	Schematic of a NN neuron.	103

5.8	Concept for CNN-based failure classification.	104
5.9	Stacked CNN layers.	106
5.10	Convolution operation.	107
5.11	Activation and pooling.	108
5.12	Fully connected layer.	108
5.13	Concept of a CAE.	109
5.14	RNN cell.	110
5.15	MVTD representations.	112
5.16	TCN concepts.	117
5.17	TCN residual blocks.	119
5.18	TCN receptive field.	120
5.19	EMRUA framework.	123
5.20	Matching batch-RUA sequence length.	126
5.21	EMRUA inference.	128
6.1	Image of the selected test EMR.	141
6.2	EMR test thresholds.	141
6.3	Image of the lab and the life-cycle test platform.	142
6.4	Simplified overview of the experimental platform.	143
6.5	Image of the developed test PCB.	144
6.6	Simplified schematic of the experimental process.	145
6.7	DA1 contact making.	147
6.8	DA2 contact making.	148
6.9	DA1 contact breaking.	148
6.10	DA2 contact breaking.	149
7.1	Correlation heatmap.	153
7.2	Accuracy score depending on decision threshold.	155
7.3	Business case.	158
7.4	Deployed API.	159
7.5	Schematic of the v_{cc} and v_{oc}	162
7.6	DA2 - Sealed EMR v_{cc}	163
7.7	DA2-03S contact making actuations.	164
7.8	DA2 - Sealed EMR v_{oc}	165
7.9	DA2-03S contact breaking actuation.	167
7.10	DA2-03S sealed EMR failed stuck-open.	168
7.11	DA2-02S CTS of sealed EMR.	169
7.12	DA2-02S rendering of the CTS.	169
7.13	DA2 - Unsealed EMR v_{cc}	170
7.14	DA2-09U contact making actuation.	171
7.15	DA2 - Unsealed EMR v_{oc}	171

7.16	DA2-09U unsealed EMR failed stuck-closed.	171
7.17	DA2-09U CTS unsealed EMR.	172
7.18	DA2-09U rendering of the CTS.	172
7.19	DA2 sealed CR.	176
7.20	DA2 sealed temperature.	177
7.21	DA2 unsealed CR.	177
7.22	DA2 unsealed temperature.	178
7.23	Kernel, dilation-base, and residual-blocks comparison.	181
7.24	Performance comparison F_T	182
7.25	Performance comparison F_S	182
7.26	Performance comparison $F_{T,S}$	183
7.27	RUA: DA2-01U - F_T - EI.	185
7.28	RUA: DA2-01U - F_S - EI.	185
7.29	RUA: DA2-01U - $F_{T,S}$ - EI.	186
7.30	Batch size effects on RUA estimation.	187
A.1	PCB design.	198
A.2	LabVIEW user interface.	199
A.3	H-PC LabVIEW main program.	201
A.4	H-PC LabVIEW main program.	203
A.5	DA1 actuations till failure.	204
A.6	DA2 actuations till failure.	204
B.1	DA2-03S - making.	205
B.2	DA2-03S - breaking.	206
B.3	DA2-09U - making.	206
B.4	DA2-09U - breaking.	206
B.5	EDX analysis.	207
C.1	DA1 - Pick-Up Time (PT).	208
C.2	DA1 - Release Time (RT).	208
C.3	DA1 - Bounce Time (BT).	209
C.4	DA1 - Arcing Time (AT) during breaking.	209
C.5	DA1 - Arcing Time (AT) during making.	209
C.6	DA1-01S F_S making.	210
C.7	DA1-01S F_S breaking.	211
C.8	DA1-01U - F_T - LI - $k = 4-d_b = 2-R = 6$	212
C.9	DA1-01U - F_S - LI - $k = 4-d_b = 2-R = 6$	212
C.10	DA1-01U - $F_{T,S}$ - LI - $k = 4-d_b = 2-R = 6$	213
C.11	DA1-01U - F_T - GI - $k = 4-d_b = 2-R = 6$	213
C.12	DA1-01U - F_S - GI - $k = 4-d_b = 2-R = 6$	214
C.13	DA1-01U - $F_{T,S}$ - GI - $k = 4-d_b = 2-R = 6$	214

C.14 DA1-02U - F_T - EI - $k = 4-d_b = 2-R = 6$.	215
C.15 DA1-02U - F_S - EI - $k = 4-d_b = 2-R = 6$.	215
C.16 DA1-02U - $F_{T,S}$ - EI - $k = 4-d_b = 2-R = 6$.	216
C.17 DA1-02U - F_T - LI - $k = 4-d_b = 2-R = 6$.	216
C.18 DA1-02U - F_S - LI - $k = 4-d_b = 2-R = 6$.	217
C.19 DA1-02U - $F_{T,S}$ - LI - $k = 4-d_b = 2-R = 6$.	217
C.20 DA1-02U - F_T - GI - $k = 4-d_b = 2-R = 6$.	218
C.21 DA1-02U - F_S - GI - $k = 4-d_b = 2-R = 6$.	218
C.22 DA1-02U - $F_{T,S}$ - GI - $k = 4-d_b = 2-R = 6$.	219
C.23 DA1-03U - F_T - EI - $k = 4-d_b = 2-R = 6$.	219
C.24 DA1-03U - F_S - EI - $k = 4-d_b = 2-R = 6$.	220
C.25 DA1-03U - $F_{T,S}$ - EI - $k = 4-d_b = 2-R = 6$.	220
C.26 DA1-03U - F_T - LI - $k = 4-d_b = 2-R = 6$.	221
C.27 DA1-03U - F_S - LI - $k = 4-d_b = 2-R = 6$.	221
C.28 DA1-03U - $F_{T,S}$ - LI - $k = 4-d_b = 2-R = 6$.	222
C.29 DA1-03U - F_T - GI - $k = 4-d_b = 2-R = 6$.	222
C.30 DA1-03U - F_S - GI - $k = 4-d_b = 2-R = 6$.	223
C.31 DA1-03U - $F_{T,S}$ - GI - $k = 4-d_b = 2-R = 6$.	223
C.32 DA1-04U - F_T - EI - $k = 4-d_b = 2-R = 6$.	224
C.33 DA1-04U - F_S - EI - $k = 4-d_b = 2-R = 6$.	224
C.34 DA1-04U - $F_{T,S}$ - EI - $k = 4-d_b = 2-R = 6$.	225
C.35 DA1-04U - F_T - LI - $k = 4-d_b = 2-R = 6$.	225
C.36 DA1-04U - F_S - LI - $k = 4-d_b = 2-R = 6$.	226
C.37 DA1-04U - $F_{T,S}$ - LI - $k = 4-d_b = 2-R = 6$.	226
C.38 DA1-04U - F_T - GI - $k = 4-d_b = 2-R = 6$.	227
C.39 DA1-04U - F_S - GI - $k = 4-d_b = 2-R = 6$.	227
C.40 DA1-04U - $F_{T,S}$ - GI - $k = 4-d_b = 2-R = 6$.	228
C.41 DA1-05U - F_T - EI - $k = 4-d_b = 2-R = 6$.	228
C.42 DA1-05U - F_S - EI - $k = 4-d_b = 2-R = 6$.	229
C.43 DA1-05U - $F_{T,S}$ - EI - $k = 4-d_b = 2-R = 6$.	229
C.44 DA1-05U - F_T - LI - $k = 4-d_b = 2-R = 6$.	230
C.45 DA1-05U - F_S - LI - $k = 4-d_b = 2-R = 6$.	230
C.46 DA1-05U - $F_{T,S}$ - LI - $k = 4-d_b = 2-R = 6$.	231
C.47 DA1-05U - F_T - GI - $k = 4-d_b = 2-R = 6$.	231
C.48 DA1-05U - F_S - GI - $k = 4-d_b = 2-R = 6$.	232
C.49 DA1-05U - $F_{T,S}$ - GI - $k = 4-d_b = 2-R = 6$.	232
C.50 DA1-06U - F_T - EI - $k = 4-d_b = 2-R = 6$.	233
C.51 DA1-06U - F_S - EI - $k = 4-d_b = 2-R = 6$.	233
C.52 DA1-06U - $F_{T,S}$ - EI - $k = 4-d_b = 2-R = 6$.	234
C.53 DA1-06U - F_T - LI - $k = 4-d_b = 2-R = 6$.	234

C.54 DA1-06U - F_S - LI - $k = 4-d_b = 2-R = 6$	235
C.55 DA1-06U - $F_{T,S}$ - LI - $k = 4-d_b = 2-R = 6$	235
C.56 DA1-06U - F_T - GI - $k = 4-d_b = 2-R = 6$	236
C.57 DA1-06U - F_S - GI - $k = 4-d_b = 2-R = 6$	236
C.58 DA1-06U - $F_{T,S}$ - GI - $k = 4-d_b = 2-R = 6$	237
C.59 DA1-16U - $N = 50$	237
C.60 DA1-16U - $N = 500$	238
C.61 DA1-16U - $N = 50000$	238

Acronyms

AE	Auto Encoder
AI	Artificial Intelligence
ALT	Accelerated Life-Cycle Testing
AMO	Assembly Maintenance Overhaul
API	Application Programming Interface
ARIMA	Auto Regressive Integrated Moving Average
ARMA	Auto Regressive Moving Average
AT	Arcing Time
AUC	Area Under the Curve
BD	Big Data
BDA	Big Data Analytics
BHA	Bottom Hole Assembly
BT	Bounce Time
CAE	Convolutional Auto Encoder
CBM	Condition Based Maintenance
CC	Coil Current
CI	Contact Current
CNN	Convolutional Neural Network
CR	Contact Resistance
CT	Closing Time
CTS	Computer Tomography Scan
CV	Contact Voltage
DA1	Data Set 1
DA2	Data Set 2
DAQ	Data Acquisition Unit
DCR	Dynamic Contact Resistance
DI	Degradation Indicator
DL	Deep Learning
DPDT	Double Pole Double Throw
DT	Decision Tree
EDX	Energy Dispersive X-Ray
EI	Exponential-Sequence Indexing
EMR	Electromagnetic Relay
EMRUA	Electromagnetic-Relay-Remaining-Useful-Actuation
EOL	End Of Life
ESP	Electrical Submersible Pump
FC	Fully Connected Layer

FMMEA	Failure Modes, Mechanisms, and Effects Analysis
GAF	Gramian Angular Field
GI	Growing-Sequence Indexing
GRU	Gated Recurrent Unit Neural Network
GUI	Graphical User Interface
H-PC	Host PC Unit
HALT	Highly Accelerated Life Cycle Testing
HTHP	High Temperature High Pressure
IGBT	Insulated Gate Bipolar Transistor
KNN	K-Nearest Neighbour
LED	Light Emitting Diode
LI	Linear-Sequence Indexing
LSTM	Long Short Term Memory Neural Network
LWD	Logging While Drilling
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MCD	Monte Carlo Dropout
ML	Machine Learning
MOSFET	Metal Oxide Semiconductor Field Effect Transistor
MTBF	Mean Time Between Failure
MVTD	Multivariate Time Series Data
MWD	Measurement While Drilling
NI	National Instruments
NN	Neural Network
NPT	Non Productive Time
OECD	Organisation for Economic Cooperation and Development
OT	Over-Travel Time
PCA	Principal Component Analysis
PCB	Printed Circuit Board
PCBA	Printed Circuit Board Assemblies
PDF	Probability Density Function
PdM	Predictive Maintenance
PHM	Prognostics and Health Management
PrM	Prescriptive Maintenance
PT	Pick-Up Time
ReLU	Rectified Linear Unit
RFC	Random Forest Classifier
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
ROC-AUC	Receiver Operating Characteristic - Area Under the Curve

ROP	Rate Of Penetration
RPM	Rotations Per Minute
RSS	Rotary Steering System
RT	Release Time
RUA	Remaining Useful Actuation
RUL	Remaining Useful Life
RVM	Relevance Vector Machine
SEM	Scanning Electron Microscopy
SKF	Stratified K-Fold Cross-Validation
SOH	State of Health
SOIC	Small Outline Integrated Circuit
SPST	Single Pole Single Throw
ST	Super-Path Time
SVM	Support Vector Machine
TCN	Temporal Convolutional Network
WOB	Weight On Bit

Publications

Journal publications

- **L. Kirschbaum, et al., (2020)** L. Kirschbaum, D. Roman, G. Singh, J. Bruns, V. Robu, and D. Flynn. AI-driven Maintenance Support for Downhole Tools and Electronics Operated in Dynamic Drilling Environments. *IEEE Access*, pages 78683-78701, vol. 8, IEEE, 2020, doi:10.1109/ACCESS.2020.2990152.
- **L. Kirschbaum, et al., (2022)** L. Kirschbaum, V. Robu, S. Swingler, and D. Flynn. Prognostics for Electromagnetic Relays using Deep Learning. *IEEE Access*, pages 4861 - 4895, vol. 10, IEEE, 2022, doi:10.1109/ACCESS.2022.3140645.

Conference proceedings

- **L. Kirschbaum, et al., (2018)** L. Kirschbaum, F. Dinmohammadi, D. Flynn, V. Robu, and M., Pecht. Failure Analysis Informing Embedded Health Monitoring of Electromagnetic Relays. In *2018 3rd International Conference on System Reliability and Safety (ICSRS)*, pages 261-267, IEEE, 2018, doi:10.1109/ICSRS.2018.8688839.
- **L. Kirschbaum, et al., (2021)** L. Kirschbaum, D. Roman, V. Robu, and D. Flynn. Deep Learning Pipeline for State-of-Health Classification of Electromagnetic Relays. In *2021 IEEE 30th International Symposium on Industrial Electronics (ISIE)*, pages 1-7, IEEE, 2021, doi:10.1109/ISIE45552.2021.9576278.

Chapter 1

Introduction

This thesis explores data-driven solutions considering challenges of industrial data analysis encountered in complex and critical systems. The focus resides on industrialisation of data-driven prognostics, as this research aims to advance current maintenance practices relating to critical oil and gas deep drilling electronics. Within the context of both, fleets of industrial assets and individual components, my research is informed by an detailed analysis of academic and industrial literature as to identify key enablers and barriers preventing the implementation of improved maintenance schemes. Inter alia, the maintenance objectives as well as the availability, volume, and resolution of monitoring data guide the selection of a suitable method. Two distinct scenarios are considered that operate on different scales: (1) fleet-level failure data collected from electronic assemblies within drilling equipment; (2) high-resolution life-cycle data of an electromechanical component. The developed Machine Learning (ML) and Deep Learning (DL) frameworks are tuned to the respective challenges to best support maintenance decision making: (1) failure detection of Bottom Hole Assembly (BHA)-Printed Circuit Board Assemblies (PCBA) by estimating failure probabilities under consideration of cost-risk scenarios; (2) computational efficient forecasting of the number of Remaining Useful Actuations (RUAs) for Electromagnetic Relays (EMRs) throughout the components life relying on sensor data collected through life-cycle experiments and readily available in industrial systems ¹.

1.1 Motivation

In spite of growing decarbonisation efforts, traditional hydrocarbons such as oil and gas continue to form one central pillar of the overall energy mix for the foreseeable future. Attributable to globally improving living standards and population growth,

¹*Parts of this chapter have already been published in my journal or conference article: L. Kirschbaum, et al., (2020) and L. Kirschbaum, et al., (2022).*

the energy demand is projected to grow as a result of rising consumerism and industrial manufacturing of goods and services [1]. This trend is led by non-Organisation for Economic Cooperation and Development (OECD) countries, albeit there is ambiguity associated with the absolute values of oil and gas contributions due to factors such as changes in global environmental policies, fluctuations of oil and gas prices, and uncertainties in the exploration of new resources. An increase in consumption by 70 per cent compared to a 15 per cent increase in OECD member-states until 2050 is predicted [2]. It is anticipated that the global share of renewable energy sources will increase significantly compared to the current energy mix in the coming decade, with an annual growth of three to six per cent [1, 2]. Though oil is expected to reach peak demand in the near future, an average annual growth of approximately 0.7 per cent is predicted over the next one to five years. The gas-share, dominated by shale gas and supported by the global extension of the Liquefied-Natural-Gas (LNG) infrastructure continues to grow with a prospected annual average of 0.9 to 1.7 per cent [1, 3]. An oil demand between 70 to 130 million barrels per day is estimated for 2040 compared to the current demand of approximately 99 million barrels per day as of 2018 [4]. Gas consumption is expected to reach 4000 billion cubic meters in 2035 and up to 5500 billion cubic meters in 2040 [1, 5]. Based on these forecasts, simply continuing to exploit existing oil and gas resources is not sufficient to meet the estimated demand. Moreover, multi-trillion US dollars of investments into oil and gas infrastructure as well as the exploration and development of existing resources and new reserves are required over the next two decades [1, 6, 7]. But mind, the costs associated with such exploration and development of new oil and gas reserves are constantly rising, since conventional and unconventional hydrocarbon deposits become increasingly remote and challenging to access while, at the same time, environmental regulations become more restrictive, i.e. with respect to environmental safety [8]. In addition, the need of accessing new reserves, requires complex wellbore geometries which can only be achieved using advanced deep drilling equipment. These systems, referred to as BHA, resemble a multi-functional assembly of interchangeable tools for subsurface navigation, analysis of the surrounding formations, power supply, and surface communication.

The growing complexity of BHAs has resulted in an ever-increasing number of assorted electronic and electromechanical components contained within such downhole tools. During drilling, downhole tools are exposed to dynamic and harsh environments which are commonly referred to as High Temperature High Pressure (HTHP) environments. In the wellbore the internal electronics are subjected to ambient temperatures exceeding 200 °C and extreme pressure regimes [9]; bending and high levels of vibrational stress are induced during drilling, further accelerating fatigue of electronics [8]. These ambient conditions lead to electronic component, circuit, or system failures which cause Non Productive Time (NPT) threatening the economic

viability of many drilling operations. Consequently, the proportion of drilling costs of the total well costs is considerable [10], with even higher costs offshore [11], e.g. NPT can cost up to \$1 million per day [12]. 15 per cent of the total annual drilling costs are due to NPT caused by downhole tool failure, more precisely, the failed subcomponents within BHA modules [13].

Current maintenance strategies of downhole tool electronics greatly rely on traditional reliability testing and historically grown maintenance practices that are often no longer able to match the encountered real-world, operational life-cycles. Furthermore, classical reliability methods require large sample sizes and run-to-failure data to derive an estimate of average life under load with sufficient accuracy - data sets which are not readily available in sufficient quantities in many industrial environments [14]. Lastly, the use of predefined life-cycle estimates for maintenance management schemes often results in a *too early* exchange or replacement of operationally viable electronics and, consequently, in cost and risk penalties [15].

This is precisely where new concepts of maintenance come into play because the reliability and operational availability of electronic-rich systems has become increasingly important [16, 17]. Such novel maintenance strategies are not limited to electronics encountered in downhole tools. Instead they aim to provide frameworks in order to extract key performance indicators that can support, guide, or automate maintenance decision making. Consider the discipline of Prognostics and Health Management (PHM) in which data analytics, both data-driven models and physical-models [18, 19], can enable the deployment of maintenance concepts such as Predictive Maintenance (PdM) or Prescriptive Maintenance (PrM), insofar as an integration of physical and digital infrastructure into joint ecosystems provides sufficient access to historical- and streamed-monitoring data, e.g. in the form of Big Data (BD) [20]. Clearly, the availability of such data and respective analytics provide the means for performance and economic enhancement of industrial systems targeting higher reliability of industrial asset and services that can lead to the desirable paradigm shift in maintenance [21, 22, 23]. Furthermore, this transition is facilitated by increasing safety compliance standards for legacy and emergent systems [24, 25, 26, 27, 28].

1.2 Research question

Electronic assemblies are present in the majority of all downhole tools. Due to the multi-component nature of downhole tools, above average failure rates are registered for relevant electronic or electromechanical components, placing them as a predominate failure cause [18]. Hereinafter, the complexity of electronic assemblies requires elaborate and costly maintenance because the prediction of electronic related failures is time demanding and insufficient with current methods while pronounced PHM for electronics is constricted in its capabilities [29]. Nonetheless, the reduction

of these electronic related failures is essential for improving reliability in industrial assets, in particular, for the oil and gas industry to be able to continue to operate economically and meet future targets.

The need for modern maintenance strategies has already been identified by other sectors, e.g. the aviation industry adopts PHM strategies to improve the reliability, reducing maintenance costs which have risen significantly over the past years [30, 16]. This is, likewise, valid for industries operating in environments comparable to the oil and gas industry presently equipping their offshore assets with PHM capabilities [31, 32, 33]. Furthermore, various sectors of the oil and gas industry have committed to the adaption of novel operational support and maintenance strategies, cf. [34, 35, 36]. However, until very recently modern data-driven maintenance strategies have been exclusively used for, e.g. *reservoir modelling* or *surface and completion* tasks. Downhole tool electronics have been widely excluded, since implementation of data-driven methods to improve reliability meets a variety of challenges. Thus, this work addresses the following research question:

How to design data-driven maintenance to improve reliability of electronics given application-specific constraints?

In general terms, I examine the industrialisation of data analysis by coupling engineering knowledge with a review of the state of the art as to investigate the prerequisites and criteria for scalable, effective and operational data analysis in the context of maintenance support for expert elicitation decision making. Here, the design and impact of targeted data analysis interlinking previously not considered data streams and isolated sources is to be scrutinised, whereas the importance of life-cycle data in order to validate derived methods is emphasised. Therefore, the thesis investigates this question based on two scenarios focusing on high-resolution data at component level and the conditional abstractions necessary at fleet-level.

1.2.1 BHA-PCBA fleet-level

The high level of specialisation of BHA tools, extended development cycles, space and hardware constraints, real-time monitoring data bandwidth limitations, insufficient digital infrastructure, and the harsh, dynamic operating environment challenges current maintenance practices. Determining failure precursors for individual electrical components is time intensive and costly for BHA electronics. Despite the digitalisation which the oil and gas industry undergoes, the transition to novel data-driven maintenance strategies is impeded, partially due to a conservative industry-specific mindset. BHAs provide sensible amounts of monitoring and performance fleet-level data which is directly and indirectly related to the failures of electronic assemblies, i.e. the PCBAs. So far, however, this data is not leveraged. Reasons are manifold, e.g.: limited data-transfer bandwidth during drilling; reduced data

acquisition and storage resources within the tools allowing only low data resolution; data-silos; limited information of failure modes and their propagation, sensor placement, interaction of various sub-modules, and integration of legacy components.

All these factors limit the development of data-driven maintenance methods for BHA electronics. Nonetheless, BHA-PCBAs hold a number of electronic components whose proper functioning is critical for the entire BHA. These mission or safety critical components may require more precise performance monitoring. Enabling data-driven maintenance at the component-level for such individual electronics in the BHA changes the requirements in terms of data quantity, data quality and data resolution. To illustrate the complexity arising at the component-level, I consider one selected component representative for the electronics in BHAs which is presented in the following section.

1.2.2 EMR component-level

By far not limited to the use in downhole tools, EMRs play a safety critical role in various industrial systems. Automation has supported an improvement in EMR manufacturing and reliability. Population based methods have proven to be a cost efficient solution to determine the reliability of bulk produced EMRs, where the predominant effort, to date, aims to quantify the degradation of EMRs using classical reliability theory [37, 38]. Thus, EMR related maintenance strategies continue to rely on traditional preventive maintenance approaches, e.g. Mean Time Between Failure (MTBF). Whilst such approaches might be acceptable in low-risk activities, high consequence environments such as deep drilling applications, require tailored maintenance strategies. Here, an EMR will be subjected to specific direct and indirect ambient loading, rather than being in conformance to generalised operational requirements expected of bulk produced quantities. In consequence, the use of pre-defined life-cycle estimates for maintenance management schemes often results in an early replacement of operationally viable EMRs [15]. Likewise, classical reliability-based testing methods have been increasingly challenged in terms of test duration by the extended durability of the current generation of EMRs [38]. Lastly, the objective of past research centres around the definition of expressive performance indicators evaluated on data aggregated only over initial actuations [39, 40]. Distinct challenges of current methods are: computationally inefficient for in-situ deployment; limited generalisation; models predominately trained on constrained data sets and feature selections; non-representative life-cycle data; reliance on classical performance indicators (Contact Resistance (CR), Bounce Time (BT), Pick-Up Time (PT), Release Time (RT), Arcing Time (AT), etc.) which are difficult to obtain within an industrial setting and suffer from significant fluctuations. All these factors render traditional, threshold-based degradation models to determine the EMR-End Of Life (EOL) inefficient.

Monitoring EMRs has been traditionally cost prohibitive in many applications. However, in line with the digital industrialisation there is an unprecedented access to large volumes of system and component monitoring data and subsequent analytics that hold the potential to improve the derivation of enhanced EMR models for maintenance purposes [22]. Nonetheless, transitioning to industrial applications entails barriers [41, 42]. Among others, uncertainty and a sensible, efficient embedding of physical, data-driven, or hybrid models into existing digital infrastructure has to be considered [27]. In this scope, enabling modern maintenance strategies by creating actionable insights from data through data-driven prognostics is confronted with high volumes of Multivariate Time Series Data (MVTSD). With the advent of Artificial Intelligence (AI), research is addressing this challenge using ML and increasingly DL for, e.g. Remaining Useful Life (RUL) prediction of electronics [43, 44, 21, 45, 46]. Standalone approaches relying on Convolutional Neural Network (CNN) architectures or in combination with Recurrent Neural Network (RNN) elements employing techniques such as MVTSD imaging [47] resonate with high volumes of data as recent publication demonstrate [48, 49, 50, 51, 52, 53, 54]. In particular, CNN based models are popular, cf. methods that link Convolutional Auto Encoder (CAE)s for automated feature extraction with RNN-based architectures [55, 56]. A substantial amount of research exploits DL for RUL estimation because these methods scale with high data volumes. However, high volumes of MVTSD still pose a challenge, in particular very long input sequences.

1.3 Research contributions

BHA-PCBA maintenance support framework: The utilisation of downhole tool memory data for data-driven maintenance of BHA electronics presents a first-time adoption in the oil and gas industry sector. This research brings a distinct contribution to the utilisation of ML on electronics for harsh environment deployments with an analysis extended into actionable decision making for operational considerations at fleet-level. In this instance, fleet-level refers to the entire fleet of BHAs made up out of the newest tool generations but also legacy tools. This fleet is distributed around the entire globe and comprises more than 10000 interchangeable units.

1. A detailed review of topical literature concerning the application of data-driven strategies in the oil and gas industry with a focus on novel, operational support and maintenance decision making.
2. Analysis of the state-of-the-art in reliability and maintenance using AI evidencing that downhole tools employing such techniques are underrepresented at the moment.
3. A comparison of current industry standard maintenance strategies with novel

data-driven strategies that assess the central barriers which prevent a wider implementation of data-driven approaches.

4. A data-driven ML support framework (Random Forest and XGBoost) for maintenance optimisation and operational support is developed and evaluated. It estimates the probability of PCBA failure within a BHA prior to an upcoming maintenance action. The algorithms are trained on aggregated downhole tool memory data; disregarding the failure mode, the method provides information of whether to conduct maintenance on the electronics or to issue a re-run of the BHA.
5. The results are extended into a commercial analysis, demonstrating the value of the methodology under various scenarios of infield failure costs and fleet reliability levels.

EMR life-cycle data set: As identified through a review of the state of the art, one central challenge for the development of data-driven prognostic strategies, in particular for the EMR, is the unavailability of high-volume and high-resolution life-cycle data, i.e. data sets containing the entire life-cycle within a controlled experimental environment. Therefore, such a life-cycle set for EMRs has been collected.

1. Development of an EMR life-cycle test platform that samples data continuously at high rates over the entire EMR life; dedicated Printed Circuit Board (PCB) to facilitate the exchange of the device under test and to enable switching a control circuit and a secondary circuit for CR measurements.
2. Aggregation of two EMR life-cycle data sets comprising unsealed and sealed EMR experiments.
3. An extensive review of failure modes and failure mechanisms of EMRs and distinct characteristics of switching EMR contacts in DC circuits.
4. A structured overview and critical analysis of commonly employed EMR Degradation Indicator (DI)s, i.e. extracted indicators which can depict the degradation trend. Among those DIs, CR is frequently applied. However, as demonstrated through the experiments, CR is subject to significant, random fluctuations due to surface oxidation processes if unsealed silver-plated EMR contacts are used. Dependent on the operating environment, this renders CR superfluous in a prognostics context.

EMR prognostics pipeline: Electromechanical components such as EMRs are omnipresent in electrical systems. A data-driven prognostics approach has the potential to improve EMR reliability in a maintenance context. Unlike previous work, high volumes of monitoring data are exploited which are commonly found in industrial systems to determine the EOL of EMRs (no invasive measures are used).

1. The first exhaustive, critical, and comparative review of research concerned with novel data-driven operation and maintenance strategies for EMRs.

2. Considering the characteristics of the Temporal Convolutional Network (TCN), a prognostic DL pipeline capable of providing actionable maintenance support is introduced. The significant qualities of the framework are its computational efficiency and its capability to take arbitrarily long MVTD inputs without an increase in the number of trainable parameters. Lastly, using Monte Carlo Dropout (MCD), a computationally efficient method to approximate the uncertainty of the forecast during inference is adopted to the specifics of the proposed approach.
3. As TCN is a novel DL architecture, the effects of distinct hyperparameters on the model performance are analysed under consideration of the number of trainable parameters and data-ingestion strategy.
4. The superiority of the statistical feature set based on Contact Voltage (CV) and Contact Current (CI) measurements is demonstrated in comparison to conventional features, e.g. CR, BT, AT, etc. which are traditionally used to assess EMR degradation.
5. A detailed analysis of the experienced failure mechanisms is performed.

1.4 Thesis structure

Figure 1.1 presents the layout of the thesis.

Chapter 2.2 provides an introductory presentation of the governing concepts of reliability and maintenance to which the field of Prognostics and Health Management (PHM) and its methods are closely linked. Relevant to PHM, key concepts of AI are presented and the respective considerations for tailoring a prognostics framework are discussed. This chapter closes with a discussion of the stimulating technological developments and trends that enable PHM based on a comparative analysis of trends in the oil and gas industry.

Chapter 3 discusses the background of the two respective case studies namely the Bottom Hole Assembly (BHA) in an oil and gas context and the Electromagnetic Relay (EMR) whilst highlighting the unique challenges.

Chapter 4 extends the prior analysis of these two respective case studies, providing an exhaustive, critical analysis of the state-of-the-art of PHM related research.

Chapter 5 presents the developed methods for BHA-PCBA maintenance classification and EMR-RUA estimation.

Chapter 6 details the developed EMR life-cycle experiment and summarises the collected data sets.

Chapter 7 highlights the findings of the two developed prognostic methods ver-

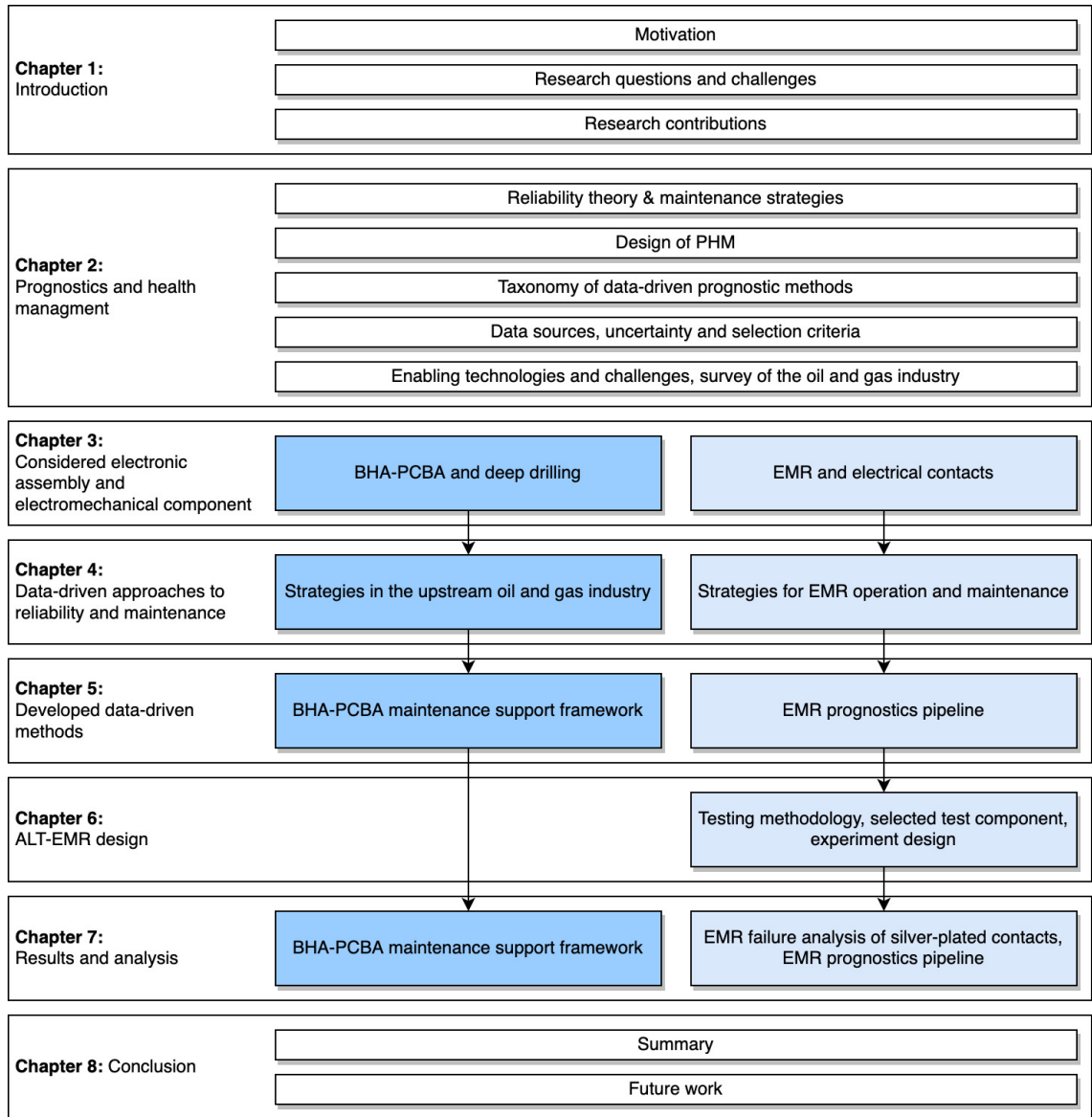


Figure 1.1: Thesis road-map.

ified on the experimental data, reviewing inter alia improvements through model hyperparameter tuning, sub-sampling strategies and feature engineering. Further, a detailed study of the EMR failure manifestation is performed.

Lastly, **Chapter 8** summarises the findings and discusses future research.

References

- [1] BP Energy. Energy outlook 2019 edition. Technical report, BP, 2019.
- [2] E Sendich. International energy outlook 2019 with projections to 2050. Technical report, U.S. Energy Information Administration, 2019.
- [3] OECD. World energy outlook 2019. Technical report, OECD, 2019.
- [4] OECD. Oil 2018: Analysis and forecasts to 2023. Technical report, OECD, 2018.
- [5] McKinsey. Global gas and LNG outlook to 2035. Technical report, McKinsey, 2018.
- [6] D Calvetti. Oil and gas industry: 2019 trends. Technical report, KPMG Argentina, 2019.
- [7] G Biscardini, D Branson, C Del Maestro Adrian, K Hawsey, N Molavi, and R Morrison. Oil and gas trends 2019-building growth strategies on shifting sands. *Proc. 22nd Annu. Global CEO Surv. Trends*, pages 1–22, 2019.

- [8] K Carter-Journet, A Kale, T Falgout, and L Heuermann-Kuehn. Drilling optimization: Utilizing lifetime prediction to improve drilling performance and reduce downtime. In *SPE Deepwater Drilling and Completions Conference*. OnePetro, 2014. doi:10.2118/170270-MS.
- [9] I Ahmad, O Akimov, P Bond, P Cairns, N Eide, T Gregg, T Heimes, A Nwosu, and F Wiese. Reliable technology for drilling operations in a high-pressure/high-temperature environment. In *IADC/SPE Drilling Conference and Exhibition*. OnePetro, 2014. doi:10.2118/167972-MS.
- [10] J Brehme and T Travis. Total BHA reliability — an improved method to measure success. In *IADC/SPE Drilling Conference*. OnePetro, 2008. doi:10.2118/112644-MS.
- [11] H Reckmann, P Jogi, F Kpetehoto, S Chandrasekaran, and J Macpherson. MWD failure rates due to drilling dynamics. In *IADC/SPE Drilling Conference and Exhibition*. OnePetro, 2010. doi:10.2118/127413-MS.
- [12] S Zhan, J Rodiek, L Heuermann-Kuehn, and J Baumann. Prognostics health management for a directional drilling system. In *IEEE Prognostics and System Health Management Conference (PHM-Harbin)*, pages 1–7. IEEE, 2011. doi:10.1109/PHM.2011.5939543.
- [13] W Aldred, D Plumb, I Bradford, J Cook, V Gholkar, L Cousins, R Minton, J Fuller, S Goraya, and D Tucker. Managing drilling risk. *Oilfield review*, 11(2):2–19, 1999.
- [14] J Liu, M Zhang, N Zhao, and A Chen. A reliability assessment method for high speed train electromagnetic relays. *Energies*, 11(3):652, 2018. doi:10.3390/en11030652.
- [15] Y Fang, Z Li, W Li, and K Li. Concerning contact resistance prediction based on time sequence and distribution character. In *50th IEEE Holm Conference on Electrical Contacts*, pages 447–452. IEEE, 2004. doi:10.1109/HOLM.2004.1353155.
- [16] S Kumar and M Pecht. Modeling approaches for prognostics and health management of electronics. *International Journal of Performability Engineering*, 6(5):467, 2010.
- [17] A Al-Mohamad, G Hoblos, and V Puig. A hybrid system-level prognostics approach with online RUL forecasting for electronics-rich systems with unknown degradation behaviors. *Microelectronics Reliability*, 111:113676, 2020. doi:10.1016/j.microrel.2020.113676.
- [18] A Prisacaru, P Gromala, M Jeronimo, B Han, and G Zhang. Prognostics and health monitoring of electronic system: A review. In *2017 18th International Conference on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems (EuroSimE)*, pages 1–11. IEEE, 2017. doi:10.1109/EuroSimE.2017.7926248.
- [19] C Bhargava, P Sharma, M Senthilkumar, S Padmanaban, V Ramachandaramurthy, Z Leonowicz, F Blaabjerg, and M Mitolo. Review of health prognostics and condition monitoring of electronic components. *IEEE Access*, 8:75163–75183, 2020. doi:10.1109/ACCESS.2020.2989410.
- [20] I Kong, S Harper, D Mitchell, J Blanche, T Lim, and D Flynn. Interactive digital twins framework for asset management through internet. In *IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT)*, pages 1–7. IEEE, 2020. doi:10.1109/GCAIoT51063.2020.9345890.
- [21] M Pecht and M Kang. *Prognostics and health management of electronics: Fundamentals, machine learning, and the internet of things*. John Wiley and Sons, Ltd., 2018. doi:10.1002/9781119515326.
- [22] E Miguelañez-Martin and D Flynn. Embedded intelligence supporting predictive asset management in the energy sector. In *Asset Management Conference*, pages 1–7. IEEE, 2015. doi:10.1049/cp.2015.1752.
- [23] W Tang, M Andoni, V Robu, and D Flynn. Accurately forecasting the health of energy system assets. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5. IEEE, 2018. doi:10.1109/ISCAS.2018.8351842.
- [24] O Zaki, M Dunnigan, V Robu, and D Flynn. Reliability and safety of autonomous systems based on semantic modelling for self-certification. *Robotics*, 10(1), 2021. doi:10.3390/robotics10010010.
- [25] X Zhao, A Banks, J Sharp, V Robu, D Flynn, M Fisher, and X Huang. A safety framework for critical systems utilising deep neural networks. In *International Conference on Computer Safety, Reliability, and Security*, pages 244–259. Springer, 2020. doi:10.1007/978-3-030-54549-9_16.
- [26] M Osborne, J Lantair, Z Shafiq, X Zhao, V Robu, D Flynn, and J Perry. UAS operators safety and reliability survey: Emerging technologies towards the certification of autonomous UAS. In *4th International Conference on System Reliability and Safety (ICSRS)*, pages 203–212. IEEE, 2019. doi:10.1109/ICSRS48664.2019.8987692.
- [27] W Tang, R Dickie, D Roman, V Robu, and D Flynn. Optimisation of hybrid energy systems for maritime vessels. *The Journal of Engineering*, 2019(17):4516–4521, 2019. doi:10.1049/joe.2018.8232.

- [28] D Roman, R Dickie, V Robu, and D Flynn. A review of the role of prognostics in predicting the remaining useful life of assets. *Safety and Reliability-Theory and Applications*, 135, 2017.
- [29] G Vogl, B Weiss, and M Helu. A review of diagnostic and prognostic capabilities and best practices for manufacturing. *Journal of Intelligent Manufacturing*, 30(1):79–95, 2019. doi:10.1007/s10845-016-1228-8.
- [30] D Nguyen, M Kefalas, K Yang, A Apostolidis, M Olhofer, S Limmer, and T Bäck. A review: Prognostics and health management in automotive and aerospace. *International Journal of Prognostics and Health Management*, 10(2):1–35, 2019. doi:10.36001/ijphm.2019.v10i2.2730.
- [31] F Dinmohammadi, D Flynn, C Bailey, M Pecht, C Yin, P Rajaguru, and V Robu. Predicting damage and life expectancy of subsea power cables in offshore renewable energy applications. *IEEE Access*, 7:54658–54669, 2019. doi:10.1109/ACCESS.2019.2911260.
- [32] W Tang, D Flynn, K Brown, R Valentin, and X Zhao. The design of a fusion prognostic model and health management system for subsea power cables. In *Oceans 2019 MTS/IEEE Seattle*, pages 1–6. IEEE, 2019. doi:10.23919/OCEANS40490.2019.8962816.
- [33] A Stetco, F Dinmohammadi, X Zhao, V Robu, D Flynn, M Barnes, J Keane, and G Nenadic. Machine learning methods for wind turbine condition monitoring: A review. *Renewable energy*, 133:620–635, 2019. doi:10.1016/j.renene.2018.10.047.
- [34] H Devold, T Graven, and S Halvorsrød. Digitalization of oil and gas facilities reduce cost and improve maintenance operations. In *Offshore Technology Conference*. OnePetro, 2017. doi:10.4043/27788-MS.
- [35] F Laborie, M Biermann, A Camp, and M Sivertsen. Prerequisites for performance-based contracts: Digitalization and data sharing changes business models in E&P supply chain. In *SPE Norway One Day Seminar*. OnePetro, 2019. doi:10.2118/195602-MS.
- [36] S Settemsdal. Highly scalable digitalization platform for oil and gas operations enables total asset visibility for predictive, condition-based fleet management across single and multiple sites. In *Abu Dhabi International Petroleum Exhibition & Conference*. OnePetro, 2019. doi:10.2118/197820-MS.
- [37] Y Fang, J Lu, J Zheng, and Z Huang. Research on the failure diagnostics parameters and the reliability prediction model of the electrical contacts. In *52nd IEEE Holm Conference on Electrical Contacts*, pages 69–72. IEEE, 2006. doi:10.1109/HOLM.2006.284067.
- [38] Y Xuerong, Y Qiong, and Z Guofu. Reliability assessment for electromagnetic relay based on time parameters degradation. In *11th International Conference on Electronic Packaging Technology & High Density Packaging*, pages 1269–1272. IEEE, 2010. doi:10.1109/ICEPT.2010.5582785.
- [39] A Wileman and S Perinpanayagam. A prognostic framework for electromagnetic relay contacts. In *PHM Society European Conference*, volume 2, pages 1–7. PHM Society, 2014. doi:10.36001/phme.2014.v2i1.1531.
- [40] J Guo, G Zhang, Y Bi, and Y Li. Life prediction of automotive electromagnetic relay based on wavelets neural network. *Chemical Engineering Transactions*, 62:1213–1218, 2017. doi:10.3303/CET1762203.
- [41] V Robu, D Flynn, and D Lane. Train robots to self-certify their safe operation. *Nature*, 553(281), 2018. doi:10.1038/d41586-018-00646-w.
- [42] X Zhao, W Huang, A Banks, V Cox, D Flynn, S Schewe, and X Huang. Assessing the reliability of deep learning classifiers through robustness evaluation and operational profiles. *arXiv preprint*, 2021. arXiv:2106.01258.
- [43] N Vichare and M Pecht. Prognostics and health management of electronics. *IEEE Transactions on Components and Packaging Technologies*, 29:222–229, 2006. doi:10.1109/TCAPT.2006.870387.
- [44] L Kirschbaum, D Roman, G Singh, J Bruns, V Robu, and D Flynn. AI-driven maintenance support for downhole tools and electronics operated in dynamic drilling environments. *IEEE Access*, 8:78683–78701, 2020. doi:10.1109/ACCESS.2020.2990152.
- [45] D Roman, S Saxena, V Robu, M Pecht, and D Flynn. Machine learning pipeline for battery state-of-health estimation. *Nature Machine Intelligence*, 3(5):447–456, 2021. doi:10.1038/s42256-021-00312-3.
- [46] D Roman, S Saxena, J Bruns, R Valentin, M Pecht, and D Flynn. A machine learning degradation model for electrochemical capacitors operated at high temperature. *IEEE Access*, 9:25544–25553, 2021. doi:10.1109/ACCESS.2021.3057959.
- [47] O Fink, Q Wang, M Svensen, P Dersin, W Lee, and M Ducoffe. Potential, challenges and future directions for deep learning in prognostics and health management applications. *En-*

- gineering Applications of Artificial Intelligence*, 92:103678, 2020. doi:10.1016/j.engappai.2020.103678.
- [48] Y Zheng, Q Liu, E Chen, Y Ge, and J Zhao. Time series classification using multi-channels deep convolutional neural networks. In *International conference on web-age information management*, pages 298–310. Springer, 2014. doi:10.1007/978-3-319-08010-9_33.
- [49] J Yang, M Nguyen, P San, X Li, and S Krishnaswamy. Deep convolutional neural networks on multichannel time series for human activity recognition. In *Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI’15*, page 3995–4001. AAAI Press, 2015. doi:10.5555/2832747.2832806.
- [50] Y Zheng, Q Liu, E Chen, Y Ge, and J Zhao. Exploiting multi-channels deep convolutional neural networks for multivariate time series classification. *Frontiers of Computer Science*, 10(1):96–112, 2016. doi:10.1007/s11704-015-4478-2.
- [51] D Verstraete, A Ferrada, E Droguett, V Meruane, and M Modarres. Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings. *Shock and Vibration*, 2017:1–17, 2017. doi:10.1155/2017/5067651.
- [52] X Ding and Q He. Energy-fluctuated multiscale feature learning with Deep ConvNet for intelligent spindle bearing fault diagnosis. *IEEE Transactions on Instrumentation and Measurement*, 66:1926–1935, 2017. doi:10.1109/TIM.2017.2674738.
- [53] G Babu, P Zhao, and X Li. Deep convolutional neural network based regression approach for estimation of remaining useful life. In *International conference on database systems for advanced applications*, pages 214–228. Springer, 2016. doi:10.1007/978-3-319-32025-0_14.
- [54] Y Chen and H Rangwala. Attention-based multi-task learning for sensor analytics. In *IEEE International Conference on Big Data (Big Data)*, pages 2187–2196. IEEE, 2019. doi:10.1109/BigData47090.2019.9006207.
- [55] Z Khan, T Hussain, A Ullah, S Rho, M Lee, and S Baik. Towards efficient electricity forecasting in residential and commercial buildings: A novel hybrid CNN with a LSTM-AE based framework. *Sensors*, 20(5):1399, 2020. doi:10.3390/s20051399.
- [56] T Le, M Vo, B Vo, E Hwang, S Rho, and S Baik. Improving electric energy consumption prediction using CNN and Bi-LSTM. *Applied Sciences*, 9(20):4237, 2019. doi:10.3390/app9204237.

Chapter 2

Prognostics and health management

Introduction

The goal to reduce operational and Assembly Maintenance Overhaul (AMO) related costs in order to increase the availability of industrial assets is the central force behind the development of novel, data-driven monitoring and maintenance strategies. The industrial digitalisation provides the necessary means to accomplish such transformation, e.g. digital industrial ecosystems or Big Data Analytics (BDA). Along these lines, the remainder of this chapter discusses the underlying concepts of reliability which relate to governing maintenance approaches. Further, the development of novel maintenance strategies within the domain of PHM, methods, and considerations are detailed. The chapter concludes highlighting the challenges and opportunities of the digital industrialisation, in particular for PHM, based on a survey of the oil and gas industry ¹.

2.1 Reliability and maintenance strategies

The notion of reliability has first been coined, when, in the 1960s, mathematical concepts of probability theory were adopted to an engineering domain [1]. Here, *reliability* is defined as the probability of a component or system to operate over a given time, within its specifications [2]. The reliability decreases over time, whereas the probability of failure increases. Reliability engineering aims to ensure low failure rates and high availability over the designated operation time [3]. Estimating reliability metrics is achieved using population-based testing methods. To this end, reliability theory incorporates a toolbox of statistical methods. For example, an

¹*Parts of this chapter have already been published in my journal or conference article: L. Kirschbaum, et al., (2020) and L. Kirschbaum, et al., (2022).*

established method used to describe population-based failure rates is known as the *bathtub*-curve, cf. Figure 2.1. The first section of the trajectory exhibits a decreasing failure frequency due to a high initial failure rate of a tested population of, e.g. components. This can be attributed to originally weak or faulty components. Once such components have failed the curve transfers to a stable phase. From now on, failures occur randomly at an almost constant rate where the actual operating time has a negligible influence. The third section is referred to as wear-out phase. The error rate increases as fatigue processes accelerate. Based on this assumption the defined useful life ends prior to the wear-out phase. An independent group of components might be subject to numerous failure modes and mechanisms, each following their own life distribution. Reliability theory makes use of a set of statistical distribution functions in order to represent these different failure distributions within populations. Commonly assumed distributions are the Exponential-, Weibull-, Normal-, or Lognormal-distribution. Thus, using a reliability centric approach an average life distribution is estimated for the component or system from a population of samples which is then projected to the individual component. Thereby, however, the actual operation conditions and environmental influences on the degradation process are neglected. This challenges traditional maintenance such as reactive and preventive strategies, in sight of increasingly complex components or system assemblies.

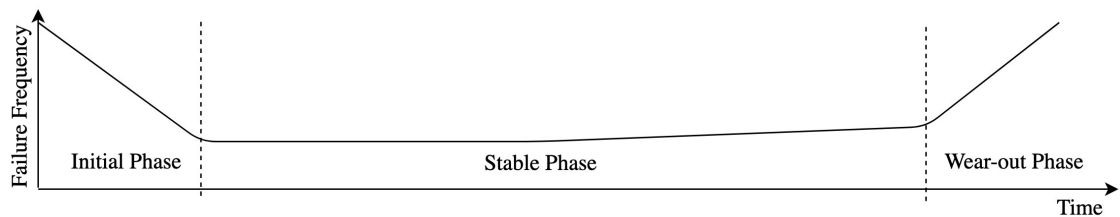


Figure 2.1: Schematic of the *bathtub*-curve, a common assumption for population-based failure behaviour in reliability engineering.

2.1.1 Reactive maintenance

Historically, a reactive or *run-to-failure* maintenance policy has been adopted under the paradigm: *an asset is operated until it fails*. Advantages of this maintenance strategy are minimal hold-ready resources, trained personal, and maintenance infrastructure. Reactive maintenance suits small, low-cost, and easily replaceable machinery [4]. However, for critical systems a failure might entail serious consequences, impacting other assets, or the safety of personnel and the environment. A significant amount of time might be required for repair causing extended periods of costly downtime.

2.1.2 Preventive maintenance

For this reason, a reactive approach may not be suitable in many instances where failure is costly or catastrophic. Preventive maintenance follows a time-based schedule, as it aims to perform a proactive, corrective action in advance in order to avoid a fault or failure as well as increasing system availability and life [5]. Such time-based maintenance intervals are predetermined. Their duration is derived through reliability testing, from historical data as well as manufacturing and expert knowledge [6]. In general, maintenance intervals are designed in such a way that only one to two per cent of all assets break down prior to the scheduled maintenance [7]. Preventive maintenance has proven to be a functioning method if degradation of components is well understood and systems are statically operated. Further, it is well suited where constant condition monitoring is not feasible in terms of complexity or costs. Thus, high levels of reliability have been achieved in the past through the employment of preventive maintenance strategies. However, preventive maintenance does not consider the real operating conditions and loads that components and systems experience over the course of their operational life. An early maintenance action risks additional costs, while the occurrence of a failure prior to a maintenance action may lead to unscheduled downtime [8]. Hence, the economic efficiency of this approach heavily depends on a well established maintenance scheme closely correlating to the true time of failure. Another disadvantage is the rigid setup of this method which is not easily adopted to dynamic operations and harsh environments. Along the lines of preventive maintenance, various metrics have been established, such as reliability- and hazard-functions as well as the Mean Time To Failure (MTTF), Mean Time To Repair (MTTR), or Mean Time Between Failure (MTBF) [3].

2.2 Overview

The Joint Strike Fighter programme of the US Airforce, in the early 1990s, has been the first to introduce the concept of Prognostics and Health Management (PHM) intended to improve reliability of US Airforce assets [9]. Unlike reliability theory, PHM prioritises the state prediction of individual system deviations over population-based statistical knowledge. The spread of PHM to sectors outside the aviation industry is largely driven by the prospect of reliability improvements, cost reduction, the availability of sensor technology, and the means of processing and analysing the collected data [10, 11, 12, 13, 14]. PHM which incorporates novel maintenance strategies is an emerging engineering discipline drawing from the fields of computer science, reliability theory, and domain specific expert knowledge. Condition Based Maintenance (CBM), PdM, or Prescriptive Maintenance are concerned with the transfer of PHM methods into actionable maintenance strategies. Figure 2.2 illustrates the relationship of PHM to classical maintenance strategies. PHM aims to provide a holistic

solution, in some instances from the design stage onwards [15]. It is recognised as an efficient systemic approach to technologies and methods aiming to optimise operational control, improve asset lifetime, reduce maintenance intervals and costs, whilst mitigating system downtime [16]. It can also support qualification testing of components to guarantee reliability requirements [13, 17]. General frameworks to support the PHM workflow are being developed and, simultaneously, established as domain specific standards [18, 19, 20, 21]. To guide the development of PHM, verification and validation methods, as well as general performance metrics have been proposed [22]. Several successful PHM implementations are reported, even though its application tailored character delays a more general dissemination [10].

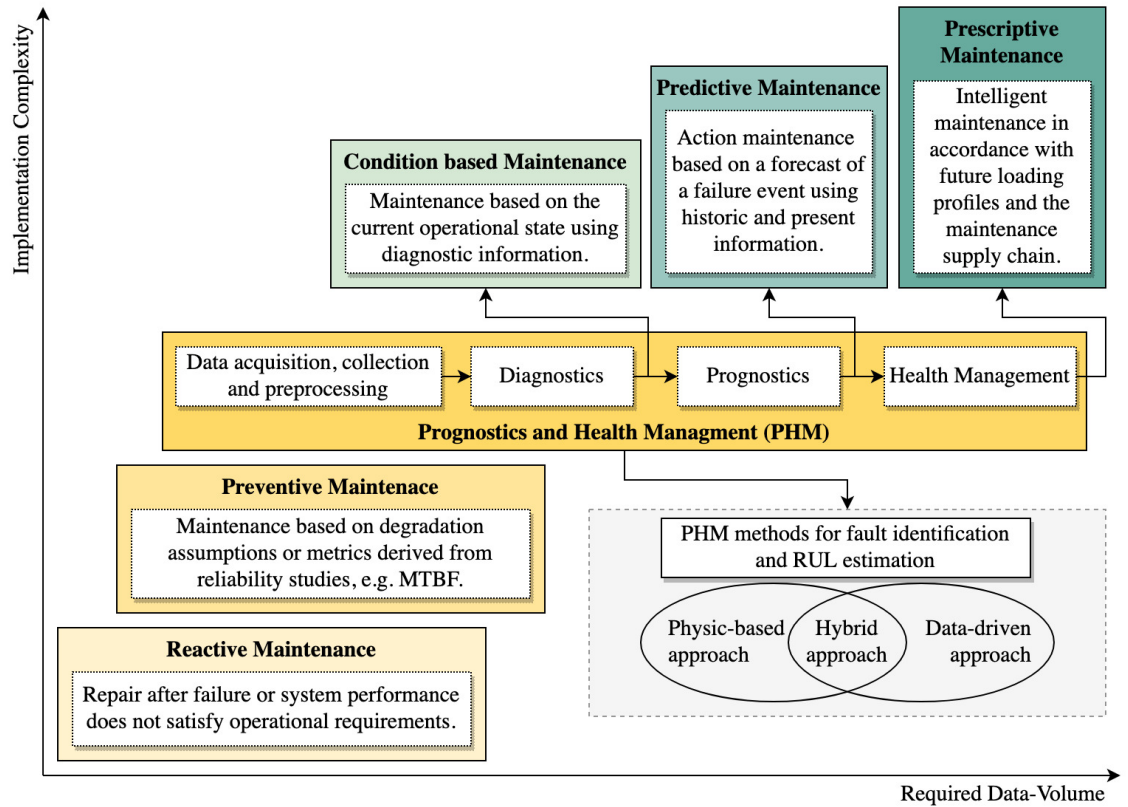


Figure 2.2: Schematic of PHM as a governing engineering discipline providing the methods for diagnostics and prognostics applications.

2.2.1 Key concepts

As illustrated in Figure 2.2, diagnostics and prognostics are important concepts within PHM. **Diagnostics** refers to the detection, isolation, and identification of a fault or failure [23]. The detection of a faulty condition is associated with an assessment of the severity of the degradation [24]. Less advanced than diagnostics [25], **prognostics** acts on the assumption that every component or system degrades over the course of its lifespan and fails eventually. Thus, it is concerned with modelling and extrapolating the process trajectory to the EOL, i.e. estimating the time to failure, on conjecture of past, current, and expected load profiles [15, 18, 26]. For the

prognostics approach, the RUL or a representative damage accumulation indicator referred to as Health Index (HI) are typical prediction targets. While the RUL provides a direct interpretation, the HI needs to be extrapolated up to a threshold [27]. RUL, as per Figure 2.3, refers to the time over which a component or system can perform its specified function. Detecting the deviation from its default operating state and extrapolating its propagation into future states allow failure mitigation and optimised maintenance intervals incorporated in, e.g. PdM.

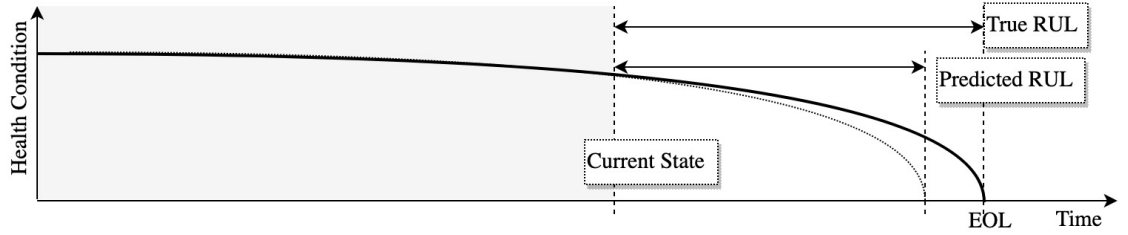


Figure 2.3: Simplified schematic of RUL estimation.

2.2.2 Frameworks

Various reference architectures have been established which describe the core elements of PHM in order to reduce ambiguity in development and implementation [28]. The National Institute of Standards (NIST) lays out the essentials of a PHM system as it distinguishes consecutive diagnostic and prognostic techniques [29]. The strategy puts a workflow forward where performance measures are continuously extracted from aggregated, preprocessed sensor data and ingested by a prognostic model. The system's degradation process is estimated and maintenance planned accordingly. This understanding of PHM is complemented by the IEEE-Std. 1856-2017 for electrical systems which elaborates on three PHM core components: *sensors*, *health-monitoring and assessment*, and *health-management* [18]. First, sensors are installed to collect system data. Second, the aggregated data is analysed to assess the system state, i.e. fault detection, isolation, and identification. Facilitated by a prognostics model, performance metrics are derived. On this basis an advisory function is interpreting the results depicting the predicted degradation or rather the development of the future system performance. Third, the health-management process aims to return the system to its default conditions by recovering from failure. It can incorporate functionality which, additionally, suggests advice for repair and maintenance as well as the mitigation of anticipated failures by adjusting the operation regime accordingly. A systemic view of PHM is proposed by [26], likewise defining a three layer PHM arrangement. *Observe* - system health data is continuously collected and preprocessed which involves de-noising and feature extraction. *Analyse* - the system state is determined to assess the level of degradation. Therefore, a diagnostic model identifies and isolates faults. Following the health assessment, a prognostic model - under consideration of the asset's history - derives a

Table 2.1: Software tools for PHM.

Ref.	Name	Developer	GUI	Description
[31]	PHM Software	CALCE	Yes	A PHM toolbox that handles data pre-processing and feature extraction. Various ML models are implemented for anomaly detection, diagnostics, and prognostics tasks.
[31]	PHM Maintenance Planning Tool	CALCE	Yes	Modelling of PHM implementation costs and maintenance planning.
[32]	Predictive Maintenance Toolbox	MATLAB	Yes	Diagnostics and prognostics modelling based on condition monitoring data using, e.g. survival models or NN.
[33]	Prognostics Algorithm Python Package	NASA	No	Physics-of-failure models for diagnostics and prognostics applications.
[34]	Pysurvival	Pysurvival	No	Various parametric, non-parametric, and ensemble algorithms for modelling the RUL.

RUL forecast with an associated measure of confidence. Once, new data is available, the diagnostic and prognostic model are updated. *Act* - based on the estimation of the prognostic model a health management function proposes the best action to transfer the system back to its default, healthy operational state. In conclusion, common among all reviewed frameworks, PHM consists of a set of structural, inter-dependent elements: sensing and processing hardware, a diagnostics or prognostics method, and an implementation of a health management function. Health management adapts the results of the prognostic approach and supports the maintenance decision making process to ensure the system integrity [26], either human-guided or automated [15]. ISO-13381-1 proposes a standardisation of prognostics taxonomy. The standard interlinks aspects of health management with prognostics which allows to use the prognostic information for maintenance decision making [30]. In particular the diagnostics and prognostics modelling aspect of these frameworks has been implemented in various software solutions used within industrial and research applications. Table 2.1 summarises a number of relevant developments.

2.2.3 Methods

The broad field of methodologies that PHM includes is commonly divided into three major areas, cf. Figure 2.2. Ultimately, the selected method depends on the physical understanding of relevant degradation processes, the availability of data and hardware as well as the requirements on model explainability [35] (refer to Chapter 2.4). **Physic-based methods** translate the physical nature of the designated system into a deterministic mathematical representation which depicts the degradation [26]. Physical models, sometimes termed physical damage models, are restricted by underlying assumptions enabling an analytical description of component or system inter-dependencies [10]. Such approaches are successful if a profound understanding of the degradation processes can be adequately mapped through a physical model

Table 2.2: Comparison of PHM methods.

	(I) Physic-based , (II) Data-driven , and (III) Hybrid
Approach	(I) The parameters of physical damage models are estimated from measured data. (II) Statistical and AI approaches extract trend information from the data. (III) Fusing physic-based with data-driven methods, either in series or in parallel.
Advantages	(I) Accuracy; provides an understanding of the present state estimation; less data required. (II) Less knowledge regarding the physical degradation processes is required; a wide range of powerful methods is available that can handle large amounts of data. (III) Compensating for drawbacks of the individual methods.
Disadvantages	(I) Deducing an adequate physical model is challenging when the system complexity increases; transfer from laboratory conditions affects the model performance; parameter estimation is affected by measurement noise and uncertainty. (II) High volumes of quality life-cycle data at different operational conditions are required; <i>black-box</i> models suffer from poor explainability. (III) Complex implementation.

[25]. Therefore, domain-specific expert knowledge as well as an extensive study of failure modes on a sub-component or even material level is required [28]. The parameters of physical models need to be estimated from collected data by methods like Kalman Filtering or Particle Filtering [17]. Unlike physical models, **data-driven approaches** are the preferred choice if a physical model cannot be obtained [36, 37]. Operational data, time series data, failure data, and reports to assess the system or component state are used to derive, e.g. the RUL without the need to explicitly model the physical processes [19]. Along traditional statistical approaches such as Gaussian Process Regression or Hidden Markov Models, AI is a popular field for data-driven PHM ranging from ML methods like Random-Forest to DL strategies [38]. A **hybrid approach** fuses *physical* and *data-driven* methods in order to compensate for the limitations of each individual approach [14]. Hybrid-series approaches use data-driven approaches to infer missing data for physical models [26]. If data-driven methods complement physical models to account for aspects not tracked by the physical models they are referred to as hybrid-parallel [35]. A comparison of the approaches is presented in Table 2.2.

2.2.4 Condition-based, predictive, and prescriptive maintenance

Industries with a paramount focus on safety and availability such as the energy or aviation industry are striving to evolve their policies from reactive or preventive maintenance towards more dynamic maintenance paradigms with the intention to improve the scheduling of the maintenance action and to reduce related costs

[39]. Such novel approaches, as illustrated in Figure 2.2, are Condition Based Maintenance (CBM) and Predictive Maintenance (PdM) situated within the realm of PHM. By collecting (real-time) condition monitoring data through sensors, a more accurate picture of the asset's deterioration can be derived. Condition monitoring through processing this data can be applied to report imminent faults. CBM and PdM adopt such strategies. Note, however, a persisting terminological inconsistency because the concept of PHM has evolved over the last decade. CBM enables maintenance work to be carried out when the current measured condition no longer meets the expectations of a regularly operating system within the defined framework conditions [40]. In PdM, using a prognostics method, a failure event is predicted. The anticipation of the failure is the main advantage of PdM over conservative maintenance strategies [22]. PdM updates the maintenance frequency and reduces unplanned downtime whilst minimising too early maintenance. Hence, its dynamic nature makes predetermined intervals superfluous and improves safety and reliability by avoiding unscheduled maintenance. Further, it simplifies logistic aspects of spare part availability [18, 15]. Based on this, Prescriptive Maintenance (PrM) extends the concept of PdM. Including the capability of making appropriate operational adjustments PrM issues maintenance actions through a strategic feedback that updates the maintenance policy [41]. The prognostics model, for example, anticipates the failure. The subsequent prescriptive method determines the optimal point for maintenance in consideration of future loading and supply chain availability considerations [42]. Research is concerned with all these aspects, developing advanced sensing technologies for condition monitoring, distinct methods to detect impending faults from the streamed data, forecast the fault propagation, and the methods used to optimise the subsequent maintenance decision making [43, 44].

2.2.5 Challenges

Regardless of the methodological approach, there are fundamental challenges opposing the ready installation of PHM in industrial settings. General guidelines are required to standardise data aggregation across the entire industrial infrastructure [45]. It is opposed by location dependent factors, a changing dynamic operation environment, and considerable signal noise [46]. Thus, the chosen approach needs to be sensitive enough to reliably detect state changes, classify known faults, detect unknown faults, but should be invariant to noise. Data availability is often limited; moreover, historical data sets might not be available in digital or standardised form. Computational efficiency plays a crucial role in order to alleviate front-end systems, also with regard to energy consumption. It poses a deployment challenge, since the volume of sensing sources is growing, whereas embedded computing resources remain limited. Industrial fleets, usually, consist of various generations of assets developed and serviced over long time spans. They require specific knowledge

for each asset generation because they may exhibit distinct fault characteristics. Therefore, the success of PHM is constrained when the required sensing technology is to be retrofitted into legacy systems [47]. The extent of these challenges depends on the industrial sector; the success of PHM is bound to the availability of data and in turn significantly influenced by the maturity of digitalisation. Thus, the key technological enabler is considered to be the *Industry 4.0* of which the Industrial Internet of Things and BDA represent core elements [48]. Lastly, the adoption of these technologies challenges existing domain expertise. Therefore, besides the dissemination of smart sensor technology and secure cloud infrastructure to facilitate BDA, the trust of stakeholders plays a central role, since the implementation of a digital ecosystem acts as backbone technology for successful PHM. In Chapter 2.5 these implementation challenges, the opportunities, and the existing reservations are illustrated based on the oil and gas industry.

2.3 Data-driven techniques

Individual electronic and electromechanical components often comprise complex electronic assemblies that interact and influence each other. Hence, a model needs to be able to depict or consider the system interactions with respective subsystems and its environment [25]. Thus, with such rising complexity, physical models might not be a viable solution due to the associated development and implementation costs, since deducing the relevant physical principles becomes infeasible [17, 49]. Data-driven approaches provide an alternative. Industrial assets are armed with abundant sensing equipment already capturing a wide range of condition monitoring data (vibration, temperature, pressure, etc.), images, videos, and maintenance reports [46]. This accumulation in sensing capabilities, enabled through improvements and availability of sensor and computing hardware, calls for data-driven methods [25]. In combination with supporting historical data, such techniques provide readily implementable, computational and cost efficient alternatives to physical models which reduces or mitigates the need for specific expert knowledge [13, 28, 36]. Generally, using some ML method, patterns and trends are first extracted from the measured signals and then correlated with their respective failure class or a cumulative degradation measure. As stated, the probability of failure prior to a maintenance action or the RUL are common prognostic targets [25]. At the same time, the sole reliance on data poses a disadvantage. The reliability of the proposed method highly depends on the complete coverage of healthy and faulty operation states and the accuracy of the aggregated data [15].

Figure 2.4 illustrates the plethora of methods that can be applied or combined. Traditionally, **statistical methods** have been used on failure data, e.g. counting the number and time of failures for a population. Increasingly, they are directly

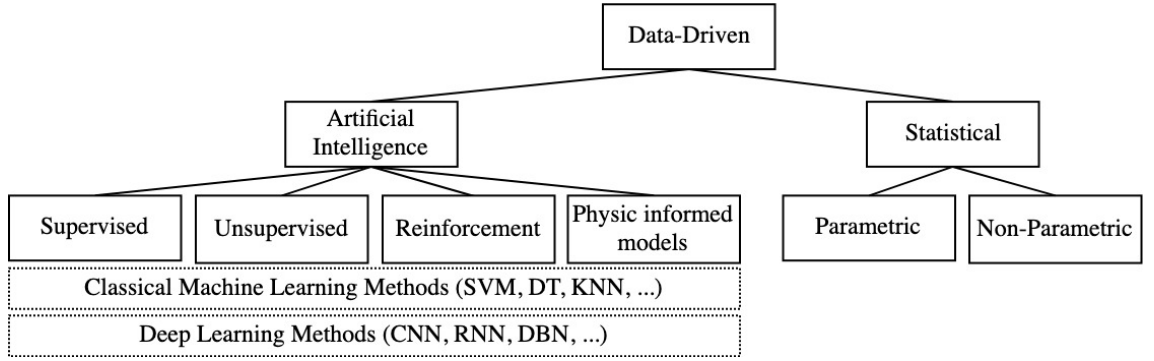


Figure 2.4: Statistical methods and ML might require higher levels of feature abstraction. In contrast, representation learning methods such as DL can learn simple and higher levels of abstracted feature representations directly from the raw data.

applied to condition monitoring data. Statistical methods derive a representation from the data (for example in the form of a Probability Density Function (PDF)) that correlates with an expected healthy operation state and, concurrently, rejects new observations if they do not match the derived statistical properties [13, 36]. Statistical methods assume underlying statistical characteristics, e.g. normally distributed data. However, this assumption does not hold true in many real world applications, where the underlying PDF is not easily obtainable, especially in the presence of high-dimensional data sets. As a rule, parametric and non-parametric approaches are distinguished. If the data follows the assumed distribution parametric methods derive their parameters on basis of the assumed distribution. If no assumption regarding the statistical characteristics of the data can be made because the underlying physics cannot be accurately identified and the evolution of the deterioration process cannot be traced, then, non-parametric methods can be applied [50]. The number of parameters used by non-parametric methods is not restricted. Like parametric methods, a wide range of non-parametric methods have found application for PHM, cf. [51].

Apart from statistical methods, more recently, several AI approaches have been established. The common defining denominator of these methods is the ability to automatically learn given tasks or patterns from high volumes of data without the need for explicit programming and annotation [52]. As to diagnostics and prognostics, this implies a direct mapping of the degradation to the target variable (for example the RUL) from processed data inputs through a learnable model, i.e. *learn-by-example* [17]. Generally spoken, an AI model aims to minimise an objective function over all training examples. Thus, it measures the error, i.e. the distance between a predicted value and a target. Improving the model's performance is achieved by iteratively adjusting the weights or parameters of the model which can be many millions in the case of DL models. Three general approaches are distinguished, namely supervised learning, unsupervised learning, and reinforcement learning. **Supervised learning** maps an input to a provided output. This mapping is learned from the labelled

data, i.e. the target [53]. Within supervised learning, classification tasks (e.g. a set of discrete values referring to individual fault classes) and regression tasks are further distinguished. While classification is used for diagnostic purposes, regression is of value to prognostics by forecasting RUL. If the desired output is unknown, e.g. various unidentified fault classes, **unsupervised learning** methods isolate groups of similar samples by identifying patterns in the input data, e.g. the correlation of an operational pattern with a distinct fault state. Unsupervised learning can be of interest for anomaly detection in systems where faults occur infrequently and, in consequence, no such labels are available [46]. Relevant to the optimisation of maintenance scheduling, **reinforcement learning** is concerned with learning to take the best course of action within a certain environment through *trial and error*, cf. [54]. In a prognostics context transfer learning may be of interest to account for changes in operating conditions, cf. [55]. Lastly, as hybrid diagnostic or prognostic solutions, ML and DL models can be interwoven with physical cause-effect information [56].

The applicability of data-driven strategies to detect deviations from the expected operational state, e.g. AI for PHM is subject of extensive research. As a rule, in a first instance research is concerned with the detection of anomalies [57]. Only in subsequent stages the classification of operational states, faults and failures, as well as predicting the trajectory of the degradation are of interest [15, 25]. Consider the example of an electrical component, namely the Insulated Gate Bipolar Transistor (IGBT). First, based on collected historical data, in [58] a method using a distance measure for anomaly detection of IGBTs is presented. Then, in [59], not only the non-nominal operating state of IGBTs is detected, but also fifteen different fault modes are classified as part of a diagnostics framework. Lastly, the authors in [60, 61] evidence how data-driven methods are used to estimate IGBT-RUL applying, e.g. Neural Network (NN) or Fuzzy-Logic relying on Accelerated Life Cycle Testing (ALT) data.

2.3.1 Machine learning and deep learning

ML algorithms such as Support Vector Machine (SVM), Decision Tree (DT), or K-Means Classifier are used for prognostic applications. However, these classical ML algorithms are limited in their capability to process raw data [52]. They often face poor generalisation, despite a variety of techniques for feature engineering that highlight discriminative aspects or suppress noise. For prognostics carefully hand-tuned, expert-engineered features depending on intrinsic domain knowledge are used. Representation learning in general (methods which learn the salient information for, e.g. classifying an input) and DL in particular circumvents this critical step automatising the feature learning from a raw, generic input [17].

DL is a modern term for the use of NN in the field of AI. A strong scientific and

industrial interest in the capabilities of DL has been initiated by the recent increase in computing performance. This, in turn, is driven by the development of Graphics Processing Units (GPU)s, BD, and an improved access to these resources [52]. Contrary to the early adoptions of NNs composed of two or three consecutive layers, nowadays NN architectures can be made up of multiple hundred layers, hence, the term *deep* learning. The many-layered models containing trainable parameters in each layer propagate an input to an output via a series of linear or non-linear mappings. Saliency of relevant features is increased with each subsequent layer which enables the model to learn multiple abstractions of the input representation [52]. Non-linear deep networks represent intricate functions able to register small relevant variations whilst suppressing irrelevant information contained in the input. DL excels on the detection of hidden structures and patterns in high-dimensional and spatially large inputs as the success of relevant research in classification and regression tasks has demonstrated. Applications are manifold such as image, text, and speech recognition [62, 63]. Other examples of Deep NNs include Restricted-Boltzman Machines in the form of Deep-Belief Networks or Deep-Boltzman Machines, Auto Encoder (AE) for denoising or feature extraction, CNN and RNN as further detailed in Chapter 5.2.1. In a prognostics context, key properties of DL can be summarised as follows: models can learn non-linear degradation processes through extraction of salient features from multi-dimensional data without explicit programming. Main disadvantages are the *black-box* character of these approaches as well as the demand for large quantities of training data.

2.4 Tailoring data-driven diagnostics and prognostics

The choice of an appropriate data-driven model is governed by various application specific factors, as detailed in Figure 2.5. One can identify fundamental, mutually related aspects such as the availability, quantity, resolution, and expressiveness of the data. In turn, this directly impacts the prognostics depth, i.e. whether the development of a fleet-wide, a system or even a component specific model is desirable. The scope of the considered application decides if an online approach is required or an offline implementation suffices. In a later instance, the tailoring of the data-driven strategy must be guided by aspects like generalisability beyond the acquired data set (transfer and adaption to different failure modes), operating conditions, or types of components and systems. The remainder of this section discusses but some selected aspects, although a multitude of factors may be considered.

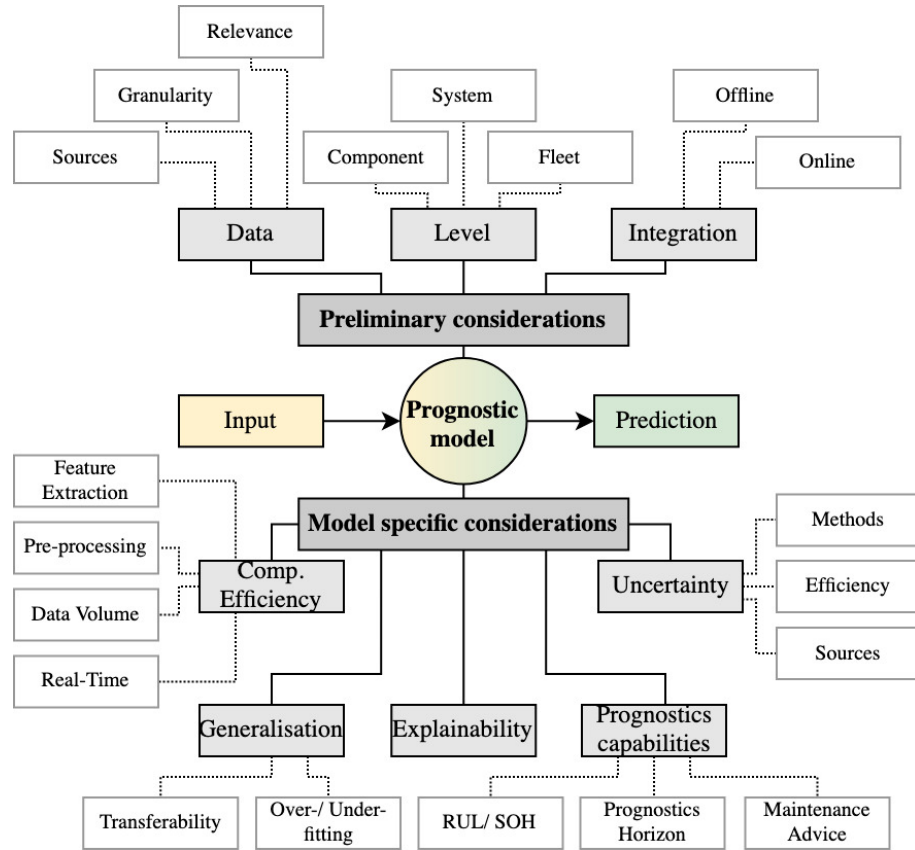


Figure 2.5: Overview of important considerations for selecting a data-driven prognostics approach.

2.4.1 Data

The selection of the prognostics approach is guided by the type of collected data [15]. To be more precise, it depends on the historic and future availability of the data (in batches or as continuous streams), its quantity, granularity or resolution, and sample size (number of individual components). Finally, the meaningfulness of the data collected in regard to the monitored wear processes has to be considered as it relates to the ability to map a monotone degradation process. Thus, it is non-trivial to select a method that operates efficiently on large, heterogeneous inputs and updates its prognostic's forecast based on recent changes. This process is further complicated by the fact that the amount of extracted data is growing exponentially along with the techniques to collect it [46].

A significant challenge of data-driven methods is the demand for large amounts of data. The data must cover the entire deterioration process under different operating regimes, cf. Chapter 6 which is concerned with the creation of such a data set. Data can be obtained from multiple sources and through different techniques. Figure 2.6 provides an overview of relevant data sources. Comprehensive sets of infield failure data that offer a complete coverage of the component's or system's operational history are often not easily to obtain and costly [15]. Further, failures might occur infrequently, since preventive maintenance actions lead to a scheduled, but early

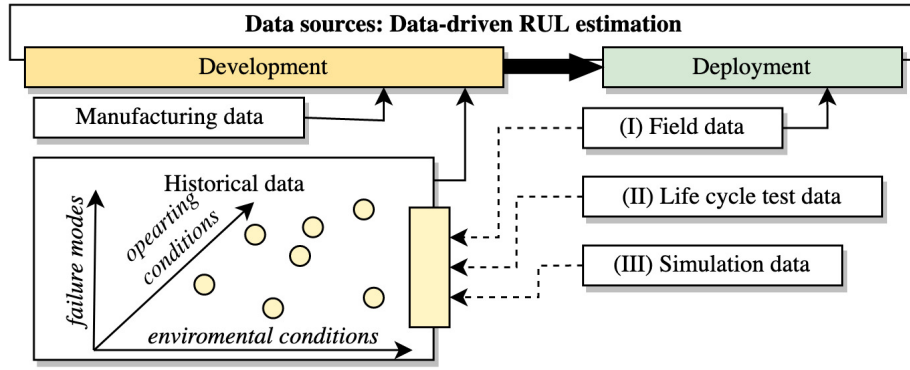
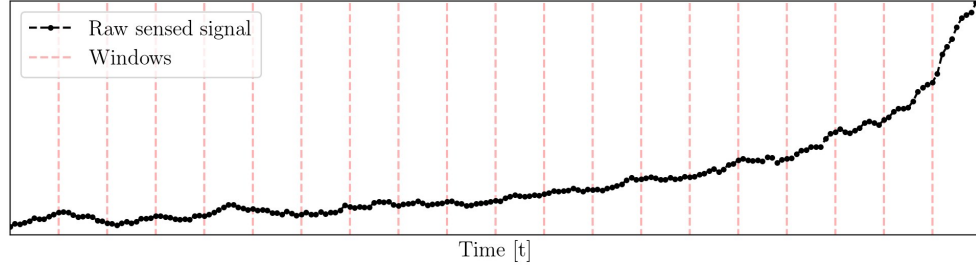


Figure 2.6: Data-driven prognostics require significant amounts of historical data in order to develop models for RUL prediction. Sources of historical data have to cover different environmental conditions, operating patterns, and failure modes which the component or system might experience: (I) field data from already deployed equipment using condition monitoring information; (II) life-cycle data which encompasses the generation of complete run-to-failure data sets for multiple components or systems. This data is collected in a laboratory setting where the exposure and influence of external factors can be controlled; (III) simulation of life-cycle data for distinct components or entire systems.

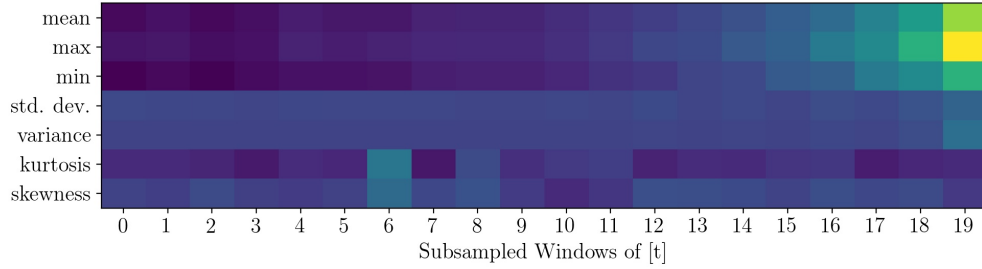
exchange. In addition, data gathered from infield sensors does not provide comprehensive monitoring if not enough sensing capabilities are installed [64]. Therefore, experimental test beds constitute an alternative to obtain run-to-failure data in order to develop data-driven prognostic models. On the downside, life-cycle test beds are expensive and require the availability of testable components or systems [17]. In many instances, acceleration of the degradation through ALT or Highly Accelerated Life Cycle Testing (HALT) is required to make components fail in reasonable time [65]. This may falsify the degradation mechanisms to the point that not all failure modes are sufficiently covered. Popular life-cycle tests and associated data sets used for benchmarking prognostic methods have been presented by [66, 67]. Lastly, simulation has been put forward as a method to obtain large quantities of life-cycle data [68]. However, simulating the degradation requires an in-depth understanding of the wear processes which - similar to a physic-based model - is difficult to obtain in case the complexity of the considered application grows.

2.4.2 Feature extraction

Prior to the development of any data-driven method, the acquired data needs to be processed. Thus, coined as feature extraction, the derivation of meaningful data representations from the raw signals and the subsequent denoising and compression of an input is an established practice [64]. Features aim to emphasise hidden trends in the raw data that are able to adequately map the component or system deterioration process [69]. The feature extraction process can be automatised as further detailed in Chapter 5.2.1.2, whereas, in prognostics it is commonly a manual, guided process informed by expert knowledge. For example, summary statistics as in Figure



(a) The raw data from a sensor measurement.



(b) Summary statistics for the sensed signal averaged over each window.

Figure 2.7: Feature extraction through waveform windowing.

2.7b can provide a downsampled feature representation of the original signal in the time-domain which is useful for diagnostic and prognostic purposes [70]. Likewise, signal processing methods inspecting the frequency-domain such as time-frequency representations, e.g. Short-Time Fourier Transform or Wavelet Transform, are techniques to derive expressive features [71, 72, 73, 74]. Prognostic applications usually deal with some form of time series data which is distinguished from other data, beside its dependence on time, by certain attributes. These components are referred to as trend, seasonality, cycle, and irregularity. Depending on the sampling rate, real-world time series data can be considered continuous [75]. In particular, prognostics is concerned with non-stationary time series data that may exhibit trends or cyclic patterns and non-regular behaviour, induced by random external influences and advancing fatigue.

While diagnostics requires features that aim to maximise the separability of nominal and non-nominal operating states from the data, i.e. faulty and non-faulty classes, prognostics is concerned with the establishment of continuous features that relate to certain intervals in the component's and system's lives, i.e. how one health state class propagates to another [76]. Thus, features are to be derived that are as monotonous as possible (either increasing or decreasing uniformly throughout the entire life). In this regard, the monotonicity of one given feature or a set of features should be verified across all samples in order to only retain such features with good predictability [64].

As shown in Figure 2.8, data preprocessing does not only involve feature extraction

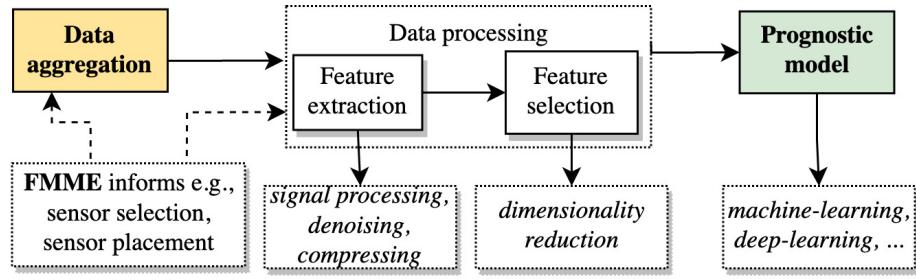


Figure 2.8: Data preprocessing steps. Failure Modes Mechanisms and Effects Analysis (FMMEA) can guide this process.

but also feature selection. Selecting a subset from the original feature set that is relevant to the degradation can improve model performance, reduce computational costs, and simplify the interpretation of results [25, 26, 77, 78]. Thus, as summarised by [79], a selected feature subset should be highly correlated with the predictive target, but the features themselves should not be correlated among each other. For example, in a supervised classification problem, the main objective would be to derive a subset of features that maximise the accuracy of the model [78]. So, the feature selection removes irrelevant and redundant features. Common measures for feature correlation are, e.g. the Correlation Coefficient based on the covariance and variance of two feature vectors, the Pearson Correlation, or the Euclidean Distance. Principal Component Analysis (PCA) is another popular method to reduce the dimensionality of the original data while retaining its relationship by ranking the features [80]. In a prognostics context, PCA has been applied for RUL estimation of bearings by [81]. Other approaches could be considered, cf. [78].

2.4.3 Depth of implementation

The explanatory context provided by the diagnostics or prognostics method follows the depth of implementation, i.e. whether an entire fleet, a single system, a sub-system, or a sole component are to be considered [17]. Table 2.3 provides some qualitative comparison. The vast majority of current research approaches prognostics from a single-unit, component-specific viewpoint rather than a system-wide perspective [82]. Besides mechanical components such as bearings, electronic components are of interest to diagnostics and prognostics research, e.g. Metal Oxide Semiconductor Field Effect Transistor (MOSFET), Light Emitting Diode (LED), IGBT, Capacitor [83]. For example, [84] presents a prognostics framework using the least square method. The authors rely on condition monitoring data to predict the MOSFET's EOL at a predetermined threshold based on changes in resistance during the *on-state*. Comparable prognostic applications for MOSFET-RUL estimation have been proposed, cf. [85, 86]. It should be noted that the majority of researched frameworks do only consider a single failure mechanism, i.e. die-attach degradation [87]. Various ML approaches such as K-Nearest Neighbour (KNN),

Table 2.3: Comparison of the requirements and the expressive power of prognostics. A *targeted maintenance action* relates to the ability of the deployed model to recommend which maintenance action to take in detail to remedy an identified failure.

Level	Data resolution	Number of individual samples	Targeted maintenance action
Fleet	+	+++	+
System/ Circuit	++	++	++
Component	+++	+	+++

Relevance Vector Machine (RVM), or SVM have been proposed for LED anomaly detection and RUL estimation [88, 89, 90]. Likewise, e.g. Long Short Term Memory Neural Network (LSTM) and RNN have been evaluated on their capabilities to predict RUL based on a decay of the initial light emission [91]. As it is the case for physical models, such data-driven models rely on measurements which relate directly to the degradation of the individual, monitored component [17]. To allow for a timely RUL estimation, measurements must be obtained in real-time and cover the degradation process [92]. Hence, an approach will face economical and technical constraints in practical application due to two major reasons. First, in electronic-rich systems, a continuously increasing number of electronic components renders the implementation of condition monitoring and subsequent prognostic modelling for each individual component infeasible. Second, as electronic components usually reside within circuits, obtained measurements such as voltage, current, and resistance are affected by other components of this circuit. To meet these challenges, data-driven models have been proposed that prioritise the cumulative deterioration state of an entire system (circuit) over the wear of its individual components. For example, a circuit’s response is considered by [93] using a SVM diagnostic model. A subsequent prognostic method deploys particle filtering triggered upon the detection of a non-nominal operation. After these explanations regarding the methodologies for component and system prognostics, finally, fleet-wide prognostics will be considered. At this scale, predicting RUL at the component level or detecting faults is not of concern. Instead, highlighting performance deviations within a fleet is of great interest [45]. Though fleet-level prognostics might support overall performance monitoring of a multitude of assets, in practice it cannot enable prescriptive maintenance, unless a system- or component-level model for a particular asset can be consulted. Due to the low data resolution the failure modes are not identifiable. In turn, the automation of necessary maintenance actions to prevent a failure is not possible. Within this realm of prognostics, concepts such as the digital twin that target the deployment of prognostic approaches are under development [94].

2.4.4 Embedding

Prognostics can be applied in an online fashion (sometimes referred to as on-board health monitoring), supported through condition monitoring data, aiding real-time mission planning [25]. For example, embedded prognostics applications enable electronics to self-test and verify their operational state [13]. Therefore, the ability of a method under consideration of its computational complexity to produce reliable RUL estimates is a guiding selection criterion for online prognostics. On the other hand, if real-time estimation of RUL is not a concern, offline prognostics can be used. Such an approach, implemented as a framework for subsequent maintenance support, is not overly concerned with aspects of computational costs. It can also tap into additional data sources that are not available in an online environment. Since a forecast will not be produced in real-time, offline prognostics is selected whenever the functioning of the monitored component or system is not predominantly mission- or safety-critical and failures are rare [15].

2.4.5 Explainability

Diagnostics and prognostics models should support maintenance and business decisions respectively. Hence, trustworthiness of these models is crucial. However, unlike physical models, data-driven ML methods are not primed with fundamental physical principles of the application, its failure modes and mechanisms [25]. And yet, if researched solutions are supposed to propagate into an industrial setting, the explainability of AI which entails transparency through the explanation of the model's reasoning is paramount [95]. The simplicity of the model remains an important consideration, as gains in performance through an increase of the model's complexity often come at the expense of explainability [84]. For example, DT models are easy to interpret through visualisation, but suffer from overfitting. By contrast, Random Forest (an ensemble of DT) improves performance at the cost of explainability. Explainability not only relates to the prediction model itself, but rather to the entire prognostics pipeline, e.g. pre-processing steps such as feature selection. An evaluation of the interaction and impact of each single feature on the model's estimate can further improve the explanatory power [96, 97]. There are pertinent methods deployed in ML such as Local Interpretable Model-Agnostic Explanation, but so far they have found little application in the domain of diagnostics and prognostics, cf. [98].

2.4.6 Uncertainty

All prognostics related applications are inevitably affected by the uncertainty of the forecast. Hence, uncertainty needs to be explicitly addressed in order to provide a meaningful RUL forecast [26]. Though, as later evidenced, cf. Chapter 4.2, prognos-

tics research ignores uncertainty in many instances. The standard classification of uncertainty into *epistemic* and *aleatic* sources can be adopted for prognostics, but classifying uncertainty based on the occurrence along the prognostics process may be more appropriate [17]. First, sources of inherent uncertainty in the prognostics pipeline must be identified and quantified. They stem from production variance within the same type of monitored system or component, unknown past loading patterns, and current operational conditions. Equally, uncertainty induced by measurement errors has to be considered. Selection of appropriate sensors can reduce this type of uncertainty [15]. In addition to sensor noise, measurement errors, and uncertain state estimations the future loading pattern might be unknown. Uncertainty from modelling errors and selection of model parameters has to be taken into account [99]. In this instance, increasing the sample size commonly reduces model uncertainty.

Model uncertainty becomes especially important if DL approaches are selected, since they are widely considered to be *black-box* models [100]. Prognostic applications using DL are only recently brought to the fore which poses a challenge for an efficient quantification of uncertainty. A general overview of methods for uncertainty estimation to be integrated in DL architectures has been presented by [101]. A Bayesian approximation, termed MCD, to estimate predictive uncertainty has been proposed by [102]. It has been evaluated for prognostics, applied to steam generators and lithium-ion batteries in order to derive confidence bounds by [103, 104]. This method is further detailed in Chapter 5.2.6.4.

2.5 Survey of the digital transformation in the oil and gas industry

The efficiency of upstream, midstream, and downstream operations is crucial for the oil and gas industry. On one hand, due to the strong volatility of oil and gas prices, on the other hand, due to overall challenging energy market conditions in combination with stricter environmental regulations [105]. Thus, in accordance with the efficiency of maintenance strategies, the industry has readily identified the need to develop application tailored methods in order to retain high levels of fleet availability [106, 107]. Yet, one has also to consider efforts taken beyond the oil and gas service industry [108]. Examples of novel, data-driven approaches to maintenance have been implemented throughout the aviation industry [109], the automotive industry [110], the manufacturing industry [14], and the offshore wind turbine industry [111]. Here, the challenges are similar, in many respects, to those faced by the oil and gas industry. At present, in comparison to other industries, the oil and gas industry, in particular the upstream sector, is conspicuous for its comparatively limited adoption

of modern digital infrastructure, industry-wide standardised digital ecosystems, and novel maintenance strategies [105, 112, 113]. Consider, for instance, the subpar deployment of already collected data to guide decision making [105, 114]. As a result, disparities between the upstream, midstream, and downstream industry and towards other industries in terms of digital maturity are evident. This issue is reflected by the growing number of publications and industrial surveys that are concerned with facilitating the digitalisation of the oil and gas industry, cf. Chapter 4.1. Such reports, commonly, identify the digital transformation as a driver for efficiency concurrently reducing costs - an essential process in order to meet the estimated energy demand [115, 116]. Moreover, in order to facilitate the deployment of data-driven solutions, extensive collaboration between key players hand in hand with less aversion towards greater reliance on data-driven solutions is imperative [114]. A common industry-wide framework to promote, align and verify digitalisation is being promoted under the term *Oil and Gas 4.0* [117]. As for the improvement of AMO through PHM, governing considerations are the implementation of the required digital infrastructure, data streamlining and aggregation, a shared digital environment, an industry wide Industrial-Internet-of-Things, and BDA [118]. Table 2.4 summarises these key findings aggregated from selected industry surveys, providing a clear picture of the main digital transformation barriers encountered in the oil and gas industry, e.g.: lacking trust in the capabilities of data-driven methods and a conservative mindset; no scalable infrastructure that facilitates data collection and analysis.

Table 2.4: Survey of oil and gas related studies that assess the state of the digital transformation.

Ref.	Key Findings (I), Opportunities (II), and Challenges (III)
[105]	<p>(I) Key enabler digitalisation; terabytes of data readily available for decision making, however, until now only small amounts of data are effectively used; estimated \$1.6 trillion potential of digital transformation; remarkable increase in application of BDA and related technologies; essential collaboration between key players.</p> <p>(II) Data-driven digital life-cycle management strategies within an industry wide digital ecosystem; data-driven, real-time cost and business models to reduce costs, improve efficiency and safety.</p> <p>(III) Conservative approach; Non-systematic, highly selective adoption of digital technology based on rudimentary data; current frameworks limit the exchange of data throughout the value chain; no industry-wide data/ sensor standards.</p>
[112]	<p>(I) Limited adaption of data-driven technologies compared to other industries; high costs, unknown scalability, marginal improvements and uncertainties are associated with the transition to digital solutions.</p> <p>(II) Declining costs for sensor technology, etc.; potential of increased efficiency and productivity; improved resilience against market fluctuations; promotion of layered digital integration.</p> <p>(III) Fast-paced digital market impedes identification of viable technologies; few flagship projects promoting the benefits; insufficient legacy digital infrastructure due to scale of operations and obsolete asset databases; lack of standards; data silos.</p>

Continued on next page

Table 2.4 – continued from previous page

Ref.	Key Findings (I), Opportunities (II), and Challenges (III)
[115]	<p>(I) Efficient technologies will be needed to meet the predicted increase in energy demand; standardised digital frameworks are key cost savers; majority of data-driven applications are used for asset monitoring.</p> <p>(II) Digital solutions speed up complex decision making processes through an evolution of operational and maintenance practices.</p> <p>(III) Costs of digital integration; risk adverse culture opposing the transfer from physical models towards data-driven approaches.</p>
[116]	<p>(I) Low level of digital transformation compared with other industries; digital technology perceived as key tool to remain successful in the face of increasingly remote and difficult to access conventional resources.</p> <p>(II) Reduction in on-site personnel through enhanced remote monitoring technologies; provision of novel revenue sources; growing levels of automation; improved capability to adapt to current market demands; optimisation of the asset life-cycle.</p> <p>(III) Cyber-security; financial constraints; aspects of remote asset connectivity; data integration and verification.</p>
[114]	<p>(I) Limited adaption of digital technologies; so far only few lighthouse projects; collaboration efforts sharing data in order to develop new digital strategies and business models.</p> <p>(II) Improving operational aspects through, e.g. PHM; efficiency and safety.</p> <p>(III) Data is only partially utilised for decision making, despite significant amounts of data being collected; data is often not suitable for consecutive analytics; cyber-security; lack of digital standards; despite the recognised potential a change opposing cultural exists; requiring fully proven solution prior to implementation; limited understanding and awareness of the potentials and pitfalls of digital technology.</p>
[113]	<p>(I) Striking lack of digital solutions in the upstream industry; automation of the drilling process enables higher levels of connectivity among subsystems of the rig; industry wide digital ecosystem to drive the transformation.</p> <p>(II) Connecting technologies among the entire supply chain; cross-organisational barriers transfer the classic linear supply chain to an integrated process entailing customers and service providers; novel data-driven strategies reduce procurement and supply chain costs.</p> <p>(III) Data silos within upstream, midstream, and downstream assets; overcoming cultural barriers and building trust.</p>
[107]	<p>(I) Costs for interconnecting industrial assets have decreased; midstream and downstream sectors adjust business models to cope with the changing energy landscape; upstream industry is asked to improve efficiency through, e.g. PHM because business model changes are limited; digitalisation is the key technology to facilitate the transformation; alignment of the digital road map with future, strategic developments.</p> <p>(II) Reducing costs and increasing revenue, through digital frameworks relying on BDA for, e.g. project design, automatisisation of the drilling process and fleet reliability.</p> <p>(III) Developments of renewable energy sources challenge and limit the potential of investments in oil and gas related technologies; digitalisation is not an independent technical consideration because it requires a corporate-wide view as well as cultural shift.</p>
[118]	<p>(I) PdM strategies replace traditional maintenance paradigms in the long term; they are conceived as central deliverables from the digitalisation; establishing the infrastructure to sustain the paradigm shift towards data-driven business models.</p> <p>(II) Cloud technology is perceived as an important enabler for supply chain integration; it is critical for future, strategic decision making.</p> <p>(III) Cultural resistance; shift into a digital cloud environment is conceived as threat because it may lead to an on-premises control loss.</p>

Summary

This chapter laid the foundation upon which prognostics methods can be formulated. The associated concepts of reliability, diagnostics, and prognostics as well as the subsequent integration of novel maintenance strategies have been highlighted. Special emphasis has been placed on the role and configuration of data-driven prognostics. Central criteria in order to select and tailor apposite prognostics solutions were examined. Subsequently, the general challenges of PHM and the backbone technology setting, realised within the digital transformation, have been discussed. It has been shown that the oil and gas industry acknowledges the potential of this digital industrialisation. However, it is facing a laborious transition to meet the stipulated objectives.

To further support this transition process within the oil and gas industry and to highlight its importance, the subsequent chapters consider two scenarios, namely the BHA-PCBA (Chapter 3.1) and the EMR (Chapter 3.2) - complex assemblies of electronics found in downhole tools and an electromechanical component are studied. First, the failure modes and mechanisms of the respective applications are discussed. Second, the state-of-the-art of PHM relating to each application is presented in detail in Chapter 4. In doing so, the specific, practical challenges when deriving prognostic strategies for upstream assets and their electronic components are extracted. Such analysis emphasises that critical applications in niche sectors benefit from the use of data-driven methods which enable and expedite the digital transition and, thus, provide a viable alternative to physic-based models. This is exactly where my research takes effect, since it explores the use of data, subject to restrictions in volume, variety, and resolution – aspects that predominantly govern the development of tailored, data-driven maintenance support strategies, cf. Chapter 5.

References

- [1] Z Ma and A Krings. Survival analysis approach to reliability, survivability and prognostics and health management (PHM). In *2008 IEEE Aerospace Conference*, pages 1–20. IEEE, 2008. doi:10.1109/AERO.2008.4526634.
- [2] J Bentley. *Introduction to reliability and quality engineering*. Prentice Hall, 1999.
- [3] P Tobias and D Trindade. *Applied reliability*. CRC Press, 2011.
- [4] L Swanson. Linking maintenance strategies to performance. *International journal of production economics*, 70(3):237–244, 2001. doi:10.1016/S0925-5273(00)00067-0.
- [5] E Zio and M Compare. Evaluating maintenance policies by quantitative modeling and analysis. *Reliability Engineering & System Safety*, 109:53–65, 2013. doi:10.1016/j.ress.2012.08.002.
- [6] J Daily and J Peterson. Predictive maintenance: How big data analysis can improve maintenance. In *Supply chain integration challenges in commercial aerospace*, pages 267–278. Springer, 2017. doi:10.1007/978-3-319-46155-7_18.
- [7] R Randall. *Vibration-based condition monitoring: industrial, automotive and aerospace applications*. John Wiley & Sons, 2021.

- [8] Mazhar Ali Khan Malik. Reliable preventive maintenance scheduling. *AIIE transactions*, 11(3):221–228, 1979. doi:10.1080/05695557908974463.
- [9] W Scheuren, K Caldwell, G Goodman, and A Wegman. Joint strike fighter prognostics and health management. In *34th AIAA/ASME/SAE/ASEE Joint propulsion conference and exhibit*, page 3710, 1998. doi:10.2514/6.1998-3710.
- [10] J Lee, F Wu, W Zhao, M Ghaffari, L Liao, and D Siegel. Prognostics and health management design for rotary machinery systems - reviews, methodology and applications. *Mechanical systems and signal processing*, 42(1-2):314–334, 2014. doi:10.1016/j.ymssp.2013.06.004.
- [11] L Liao and F Köttig. Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction. *IEEE Transactions on Reliability*, 63(1):191–207, 2014. doi:10.1109/TR.2014.2299152.
- [12] S Rezvanizani, Z Liu, Y Chen, and J Lee. Review and recent advances in battery health monitoring and prognostics technologies for electric vehicle (EV) safety and mobility. *Journal of Power Sources*, 256:110–124, 2014. doi:10.1016/j.jpowsour.2014.01.085.
- [13] T Sutharssan, S Stoyanov, C Bailey, and C Yin. Prognostic and health management for engineering systems: a review of the data-driven approach and algorithms. *The Journal of Engineering*, 2015(7):215–222, 2015. doi:10.1049/joe.2014.0303.
- [14] T Xia, Y Dong, L Xiao, S Du, E Pan, and L Xi. Recent advances in prognostics and health management for advanced manufacturing paradigms. *Reliability Engineering & System Safety*, 178:255–268, 2018. doi:10.1016/j.res.2018.06.021.
- [15] K Goebel, M Daigle, A Saxena, I Roychoudhury, S Sankararaman, and J Celaya. *Prognostics: The science of making predictions*. CreateSpace Independent Publishing Platform, 2017.
- [16] I Roychoudhury, A Saxena, J Celaya, and K Goebel. Distilling the verification process for prognostics algorithms. In *Annual Conference of the PHM Society*, volume 5. PHM Society, 2013. doi:10.36001/phmconf.2013.v5i1.2318.
- [17] N Vichare and M Pecht. Prognostics and health management of electronics. *IEEE Transactions on Components and Packaging Technologies*, 29:222–229, 2006. doi:10.1109/TCAPT.2006.870387.
- [18] M Pecht. IEEE standard framework for prognostics and health management of electronic systems. Standard, IEEE Standards Association, NJ, USA, 2017. doi:10.1109/IEEESTD.2017.8227036.
- [19] O Dragomir, R Gourtevau, N Zerhouni, and F Dragomir. Framework for a distributed and hybrid prognostic system. *IFAC Proceedings Volumes*, 40(18):431–436, 2007.
- [20] C Teubert, M Daigle, S Sankararaman, K Goebel, and J Watkins. A generic software architecture for prognostics (GSAP). *International journal of prognostics and health management*, 8(2), 2017. doi:10.36001/ijphm.2017.v8i2.2618.
- [21] D Kozjek, D Kralj, and P Butala. A data-driven holistic approach to fault prognostics in a cyclic manufacturing process. *Procedia CIRP*, 63:664–669, 2017. doi:10.1016/j.procir.2017.03.109.
- [22] A Saxena, J Celaya, B Saha, S Saha, and K Goebel. Metrics for offline evaluation of prognostic performance. *International Journal of Prognostics and health management*, 1(1):4–23, 2010. doi:10.36001/ijphm.2010.v1i1.1336.
- [23] S Yang, D Xiang, A Bryant, P Mawby, L Ran, and P Tavner. Condition monitoring for device reliability in power electronic converters: A review. *IEEE transactions on power electronics*, 25(11):2734–2752, 2010. doi:10.1109/TPEL.2010.2049377.
- [24] S Uckun, K Goebel, and P Lucas. Standardizing research methods for prognostics. In *2008 International Conference on Prognostics and Health Management*, pages 1–10. IEEE, 2008. doi:10.1109/PHM.2008.4711437.
- [25] S Kumar and M Pecht. Modeling approaches for prognostics and health management of electronics. *International Journal of Performability Engineering*, 6(5):467, 2010.
- [26] K Javed, R Gouriveau, and N Zerhouni. State of the art and taxonomy of prognostics approaches, trends of prognostics applications and open issues towards maturity at different technology readiness levels. *Mechanical Systems and Signal Processing*, 94:214–236, 2017. doi:10.1016/j.ymssp.2017.01.050.
- [27] J Sim, S Kim, H Park, and J Choi. A tutorial for feature engineering in the prognostics and health management of gears and bearings. *Applied Sciences*, 10(16):5639, 2020. doi:10.3390/app10165639.
- [28] H Elattar, H Elminir, and A Riad. Prognostics: a literature review. *Complex & Intelligent Systems*, 2(2):125–154, 2016. doi:10.1007/s40747-016-0019-3.

- [29] G Vogl, B Weiss, and M Helu. A review of diagnostic and prognostic capabilities and best practices for manufacturing. *Journal of Intelligent Manufacturing*, 30(1):79–95, 2016. doi:10.1007/s10845-016-1228-8.
- [30] ICS. Condition monitoring and diagnostics of machines — prognostics — part 1: General guidelines. Standard, International Organization for Standardization, Geneva, CH, 2015.
- [31] Michael Pecht et al. Prognostics and health management, 2022. CALCE, University of Maryland. <https://calce.umd.edu/prognostics-and-health-management>.
- [32] Matlab. Predictive maintenance toolbox, 2022. The Mathworks Inc. <https://www.mathworks.com/products/predictive-maintenance.html>.
- [33] Christopher Teubert, Matteo Corbetta, and Chetan Kulkarni. Prognostics algorithm python package, 2022. https://github.com/nasa/prog_algs.
- [34] Stephane Fotso et al. PySurvival: Open source package for survival analysis modeling, 2019. <https://www.pysurvival.io/>.
- [35] H Meng and Y Li. A review on prognostics and health management (PHM) methods of lithium-ion batteries. *Renewable and Sustainable Energy Reviews*, 116:109405, 2019. doi:10.1016/j.rser.2019.109405.
- [36] K Tsui, N Chen, Q Zhou, Y Hai, and W Wang. Prognostics and health management: A review on data driven approaches. *Mathematical Problems in Engineering*, 2015, 2015. doi:10.1155/2015/793161.
- [37] V Nguyen, M Kefalas, K Yang, A Apostolidis, M Olhofer, S Limmer, and T Baeck. A review: Prognostics and health management in automotive and aerospace. *International Journal of Prognostics and Health Management*, 10, 2019. doi:10.36001/ijphm.2019.v10i2.2730.
- [38] D An, N Kim, and J Choi. Practical options for selecting data-driven or physics-based prognostics algorithms with reviews. *Reliability Engineering & System Safety*, 133:223–236, 2015. doi:10.1016/j.ress.2014.09.014.
- [39] I de Pater and M Mitici. Predictive maintenance for multi-component systems of repairables with remaining-useful-life prognostics and a limited stock of spare components. *Reliability Engineering & System Safety*, 214:107761, 2021. doi:10.1016/j.ress.2021.107761.
- [40] R Ahmad and S Kamaruddin. An overview of time-based and condition-based maintenance in industrial application. *Computers & industrial engineering*, 63(1):135–149, 2012. doi:10.1016/j.cie.2012.02.002.
- [41] K Nguyen and K Medjaher. A new dynamic predictive maintenance framework using deep learning for failure prognostics. *Reliability Engineering & System Safety*, 188:251–262, 2019. doi:10.1016/j.ress.2019.03.018.
- [42] H Marques and A Giacotto. Prescriptive maintenance: Building alternative plans for smart operations. In *10th Aerospace Technology Congress*, pages 231–236. Linköping University Electronic Press, 2019.
- [43] A Bousdekis, K Lepenioti, D Apostolou, and G Mentzas. Decision making in predictive maintenance: literature review and research agenda for industry 4.0. *IFAC*, 52(13):607–612, 2019. doi:10.1016/j.ifacol.2019.11.226.
- [44] A Elwany and N Gebraeel. Sensor-driven prognostic models for equipment replacement and spare parts inventory. *IIE Transactions*, 40(7):629–639, 2008. doi:10.1080/07408170701730818.
- [45] A Voisin, G Medina-Oliva, M Monnin, J Leger, and B Iung. Fleet-wide diagnostic and prognostic assessment. In *Annual Conference of the Prognostics and Health Management Society*. PHM Society, 2013. doi:10.36001/phmconf.2013.v5i1.2311.
- [46] O Fink, Q Wang, M Svensen, P Dersin, W Lee, and M Ducoffe. Potential, challenges and future directions for deep learning in prognostics and health management applications. *Engineering Applications of Artificial Intelligence*, 92:103678, 2020. doi:10.1016/j.engappai.2020.103678.
- [47] A López, A Márquez, J Fernández, and A Bolaños. Towards the industrial application of PHM: Challenges and methodological approach. In *PHM Society European Conference*, volume 2. PHM Society, 2014. doi:10.36001/phme.2014.v2i1.1563.
- [48] A Ustundag and E Cevikcan. *Industry 4.0: managing the digital transformation*. Springer, 2017. doi:10.1007/978-3-319-57870-5.
- [49] K Medjaher and N Zerhouni. Hybrid prognostic method applied to mechatronic systems. *The International Journal of Advanced Manufacturing Technology*, 69(1-4):823–834, 2013. doi:10.1007/s00170-013-5064-0.

- [50] R Zhou, N Serban, and N Gebraeel. Degradation modeling applied to residual lifetime prediction using functional data analysis. *The Annals of Applied Statistics*, pages 1586–1610, 2011.
- [51] A Prisacaru, P Gromala, M Jeronimo, B Han, and G Zhang. Prognostics and health monitoring of electronic system: A review. In *2017 18th International Conference on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems (EuroSimE)*, pages 1–11. IEEE, 2017. doi:10.1109/EuroSimE.2017.7926248.
- [52] Y LeCun and G Bengio, Yand Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015. doi:10.1038/nature14539.
- [53] I Goodfellow, Y Bengio, and A Courville. *Deep learning*, volume 1. MIT press Cambridge, 2016.
- [54] M Knowles, D Baglee, and S Wermter. Reinforcement learning for scheduling of maintenance. In *International Conference on Innovative Techniques and Applications of Artificial Intelligence*, pages 409–422. Springer, 2010. doi:10.1007/978-0-85729-130-1_31.
- [55] W Zhang, X Li, H Ma, Z Luo, and X Li. Transfer learning using deep representation regularization in remaining useful life prediction across operating conditions. *Reliability Engineering & System Safety*, 211:107556, 2021. doi:10.1016/j.ress.2021.107556.
- [56] R Nascimento and F Viana. Fleet prognosis with physics-informed recurrent neural networks. *arXiv preprint*, 2019. arXiv:1901.05512.
- [57] V Sotiris, W Peter, and M Pecht. Anomaly detection through a bayesian support vector machine. *IEEE Transactions on Reliability*, 59(2):277–286, 2010. doi:10.1109/TR.2010.2048740.
- [58] N Patil, D Das, and M Pecht. Anomaly detection for IGBTs using mahalanobis distance. *Microelectronics Reliability*, 55(7):1054–1059, 2015. doi:10.1016/j.microrel.2015.04.001.
- [59] M Fei, L Ning, M Huiyu, P Yi, S Haoyuan, and Z Jianyong. On-line fault diagnosis model for locomotive traction inverter based on wavelet transform and support vector machine. *Microelectronics Reliability*, 88:1274–1280, 2018. doi:10.1016/j.microrel.2018.06.069.
- [60] M Ahsan, S Stoyanov, and C Bailey. Data driven prognostics for predicting remaining useful life of IGBT. In *39th International Spring Seminar on Electronics Technology (ISSE)*, pages 273–278. IEEE, 2016. doi:10.1109/ISSE.2016.7563204.
- [61] A Alghassi, P Soulatiantork, M Samie, S Perinpanayagam, and M Faifer. Reliability enhance powertrain using fuzzy knowledge base prognostics model. In *17th European Conference on Power Electronics and Applications*, pages 1–9. IEEE, 2015. doi:10.1109/EPE.2015.7311768.
- [62] A Dhillon and G Verma. Convolutional neural network: a review of models, methodologies and applications to object detection. *Progress in Artificial Intelligence*, 9(2):85–112, 2020. doi:10.1007/s13748-019-00203-0.
- [63] A Odo, S McKenna, D Flynn, and J Vorstius. Towards the automatic visual monitoring of electricity pylons from aerial images. In *15th International Conference on Computer Vision Theory and Applications*, pages 566–573. SciTePress, 2020.
- [64] K Javed, R Gouriveau, R Zemouri, and N Zerhouni. Features selection procedure for prognostics: An approach based on predictability. *IFAC Proceedings Volumes*, 45(20):25–30, 2012. doi:10.3182/20120829-3-MX-2028.00165.
- [65] Y Xuerong, Y Qiong, and Z Guofu. Reliability assessment for electromagnetic relay based on time parameters degradation. In *11th International Conference on Electronic Packaging Technology & High Density Packaging*, pages 1269–1272. IEEE, 2010. doi:10.1109/ICEPT.2010.5582785.
- [66] A Saxena and K Goebel. Turbofan engine degradation simulation data set. *NASA Ames Prognostics Data Repository*, pages 1551–3203, 2008.
- [67] P Nectoux, R Gouriveau, K Medjaher, E Ramasso, B Chebel-Morello, N Zerhouni, and C Varnier. PRONOSTIA: An experimental platform for bearings accelerated degradation tests. In *IEEE International Conference on Prognostics and Health Management*, pages 1–8. IEEE, 2012.
- [68] M Djeziri, S Benmoussa, and M Benbouzid. Data-driven approach augmented in simulation for robust fault prognosis. *Engineering Applications of Artificial Intelligence*, 86:154–164, 2019. doi:10.1016/j.engappai.2019.09.002.
- [69] F Ahmadzadeh and J Lundberg. Remaining useful life estimation: review. *International Journal of System Assurance Engineering and Management*, 5(4):461–474, 2014. doi:10.1007/s13198-013-0195-0.

- [70] T Gao, Y Li, X Huang, and C Wang. Data-driven method for predicting remaining useful life of bearing based on bayesian theory. *Sensors*, 21(1):182, 2021. doi:10.3390/s21010182.
- [71] X Ding and Q He. Energy-fluctuated multiscale feature learning with Deep ConvNet for intelligent spindle bearing fault diagnosis. *IEEE Transactions on Instrumentation and Measurement*, 66:1926–1935, 2017. doi:10.1109/TIM.2017.2674738.
- [72] D Verstraete, A Ferrada, E Droguett, V Meruane, and M Modarres. Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings. *Shock and Vibration*, 2017:1–17, 2017. doi:10.1155/2017/5067651.
- [73] Y Liao, X Zeng, and W Li. Wavelet transform based convolutional neural network for gearbox fault classification. In *IEEE Prognostics and System Health Management Conference (PHM-Harbin)*, pages 1–6. IEEE, 2017. doi:10.1109/PHM.2017.8079274.
- [74] L Jing, T Wang, M Zhao, and P Wang. An adaptive multi-sensor data fusion method based on deep convolutional neural networks for fault diagnosis of planetary gearbox. *Sensors*, 17(2):414, 2017. doi:10.3390/s17020414.
- [75] G Nason. Stationary and non-stationary time series. *Statistics in volcanology*, 60, 2006.
- [76] K Medjaher, F Camci, and N Zerhouni. Feature extraction and evaluation for health assessment and failure prognostics. In *Proceedings of First European Conference of the Prognostics and Health Management Society*, pages 111–116. PHM Society, 2012. doi:10.36001/phme.2012.v1i1.1443.
- [77] Z Zhao, F Morstatter, S Sharma, S Alelyani, A Anand, and H Liu. Advancing feature selection research. *ASU feature selection repository*, pages 1–28, 2010.
- [78] J Cai, J Luo, S Wang, and S Yang. Feature selection in machine learning: A new perspective. *Neurocomputing*, 300:70–79, 2018. doi:10.1016/j.neucom.2017.11.077.
- [79] M Hall. *Correlation-based feature selection for machine learning*. PhD thesis, University of Waikato, 1999.
- [80] J Jackson. *A user’s guide to principal components*, volume 587. John Wiley & Sons, 2005.
- [81] A Tayade, S Patil, V Phalle, F Kazi, and S Powar. Remaining useful life (RUL) prediction of bearing by using regression model and principal component analysis (PCA) technique. *Vibroengineering Procedia*, 23:30–36, 2019. doi:10.21595/vp.2019.20617.
- [82] G Vogl, B Weiss, and M Helu. A review of diagnostic and prognostic capabilities and best practices for manufacturing. *Journal of Intelligent Manufacturing*, 30(1):79–95, 2019. doi:10.1007/s10845-016-1228-8.
- [83] C Bhargava, P Sharma, M Senthilkumar, S Padmanaban, V Ramachandaramurthy, Z Leonowicz, F Blaabjerg, and M Mitolo. Review of health prognostics and condition monitoring of electronic components. *IEEE Access*, 8:75163–75183, 2020. doi:10.1109/ACCESS.2020.2989410.
- [84] C Alonso-González, B Pulido, M Cartón, and A Bregon. A big data architecture for fault prognostics of electronic devices: Application to power MOSFETs. *IEEE Access*, 7:102160–102173, 2019. doi:10.1109/ACCESS.2019.2929111.
- [85] J Celaya, A Saxena, C Kulkarni, S Saha, and K Goebel. Prognostics approach for power MOSFET under thermal-stress aging. In *2012 Proceedings Annual Reliability and Maintainability Symposium*, pages 1–6. IEEE, 2012. doi:10.1109/RAMS.2012.6175487.
- [86] S Dusmez, H Duran, and B Akin. Remaining useful lifetime estimation for thermally stressed power MOSFETs based on on-state resistance variation. *IEEE Transactions on Industry Applications*, 52(3):2554–2563, 2016. doi:10.1109/TIA.2016.2518127.
- [87] J Celaya, A Saxena, S Saha, and K Goebel. Prognostics of power MOSFETs under thermal stress accelerated aging using data-driven and model-based methodologies. In *Annual Conference of the Prognostics and Health Management Society*. PHM Society, 2011. doi:10.36001/phmconf.2011.v3i1.1995.
- [88] M Chang, C Chen, D Das, and M Pecht. Anomaly detection of light-emitting diodes using the similarity-based metric test. *IEEE Transactions on industrial informatics*, 10(3):1852–1863, 2014. doi:10.1109/TII.2014.2332116.
- [89] M Chang, M Kang, and M Pecht. Prognostics-based LED qualification using similarity-based statistical measure with RVM regression model. *IEEE Transactions on Industrial Electronics*, 64(7):5667–5677, 2017. doi:10.1109/TIE.2017.2677301.
- [90] M Ibrahim, J Fan, W Yung, A Prisacaru, W van Driel, X Fan, and G Zhang. Machine learning and digital twin driven diagnostics and prognostics of light-emitting diodes. *Laser & Photonics Reviews*, 14(12):2000254, 2020. doi:10.1002/lpor.202000254.

- [91] Z Jing, J Liu, M Ibrahim, J Fan, X Fan, and G Zhang. Lifetime prediction of ultraviolet light-emitting diodes using a long short-term memory recurrent neural network. *IEEE Electron Device Letters*, 41(12):1817–1820, 2020. doi:10.1109/LED.2020.3034567.
- [92] A Al-Mohamad, G Hoblos, and V Puig. A hybrid system-level prognostics approach with online RUL forecasting for electronics-rich systems with unknown degradation behaviors. *Microelectronics Reliability*, 111:113676, 2020. doi:10.1016/j.microrel.2020.113676.
- [93] A Vasan, B Long, and M Pecht. Diagnostics and prognostics method for analog electronic circuits. *IEEE Transactions on Industrial Electronics*, 60(11):5277–5291, 2012. doi:10.1109/TIE.2012.2224074.
- [94] F Tao, M Zhang, Y Liu, and A Nee. Digital twin driven prognostics and health management for complex equipment. *Cirp Annals*, 67(1):169–172, 2018. doi:10.1016/j.cirp.2018.04.055.
- [95] P De Laat. Algorithmic decision-making based on machine learning from big data: Can transparency restore accountability? *Philosophy & technology*, 31(4):525–541, 2018. doi:10.1007/s13347-017-0293-z.
- [96] J Gu, D Barker, and M Pecht. Uncertainty assessment of prognostics of electronics subject to random vibration. In *AAAI Fall Symposium: Artificial Intelligence for Prognostics*, pages 50–57. AAAI, 2007.
- [97] S Lundberg, G Erion, H Chen, A DeGrave, J Prutkin, B Nair, R Katz, J Himmelfarb, N Bansal, and S Lee. From local explanations to global understanding with explainable AI for trees. *Nature machine intelligence*, 2(1):56–67, 2020. doi:10.1038/s42256-019-0138-9.
- [98] T Chen, M Song, H Hui, and H Long. Battery electrode mass loading prognostics and analysis for lithium-ion battery - based energy storage systems. *Energy Res*, 9:754317, 2021. doi:10.3389/fenrg.2021.754317.
- [99] D Draper. Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):45–70, 1995.
- [100] A Brando, D Torres, J Rodriguez-Serrano, and J Vitria. Building uncertainty models on top of black-box predictive APIs. *IEEE Access*, 8:121344–121356, 2020. doi:10.1109/ACCESS.2020.3006711.
- [101] B Lakshminarayanan, A Pritzel, and C Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *arXiv preprint*, 2016. arXiv:1612.01474.
- [102] Y Gal and Z Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *International conference on machine learning*, pages 1050–1059. MLR press, 2016.
- [103] H Nguyen, J Liu, and E Zio. A long-term prediction approach based on long short-term memory neural networks with automatic parameter optimization by tree-structured Parzen estimator and applied to time-series data of NPP steam generators. *Applied Soft Computing*, 89:106116, 2020. doi:10.1016/j.asoc.2020.106116.
- [104] M Wei, H Gu, M Ye, Q Wang, X Xu, and C Wu. Remaining useful life prediction of lithium-ion batteries based on Monte Carlo Dropout and gated recurrent unit. *Energy Reports*, 7:2862–2871, 2021. doi:10.1016/j.egy.2021.05.019.
- [105] M Spelman, B Weinelt, P Gomez, R van Heusden, and R Siyam. White paper: Digital transformation initiative – oil and gas industry. Technical report, World Economic Forum, 2017.
- [106] L Kirschbaum, D Roman, G Singh, J Bruns, V Robu, and D Flynn. AI-driven maintenance support for downhole tools and electronics operated in dynamic drilling environments. *IEEE Access*, 8:78683–78701, 2020. doi:10.1109/ACCESS.2020.2990152.
- [107] M Fraser, T Anastaselos, and R Kumar. Whitepaper: The disruption in oil and gas upstream buisness by industry 4.0. Technical report, Infosys, 2018.
- [108] A Gezdur and J Bhattacharjya. Digitization in the oil and gas industry: Challenges and opportunities for supply chain partners. In *Working Conference on Virtual Enterprises*, pages 97–103. Springer, 2017. doi:10.1007/978-3-319-65151-4_9.
- [109] Z Zhao, B Liang, X Wang, and W Lu. Remaining useful life prediction of aircraft engine based on degradation pattern learning. *Reliability Engineering & System Safety*, 164:74–83, 2017. doi:10.1016/j.ress.2017.02.007.
- [110] D Nguyen, M Kefalas, K Yang, A Apostolidis, M Olhofer, S Limmer, and T Bäck. A review: Prognostics and health management in automotive and aerospace. *International Journal of Prognostics and Health Management*, 10(2):1–35, 2019. doi:10.36001/ijphm.2019.v10i2.2730.

- [111] P Lu, H Liu, C Serratella, and X Wang. Assessment of data-driven, machine learning techniques for machinery prognostics of offshore assets. In *Offshore Technology Conference*. OnePetro, 2017. doi:10.4043/27577-MS.
- [112] A Mittal, A Slaughter, and V Bansal. From bytes to barrels: The digital transformation in upstream oil and gas. Technical report, Deloitte-Insights, 2017.
- [113] C Sackschewsky, E Yavar, and J Cascini. How industry 4.0 is transforming the oil and gas supply chain. Technical report, BDO International, 2018.
- [114] C Procaccini. Application of data analytics technologies to improve asset operations and maintenance: Digital landscaping study of the oil and gas sector. Technical report, The Oil and Gas Technology Centre, 2018.
- [115] Aspentech. White paper: Beyond oil digitalization: The roadmap to upstream profitability. Technical report, Aspentech, 2017.
- [116] A Martinez-Mere and F Castagna. Digital transformation in the oil and gas industry - panorama. Technical report, International Atomic Energy Agency, 2017.
- [117] H Lu, L Guo, M Azimi, and K Huang. Oil and gas 4.0 era: A systematic review and outlook. *Computers in Industry*, 111:68–90, 2019. doi:10.1016/j.compind.2019.06.007.
- [118] Ma Richards and R de Silvia. Digital transformation for oil and gas: What’s important to the industry? Technical report, Oil and Gas IQ, 2018.

Chapter 3

Considered electronic assembly and electromechanical component

Introduction

As explained in the previous chapter, the selection of diagnostic or prognostic strategies in the context of PHM follows a variety of criteria. Foremost, the application nature and the available data are to be considered. Thus, the understanding of the respective application is a key prerequisite.

In order to provide the necessary technical context, Chapter 3.1 discusses the importance of the BHA during the deep drilling process and its built-in electronic assemblies, i.e. BHA-PCBAs. Subsequently, the relevant failure modes and current maintenance practices are studied. Likewise, the design and construction of EMRs as well as associated failure modes and mechanisms are presented in Chapter 3.2, where the focus is placed on DC electrical arcing.

3.1 BHA-PCBA and deep drilling

The Bottom Hole Assembly (BHA) is a multi-functional assembly that allows a precise navigation of the drill bit through the subsurface and reservoir. The drivers of this technical capability relate to the need to improve the drilling efficiency through a more detailed understanding of the ambient geological formations ¹.

3.1.1 Synopsis of the deep drilling process

In order to exploit oil and gas reservoirs, a well is drilled into the subsurface cutting through various geological formations, cf. Figure 3.1-(a). To facilitate the drilling

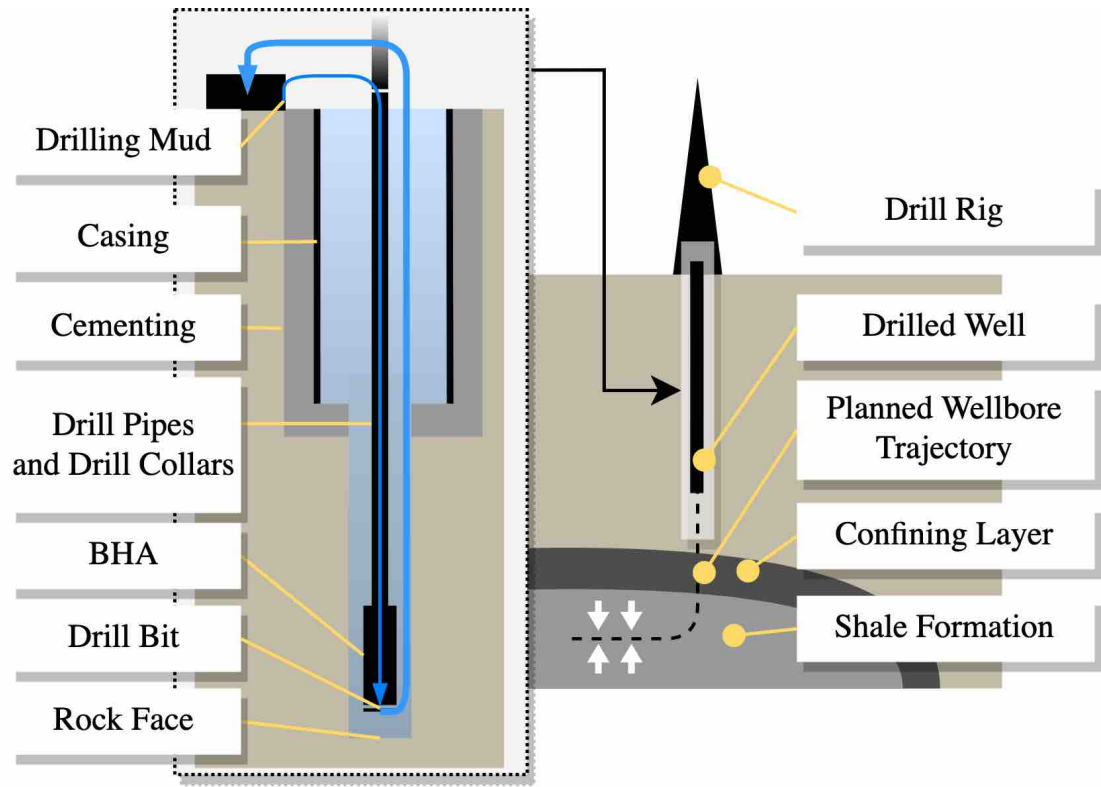
¹*Parts of this chapter have already been published in my journal or conference article: L. Kirschbaum, et al., (2020).*

process a rig is installed at the surface. The well is drilled following a mapped-out wellbore trajectory. Its design is guided by information obtained through geophysical measurements. At the lowest part of the drill string the drill bit cuts the adjacent rock face. The BHA controls the drilling process and is situated just behind the drill bit, cf. Figure 3.1-(b). The modules of the BHA are followed by a set of drill collars, i.e. heavy drill pipes which increase the Weight On Bit (WOB). This force pushes the drill bit against the rock face. As the wellbore becomes deeper more and more drill pipes are added to the drill string from the surface. Drilling mud is pumped continuously through the drill pipes down to the drill bit and up the drilled wellbore at high pressures. The drilling mud exits into the surrounding wellbore at nozzles placed within the drill bit. Then, it transports the rock cuttings back to the surface, where the cuttings are removed and the cleaned mud is pumped back down again. Further functions of the drilling mud (some kind of bentonite-water mixture) are: firstly, stabilisation of the wellbore preventing it to collapse onto the drill string by exerting counter pressure on the formation; secondly, a sealing layer on the wellbore's wall is formed by the drilling mud termed *filter-cake*. It limits the loss of drilling fluid into the surrounding formations which becomes important when drilling through highly soluble or permeable formations, e.g. salt or clay. As the well becomes deeper, the drilling process is interrupted at regular intervals in order to stabilise the newly drilled section of the well using metal tube casings. The casing is then cemented. Hence, with each completed section the borehole becomes deeper, but also narrower. Once in the reservoir, a horizontal section is drilled in order to maximise the inflow area for oil or gas. The well is completed for production through the installation of production tubing, sand screens, pumps, and a blow-out-preventer at the surface's wellhead.

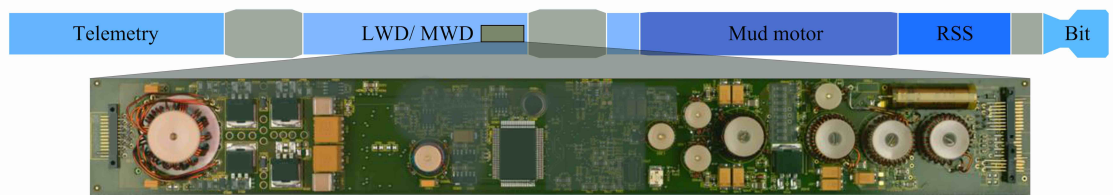
3.1.2 BHA and downhole data transfer

The BHA provides measurement, steering and communication capabilities during drilling; it supplies power to the downhole tools and allows control of the wellbore trajectory, while surveying the surrounding formation. As illustrated in Figure 3.1-(b), multiple interchangeable tools are assembled, based on the requirements of the drilling operation. The BHA commonly exceeds a length of 9 m.

The Rotary Steering System (RSS) has been developed to allow drill-bit steering and subsequently to control the well path. Today, two technologies are established: *push-the-bit* in the desired direction by extending hydraulic pads; *point-the-bit* by bending the shaft above the drill bit [1]. Measurement While Drilling (MWD) tools determine the position and orientation of the drill string. Common measurements during drilling are the *inclination*, the *azimuth*, and the *rotational speed* of the drill string. Logging While Drilling (LWD) tools measure the characteristics of the surrounding formations, such as resistivity, porosity, or formation pressure. A



(a) Schematic of the drilling process.



(b) Schematic diagram of a BHA consisting of various modules such as Logging While Drilling (LWD), Measurement While Drilling (MWD), or the Rotary Steering System (RSS). The BHA modules contain complex electronic and electromechanical systems to provide advanced functionality; PCBAs hold a multitude of electric components and are embedded into the downhole tools.

Figure 3.1: Oil and gas drilling.

turbine situated within the BHA power module through which the drilling mud flows is used to drive a generator that supplies power to the BHA electronics. However, in order to drive the drill bit at the rock face a so called mud motor is used - an additional BHA module. Unlike the turbine, the mud motor follows the principles of a progressive cavity displacement pump. The telemetry unit is usually located at the top of the BHA. This module serves as communication link between the BHA and the surface. Via mud pulse telemetry, data is sent to and received from the surface. This is the only reliable technique for communication to greater drilling depths and is standard throughout the oil and gas industry [2]. However, data transfer rates are slow at only 10 bit/s [3].

Prior to each drilling operation, drilling service customers specify their survey requirements for the expected geological formation, i.e. type of data and sampling rate. Hereinafter, the service provider designs the BHA. There are a multitude of mission possibilities and combinations of BHA modules. Data transmission bandwidth is utilised to operate the BHA and transmit formation data. Spare bandwidth may not be available to transmit tool health data while drilling. However, three general data sources relevant to BHA operation and maintenance can be distinguished. **Field data** is acquired during drilling and serves as the predominant source for real-time operational decision making. It contains information such as tool azimuth and inclination. The data received from downhole tools during drilling is truncated due to slow data transfer rates. After completion of a run, high resolution downhole tool-memory data is available on the surface and manually downloaded from the BHA tools. This data may contain various sensor measurements as well as tool-specific diagnostic information and event-logs. Personnel might enter supplementary notes. In practice this **post-run data** is often not readily available, since it is only occasionally downloaded during any follow-up maintenance. Furthermore, an insufficient digital infrastructure or data regulations imposed by local authorities confine access to this data. **Maintenance data** is obtained during the process of AMO. It can provide detailed insights of BHA failure root causes. However, it is subjective, and formats may vary widely. Like post-run data, aggregating complete maintenance reports and converting those to a practical format is expensive and time demanding due to the lack of standardisation. Table 3.1 summarises the advantages and disadvantages of various drilling data sources.

3.1.3 Failure modes of downhole tool electronics

The growing complexity of downhole tools has resulted in an ever-increasing count of electrical subsystems and components placed on PCBAs, larger memory storage, and faster processors. Recall, the downhole tools and the electronic assemblies experience a HTHP environment during drilling. Depending on the drilling method and the

Table 3.1: Downhole tool data sources.

Data type	Field data	Post-run data	Maintenance data
Advantages	Real-time; availability; supports onsite decisions	Tool-memory; field-reports; event-logs	Detailed failure analysis; cost data
Disadvantages	Low sampling rate; sparse data; additional effort to collect and store; quantity and quality varies	Availability, as not commonly stored; advanced expertise needed for analysis; ignored as no immediate payoff; quantity and quality varies	Low sampling rate; textual data; no concise reporting standard; often subjective

depth of the borehole this exposure can range from hours to several days. Whilst in the wellbore the electronics are subjected to ambient temperatures exceeding 200 °C, extreme pressures, bending stresses, and high levels of vibration [4, 5]. Thus, the variety of electronics deployed, e.g. switches and relays, Small Outline Integrated Circuits (SOIC)s (typically as dual in line package in legacy tools), multi-chip modules, capacitors, semiconductors, and resistors are susceptible to a variety of failure mechanisms. Bulk quantities of these off-the-shelf components are sourced through external suppliers. All components must be rated as high-temperature electronics.

In order to mitigate failure and avoid accelerated ageing of BHA electronics, various techniques have been established. Electronics are kept under atmospheric pressure by the means of pressure barrels. Likewise, thermal effects are opposed through the use of flasks, heat sinks, and thermal paste that improve heat dissipation and defer the inevitable temperature increase of electronic components during drilling. However, the heat produced by the electronic components themselves will eventually exceed the rising ambient temperature which significantly affects their durability and performance. Sealing components are an established strategy to suppress stress from vibration acting as a damper. Further factors that should be considered are the component orientation, the number of components, the type of the electronic packaging, the soldering type, and the board geometry. In addition, the PCBA placement and the orientation within the housing of the BHA modules are decisive factors in order to suppress effects of vibrations. Typical PCBA failures can be functional, software related, physical, or a combination of the three. Figure 3.2 presents examples of a connector damage, a semiconductor failure, and a capacitor failure. Connection damages can be due to mechanical load, e.g. excessive vibrations. Semiconductor and capacitor failures may be caused by shorted circuits or overheating which then leads to shorted circuits. Effects to consider are, e.g. electromigration or corrosion. Various additional factors are reported that affect the reliability of electrical components, e.g. unsteady power supply and drill string rotations per minute [5]. However, exposure to high temperature and subsequent overheating of

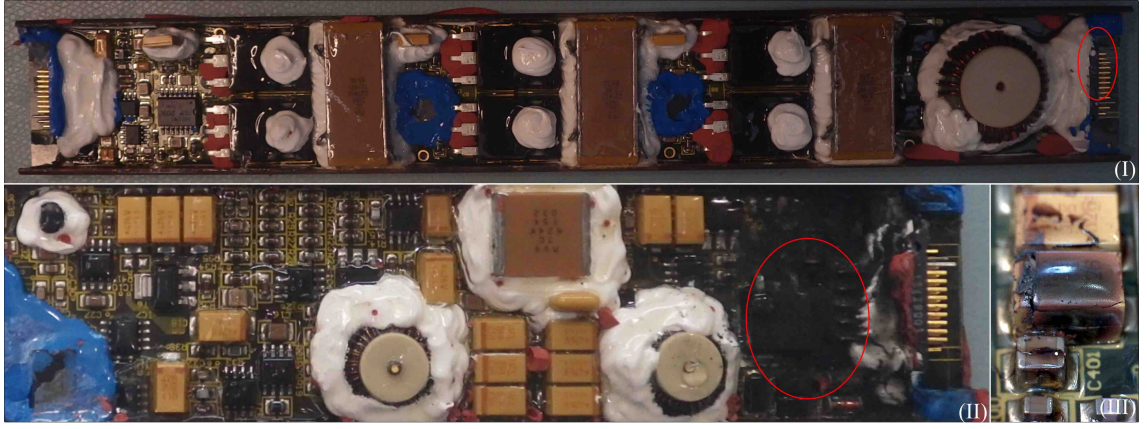


Figure 3.2: Failed BHA-PCBAs due to: (I) Connector damage, (II) Semiconductor Failure, (III) Capacitor failure.

components is identified as the predominant cause for electrical component failure [6]. The failure rate of downhole electronics doubles with every 10 °C increase in ambient temperature [3]. High temperature reduces the strength of connections and components. Furthermore, SOICs are especially liable to wire bond fatigue and breakage due to the thermo-mechanical cycling the BHA undergoes during drilling. This thermo-mechanical cycling is due to the repetitive process of drilling and removal of the drill string from the wellbore in order to change the configuration of the BHA or the drill bit. At the surface the components cool down to ambient temperature, while in the wellbore the temperature may exceed 200 °C. This can lead to failure of entire circuits due to fatigue. Moreover, lateral, axial, and torsional vibrations predominantly impact the electronics reliability [7]. Note, lateral vibrations have been identified as the principal factor contributing to the PCBA degradation accounting for up to 29 % of vibration related MWD tool failures [8].

3.1.4 Data-driven downhole tool maintenance strategies

Current BHA electronic maintenance strategies greatly rely on reliability testing. Though, due to the large number of subsystems and their individual components coupled with harsh operating conditions, traditional maintenance strategies based on offline reliability analysis are not able to match the encountered operational life-cycle of downhole tools. Moreover, significant efforts, time, and expenditures are needed to derive reliability metrics via traditional methods, e.g. tools like Failure Modes Mechanisms and Effects Analysis (FMMEA) due to the diverse range of failure modes. The complexity of the tools imposes arduous procedures to adequately reproduce failures and perform root cause analyses. Simultaneously, the required time for AMO services of downhole tool electronics increases – concurrently reducing fleet availability. Ultimately, an increased fleet volume is required to serve the same customer base. The inability to find failure root causes often leads to bulk removal of parts. Since the 1980s, a MTBF approach is widely used throughout the oil and

gas industry in order to determine the statistical downhole tool reliability. Depending on the wellbore, the formation, and previous exposure to high temperature and vibration levels the electronics life is typically in the range of around 100 hours. However, MTBF is misleading if applied as a metric for downhole tool reliability [8]. The dynamic environmental and operational parameters are not consistent within the wellbore due to HTHP and, therefore, do not fulfil the assumption of a static operation behind the MTBF approach. Furthermore, notwithstanding an overall improvement of downhole tool reliability, since the introduction of these metrics every third BHA continues to remain the cause for NPT [9]. In addition to insufficient reliability metrics, no concise definition of a *failure* has been issued throughout the industry. For example, [10] lists a large number of BHA states which, depending on the respective source, are sometimes classified as still-operational, at other times as faulty. Moreover, in practice, determining the failure mode is often not feasible and circuits are replaced as a whole, since NPT on the drilling rig is significantly more expensive than such immediate, corrective maintenance action. Under these very aspects MTBF – a solely statistical reliability-based measure – turns out to be an overall ill-defined metric for assessing the true reliability of downhole tools and its components. Likewise, it is difficult to make an informed maintenance decision on this basis. Moreover, with the ongoing automation of oil and gas industry assets, the technical and operational complexity evolves, questioning the capability of current strategies to meet stringent maintenance requirements in the future.

3.2 EMR and electrical contacts

3.2.1 EMR operation principle

The EMR is of similar importance to downhole tool electronics, but also of general interest within applications where it plays a safety or mission critical role. The main task of an EMR is the electrical separation of the control- from the load-circuit [11]. It consists of a magnetic coil, a travel armature, a spring, and a contact pair, cf. Figure 3.3. Although the EMR has been subjected to considerable design improvements over the past decades, the core components and working principle remain essentially unchanged ².

Even though alternative electrical switching devices are available, e.g. MOSFETs or Solid-State-Relays without mechanical parts that exhibit an improved reliability, the EMR puts a set of distinct characteristics forward [12]. It features an overall low CR in the $m\Omega$ range which reduces switching losses; high breakdown voltage of up to 1000 V; total isolation of the switching and control. The latter is not the case for

²Parts of this chapter have already been published in my journal or conference article: L. Kirschbaum, et al., (2022) and L. Kirschbaum, et al., (2021).

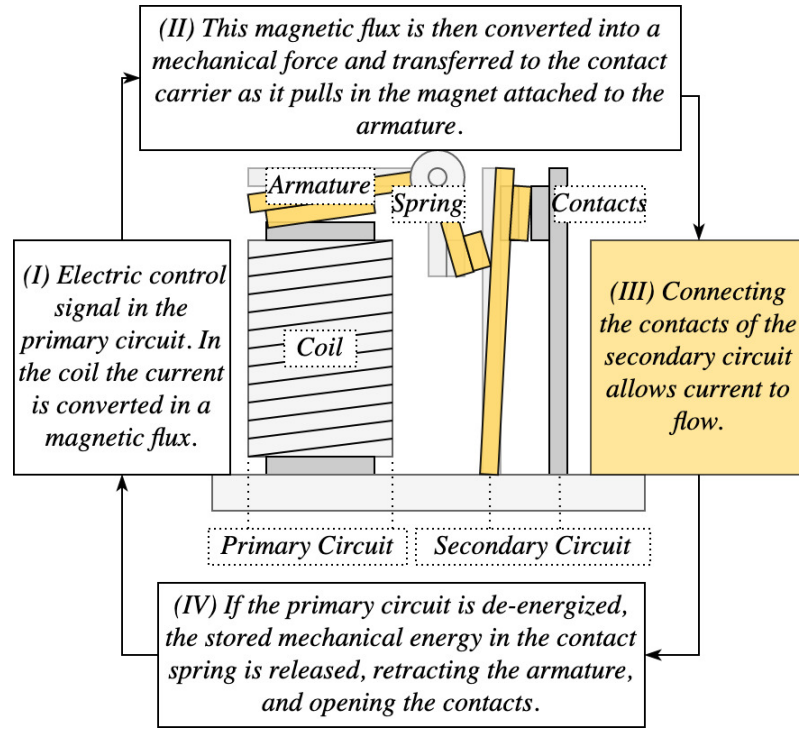


Figure 3.3: Schematic, simplified working diagram of the central functions and main components of an EMR (cf. [11]).

most semiconductor-based switching components. EMRs can be employed where switching is independent from the current direction. For example, EMRs are commonly the preferred choice in safety critical applications within nuclear power plants as they can be run in a *fail-open* fashion [13]. However, despite miniaturisation efforts, EMRs have a comparatively large form factor. In addition, the switching is slow - in the *ms* range, cf. MOSFET in the *ns*, or Solid-State-Relays < 0.2 ms range [14].

3.2.2 Failure modes

Due to the electromechanical nature of the EMR, its life is dependent on the mechanical life and the electrical life of the individual subcomponents. Generally, the mechanical life is in the order of 10^7 actuations compared to the electrical life at 10^6 actuations [14]. An EMR is considered failed if it can no longer perform the required switching function. Electrical, contact related failures prevail. Contact making or breaking related failures are the predominant failure modes [15, 16, 17, 18]. Beside the electrical load, the EMR life is affected by the operating environment, e.g. operating in environments with high humidity will lead to increasing rates of contact corrosion. Degradation is further accelerated through elevated temperatures. Continuous wear of the contact surface causes poor conductivity and high CR [19]. Atmospheric contamination, e.g. particles or vapours of silicon, contaminate the contact surface and affect EMR performance. Vibrations can cause contact chattering or damage mechanical parts depending on the EMR design. Though one

can distinguish failure modes by type, multiple of the respective failure mechanisms interact. The following sections discuss the relevant failure modes and failure mechanisms.

3.2.2.1 Contacts

As a rule, contact related failures occur over long duration. Such failures depend on the applied voltage and current, load type, the operating temperature, and the pollution of the operating environment. The root-causes for contact failures are excessive material transfer and material loss due to electrical arc discharge and contact bouncing [15]. A symptom of contact erosion is reduced contact force and increased CR. Further, welding, bridging, and sticking of contacts or corrosion and contamination (through deposition of isolating and semiconducting films that stem from eroded and worn contact material or carbides dissolved from organic gases) are the governing failure mechanisms. They lead to making and breaking failures, impermissible operate and release times, respectively, an increase of CR beyond an acceptable threshold, and high levels of contact noise [11]. An overview of the contact related failure modes is provided in Table 3.2. A comprehensive list of contact related failure modes and mechanisms can be found in [14].

Making failure: *The inability to make a connection whilst the CR remains below the maximum permissible CR.* If the application setting is favourable for electrical arcing, the likelihood of a making failure due to erosion and subsequent material transfer is high [17]. In some instances an increasing CR can point to an impending making failure. However, the actual failure starts to emerge when erosion has well advanced and the Over-Travel Time (OT) and contact force have become too small to enforce sufficient contact making. In general, a conjoint increase in BT can be observed [14]. However, making failure can also be caused by preceding mechanical failures, such as spring failure, bent armature, and coil failure.

Breaking failure: *The inability to break a connection and interrupt the current flow within the specified maximum opening time.* If contacts are switched under load, contact bridging due to contact welding may occur. Contact welding is common during contact making if bouncing is present because of the intermittent arc discharge. If the weld is strong enough, it prevents the contacts from separating. If contacts are welded together, CR is lower than the specified minimum CR for open contacts. Mechanical failures preventing the contacts from opening relate to spring or armature failures.

Operation time failure: *The duration of contact making and contact breaking exceeds a specified threshold.* If the RT increases, this can be due to the spring

degradation which may reduce the contact pull force. However, more likely are local micro-welds across the contact surfaces that cause the contacts to stick. Both, PT and RT are likely to increase if the coil operating voltage changes due to coil deterioration [20].

Elevated contact resistance failure: *Unacceptably high CR while contacts are closed.* Manufacturers specify the maximum acceptable CR for closed contacts (in the $m\Omega$ range). In general, different interacting factors influence the CR which can be distinguished by having either a decreasing or increasing effect. EMRs in storage are subject to chemical reactions on the contact surface, such as the formation of silver sulphides that deposit as insulating layers and, in turn, cause an increased CR [11]. During operation under low-load changes of CR are dominated by mechanical effects, despite occasional increases of CR through, e.g. polymerisation or corrosion on the contact surface. However, if the amount of carried current increases - and the contact temperature respectively - continuous electrical fatigue due to arc erosion inevitably reduces the contact force. In such instances, film formation in combination with corrosion becomes the dominating degradation regime. A subsequent reduction of the effective contact area leads to a significant increase of CR. Unsealed EMRs operating in high temperature environments are liable to serious rates of contact oxidation. Mechanical actuation may rupture deposited films and, thereby, reduce CR [21]. In addition, effects like ion-sputtering can temporally clean the contact surface and reduce CR. One should note that an increase in CR is, in general, accompanied by increasing levels of contact noise.

3.2.2.2 Coil and mechanical parts

Coil Long-term switching impacts the coil resistance. Deposition of evaporated contact material particles on the coil wire or combustion of the insulation material due to excessive heat reduce the coil's insulation resistance [22]. Further, poorly welded coil wires might be a failure root-cause. In general, coil failure is most likely to happen if the ambient temperature is high which causes the coil to overheat. The encountered failure modes are a shorted coil or changes in the pick-up or release voltage. However, it should be noted that the likelihood of these failures is very low in comparison to contact related failures.

Mechanical parts The wear of mechanical parts, e.g. the armature or spring, causes a reduction in contact force or variations in contact velocity. This wear stems from material fatigue due to vibrations or excessive heat, e.g. from high-current arc discharges.

Table 3.2: Contact related failure modes and mechanisms.

Modes & Mechanisms	Causes
Making Failure	
Reduced Contact Force	Excessive contact material loss due to electrical erosion.
Mechanical Fatigue	Reduction in spring force; bent or stuck armature.
Breaking Failure	
Welding and Bridging	Arc discharge during bouncing and a reduction in contact force.
Mechanical Fatigue	Stuck armature.
High Contact Resistance	
Contact Erosion	Contact surface roughness increases due to ongoing material loss and material re-deposition from electrical arcing reducing the effective contact area.
Fretting	Deposition of material on the contact surface (partially insulating) due to mechanical wear.
Contamination	Corrosion; film-formation from various sources, e.g. organic particles, oxides, carbides.
Extended Operate and Release Time	
Micro-Welding	Formation of small, weak welds during contact bouncing cause the contacts to stick together.
Coil Fatigue	Increase in CR due to wear of insulation material from long exposure to elevated temperature, subsequently increasing the required pick-up voltage.
Spring Fatigue	A reduction in spring force that cause a reduced pull-in force and, therefore, a reduction in contact velocity.

3.2.3 Contact resistance

In order to understand electrical arcing and the deteriorating effects that act on the contact surfaces during switching, a brief explanation of CR is beneficial.

3.2.3.1 Constriction resistance

Though not obvious to the eye, the nominal contact area is not the true conductive contact area. The inherent microscopic surface roughness restricts the path of the current flow [14, 23]. The *actual* contact points are referred to as *a-spots*. The observed increase of CR through this limited interface compared to the resistance via the apparent contact surface is termed constriction resistance $R_{constriction}$. It is subject to the elastic and plastic properties of the contact material [11, 24]. Different shapes of *a-spots* can be distinguished based on the contact surface texture because the shape of the *a-spot* affects $R_{constriction}$. Usual simplifications assume square, rectangular, or circular shapes. $R_{constriction}$ of a single, circular *a-spot* interface is defined in Equation 3.1.

$$R_{constriction} = \frac{(\rho_1 + \rho_2)}{4a} \quad (3.1)$$

where ρ defines the resistivity of each contact body, a the radius of the *a-spot*. As [14] reports, the actual contact making surface is not required to be large to cause the constriction resistance to be low (an effective contact area of 10 μm causes a resistance around 1 $m\Omega$). Metallic contacts embody multiple *a-spots* that are

found to be arranged in clusters. The distribution of these clusters has a negligible effect on CR in practical applications. Here, the assumption that contact is made uniformly across the nominal contact area is common [23]. Equation 3.2 gives an approximation of the $R_{constriction}$ for n circular a -spots in a cluster

$$R_{constriction} = \rho \left(\frac{1}{2na} + \frac{1}{2\alpha} \right) \quad (3.2)$$

$$\alpha = \left(\frac{A}{\pi} \right)^{\frac{1}{2}} \quad (3.3)$$

where α denotes the cluster radius and A represents the nominal contact area [25]. An empirically supported approximation for $R_{constriction}$ is given by [23] in Equation 3.4 if only plastic deformation is assumed under a contact force F . H denotes the *Brinell*-material-hardness; θ represents an empirical coefficient describing the contact surface cleanliness.

$$R_{constriction} = \left(\frac{\rho^2 \theta \pi H}{4F} \right)^{\frac{1}{2}} \quad (3.4)$$

Equation 3.5 provides an alternative, simplified approximation that is found to be used in practice assuming clean contacts, cf. [11].

$$R_{constriction} = \frac{\rho_1 + \rho_2}{2} \sqrt{\frac{H}{F}} \quad (3.5)$$

As becomes apparent from Equation 3.4 or Equation 3.5, low contact forces result in high CR. In general, CR is deemed unacceptable if the contact force is smaller than 0.05 N [11].

As a rule, CR will increase if the power increases up until the material softening voltage is reached. A temperature differential can be observed between the a -spots and the contact body. The contact material is subjected to plastic deformation at the conducting a -spots, since Joule heating causes a local temperature increase. Plastic deformation increases the effective contact area. This, in turn, causes an observable drop in CR. However, if the voltage continues to rise, the CR will continue to increase until the melting point of the contact material is reached. Now, a rapid increase in the effective contact area can be observed going hand in hand with a secondary decrease in CR [11]. This effect is sometimes referred to as *self-healing* of contacts [14]. Concluding, the CR is highly dependent on the contact force, but essentially independent of the nominal contact area A due to plastic deformation of the contacts upon closure.

3.2.3.2 Film resistance

The CR is affected not only by the contact shape and the $R_{constriction}$, but also by a problematic thin film layer build up on the contacts. This causes the actual number of conducting *a-spots* to be further limited because such layers may form an isolating barrier. Thus, electrical conduction only takes place at the spots where the film ruptures during contact making. As [11] reports, surface films as thin as 10 nm can result in high CR. Equation 3.6 gives an approximation for films of a thickness l and an area πa^2 ; ρ_{film} denotes the resistivity of the film.

$$R_{film} = \frac{\rho_{film} l}{\pi a^2} \quad (3.6)$$

However, depending on the type of the film, the resistance of the film may be neglectable. In some instances, it might even further decrease the constriction resistance, as the effective contact area is increased through a conducting film.

3.2.4 Failure mechanisms

This section elaborates failure mechanisms most relevant to EMR contacts. Essentially, all failure mechanisms contribute to a change in CR, ultimately leading to one of the listed failure modes.

3.2.4.1 DC arc erosion

Electrical arcing leads to electrical erosion across the contact surface [14, 26, 27, 28, 29]. Electrical arcs are the result of a voltage applied across a pair of open electrical contacts that cause an electric field. An arc consists of plasma which is the ionised gas in the contact gap. Thus, the arc is a high current gaseous discharge. A comprehensive discussion of the phenomenon of electrical arcing can be found in [11, 14, 30]. In DC circuits, a material redistribution process due to electrical erosion manifests as a continuous net transfer from one contact to the other. A *pip* and *crater* structure on the contact surfaces forms. Whether the *pip* and *crater* are located on the cathode or the anode depends on the duration and energy of the arc, the circuit's inductance, the contact material, the switching speed of the contacts, the cleanliness of the contact surface, and the contact dimensions [14]. The rate of material erosion will increase as the energy and duration of the arc increases.

Two phases of the erosion process can be distinguished: the initial *metallic-phase* arcing and the following *gaseous-phase* arcing. The physical processes governing these two stages are schematically illustrated in Figure 3.4. During the metallic-phase arc, after the rupture of the molten metal bridge (cf. Chapter 3.2.5.2), the gross of current is transferred by metal-ions. Hence as [31, 32] demonstrate, material accumulates at the cathode, stemming from vaporised and subsequently ionised

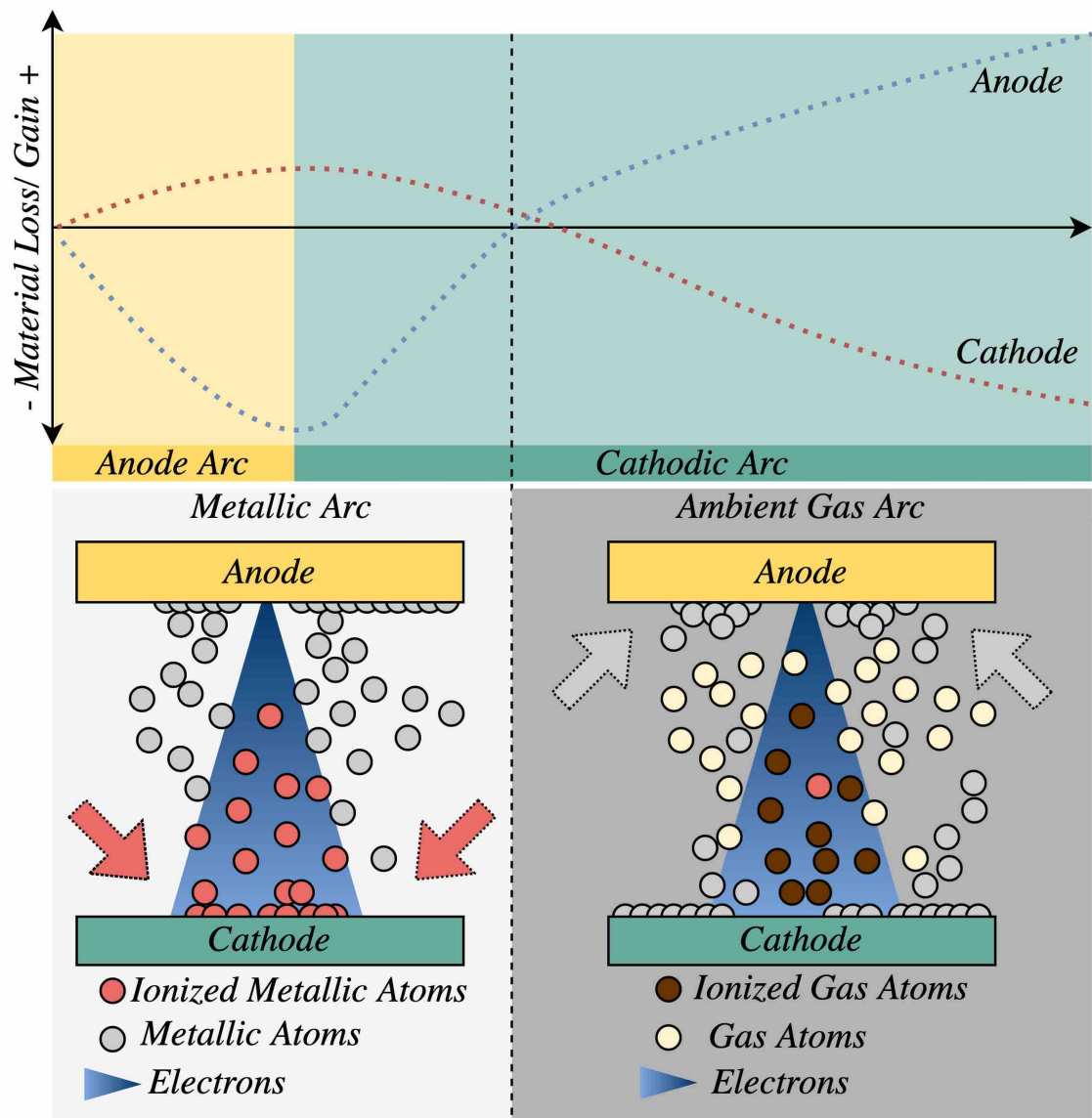


Figure 3.4: Schematic display of the erosion processes at cathode and anode during the phase of electrical arcing in metallic vapour and in ambient air.

metal atoms. Simultaneously at the anode, electron bombardment leads to disintegration of anode material. Hereinafter, the electrical arc will transfer to an arc operating in ambient air as the density of the metallic vapour decreases. The arc is now predominantly ionising the ambient gas atoms. The impact of the ionised gas atoms further erodes material on the cathode. This process is termed gaseous ion sputtering [33]. As one can see in the upper section of Figure 3.4, with the transfer to an gaseous-phase arc an increasing net gain at the anode is observed because the metallic atoms - separated from the cathode contact surface through the impacting gaseous ions - aggregate at the anode region. In DC circuits a cathode gain - material build up (i.e. *pip*) - is likely in circuits with short arcing times because the duration of the arc operating in ambient air is relatively short in comparison to the metallic-phase arc, cf. Chapter 7.2. With increasing arcing duration, an anode *pip* and a cathode *crater* will become more likely.

Bouncing during contact making is common because kinetic energy is preserved in the closing contacts, cf. Chapter 3.2.5.1. As [34] report, high frequent bouncing, i.e. many short bounces in close succession accelerate erosion through arcing and the formation of *pip* and *crater* structures. Hence, adjusting the bounce to a lower frequency is an important design consideration for switching contacts. However, if bouncing is neglectable, the bulk of contact erosion will take place during contact breaking. In general, a lower erosion rate can be achieved through higher opening speeds shortening the overall arc duration, since the rate of erosion depends on the duration of the arc [35]. Though, this may again lead to more bouncing and, therefore, erosion during contact making respectively.

To conclude, electrical arcing in DC circuits causes contact material vaporisation or ejection and re-deposition of vaporised and ejected material. Whether anodic or cathodic erosion is dominant depends on the contact material, the surface contamination or film, the duration of the arc, and the electrical current. Contact erosion leads to a deterioration of the contact surface and a reduced effective contact area of closed contacts. This is reflected by an increase in CR. The problem is aggravated whenever the contact force is reduced due to such material loss. General methods to mitigate the effects of contact erosion through electrical arcing are presented in Chapter 6.1.

3.2.4.2 Contact welding

When the contacts part, local Joule heating heats up the contact material, leading to a locally constrained melt and a sequential weld, illustrated in Figure 3.5. Just like electrical erosion, welding of contacts depends on a multitude of factors. It is, indeed, critical if the contacts can no longer part, i.e. welds cannot be separated and the contacts continue to stick together [36]. Although, welding during contact

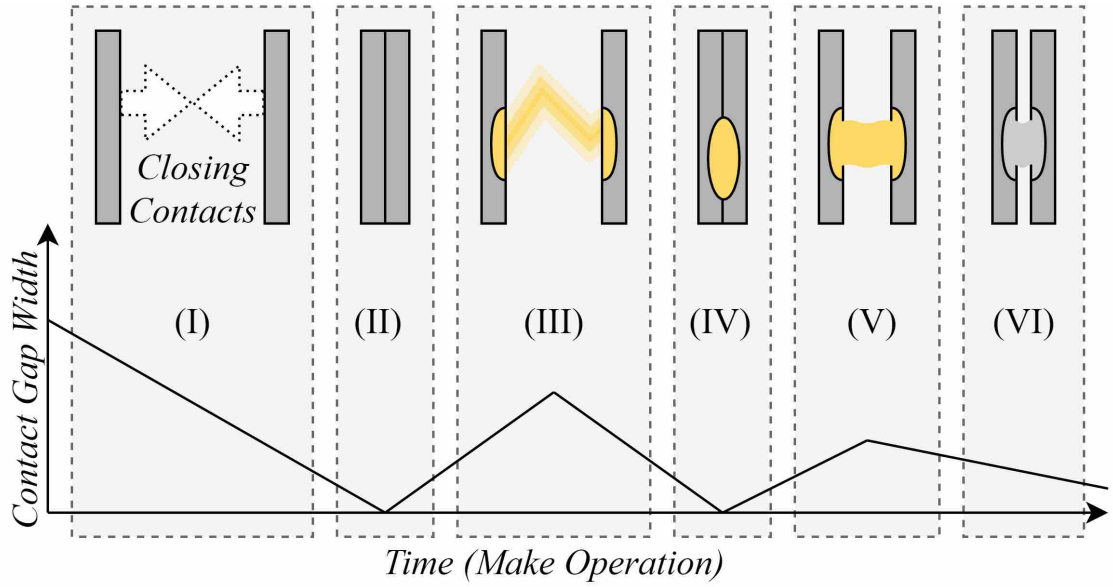


Figure 3.5: Schematic display of the welding process of closing contacts subjected to arcing. (I) Contacts close; (II) due to the preserved kinetic energy the contacts bounce for the first time; (III) the voltage drops to the minimum arc voltage and an arc forms, heating up the contact material on the anode fall and cathode fall region; (IV) the contacts close on the molten metal pools; (V) as the contacts open for the second bounce, a weld forms; (VI) if a weld establishes and cools down quickly, its strength might be greater than the remaining kinetic contact force. This prevents the contacts from closing or opening again.

breaking is possible, it is not very likely [37, 38]. On the contrary, the problem of contact welding during contact making is exacerbated if contact bouncing and arcing are present. Depending on the current and the type of the contact material, molten metal pools develop on both contacts. As the contacts open and close upon these interfaces a weld is formed. Such weld becomes problematic if the contact force is reduced prior to the contacts being fully closed, since the simultaneous reduction in contact velocity provides more time for the weld to cool down and harden [39].

Relevant to the welding strength is the load current and the duration of the bounce. The strength of the weld increases with the amplitude of the applied current [40]. As bouncing events become shorter with each subsequent bounce an increase in the weld strength for late bounce events, somewhat growing exponentially past the 4th bounce can be observed [41]. Further effects of welding during bouncing are discussed in [36, 14].

3.2.4.3 Contact contamination

In addition to the effects of contact shape and constriction resistance, the CR is affected by the film resistance. Isolating or semiconducting layers can be deposited on the contact surfaces through, e.g. outgassing of plastic sealings or insulation materials, material abrasion, and contamination from the ambient air. Thus, the actual number of conducting *a-spots* is further reduced which may lead to a CR build-up [23]. The extent of film formation depends on the storage duration, the

environment, the operational conditions, and alterations of the contact surfaces from electrical arcing.

A major source for surface film contamination is silicon. Silicon compounds are commonly found in, e.g. lubrication, insulation material, paints, and plastic components such as EMR enclosures. Vapours emitted from those silicon containing materials can form insulating films that deposit on the contact surface [42, 43]. During switching silicon breaks down to silica (silicon-dioxide), compromising contact performance, significantly increasing CR towards the EMR-EOL [44]. Highly contaminated environments may also contain other particles that are deposited on the contact surfaces. These particles sometimes have a higher hardness than the contact material itself, e.g. gold [14]. Hence, if contacts are closed, such particles are embedded into the contact surface, increasing the surface roughness in turn reducing the number of current carrying *a-spots*. As the authors in [45] point out, depending on the particle orientation and size, a temporary reduction of the increased CR can be achieved by applying a sufficiently large contact force. Further, one should consider *fritting* in the presence of contaminating films. Fritting, extensively discussed in [14, 45], refers to the electrical destruction of the isolating layer when the switching voltage increases and an instantaneous reduction of CR is observed.

3.2.4.4 Fretting

Lastly, not to be confused with fritting, contact surfaces are liable to *fretting* [47, 48]. Fretting contributes to the wear of the contacts as the mutual displacement of the contacts against each other leads to the abrasion of the surface as it produces debris. External vibrations or different rates of thermal expansion of the contact materials are responsible for fretting. It increases CR and promotes other degradation mechanisms that contribute to CR fluctuations over the EMR life as well as an increase in contact noise [11]. The process stages of fretting are detailed in Figure 3.6.

3.2.5 Contact operation

3.2.5.1 Contact making

An arc establishes if the voltage across two contacts is higher than the breakdown voltage and the travel time to make contact is longer than the minimum time necessary for the discharge. This type of arc is sometimes termed *pre-strike* arc and might be of very short duration, i.e. *ns*-range. It can be observed that a voltage increase relates to a decreased time to discharge which allows sufficient arcing already in closely spaced, fast-closing contacts.

In Figure 3.7 the first contact is established at 0.5 ms. The voltage drops and the

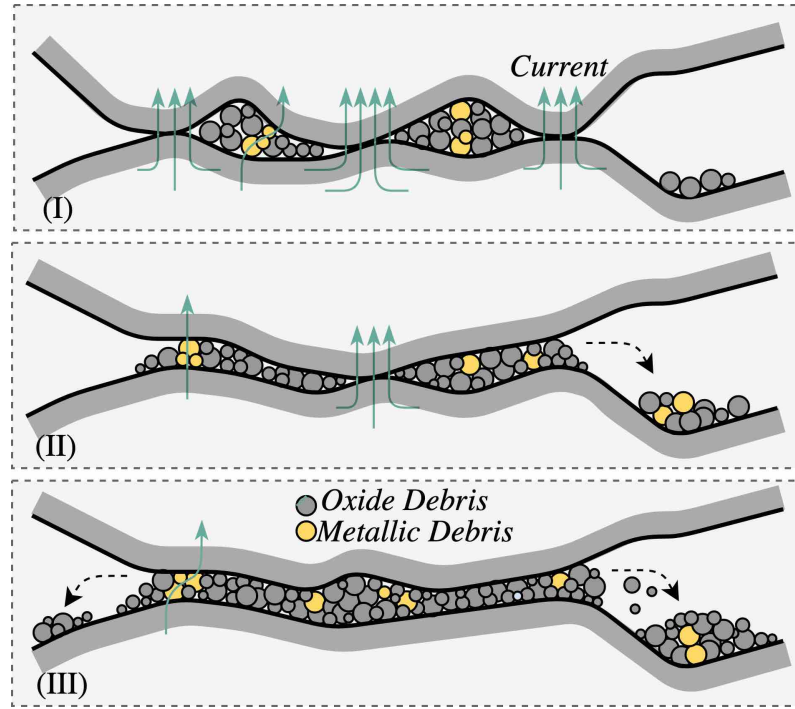


Figure 3.6: Schematic display of the process of closed contacts subjected to fretting. (I) Partially and fully oxidised metallic particles on the rough contact interfaces restrict the current flow to the current carrying *a-spots* that are affected by the distribution of surface asperities [46]. Some temporary current carrying paths may establish through metallic and partially oxidised particles in the debris matrix [14]. (II) Over the course of operation the contact surface further degrades, reducing the effective contact interface accumulating more debris and increasing CR. (III) A significant increase in CR can be detected, since no direct contact is made. Current conduction may only take place along isolated current carrying paths in the debris matrix that is made up of non-oxidised particles and subject to immediate changes whenever debris is displaced.

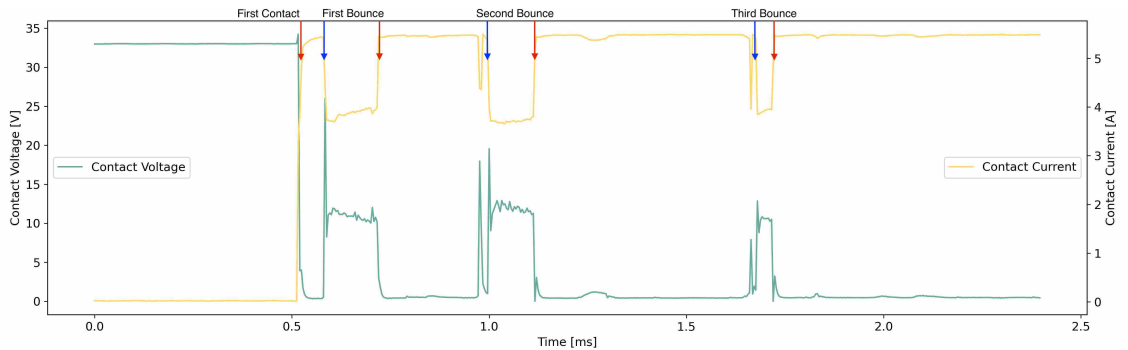


Figure 3.7: Voltage and current waveform for a making actuation; EMR ($AgSnO_2In_2O_3$)-plated copper contacts; measurements obtained during the experiment presented in Chapter 6.

current increases as the contact is made among the conducting *a-spots*. However, due to the kinetic energy preserved in the contacts, the moving contact carrier bounces back. This can be observed around 0.6 ms. A molten metal bridge forms and then ruptures which is followed by a very short metallic-phase arc. As the contact gap widens, the arc passes into an ambient-gaseous-arc. This is accompanied by a voltage spike prior to it settling around the minimum arc voltage of 13 V at 0.65 ms. This very voltage spike is due to the high pressure metal vapour region between the closely spaced contacts which initially does not allow conductivity within the metal particle cloud [30, 49, 50]. The arc reaches its maximum length approximately at 0.7 ms as the contacts start to close again. The contacts touch the 2nd time at 0.8 ms; the arc is extinct and the voltage drops to the closed-circuit voltage. During this second bounce the contacts essentially close on a molten metal surface because of the arc's heat up. This damps the impact of the contacts and further reduces the kinetic energy. However, it is here where contact welding might occur. The duration of the subsequent bounces decreases - one can observe a second and third arc, each shorter in duration than the previous one - as the preserved kinetic energy of the contacts is further reduced. Due to the effects of electric arcs, the decrease of kinetic energy of the contacts is accelerated compared to a decrease by purely mechanical bouncing. Lastly the contacts settle at 1.8 ms because the contact force overcomes the remaining kinetic energy.

3.2.5.2 Contact breaking

When metallic, current carrying contacts separate, an instantaneous sequence commences. A large number of research projects have dealt with this topic, trying to empirically understand the processes involved and to determine the underlying physical phenomena [31, 51, 52]. In general, EMR contacts break as follows: 1st, the contact force decreases and the contacts part. As the effective contact area diminishes the voltage increases because the number of current carrying *a-spots* is further reduced. This process of opening is accelerated by the so called *blow-off force* which is a result of the increasingly restricted current flow through the diminishing effective contact area [53]. This force will reduce quickly as the contacts further open and is substituted by forces stemming from the electrical arc. During the initial phase the voltage increases above the static voltage of the closed contacts; as the contact surface decreases further, the local restriction of the current flow heats up the remaining contact spots. Reaching the melting temperature at the contact spots, a bridge of molten metal will form and span between the parting contacts. Meanwhile, a steady voltage increase can be observed whereas the melting voltage exhibits a quasi-static behaviour only for currents below 100 A [50]. 2nd, during the stable phase, the voltage increases. This can be observed for all current levels, in an air environment as well as in vacuum [49]. Different mechanisms contribute to the

material transfer as the contacts separate, some being attributed to the *Thomson-Effect* [11]. However, the author in [54] argues that the majority of material transfer is due to electromigration. It is reasoned that electromigration will be predominant, since molten metal bridges are of small diameters. Hence, high current densities and elevated temperatures increase the rate of diffusing ions. The temperature varies between the melting and the boiling point of the metal. As soon as the contacts are further separated, the 3rd regime commences. The bridge becomes increasingly unstable which ultimately leads to its rupture. This phase distinguishes itself by its oscillating voltage fluctuations, spiking up to the minimum arc voltage and dropping down to the melting voltage. Note, however, such voltage spikes can have a stabilising effect on the bridge. Due to the increased power more metal is molten at the bridge root and sustains the elongating bridge by increasing its diameter. Vice versa, the current density is reduced which minimises thermal stress. However, a set of interacting processes excites the molten metal bridge rupture, e.g. the temperature in the bridge might reach the boiling temperature of the material, hydrodynamic instabilities in the material, dynamic changes in surface tension as the bridge stretches, and magnetic pinch forces depending on the carried current [50, 55]. Following the rupture of the molten metal bridge, the 4th phase commences as an initial arc forms, also referred to as *bridge-column* arc or *pseudo-arc* [55]. However, it is important to emphasise that the arc will only form once the bridge has ruptured [49]. Metal vapour remains in the contact gap that consists of $\pm 5\%$ of the particles from the molten bridge rupture. At this very initial stage and prior to the bridge-column arc a non-equilibrium high pressure zone is established. It is characterised by the high density metal vapour and very low conductance [49, 56]. Thereby, the voltage rapidly increases between the contacts and peaks as soon as the pressure begins to fall to 2 – 3 bar. Along with this ongoing pressure decrease, the voltage decays just as quickly [14]. Now, the area between the contacts acts like a capacitor with a very small capacitance [49]; remaining charges from the circuit inductance, which prevent an instantaneous change in current, flow into this capacitor. At last, the bridge-column arc is established as the current carrying ions impact the cathode at the origin of the molten metal bridge. High erosion rates resulting in material transfer from anode to cathode can be observed during this phase as most current is carried by ions. Because the pressure continues to fall, the bridge column arc changes into a normal arc operating predominantly in the ambient gas rather than in the metal vapour, at a voltage near the material dependent minimum arc voltage. Material transfer continues from anode to cathode, though the transfer rate decreases.

Following the breakdown of the molten metal bridge, detailed above, one can comprehend the process of ambient air arc establishment, sustainment, and extinguishment in Figure 3.8. First, the contact gap increases, the voltage spikes, then settles

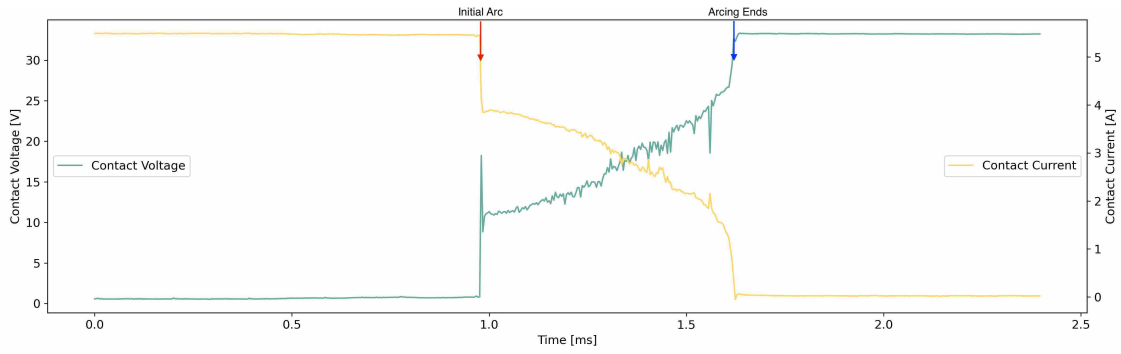


Figure 3.8: Voltage and current waveform for a breaking actuation; EMR ($AgSnO_2In_2O_3$)-plated contacts; measurements obtained during the experiment presented in Chapter 6.

around 13 V across the contacts, cf. [14] for an analysis of the required minimum arc voltage. An instantaneous current drop to 5 A follows; the initial arc establishes at 0.95 ms. Now, the metallic-phase arc transfers into an arc burning in ambient air. The arc begins to lengthen due to the contacts parting. Simultaneously, the arc diameter shrinks. This causes the reported increase in voltage. The waveform of the current and the voltage behave increasingly linear and smooth at higher current levels during contact breaking. Though, at relatively low current-levels, as in Figure 3.8, a distinct sequence of voltage steps and fluctuation can be differentiated. [57] report the occurrence of those distinct steps. The initial voltage step can always be observed, though, the probability of subsequent steps decreases as the circuit voltage is increased. However, the steps observed in Figure 3.8 are less distinct, despite the relatively low switching voltage. This can be attributed to the fast switching of the EMR contacts (around 0.6 ms to complete the opening of the contacts). The voltage spike at 1.0 ms may be attributed to the ongoing transfer from metallic to ambient air arc because the metal vapour in the contact gap is still being diffused by the molecules of the ambient air, but no longer able to maintain the discharge of the electric arc [58]. However, [59] shows that such behaviour can also be observed in vacuum and, therefore, the explanation of this phenomenon given by [58] does not yet provide a satisfactory answer. Prior to 1.6 ms, one can notice a sharp decrease in current turning the energy balance of the arc negative, i.e. the arc loses more energy than supplied through the cathode. Hence, it becomes unstable as soon as the current drops below 1 A and the voltage reaches 30 V. Subsequently the arc is extinct at 0.5 A, which agrees with measurements reported by [60].

Summary

In this chapter, the background of the two case studies considered has been discussed. BHA-PCBAs were examined from the point of view of the upstream oil and gas industry. The distinct failure modes and mechanisms of EMRs have been presented from a general viewpoint that is not limited to the application within

downhole tools. Based upon this analysis, the following Chapter 4 assesses the industry and component specific state of the art in the realm of data-driven approaches to maintenance and reliability.

References

- [1] S Menand, J Simon, Cand Gaombalet, L Macresy, L Gerbaud, M Ben Hamida, Y Amghar, H Denoix, B Cuiller, and H Sinardet. PDC bit steerability modeling and testing for push-the-bit and point-the-bit RSS. In *IADC/SPE Drilling Conference and Exhibition*. OnePetro, 2012. doi:10.2118/151283-MS.
- [2] S Mwachaka, A Wu, and Q Fu. A review of mud pulse telemetry signal impairments modeling and suppression methods. *Journal of Petroleum Exploration and Production Technology*, 9(1):779–792, 2019. doi:10.1007/s13202-018-0483-y.
- [3] R Beckwith. Downhole electronic components: Achieving performance reliability. *Journal of Petroleum Technology*, 65(08):42–57, 2013. doi:10.2118/0813-0042-JPT.
- [4] I Ahmad, O Akimov, P Bond, P Cairns, N Eide, T Gregg, T Heimes, A Nwosu, and F Wiese. Reliable technology for drilling operations in a high-pressure/high-temperature environment. In *IADC/SPE Drilling Conference and Exhibition*. OnePetro, 2014. doi:10.2118/167972-MS.
- [5] K Carter-Journet, A Kale, T Falgout, and L Heuermann-Kuehn. Drilling optimization: Utilizing lifetime prediction to improve drilling performance and reduce downtime. In *SPE Deepwater Drilling and Completions Conference*. OnePetro, 2014. doi:10.2118/170270-MS.
- [6] K Bybee. Reliable electronics for high-temperature downhole applications. *Journal of Petroleum Technology*, 52(7):56–57, 2000. doi:10.2118/0700-0056-JPT.
- [7] S Zhan and J Zhao. Advanced prognostic technique for improving the drilling performance of downhole tools. In *2015 First International Conference on Reliability Systems Engineering (ICRSE)*, pages 1–9. IEEE, 2015. doi:10.1109/ICRSE.2015.7366418.
- [8] H Reckmann, P Jogi, F Kpetehoto, S Chandrasekaran, and J Macpherson. MWD failure rates due to drilling dynamics. In *IADC/SPE Drilling Conference and Exhibition*. OnePetro, 2010. doi:10.2118/127413-MS.
- [9] J Brehme and T Travis. Total BHA reliability — an improved method to measure success. In *IADC/SPE Drilling Conference*. OnePetro, 2008. doi:10.2118/112644-MS.
- [10] F Ng. Recommendations for MWD tool reliability statistics. In *SPE Annual Technical Conference and Exhibition*. OnePetro, 1989. doi:10.2118/19862-MS.
- [11] H Sauer. *Modern Relay Technology*. Huethig, 1986.
- [12] Q Yu, M Qi, S Wang, and G Zhai. Research on life prediction based on wavelet transform and ARMA model for space relay. In *4th IEEE Conference on Industrial Electronics and Applications*, pages 1275–1280. IEEE, 2009. doi:10.1109/ICIEA.2009.5138407.
- [13] S Sravanthi, R Dheenadhayalan, M Sakthivel, K Devan, and K Madhusoodanan. A method for online diagnostics of electromagnetic relays against contact welding for safety critical applications. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 5(12):1734–1739, 2015. doi:10.1109/TCPMT.2015.2498624.
- [14] P Slade. *Electrical contacts: principles and applications*. CRC press, 2017.
- [15] Y Fang, Z Li, W Li, and K Li. Concerning contact resistance prediction based on time sequence and distribution character. In *50th IEEE Holm Conference on Electrical Contacts*, pages 447–452. IEEE, 2004. doi:10.1109/HOLM.2004.1353155.
- [16] A Wileman and S Perinpanayagam. Integrated vehicle health management: An approach to dealing with lifetime prediction considerations on relays. *Microelectronics Reliability*, 55(9-10):2165–2171, 2015. doi:10.1016/j.microrel.2015.06.013.
- [17] Y Sun, Y Cao, Y Zhang, and C Xu. A novel life prediction method for railway safety relays using degradation parameters. *IEEE Intelligent Transportation Systems Magazine*, 10(3):48–56, 2018. doi:10.1109/MITS.2018.2842048.
- [18] IEC. Electromechanical elementary relays - part 1: General and safety requirements. Standard, International Electrotechnical Commission, Geneva, CH, 2015.
- [19] J Liu, M Zhang, N Zhao, and A Chen. A reliability assessment method for high speed train electromagnetic relays. *Energies*, 11(3):652, 2018. doi:10.3390/en11030652.
- [20] Y Xuerong, M Yue, M Hang, and Z Guofu. Degradation failure model of electromagnetic relay. In *26th International Conference on Electrical Contacts (ICEC)*, pages 116–123. IET, 2012. doi:10.1049/cp.2012.0633.

- [21] Z Chen, H Mizukoshi, and K Sawa. Contact resistance characteristics of Ag material in breaking low-load DC arcs. *IEEE Transactions on Components, Packaging, and Manufacturing Technology: Part A*, 17(1):113–120, 1994. doi:10.1109/95.296376.
- [22] Y Xuerong, R Fu, Y Wu, Y Lin, and G Zhai. Simulation and diagnosis of degradation failure on electromagnetic relay. In *IEEE Prognostics and System Health Management Conference (PHM-Harbin)*, pages 1–9. IEEE, 2017. doi:10.1109/PHM.2017.8079301.
- [23] R Holm. *Electric contacts: theory and application*. Springer Science & Business Media, 2013.
- [24] J Swingler and A Sumption. Arc erosion of AgSnO₂ electrical contacts at different stages of a break operation. *Rare Metals*, 29(3):248–254, 2010. doi:10.1007/s12598-010-0043-1.
- [25] J Greenwood. Constriction resistance and the real area of contact. *British Journal of Applied Physics*, 17(12):1621, 1966. doi:10.1088/0508-3443/17/12/310.
- [26] Z Chen and K Sawa. Effect of oxide films and arc duration characteristics on Ag contact resistance behavior. *IEEE Transactions on Components, Packaging, and Manufacturing Technology: Part A*, 18(2):409–416, 1995. doi:10.1109/95.390325.
- [27] J Swingler and J McBride. The erosion and arc characteristics of AgCdO and AgSnO₂ contact materials under DC break conditions. *IEEE Transactions on Components, Packaging, and Manufacturing Technology: Part A*, 19(3):404–415, 1996. doi:10.1109/95.536842.
- [28] D Grogg and C Schrank. Impact of the gas environment on the electric arc. In *62nd IEEE Holm Conference on Electrical Contacts*, pages 125–128. IEEE, 2016. doi:10.1109/HOLM.2016.7780019.
- [29] L Yi and N Lu. Evaluation and comparison of the arc erosion characteristic of Ag/SnO₂/In₂O₃ contact materials used for an automotive relay. In *64th IEEE Holm Conference on Electrical Contacts*, pages 387–393. IEEE, 2018. doi:10.1109/HOLM.2018.8611673.
- [30] P Slade. The transition from the molten bridge to the metallic phase bridge column arc between electrical contacts opening in vacuum. In *23rd International Symposium on Discharges and Electrical Insulation in Vacuum*, volume 1, pages 198–201. IEEE, 2008. doi:10.1109/DEIV.2008.4676753.
- [31] P Slade. An investigation of current interruption in low-voltage circuits. *IEEE Transactions on Parts, Materials and Packaging*, 5(1):56–61, 1969. doi:10.1109/TPMP.1969.1136052.
- [32] W Rieder. Low current arc modes of short length and time: a review. *IEEE Transactions on Components and Packaging Technologies*, 23(2):286–292, 2000. doi:10.1109/6144.846766.
- [33] Z Chen and K Sawa. Particle sputtering and deposition mechanism for material transfer in breaking arcs. *Journal of applied physics*, 76(6):3326–3331, 1994. doi:10.1063/1.358501.
- [34] G Witter and L Polevoy. Contact erosion and material transfer for contacts in automotive relays. In *42nd IEEE Holm Conference on Electrical Contacts*, pages 223–228. IEEE, 1996. doi:10.1109/HOLM.1996.557201.
- [35] Z Chen and G Witter. A comparison of contact erosion for opening velocity variations for 13 volt circuits. In *52nd IEEE Holm Conference on Electrical Contacts*, pages 15–20. IEEE, 2006. doi:10.1109/HOLM.2006.284058.
- [36] W Rieder and A Neuhaus. Contact welding influenced by anode arc and cathode arc, respectively. In *50th IEEE Holm Conference on Electrical Contacts*, pages 378–381. IEEE, 2004. doi:10.1109/HOLM.2004.1353144.
- [37] L Zhao, Z Li, H Zhang, and M Hasegawa. Random occurrence of contact welding in electrical endurance tests. *IEICE transactions on electronics*, 94(9):1362–1368, 2011. doi:10.1587/transele.E94.C.1362.
- [38] H Jiang and L Zhenbiao. A review on welding in make and break operation of electrical contacts. In *3rd International Conference on Electric Power Equipment-Switching Technology (ICEPE-ST)*, pages 1–5. IEEE, 2015. doi:10.1109/ICEPE-ST.2015.7368324.
- [39] X Yang, J Huang, Z Li, J Liu, Q Wang, and M Hasegawa. The preceding voltage pulse and separation welding mechanism of electrical contacts. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 6(6):846–853, 2016. doi:10.1109/TCPMT.2016.2552202.
- [40] A Ksiazkiewicz and J Janiszewski. Welding tendency for selected contact materials under different switching conditions. *Eksploatacja i Niezawodność – Maintenance and Reliability*, 21:237–245, 2019.
- [41] W Ren, Y He, J Jin, and S Man. Investigations of the contact bounce behaviors and relative dynamic welding phenomena for electromechanical relay. *Review of Scientific Instruments*, 87(6):065111, 2016. doi:10.1063/1.4953838.
- [42] T Tamai, K Miyagawa, and M Furukawa. Effect of switching rate on contact failure from contact resistance of micro relay under environment containing silicone vapor. In *43rd IEEE*

- Holm Conference on Electrical Contacts*, pages 333–339. IEEE, 1997. doi:10.1109/HOLM.1997.638059.
- [43] N Kitchen and C Russell. Silicone oils on electrical contacts-effects, sources, and countermeasures. *IEEE Transactions on Parts, Hybrids, and Packaging*, 12(1):24–28, 1976. doi:10.1109/TPHP.1976.1135104.
- [44] M Hasegawa, N Kobayashi, and Y Kohnno. Contact resistance characteristics of relays operated in silicone-vapor-containing and non-silicone atmospheres with different electrical load conditions. In *57th Holm Conference on Electrical Contacts (Holm)*, pages 1–7. IEEE, 2011. doi:10.1109/HOLM.2011.6034787.
- [45] J Zhang and X Wen. The effect of dust contamination on electric contacts. *IEEE transactions on components, hybrids, and manufacturing technology*, 9(1):53–58, 1986. doi:10.1109/TCHMT.1986.1136620.
- [46] R Malucci. Impact of fretting parameters on contact degradation. In *42nd IEEE Holm Conference on Electrical Contacts*, pages 395–403. IEEE, 1996. doi:10.1109/HOLM.1996.557220.
- [47] P Hurricks. The mechanism of fretting — a review. *Wear*, 15(6):389–409, 1970. doi:10.1016/0043-1648(70)90235-8.
- [48] A Siddaiah, A Kasar, V Khosla, and P Menezes. In-situ fretting wear analysis of electrical connectors for real system applications. *Journal of Manufacturing and Materials Processing*, 3(2):47, 2019. doi:10.3390/jmmp3020047.
- [49] P Slade. Opening electrical contacts: The transition from the molten metal bridge to the electric arc. *IEICE transactions on electronics*, 93(9):1380–1386, 2010.
- [50] P Koren, M Nahemow, and P Slade. The molten metal bridge stage of opening electrical contacts. *IEEE Transactions on Parts, Hybrids, and Packaging*, 11(1):4–10, 1975. doi:10.1109/TPHP.1975.1135030.
- [51] C Leung and A Lee. Contact erosion in automotive DC relays. *IEEE transactions on components, hybrids, and manufacturing technology*, 14(1):101–108, 1991. doi:10.1109/HOLM.1990.113001.
- [52] J Swingler, A Sumption, and J McBride. The evolution of contact erosion during an opening operation at 42V. In *51st IEEE Holm Conference on Electrical Contacts*, pages 346–351. IEEE, 2005. doi:10.1109/HOLM.2005.1518268.
- [53] P Slade and E Taylor. The repulsion or blow-off force between closed contacts carrying current. In *62nd IEEE Holm Conference on Electrical Contacts*, pages 159–162. IEEE, 2016. doi:10.1109/HOLM.2016.7780025.
- [54] R Timsit. Electromigration in a liquid bridge before contact break. In *56th IEEE Holm Conference on Electrical Contacts*, pages 1–6. IEEE, 2010. doi:10.1109/HOLM.2010.5619474.
- [55] A Logachev, P Tenitskiy, and A Vykhotsev. Analysis of voltage oscillograms at disconnecting contacts. In *International Symposium on Discharges and Electrical Insulation in Vacuum (ISDEIV)*, pages 225–228. IEEE, 2014. doi:10.1109/DEIV.2014.6961660.
- [56] Z Chen and K Sawa. Effect of arc behavior on material transfer: A review. In *42nd IEEE Holm Conference on Electrical Contacts*, pages 238–251. IEEE, 1996. doi:10.1109/HOLM.1996.557203.
- [57] S Sharkh and J McBride. Voltage steps in atmospheric low current arcs between opening silver metal oxide contacts. In *43rd IEEE Holm Conference on Electrical Contacts*, pages 233–237. IEEE, 1997. doi:10.1109/HOLM.1997.638047.
- [58] P Boddy and T Utsumi. Fluctuation of arc potential caused by metal-vapor diffusion in arcs in air. *Journal of Applied Physics*, 42(9):3369–3373, 1971. doi:10.1063/1.1660739.
- [59] N Vogel. Burning voltage in atmospheric and vacuum short arcs of vanishing length. *Contributions to Plasma Physics*, 29(1):11–24, 1989. doi:10.1002/ctpp.2150290104.
- [60] M Hasegawa, Y Tamaki, and Y Kamada. An experimental study on minimum arc current of relay contacts and possible re-interpretation of the meaning thereof. In *52nd IEEE Holm Conference on Electrical Contacts*, pages 153–158. IEEE, 2006. doi:10.1109/HOLM.2006.284080.

Chapter 4

Data-Driven approaches to reliability and maintenance

Introduction

In the previous chapter aspects of the two considered case studies have been outlined. This chapter offers a review of recent scientific literature in respect to reliability and maintenance paradigms for such systems and components. It aims first at identifying the salient research areas; second to critically assess the results achieved so far; third to characterise the desiderata in both theoretical approaches and methods applied. Recent developments in the oil and gas industry driven by digitalisation in combination with upstream specific approaches to improve asset reliability are reviewed in Chapter 4.1. Subsequently, in Chapter 4.2 the state of the art applicable to EMR reliability, diagnostics, and prognostics is discussed.

4.1 Strategies in the upstream oil and gas industry

4.1.1 Review of topical literature

A thorough bibliometric study of research published in the relevant field has been conducted in order to determine the state of the art of data-driven applications. Using *key-word* search, publications in various literature data-bases have been analysed based on Google-Scholar [1], OnePetro [2], Elsevier [3], Scopus [4], and the Web-of-Science [5]. The search count for different combinations of key-words, e.g. "NN" & "PdM" & "BHA" & "PCBA", has been aggregated over the considered literature data-bases. As shown in Figure 4.1, over the past two decades the utilisation of data-driven applications has increased throughout the sectors of the oil and gas

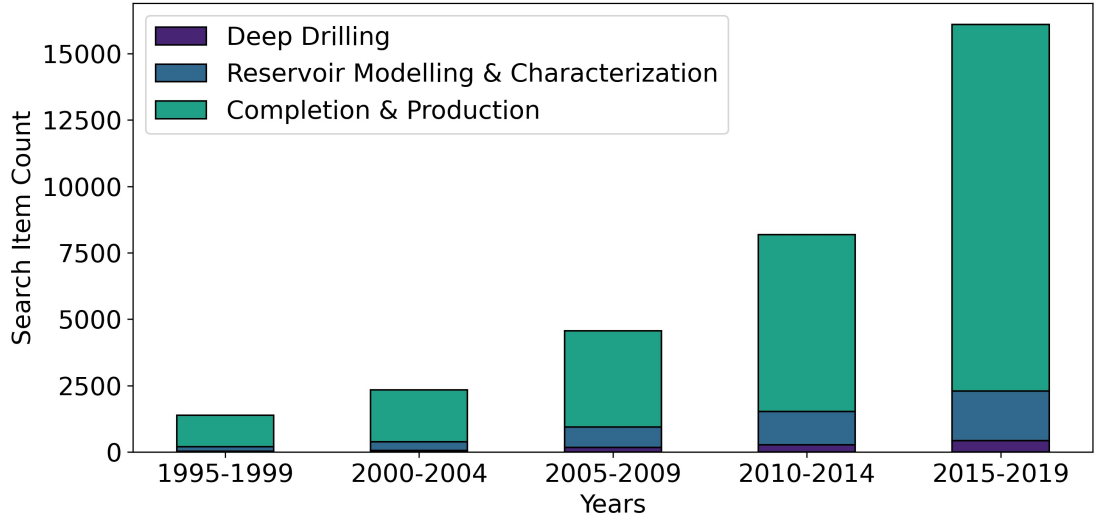


Figure 4.1: Application of data-driven methods throughout the oil and gas industry classified into three major sub-disciplines. Searched research data-bases: [1, 2, 3, 4, 5].

industry. However, the focus resides on *Reservoir Modelling & Characterisation*, cf. [6, 7, 8], and on *Completion & Production* related applications, cf. [9, 10, 11, 12]. *Drilling* as part of the upstream sector has only recently received research interest. The remainder of this section analyses the contribution of relevant data-driven approaches, barriers, and enabling technologies ¹.

Over the last two decades academia and industry have turned their interest towards the application of data-driven methodologies leveraging, e.g. BD in the context of deep drilling [13]. Many deep drilling related applications focus on performance optimisation during the drilling process. Conventional drilling performance metrics such as WOB, Rate Of Penetration (ROP)², or torque have been used to identify critical issues during the drilling processes. The authors in [14] present an approach using a NN to identify bit balling³ based on such features. The NN is trained on time series data in order to predict the ROP which is then compared to the measured ROP during drilling. The authors conclude that bit balling is likely whenever the experienced ROP deviates from the model's prediction. Likewise, a DL method for ROP prediction is presented in [15]. A LSTM model is deployed which incorporates additional features for training to improve forecasting performance. These features are: type of the drill bit, formation properties, rheological properties of the drilling mud. Alternatively, a recent case study by [16] demonstrates how a ML method employing DT can be used to forecast ROP based on surface measurements such as WOB, Rotations Per Minute (RPM), torque, and drilling depth. However, due to

¹Parts of this chapter have already been published in my journal or conference article: L. Kirschbaum, et al., (2020).

²Rate Of Penetration - the speed at which the drill bit advances through the formation.

³In shale formation clay particles might stick to the tool face blocking the bits' nozzles and reducing the ROP which is known as bit balling.

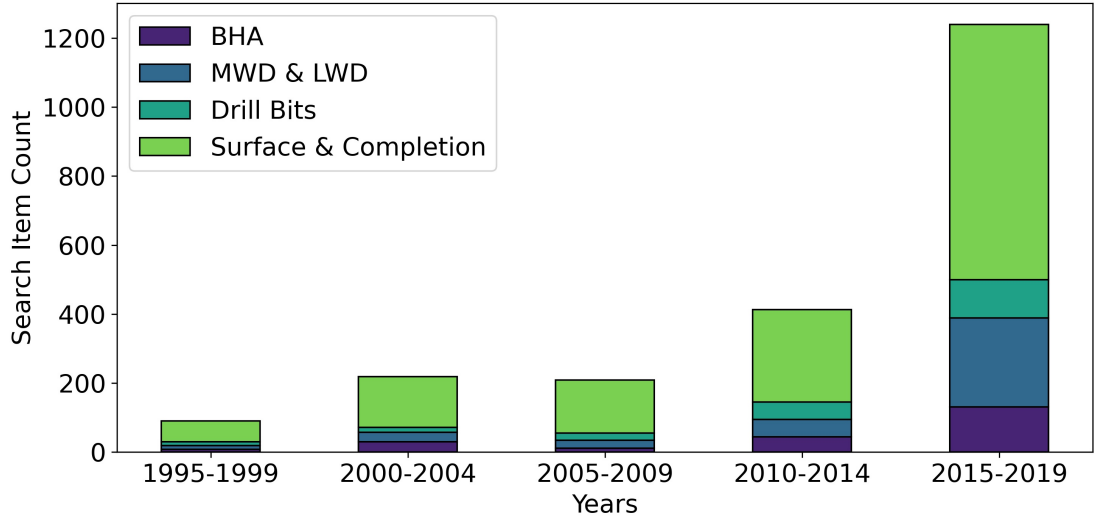


Figure 4.2: Application of data-driven methods for different downhole tools, i.e. BHA modules in general, MWD & LWD, and Drill Bits compared to general Surface & Completion related applications. Searched research data-bases: [1, 2, 3, 4, 5].

a tendency towards overfitting, [17] discards DT as suitable method for ROP prediction. Instead, the authors utilise a Random Forest Regressor that is less prone to overfitting and further improves performance over various considered ML algorithms such as Support Vector Machines, K-Nearest Neighbours or NN. In [18] the authors propose a deployable real-time solution for ROP prediction. Its performance is evaluated by comparing eight different algorithms. The algorithms are trained on a data set aggregated from drilling operations at 50 different wellbore locations. In addition to the above-mentioned features, additional features like flow-rate and formation property indicators are being used to boost performance. In order to improve interpretability and increase the reliability of ROP prediction a hybrid approach combining physical models with ML is suggested in [19]. Further examples using AI for ROP estimation can be found in [20, 21].

As evidenced in Figure 4.2, a more detailed analysis of data-driven methods for oil and gas industry assets reveals an increasing interest in the application of these approaches for downhole tools. Only a small proportion of relevant publications considers data-driven methods for optimisation of operation and maintenance applied to downhole tools, e.g. for the BHA amounting to 8.4 % of the considered research body between 2015 to 2019. However, recent trends that relate to the increasing economical and operational challenges the oil and gas industry is facing, suggest significant future investments into novel maintenance related applications. Such investments have a technology-enhancing industry-wide effect on the oil and gas industry, but in detail they have a particular effect on the development of data-driven maintenance strategies for the upstream industry, especially for deep drilling tools. To further illustrate such research efforts, the remainder of this section reviews

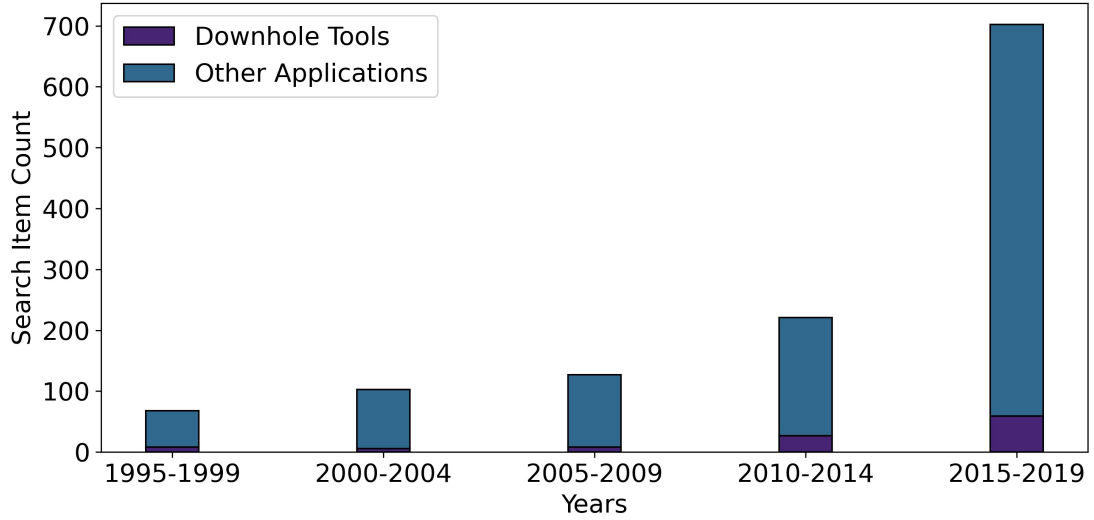


Figure 4.3: Data-driven methods for novel maintenance strategies distinguishing between downhole tool applications and the remained of oil and gas sectors. Searched research data-bases: [1, 2, 3, 4, 5].

approaches relating to the upstream industry.

Predictive maintenance of artificial lift systems is an active field of research. These lift systems are used in the completed wellbore to maximise production from the well once the initial pressure of the reservoir has depleted. In [22] Principal Component Analysis is applied to identify deviation from nominal operation conditions of Electrical Submersible Pumps (ESP)s which are used as artificial lift system. This promising approach identifies impending pump failures. It coincides with findings reported in [23]. A Support Vector Machine based approach is developed by [24]. Thereby, the authors demonstrate an alternative ML application for the prediction of ESP failures. Besides ESP, a system to prolong oil production is known as beam- or rod-pumping. Traditionally the working conditions of these pumps are assessed by the use of *dynamometer cards*. The authors in [25] deploy a CNN for classification of the operating stages of rod-pumps from such *dynamometer cards*. The method achieves over 90 % classification accuracy of rod-pump failures during a field deployment test. A different approach for rod-pump diagnostics is presented in [26]. Here, the authors identify DT to be well suited for classifying the operational state of rod-pumps. Further examples can be found in [27, 28, 29, 30, 31]. Novel maintenance is of increasing interest in other upstream sectors as well. Offshore assets such as platforms, generators, or turbines are the focus of multiple AI-driven CBM, PdM, or digital-twin applications [32, 33, 34]. For example, the detection, analysis, and prediction of corrosion of offshore infrastructures with state-of-the-art ML algorithms is an actively pursued area of research, cf. [35, 36].

Contrary to the application scenarios outlined above, research related to the BHA or its modules, e.g. the RSS, MWD, and LWD only accounts for a small portion

of the overall research efforts, cf. Figure 4.3. Multiple reasons explaining this lack in research can be identified. Besides the constraints imposed by the harsh and uncertain environments in which downhole tools operate, the challenges associated with a data-driven maintenance approach need to be understood. Restrictions in terms of hardware availability and a lack in data availability have – so far – impeded the broad application of data-driven maintenance strategies. While the latter one is by far the more pressing issue, hardware constraints in downhole tools have proven to be a major hurdle. In [37], the author discusses various reasons which prevent the installation of additional sensing capabilities required for more elaborate data-driven maintenance applications: space constraints, long lasting development cycles, and hardware availability. Furthermore, any computational process needs to be designed to use as little power as possible to avoid any additional heat dissipation from the electronic components. Hence, the computational resources for downhole data processing, e.g. high-volume data sampling, are limited. Along these lines, the authors in [13] state the need for improved electronic resilience in combination with extended functional capabilities to be able to continue to cope with those challenges. This insight is supported by the advancements in enabling sensors and electronics.

As evidenced in Chapter 2.5, the oil and gas industry has begun to adopt modern digital infrastructures to streamline and centralise data from drilling operations. Admittedly more data is becoming available, however, the limited access to this data still represents a significant challenge for the implementation of data-driven maintenance for downhole tools. Here, the global scale of drilling operations, the volume, veracity, variety, velocity, and value of the data along the manifold value chain towards data aggregation and management have to be taken into account. For example, [38] highlight the need of access to complete and concise data throughout the entire life of a downhole tool, essential for robust lifetime prediction. This rationale is supported by [39], stipulating the importance of high quality/ high frequency data for detection of tool failures and a complementary tool database capturing the tool history from various data sources. Further case studies of the challenges the oil and gas industry is facing in centralising and streamlining drilling data are presented in [40, 41].

4.1.2 Comparative assessment

The success of upstream data-driven maintenance, therefore, depends heavily upon future data consistency and availability. In addition, increasingly complex well geometries and harsh drilling environments are to be taken into account. According to the analysis of scientific literature and personal consultation with experts the following five major challenges can be formulated:

1. Increasingly complex electronic assemblies render current maintenance strate-

gies inefficient.

2. Extreme environmental, dynamic operation conditions restrict the applicability of metrics such as MTBF.
3. Real-time health monitoring is constrained due to the limited data transfer during drilling.
4. Space, design, and hardware constraints impose restrictions on retrofitting required sensing capabilities, e.g. canary-sensors.
5. Governmental data restrictions, local data-silos, and industry policies limit access to data required to develop data-driven solutions.

This list makes the interdependencies and constraints as well as the desiderata in research paramount. The following section presents research approaches that directly tackle these issues. Already in 1999, Aldred et al. [42] stressed the importance of a holistic data approach in order to minimise unplanned downtime due to BHA failure. The concept of a holistic BHA reliability method in order to improve the decision making for BHA maintenance is then discussed in [43]. BHA reliability is improved by implementing a streamlined approach to capture tool-data and failure-reports in order to reduce the number of systematic downhole tool failures. However, again, the authors note that maintaining such an approach on a global scale and throughout many departments participating in the development, maintenance, and operation of the tools requires considerable efforts. In order to illustrate the potential, the authors in [44] propose - as one of the first - a data-driven model to determine the stress level a BHA has been exposed to. Cumulative stress-time functions are calculated using lateral vibration run-to-failure data. This resembles the encountered stress history of the BHA and reflects various stress scenarios. Subsequently, approximation functions are used to determine the stress-history of an operating BHA. This new curve is then compared to the nominal stress-time type curves using kernel regression. Via a similarity measure a RUL prediction for the operating BHA is derived. The authors achieve an accuracy range of 2 – 20 % of the total tool lifetime. The authors further refine their methodology in [38] to address the common issue of unavailability of complete BHA stress-history data records. The proposed health state estimation model determines the stress-time curves based on parts of the data rather than the complete history. The overall life of the BHA is then estimated based on these results.

The requirements for an offline RUL estimation model supporting the operator's decision whether to rerun a downhole tool is investigated in [39]. In order to reduce NPT, [45] present a RNN based approach to estimate the RUL of composite drill pipes. Failure of drill pipes often results in long downtime as recovery from the wellbore is difficult. With a similar aim, the authors in [40] develop a data-driven model for anomaly detection of the RSS using a non-parametric Fuzzy Inference system that relies on field-data measurements (RPM, electrical current of the mod-

Table 4.1: Overview of data-driven maintenance applications for downhole tools.

Ref.	Year	Tool	Approach
[43]	2008	BHA	Establishing a reliable tool data base to aid operational and maintenance decision support respectively improving BHA fleet reliability.
[44]	2009	BHA	Cumulative vibration stress profiles reflecting tool degradation; RUL forecast based on stress profile similarity measure.
[48]	2010	MWD	Identification of meaningful features from field-data relating to MWD failures.
[38]	2010	BHA	Extending the approach developed in [44] to reduce the required data volume.
[40]	2011	RSS	Anomaly detection using non-parametric Fuzzy Inference System and a sequential probability ratio test to identify possible degradation trends utilising field-data.
[45]	2014	BHA	RNN for RUL prediction based on field-data for composite drill pipes.
[49]	2014	BHA-PCBA	Field-data based cumulative degradation model for RUL estimation.
[47]	2015	BHA-PCBA	Retrofittable health management solution for RUL estimation using canary sensors.
[50]	2015	RSS	Maintenance framework utilising field-data and failure-reports.
[51]	2018	LWD	Fault classification using DT.
[39]	2018	BHA	Framework for holistic downhole tool data aggregation to derive fleet-level reliability information in order to reduce NPT.

ule, steering-pad pump pressure). Based on BHA-PCBA failure reports, in [46] an elaborate hybrid method for the development of a cumulative stress model for RUL forecasting and maintenance decision support is presented. In [47] a retrofittable health management solution for real-time BHA-PCBA prognostics is proposed. The authors suggest implementing a set of canary sensors to be embedded in the PCBAs recording different sources of downhole tool vibration. Against the background of pre-defined stress-levels a cumulative stress-score is attained. A data-driven degradation model correlates the canary sensor failure times with the encountered stress-levels during drilling in order to provide a RUL forecast.

In [48], the authors establish a relation between drilling dynamics and MWD tool failures. Sources of downhole vibration data are consulted to obtain failure probabilities using a Logistic Regression approach. These probabilities reflect the cumulative stress the MWD module has been exposed to. However, the authors state that their approach operates under the assumption that each module is considered new at the beginning of the drilling operation. This assumption clearly limits the applicability of the proposed method because the complete stress history of the tool is no longer considered. Lastly, a data-driven method for the assessment of the health state of the neutron pulse generator of the LWD module is presented in [51]. Therefore, tool memory data is downloaded after each run. Relevant data is selected based on expert knowledge. A robust health state indicator is constructed using three input features, namely: the output of the neutron detector, the drawn current, and

voltage. Based on these features a DT Classifier is trained, determining a healthy or non-healthy operating state of the neutron pulse generator during drilling. Further case studies that look at the potential of data-driven maintenance for downhole tools are presented in [50, 52, 49]. An overview of the reviewed applications is provided in Table 4.1.

4.2 Strategies for EMR operation and maintenance

Various measures have been developed to determine and represent the degradation in EMR applications. Of particular interest is the electrical lifetime subject to contact wear [53, 54]. Hence, as a first step this chapter discusses the advantages and disadvantages of EMR performance indicators that aim to quantify the degradation by introducing CR in Chapter 4.2.1 and commonly deployed reference Degradation Indicators (DI)s in Chapter 4.2.2. Thereafter, Chapter 4.2.3 reviews the reliability and maintenance related strategies for a number of critical EMR applications; it concludes with a critical assessment of the examined approaches presented in Chapter 4.2.4 ⁴.

4.2.1 Contact resistance

The most popular measure among the developed DIs is CR [55]. However, as the initial EMR-CR is already very small, its increase until failure is typically only within the $m\Omega$ -range. This poses a challenge for accurately measuring changes which can be achieved reliably only with a *4-Wire* setup [56]. Therefore, presupposing accurate CR measurements in embedded online health management is not a viable solution due to the required sensing hardware and associated costs. In addition, depending on the type of EMR, CR is subject to more or less random fluctuations ranging well above the rated maximum permissible CR and masking underlying trends. This renders the definition of a static CR-based EOL threshold unfeasible. The rate and intensity of fluctuations depends inter alia on the operating pattern and the load, but foremost on the contact material and environment [57]. Hence, CR measurements are often not capable to capture the overall EMR degradation [58, 53]. Chapter 7.2 discusses the aspects of CR fluctuations in detail.

4.2.2 Derived reference DIs

The above mentioned challenges in combination with the need to distinguish failure modes drove the search for alternative DIs. Derived DIs aim to depict EMR wear and can be split in two groups namely: (1) *non-intrusive* time-based reference DIs

⁴*Parts of this chapter have already been published in my journal or conference article: L. Kirschbaum, et al., (2022).*

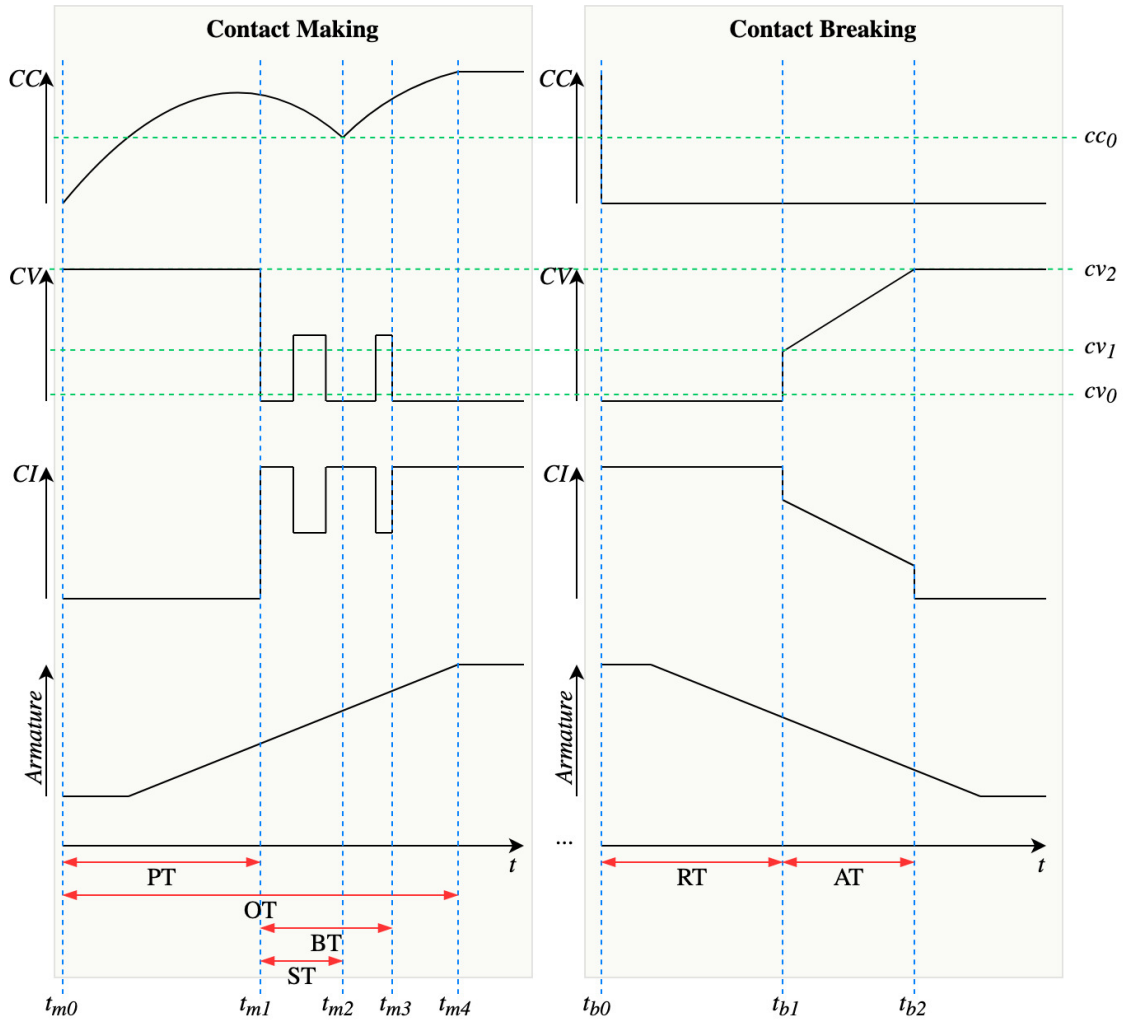


Figure 4.4: Selected time-based reference DIs derived for contact making of a normally open EMR. Pick-Up Time (PT), Over-Travel Time (OT), Bounce Time (BT), Super-Path Time (ST), Release Time (RT), Arcing Time (AT); cv_0 closed circuit voltage, cv_1 minimum arcing voltage, cv_2 open circuit voltage, cc_0 inductive nick current.

and amplitude-based reference DIs; (2) *intrusive* DIs which require disassembling the EMR or ancillary sensing capabilities.

4.2.2.1 Non-intrusive DIs

Such DIs predominately rely on the measurement of CV and Coil Current (CC) waveforms during contact making and contact breaking. By processing CV and CC waveforms one can derive a set of measures as displayed in Figure 4.4 to detect rising or falling edges. Therefore, threshold values for the closed and open circuit voltage, the minimum arcing voltage (e.g. silver-plated copper contacts approximately 13 V) and the coil current are used that are further detailed in Table 4.2. However, note that for time-based reference DIs naming conventions throughout literature are somewhat ambiguous and depend on the type of contact configuration [59].

The trajectory of time-based reference DIs over the EMR life time varies in respect to the experienced failure mechanism, the design of the EMR, and the operational

Table 4.2: An overview of classical EMR DIs.

Time-based reference DI		
Pick-Up Time		The time between the increase of coil voltage and the first CV drop during making. BT is not considered [63]. For example, [64] report that the magnetic force decreases if the coil wire resistance increases, causing a slower movement of the armature and therefore an increase in PT. PT decreases if the spring force decreases due to stress relaxation.
Over-Travel Time		The time between the armature/ contact start of travel and the complete closure of the armature [62]). For example, [57] report decreasing OT till stuck-open failure as the contacts erode and the subsequent armature over-travel becomes too small to force proper contact making. Likewise [60] reports decreasing OT for eroding contacts. However, if contact sticking/ welding is predominant OT is likely to increase.
Arcing Time		The time when CV is between the limits of 10 % and 90 % of the open circuit voltage during contact breaking or contact making [54]. The effective start of the arcing duration depends inter-alia on the minimum arcing-voltage of the contact material.
Bounce Time		The interval between the CV drop during contact making and the last bounce, i.e. before the CV settles. The start of BT is sometimes specified at 90 % of the open circuit voltage; the bounce pulse has to be longer than 10 μ s [59]. The entropy of the BT appears to correlate with the EMR-EOL [65]. BT significantly depends on the amount of contact surface wear [66].
Release Time		The time interval between the coil voltage drop/ de-energization and the initial CV increase. BT and AT are not considered [59]. RT increases towards EOL when contacts are subject to erosion. A decreasing trend is reported in [60] if contact welding is the governing failure mechanism.
Super-Path Time		The time interval between the CV drop at the contacts and the <i>inductive nick</i> of the pick-up CC at the coil during contact making [63]. [67] finds this to be a significant DI.
Amplitude-based reference DI		
Dynamic Contact Resistance (DCR)		The resistance across the closed contacts just after the contacts have settled. This can be measured by recording CV and CI, though the accuracy of the used sensing equipment needs to be considered. For an extensive discussion on Dynamic Contact Resistance (DCR) as DI refer to [66].
Pull-In Voltage		The coil voltage at which the armature movement commences. This DI has relevance for coil failures as an increase in coil resistance can be observed via an increase in the pull-in voltage [62].
Dynamic CV		The CV drop upon first contact making.
Static CV		The CV as soon as contact bouncing has settled and contact is made.

environment [60, 61, 62]. However, as Section 4.2.3 will demonstrate in detail, no generally valid non-intrusive DI has been established so far within the body of reviewed research. Nevertheless, DIs can provide application-specific, valuable information regarding the state of EMR degradation. An overview of EMR time-based reference DIs and amplitude-based reference DIs is provided in Table 4.2.

4.2.2.2 Intrusive DIs

Various research approaches have used intrusive measurements to describe and quantify EMR contact degradation, e.g. the mass transfer of the contacts determined by detaching and weighing the contacts at regular intervals. Alternatively, radioactive tracers have supported the understanding of material transfer among switching,

arcing contacts [68]. Further analysis of contact surfaces at different stages using Energy Dispersive X-Ray (EDX) and Scanning Electron Microscopy (SEM) have been employed [69, 70]. In [71] the arc discharge is optically monitored. Lastly, measurements of static and dynamic contact pressure have allowed an assessment of the contact health [72]. While all these DIs are able to depict the degradation in some way, obtaining these measurements at scale outside controlled laboratory environments is impractical [59, 73].

4.2.3 Review of topical literature

This chapter reviews research approaches that consider the plethora of discussed DIs and CR in order to derive data-driven methods that support EMR maintenance decision making. The following section is structured by the EMR application domain since the reviewed research relies on previous findings from those industrial sectors. General DC EMRs are found in consumer electronics. They rarely play a mission or safety critical purpose. However, as the review shows researchers actively explore the capabilities of data-driven maintenance in domains where EMRs play such a critical role: for example, the aviation and aerospace industry (Chapter 4.2.3.2, cf. the ISO 5867-1:1996 [74]); the railway industry - here EMRs are used, e.g. track circuit monitoring to detect whether a certain part of the railway track is currently occupied by a train (Chapter 4.2.3.4); or the automotive industry (Chapter 4.2.3.5) that requires low manufacturing costs - but the EMRs also have to withstand high levels of vibrations and shocks, varying temperatures, contamination whilst also being able to switch circuits with inrush currents of up to 100 A.

4.2.3.1 General DC EMR

In [75], the authors present results of life-cycle tests and a failure mode analysis, studying silver-plated copper contact rivets. A high, repeating rate of random, but non-critical switching failures (i.e. relatively high CR while closed) is observed, after which the EMR returns to nominal operation. Different CR stages can be distinguished over the course of the EMR life. An increase of the CR beyond the manufacturer's maximum rating and CR fluctuations are reported. Towards the EOL the amplitude and frequency of the fluctuations increases, though, an underlying increasing trend of CR is documented. The authors attribute this to contact contamination which deteriorates the connection between the contact surfaces. Just before the EMR's EOL a rapid CR increase is observed. The operational failure is traced back to an excessive build up of oxide layers on the contact surface. Reaching a critical thickness, the contact force is too low to cause a rupture of the contamination layer, nor can a conducting path be established by electrical arcing. Apart from unacceptable high CR failures, mechanical failure of the contact carrier (spring-type) is discussed. The authors extend their research towards a predictive maintenance

approach in [58], focusing on prediction of the CR degradation trajectory. A Moving Average Model, an Auto Regressive Integrated Moving Average (ARIMA)-Model, an Exponential Smoothing Model, and a NN-Model are compared in terms of predictive performance. The last one is found to be best suited to predict EMR-EOL. However, the authors state that predicting the EOL solely on CR is highly dependent on the type of application and, thus, does not generalise well.

The authors of [76] are concerned with improving EMR maintenance schemes. The authors point out that in many instances EMRs are exchanged early, often in accordance with pre-determined life estimates, cf. preventive maintenance approaches. Following up on the arguments given by [58], [76] emphasise that dynamic CR measurements should be used in addition to static CR measurements for EMR-RUL estimation. The authors demonstrate that Dynamic Contact Resistance (DCR), recorded during the closing actuation, can be a valuable DI. DCR shows a comparatively more pronounced trend. A statistical regression model to estimate the EMR life that forecasts CR is then presented. Even though CR is one of the most prominent indicators of contact life, in [77] the authors refine their prior work stressing the importance of alternative DIs, since an accurate analytical solution describing CR is yet to be found. Therefore, to enable data-driven methods, meaningful EMR DIs need to have a low correlation among each other, while highly correlating with the EMR's life. Using factor-analysis, the authors identify promising indicators from the sensed waveform, being the DCR, CR_{max} , and BT. Based on the extracted features a fuzzy model is used to evaluate contact reliability. However, no further insight into the concrete nature of the measurements, nor the contact failure modes is provided. The authors revisit their approach in [65], addressing the use of BT for estimation of the EMR performance. Analysing the entropy of the BT using sequence encoding, the extracted trend relates to the EMR life and its EOL, although the analysis is performed on a fairly small data set.

The research in [60] is motivated by the need for a novel method to evaluate the EMR life, since traditional reliability methods have become too time demanding no longer yielding failures in reasonable test-time due to improvements in design and quality of EMRs. The experiments conducted within the study further reveal that different failure mechanisms change the shape of DI degradation trajectories, demonstrated for Closing Time (CT), RT, and OT. It is found that OT decreases, and RT increases as the contacts erode; however, if contacts fail due to contact welding OT increases and RT decreases; contact failure due to contamination and unacceptable high CR did not exhibit a clear trend. As in [75], the authors attribute this to alterations of the mass transfer between the contacts and the resulting random contact morphology. The potential of a regression model combining the effects of the identified failure mechanisms to predict the EMR degradation process for reliability purposes is explored. In [61], the authors confirm prior findings, high-

lighting different characteristic trajectories of DIs for various failure mechanisms, e.g. contact erosion, contact welding, and contact contamination. A trendable DI is presented, referred to as the fluctuation coefficient, i.e. the correlation between the changes in CR, PT, RT, OT, BT, and AT. After preprocessing features using Wavelet Decomposition a linear-model is derived for DI trajectory forecasting. The best performance is achieved for either OT and BT related degradation trajectories if the EMR's degradation is dominated by contact erosion. The subsequent model does not perform as well for contact welding and contamination failures. While the proposed fluctuation coefficient improves the performance for these failure mechanisms the reported accuracy, however, remains low. In addition, the authors study the performance of a physical model considering arc energy and contact form factors mapping the degradation of contacts to different environmental factors and materials.

A recent analysis highlights two key challenges for a broad application of PHM to EMR [62]: (1) the lack of life-cycle data, since, currently, there are only very few deployed online monitoring health management solutions for EMR; (2) the uncertainty associated with the DIs. To address the former issue, simulation of EMR life-cycle data is proposed. Coil and armature related failure mechanisms are simulated and compared to the measured CC waveform. This allows to derive a reference for healthy or unhealthy CC waveform shapes. A diagnostic framework is proposed using the Mahalanobis Distance to discriminate between EMR operation states and failure modes respectively. [78] investigates DIs of EMRs subjected to contact welding under a Direct Current (DC) lamp-load setting. The representative PDFs of CC and CV for different states throughout the EMR life are determined.

4.2.3.2 Aviation and Aerospace DC EMR

In [53] the authors propose a methodology for aerospace EMR life estimation. Key to this methodology is the decomposition of the derived DIs into a trend, cyclic terms using Wavelet decomposition, and a stochastic term. Forecasting models are developed respectively, using a Polynomial Model and Auto Regressive Moving Average (ARMA) Model. The authors find that such preceding decomposition might improve accuracy and point out that previous work has focused solely on the trend term of degradation parameters, e.g. [79], ignoring the influence of repeating/ fluctuating patterns in the waveform, i.e. the cyclic term, that can deteriorate the performance of the forecasting model. The authors affirm that CR is not a robust DI due to the encountered significant fluctuations. However, PT, OT and RT are found to be good indicators reflecting the degradation behaviour of the EMR contacts. PT increases throughout the EMR life which is attributed to a growth of the contact gap due to contact erosion. PT returns the most accurate forecasts for EOL estimation among all tested DIs. Though, as previously pointed out by [58, 80], the success of such ap-

proach highly depends on the specific EMR design and the distinct application. As in [53], in [81] the authors present a time series based EMR life prediction method identifying CR, the CV peak voltage, BT, RT, and Super-Path Time (ST) as valuable DIs. Relying on ST, a Volterra Series based Regression Model in combination with optimised Wavelet Decomposition, significant frequencies are derived to forecast the ST trajectory. The model accurately predicts ST through the second half of an EMR's life. However, no EOL threshold is presented making an evaluation of the performance of the proposed model challenging. In [82] the authors show that the initial mean and variation of CR time series measurements can be used to estimate the life of the EMR. The value of the proposed approach lies in the capability to determine whether a new EMR will perform within its specification throughout the rated life under known operational conditions based only on its initial state. The initial parameters are determined over the course of the first 1000 actuations. To estimate robust parameters, the authors derive the probability-weighted average of the selected DI. The life of the EMR is modelled using polynomial-fitting. The authors, using EMR life-cycle data, determine the relationship between the EMR life and the sum of the weighted distance of the initial mean and variation of CR. The proposed model shows a tendency to underestimate the EOL, yielding an early prediction of around 15 %. Since only a small data set was used, it is difficult to assess the applicability of the proposed approach.

[83] combine a physical-model of erosion related contact degradation with the findings of an ALT performed on EMRs used by the aviation industry. Using Kalman Filtering, the measured CR is fused with the physical-model to estimate the degradation dependent trajectory of the CR. However, no measure of performance for this approach is presented. As emphasised before, the feasibility of CR as the sole measure for the advancing EMR degradation is conditional on the application chosen [75, 60]. The authors themselves present evidence of the shortcomings when predicting CR as they extend on their initial research in [84]. Results from EMR life-cycle tests are introduced which focus on contact welding related failures. As in [75], [84] find that the CR degradation trajectory can be separated in various stages. In some instances it is shown that CR increases linearly until the manufacturers' stated minimum operating life. Beyond this, CR is subject to significant fluctuations. However, the authors disregard CR, since it lacks the necessary expressive power in order to be considered as robust DI (high variance among the recorded samples). It is further limited by the anomalous short-time fluctuations throughout its operation life. Nonetheless, the authors postulate a regression model using Grey-System theory relying on the CR EOL-threshold. And yet, these very fluctuations in combination with a static threshold cause the proposed model to predict the EOL too early.

4.2.3.3 High-Voltage DC EMR

The authors in [73] review the failure modes of High-Voltage DC EMRs. Here, the predominant failure mechanisms are contact erosion and contact welding. To judge the severity of arcing the authors point out that contact mass variations are unsuitable for online life time prediction models as they are difficult to obtain without dismantling the EMR for each measurement. However, the arc charge directly relates to the contact mass loss from electrical erosion. It can be calculated by measuring the current of the arc discharge. In turn, this serves as a DI, since the cumulative variation of the arc charge increases linearly towards the EOL, whereby the failure threshold is experimentally determined for the presented application. Furthermore, the effects of the arc charge under the influence of a permanent magnet used for arc suppression are studied. As expected, with increasing magnetic flux density the arc duration is shortened. Likewise, the cumulative variation of the arc charge exhibits a smaller slope. The authors derive the linear relationship between flux density and the EMR-life. Using the proposed method, the minimal flux density needed to reach a specified number of actuations is estimated. The authors state that their method could decrease EMR development time. The applicability of the proposed approach to low-voltage DC EMR applications and the effects of contact contamination on the arc charge are to be discussed. Building on [73], the authors in [80] analyse the correlation between contact velocity and electrical life. Experimentally it is demonstrated that higher breaking velocities decrease arc erosion as the arc is extinguished earlier due to the widening contact gap. The authors show that the cumulative arc erosion mass under different breaking velocities increases linearly with decreasing breaking velocity. A mean EOL threshold based on the arc erosion mass is experimentally determined. It is pointed out that contact velocity and the magnetic flux density are both subject to variations due to manufacturing tolerances which may significantly impact the life estimation. A prerequisite for the proposed method is an extensive experimental study of the parameters contributing to electrical erosion in order to determine the linear relationship between the increasing erosion rate, the magnetic flux density, and the contact velocity. Therefore, as it is the case in [73], the generalisability of this approach is at issue.

4.2.3.4 Railway EMR

[54, 67] present diagnostic methods for railway EMRs using Wavelet Decomposition for denoising of extracted DIs. RUL prediction is performed with ST, BT, AT, and RT as features. To further constrain the feature space, reliable degradation features based on the Fisher-Discrimination-Criterion are selected. ST and BT are identified as key DIs. Contact erosion, contact welding, and contact contamination are the dominant failure mechanisms important to railway EMR as the study suggests. To account for the variance in the derived DIs, the authors propose the use of the

Mahalabonis-Distance to classify failure modes, cf. [62]. For RUL prediction, NN is employed, achieving 84 % forecasting accuracy. However, the study does not provide details on the extent of the prediction horizon. As before, the authors emphasise the difficulties when selecting appropriate EOL thresholds for respective EMR-DIs.

In [85], the authors address the previously encountered challenges. A novel EMR life forecasting strategy is proposed, allowing a prediction horizon of up to 500 actuations ahead. The authors point out how this strategy can be beneficial for railway related PdM. RT is identified as key DI using the method previously presented in [54]. The non-linear and non-stationary behaviour of the degradation process is emphasised in this research context. An ensemble of Empirical Mode Decomposition in combination with improved Variational Mode Decomposition is proposed in order to decompose the RT time series. Eight Intrinsic Mode Functions and a Residual Component are then derived. These features act as input to a multi-layer Radial Basis Function NN. The NN predicts the trajectory of RT. The authors demonstrate that the proposed preprocessing steps to prime the input features for the NN improve model performance. Nevertheless, the meaningfulness of predicting a DI trajectory to infer EOL is debatable, since the DIs are prone to fluctuations and high levels of variance as reported in previous works. Accurately predicting such fluctuations is not of immediate benefit to EMR-RUL forecasting and underpins the need for uncertainty quantification with any prognostics methodology.

[86] propose a method for PdM and the reduction of test-time for railway EMRs. CR and CT are recorded during ALT experiments at various elevated temperature levels in order to shorten the required time-to-failure. Two physical models are derived: (1) with regard to a CR increase that is attributed to contact corrosion; (2) with regard to the observed changes in CT. To determine the coefficients, the models are fitted to life-cycle data using the Least Square Method. The fitted models are evaluated on the EMR life at lower temperatures, yielding a low prediction error against the observed EMR life.

4.2.3.5 Automotive EMR

[63] consider CR, PT, ST, BT, AT, and RT as DIs for automotive EMRs. The authors emphasise the degradation behaviour of the EMR motivates the selection of a NN model due to its capability of learning non-linear relationships. Using the aforementioned DIs as input, the authors evaluate the performance considering various NN topologies and training sets to predict the EMR-RUL. They deploy a NN using a Wavelet Function as activation function, since it yields superior results. The failure mode of the tested automotive EMR is not further specified in the publication. In [66] an alternative model to predict EMR life in automotive applications is proposed, using the Improved Fireworks Algorithm Grey NN, i.e. a swarm opti-

misation based algorithm. The method is evaluated with life-cycle tests at different temperatures predicting the EMR-RUL based on the initial state of the EMR where measurements of DCR, closed-circuit and open-circuit contact voltage, and BT are taken at regular intervals. Finally, the authors stress that the model could be further refined if more parameters would be considered, e.g. the operating environment.

4.2.3.6 Nuclear EMR

In [87] an embedded, non-intrusive method to determine welded EMR contacts in safety critical application is presented. Contact welding poses a significant risk, since some circuitry could no longer be decoupled during an emergency. This insight drives the research for an online health monitoring scheme in the context of nuclear power plants. Recall, the armature movement produces a distinct CC waveform (the so called *inductive-nick* and *inductive-peak*), since the travelling armature affects the inductance of the coil [88]. [87] demonstrate that welded contacts prevent the EMR armature from moving in a similar fashion. Thus, the CC waveform does not exhibit such distinct characteristics. To not adversely affect the nominal operation of an application through contact making and breaking a short de-energisation of the coil is followed by an immediate re-energisation of the coil (prior to contact breaking). The CC waveform is recorded. It is concluded that the distinction between welded and non-welded contacts is possible solely based on the shape of the CC waveform. Therefore, the EMR contacts are not required to be opened in order to determine the presence of contact welding. The authors are aware that the application of the proposed method heavily depends on the type and design of the EMR. In [89] they expand their research. An integrated circuit for online diagnostic of EMR contact welding detection referred to as *Relay Output Card* is presented. Using a Markov Model and the CC waveform shape characteristics deduced in [87], the capabilities of the proposed circuit are demonstrated by improving the reliability of safety critical EMRs with low failure rates in nuclear application. Similar research has been conducted by [90]. The authors extend on the non-intrusive contact welding detection proposed by [87, 89]. A range of failure criteria aiding the automatic detection of contact welding is reviewed and the robust determination of a healthy CC waveform using an embedded circuit is further refined.

4.2.3.7 Storage of EMR

The degradation of EMR when stored, in particular, the deterioration of the contacts during storage has been subject to an evolving field of research over the past decade [91, 92, 93, 64, 94, 95, 96] and [69, 97]. In [91], the authors expose aerospace EMRs to accelerated degradation testing without contact switching under elevated temperature conditions. Performance measures are taken, namely CR. It is shown that temperature stressing of EMRs leads to a faster increase in CR. This behaviour

becomes significant beyond 100 °C. The authors identify temperature accelerated corrosion as a principal cause for high CR and low conductivity due to surface film formation of oxide corrosion films in silver-plated contacts. These effects are further analysed in [96, 95], attributing the increase of CR during extended storage time not only to oxides but also to sulphides and carbides depositing on the contact surface. The authors demonstrate that measurements of PT, RT, OT, and BT exhibit distinct trends [91]. Exploiting these findings, a method to predict CR for EMR in storage is proposed in [92]. The selection of Grey-NN as prediction model is motivated by the non-linearity of the degradation process, cf. [63]. The proposed, combined Grey-NN is superior to using either standalone Grey-System-Theory or NN. A similar approach, employing Grey-NN for aerospace EMR in storage is presented in [94]. However, CT is chosen as prediction target instead of CR. In [93], the authors link their previous research of elevated temperature testing to storage life prediction at ambient temperature. To facilitate this, the authors pre-process the CR measurements obtained at elevated temperature using Wavelet Decomposition. An exponential model is found to exhibit the best predictive performance for CR. Then, the maximal storage life under ambient conditions is obtained using the Arrhenius Model. However, it is important to point out that no experimental validation for the estimated EMR storage life under ambient temperature is provided.

In [64] the authors develop a degradation model for aerospace EMR using PT. Two major factors affecting PT are identified. If the coil wire resistance increases, the magnetic force decreases causing a slower movement of the armature and, therefore, an increase in PT. On the contrary if the spring force decreases due to ageing stress relaxation, PT decreases. The latter one is the dominating degradation mechanism in the examined study. It exhibits a direct linear relationship with the spring force relaxation. Based on the Larson Miller Model a method is proposed that establishes the relationship between spring force decrease and EMR storage life. Since no assessment is provided whether the decrease in PT and the failure threshold based on CR resemble similar trajectories for the sampled EMR population, the question for EOL-threshold selection is left unanswered.

[69] presents a method to predict EMR degradation during storage based on CR increase. Accelerated storage degradation tests have been conducted and failure mechanisms determined with SEM and EDX analysis. Using Particle Filtering the unknown parameters of a physical-model are determined from the experimental data. This research has been extended in [97]. The life shortening effects under the influence of elevated temperature leading to excessive fretting corrosion are revisited by the authors. A refined model to forecast the storage life based on estimating the CR trajectory is proposed. As the results indicate, the proposed model is capable of forecasting the storage life accurately for the considered EMR. The forecasting performance improves closer to the actual EOL.

4.2.4 Comparative assessment

In Table 4.3 a summary of the reviewed literature is presented, comparing various approaches and objectives to diagnose the operating state of the EMR in order to forecast the EOL or a DI trajectory. Figure 4.5 illustrates the relation of measurements, derived features, models, and research objectives. The line thickness between the nodes corresponds with the usage frequency within the reviewed body of literature. For example, CR is a prominent feature, derived either directly using an Ohmmeter or via CV and CI measurements. Likewise, data-driven models are frequently employed to assess the EMR-EOL through an estimation of the future state of certain time-based reference DIs.

From the analysis of the literature several challenges in monitoring and maintaining EMR using data-driven techniques become evident. Classical DIs do not generalise across different EMR designs, contact types, contact material, operational environments, or loading scenarios, since each application results in distinctly different degradation trajectories, cf. [98, 53, 62, 54, 85, 86]. In addition, different failure modes cause classical DIs to exhibit incoherent, opposing trends often even changing within the same batch of EMRs, cf. [60, 62]. This inevitably affects the direct use of DIs as performance metrics. Predicting or forecasting the trajectory of such DIs is flawed, because they are subject to high levels of variance and fluctuations in the switching process, cf. the build-up and destruction of oxide films on the contact surface (particularly problematic for CT and OT) as adverted by the above review. Furthermore, CR is disregarded as DI for a wide variety of EMRs (silver-plated contacts), because in most instances any noticeable trend is masked by significant fluctuations, cf. [53, 84] and Chapter 7.2. Research is further affected by the lack of sufficient EMR life-cycle data sets to validate and benchmark the proposed data-driven approaches. Moreover, uncertainty of the forecast is not of concern in any of the reviewed approaches. Collectively, this impedes the development of applicable data-driven diagnostics and prognostics for EMRs.

Table 4.3: Summary of reviewed methodologies.

No.	Ref.	Year	Signals	Features	Method(s)	Objectives
General DC EMR						
0	[58]	2004	CV, CI	CR	ARIMA, Exp. Reg. Model	Predicting EOL based on CR trajectory.
1	[76]	2004	CV, CI	DCR, CR, max.-CR, BT	Reg. Model	Predicting CR.
2	[77]	2006	CV, CI	DCR, CR, max.-CR, BT	Fuzzy Model	EMR contact reliability evaluation.
3	[60]	2010	CR, CC, CV	PT, CT, BT	Reg. Model	Improving EMR reliability estimation.
Continued on next page						

Table 4.3 – continued from previous page

No.	Ref.	Year	Signals	Features	Method(s)	Objectives
4	[65]	2012	CC, CV	BT	Symbolic-Sequence-Analysis	Estimating entropy of BT.
5	[61]	2012	CV, CI	CR, OT, RT, CT, BT, AT, Fluctuation Coefficient	Wavelet Dec., Linear-Model, Physical Model	Predicting the DI trajectory.
6	[78]	2017	CV, CC	CV, CC	Weibull Distribution	Classifying EMR state based on distribution of CV and CC measurements.
7	[62]	2017	Simulation	PT, OT, CC, Spring Force	Mahalanobis Distance	Classifying EMR failure modes.
Aviation and Aerospace DC EMR						
8	[53]	2009	CC, CV	PT, OT, RT	Wavelet-Dec., ARIMA	Predicting EOL based on DIs.
9	[83]	2014	CR	CR	Kalman Filter, Physical Model	Contact degradation modelling based on CR.
10	[84]	2015	CV, CR	CR	Grey System Theory	Predicting EOL based on CR.
11	[82]	2016	CV, CI	CR	Polynomial Model	Predicting EMR life based on initial measurements of CR.
12	[81]	2016	CV, CC, CR	CR, BT, RT, ST	Wavelet Dec., Reg. Model	Predicting EOL based on DIs.
High-Voltage DC EMR						
13	[73]	2016	CI	Arc-Charge, Magnetic-Flux	Linear Model	Predicting EOL based on variations in DI.
14	[80]	2017	CI, Erosion Mass, Velocity	Velocity	-	Predicting EOL based on changes in Contact-Velocity.
Railway EMR						
15	[99]	2016	CV, CC	ST	PCA, NN, Mahalanobis Distance	Failure classification and RUL prediction.
16	[86]	2018	CR, CV, CC	CR, CT	Physical Model	Correlating elevated temperature tests to predict contact degradation at ambient temperature.
17	[85]	2019	CV, CC	RT	NN	Predicting changes in EMR RT.
Automotive EMR						
18	[63]	2017	CR, CV, CC	CR, PT, RT, BT, ST, AT	Wavelet Dec. NN	Life prediction utilising the proposed features.
19	[66]	2020	CV, CI	DCR-duration, BT, CV-static, -dynamic	Particle Swarm Optimisation, Grey NN	Life prediction based on the initial state of the measured DIs.
Nuclear EMR						

Continued on next page

Table 4.3 – continued from previous page

No.	Ref.	Year	Signals	Features	Method(s)	Objectives
20	[89]	2016	CC	CC waveform shape	Threshold	Classification of contact welding (cf. [87]).
21	[90]	2017	CC	CC waveform shape	Threshold	Diagnostic circuit to detect the contact welding.
Storage of EMR						
22	[91]	2012	CR, CV, CC	CR, PT, OT, RT, BT	-	DI derived at elevated temperature, cf. [96, 95].
23	[92]	2013	CR	CR	Grey NN	Predicting the trajectory of CR.
24	[93]	2014	CR	CR	Wavelet Dec., Exponential Model	Predicting the trajectory of CR.
25	[64]	2015	CV, CC	PT	Miller Larson Model	Utilising the linear relationship between spring relaxation and PT to predict EMR-EOL.
26	[94]	2016	CV, CC	CT	Grey NN	Predicting the trajectory of CT.
27	[69]	2017	CR	CR	Physical Model, Particle Filtering	Predicting CR.
28	[97]	2019	CR	CR	Physical Model	Predicting CR.

Summary

This chapter recapitulated the state of the art. The unique challenges impeding the dissemination of novel maintenance solutions and, therefore, the industrialisation of data-driven maintenance to inform decision making for such applications have been identified. Taking these findings into account, data-driven maintenance frameworks can be tailored for BHA-PCBA failure classification and EMR-RUA estimation respectively. Thus, the following Chapter 5 integrates the distinct characteristics of each developed method in the context of the relevant areas of research.

References

- [1] Google scholar. <https://scholar.google.com/>, 2020. Accessed: 2020-03-22.
- [2] Onepetro. <https://www.onepetro.org/>, 2020. Accessed: 2020-03-22.
- [3] Elsevier. <https://www.elsevier.com/>, 2020. Accessed: 2020-03-22.
- [4] Scopus. <https://www.scopus.com/>, 2020. Accessed: 2020-03-22.
- [5] Web-of-knowledge. <https://www.webofknowledge.com/>, 2020. Accessed: 2020-03-22.
- [6] B Bageri, F Anifowose, and A Abdulraheem. Artificial intelligence based estimation of water saturation using electrical measurements data in a carbonate reservoir. In *SPE Middle East Oil & Gas Show and Conference*. OnePetro, 2015. doi:10.2118/172564-MS.

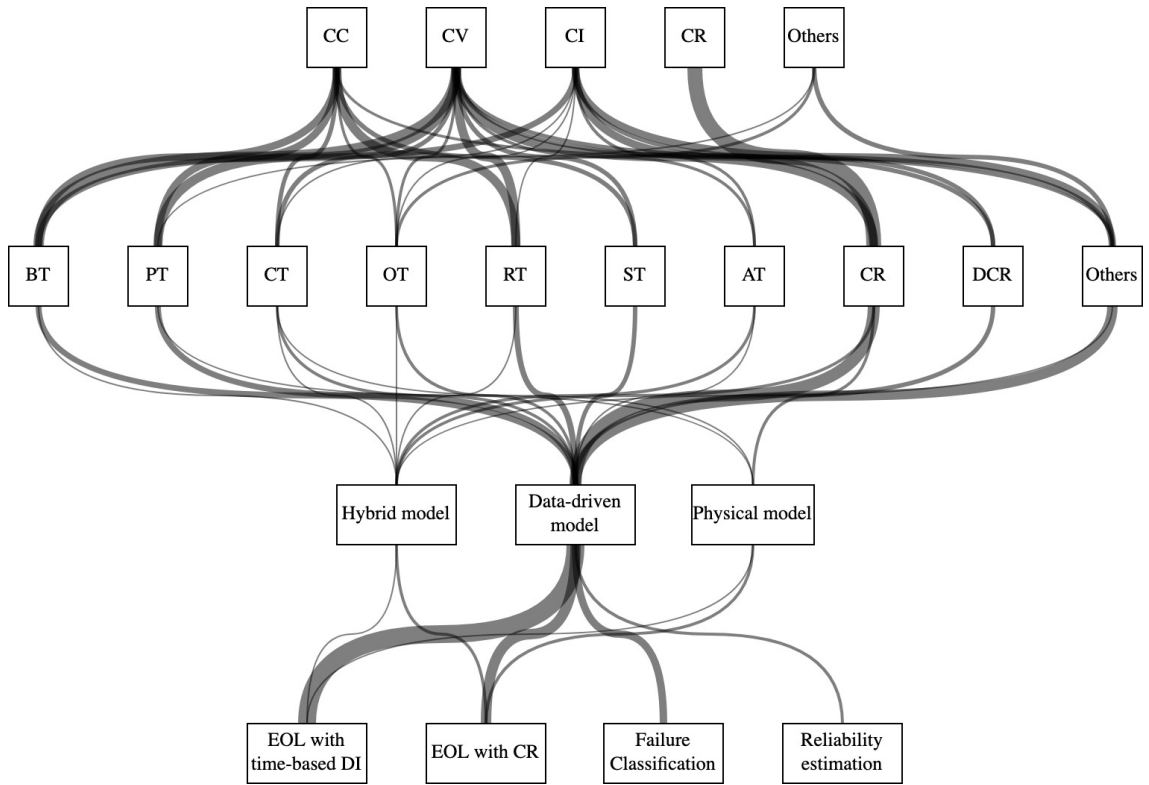


Figure 4.5: A network graph to illustrate the dependencies and frequency of use within the body of reviewed literature; signals (top), features (second from the top), models (third from the top), and objectives (bottom); cf. Table 4.3.

- [7] T Jobe, E Vital-Brazil, and M Khait. Geological feature prediction using image-based machine learning. *Petrophysics - The SPWLA Journal of Formation Evaluation and Reservoir Description*, 59(06):750–760, 2018. doi:10.30632/PJV59N6-2018a1.
- [8] D Ferraretti, G Gamberoni, and E Lamma. Unsupervised and supervised learning in cascade for petroleum geology. *Expert Systems with Applications*, 39(10):9504–9514, 2012. doi:10.1016/j.eswa.2012.02.104.
- [9] C Chan. An expert decision support system for monitoring and diagnosis of petroleum production and separation processes. *Expert Systems with Applications*, 29(1):131–143, 2005. doi:10.1016/j.eswa.2005.01.009.
- [10] C Noshi, A Assem, and J Schubert. The role of big data analytics in exploration and production: A review of benefits and applications. In *SPE International Heavy Oil Conference and Exhibition*. OnePetro, 2018. doi:10.2118/193776-MS.
- [11] J Xiao and X Sun. Big data analytics drive EOR projects. In *SPE Offshore Europe Conference & Exhibition*. OnePetro, 2017. doi:10.2118/186159-MS.
- [12] M Neuroth, P MacConnell, F Stronach, and P Vamplew. Improved modelling and control of oil and gas transport facility operations using artificial intelligence. In *Applications and Innovations in Intelligent Systems VII*, pages 119–136. Elsevier, 2000. doi:10.1016/S0950-7051(00)00049-6.
- [13] K Carter-Journet, A Kale, T Falgout, and L Heuermann-Kuehn. Drilling optimization: Utilizing lifetime prediction to improve drilling performance and reduce downtime. In *SPE Deepwater Drilling and Completions Conference*. OnePetro, 2014. doi:10.2118/170270-MS.
- [14] A Takbiri-Borujeni, E Fathi, T Sun, and M Rahmani, Rand Khazaeli. Drilling performance monitoring and optimization: a data-driven approach. *Journal of Petroleum Exploration and Production Technology*, 9(4):2747–2756, 2019. doi:10.1007/s13202-019-0657-2.
- [15] J Han, Y Sun, and S Zhang. A data driven approach of ROP prediction and drilling performance estimation. In *International Petroleum Technology Conference (IPTC)*. OnePetro, 2019. doi:10.2523/IPTC-19430-MS.
- [16] M Rashidi, A Asadi, A Abbasi, and E Asadi. Machine learning’s application in estimation of the drilling rate of penetration - a case study from a wellbore in iran. In *ARMA-CUPB Geothermal International Conference*. OnePetro, 2019.

- [17] C Noshi. Application of data science and machine learning algorithms for ROP optimization in west texas: Turning data into knowledge. In *Offshore Technology Conference (OTC)*. OnePetro, 2019. doi:10.4043/29288-MS.
- [18] K Singh, S Yalamarty, M Kamyab, and C Cheatham. Cloud-based ROP prediction and optimization in real time using supervised machine learning. In *Unconventional Resources Technology Conference (URTEC)*. OnePetro, 2019. doi:10.15530/urtec-2019-343.
- [19] C Hegde, C Soares, and K Gray. Rate of penetration (ROP) modeling using hybrid models: deterministic and machine learning. In *Unconventional Resources Technology Conference (URTEC)*. OnePetro, 2018. doi:10.15530/URTEC-2018-2896522.
- [20] C Hegde, S Wallace, and K Gray. Using trees, bagging, and random forests to predict rate of penetration during drilling. In *SPE Middle East Intelligent Oil and Gas Conference and Exhibition*. OnePetro, 2015. doi:10.2118/176792-MS.
- [21] C Hegde, M Pyrcz, H Millwater, H Daigle, and K Gray. Fully coupled end-to-end drilling optimization model using machine learning. *Journal of Petroleum Science and Engineering*, 186:106681, 2020. doi:10.1016/j.petrol.2019.106681.
- [22] S Sherif, O Adenike, E Obehi, A Funso, and B Eytuyo. Predictive data analytics for effective electric submersible pump management. In *SPE Nigeria Annual International Conference and Exhibition*. OnePetro, 2019. doi:10.2118/198759-MS.
- [23] M Abdelaziz, R Lastra, and J Xiao. ESP data analytics: Predicting failures for improved production performance. In *Abu Dhabi International Petroleum Exhibition & Conference*. OnePetro, 2017. doi:10.2118/188513-MS.
- [24] D Guo, C Raghavendra, K Yao, M Harding, A Anvar, and A Patel. Data driven approach to failure prediction for electrical submersible pump systems. In *SPE Western Regional Meeting*. OnePetro, 2015. doi:10.2118/174062-MS.
- [25] X Wang, Y He, F Li, X Dou, Z Wang, H Xu, and L Fu. A working condition diagnosis model of sucker rod pumping wells based on big data deep learning. In *International Petroleum Technology Conference (IPTC)*. OnePetro, 2019. doi:10.2523/IPTC-19242-MS.
- [26] P Bangert and S Sharaf. Predictive maintenance for rod pumps. In *SPE Western Regional Meeting*. OnePetro, 2019. doi:10.2118/195295-MS.
- [27] S Liu, C Raghavendra, Y Liu, K Yao, O Balogun, L Olabinjo, R Soma, J Ivanhoe, B Smith, and B Seren. Automatic early fault detection for rod pump systems. In *SPE annual technical conference and exhibition*. OnePetro, 2011. doi:10.2118/146038-MS.
- [28] Y Liu, K Yao, C Raghavendra, A Wu, D Guo, J Zheng, L Olabinjo, O Balogun, and I Ershaghi. Global model for failure prediction for rod pump artificial lift systems. In *SPE Western Regional & AAPG Pacific Section Meeting 2013 Joint Technical Conference*. OnePetro, 2013. doi:10.2118/165374-MS.
- [29] D Gold and T Treiberg. Implementation of automated expert diagnostics to provide enhanced rod pump surveillance at S. Belridge Field, CA. In *SPE Annual Technical Conference and Exhibition*. OnePetro, 2001. doi:10.2118/71527-MS.
- [30] S Gupta, L Saputelli, and M Nikolaou. Applying big data analytics to detect, diagnose, and prevent impending failures in electric submersible pumps. In *SPE Annual Technical Conference and Exhibition*. OnePetro, 2016. doi:10.2118/181510-MS.
- [31] E Martinez, W Moreno, V Castillo, and J Moreno. Rod pumping expert system. In *Petroleum Computer Conference*. OnePetro, 1993. doi:10.2118/26246-MS.
- [32] S Cho, H Jun, J Shin, Ho Hwang, and C Ha. A study on the development of prognosis system for offshore plant equipment. In *The Twenty-Fifth International Ocean and Polar Engineering Conference*. OnePetro, 2015.
- [33] P Lu, H Liu, C Serratella, and X Wang. Assessment of data-driven, machine learning techniques for machinery prognostics of offshore assets. In *Offshore Technology Conference*. OnePetro, 2017. doi:10.4043/27577-MS.
- [34] U Tygesen, K Worden, T Rogers, G Manson, and E Cross. State-of-the-art and future directions for predictive modelling of offshore structure dynamics using machine learning. In *Dynamics of Civil Structures*, volume 2, pages 223–233. Springer, 2019. doi:10.1007/978-3-319-74421-6_30.
- [35] W Nash, T Drummond, and N Birbilis. Deep learning AI for corrosion detection. In *NACE International Corrosion Conference Proceedings*, pages 1–11. OnePetro, 2019.
- [36] C Ejimuda and C Ejimuda. Using deep learning and computer vision techniques to improve facility corrosion risk management systems. In *SPE Western Regional Meeting*. OnePetro, 2018. doi:10.2118/190036-MS.

- [37] K Bybee. Reliable electronics for high-temperature downhole applications. *Journal of Petroleum Technology*, 52(7):56–57, 2000. doi:10.2118/0700-0056-JPT.
- [38] D Garvey, M John, and J Baumann. Nonparametric life consumption modeling of high end drilling tools. In *2010 Proceedings-Annual Reliability and Maintainability Symposium (RAMS)*, pages 1–6. IEEE, 2010. doi:10.1109/RAMS.2010.5448034.
- [39] H Reckmann, A Hohl, H Oueslati, and O Akimov. Step change in BHA reliability and performance due to utilization of downhole drilling dynamics services. In *IADC/SPE Asia Pacific Drilling Technology Conference and Exhibition*. OnePetro, 2018. doi:10.2118/191096-MS.
- [40] S Zhan, J Rodiek, L Heuermann-Kuehn, and J Baumann. Prognostics health management for a directional drilling system. In *IEEE Prognostics and System Health Management Conference (PHM-Harbin)*, pages 1–7. IEEE, 2011. doi:10.1109/PHM.2011.5939543.
- [41] B Alotaibi, W Contreras, and M Nefai. Bottom-hole assembly management system (BHAMS). In *SPE Saudi Arabia Section Technical Symposium and Exhibition*. OnePetro, 2014. doi:10.2118/172175-MS.
- [42] W Aldred, D Plumb, I Bradford, J Cook, V Gholkar, L Cousins, R Minton, J Fuller, S Goraya, and D Tucker. Managing drilling risk. *Oilfield review*, 11(2):2–19, 1999.
- [43] J Brehme and T Travis. Total BHA reliability — an improved method to measure success. In *IADC/SPE Drilling Conference*. OnePetro, 2008. doi:10.2118/112644-MS.
- [44] D Garvey, J Baumann, J Lehr, and J Hines. Pattern recognition based remaining useful life estimation of bottom hole assembly tools. In *SPE/IADC Drilling Conference and Exhibition*. OnePetro, 2009. doi:10.2118/118769-MS.
- [45] A Lahmadi, L Terrissa, and N Zerhouni. A data-driven method for estimating the remaining useful life of a composite drill pipe. In *2018 International Conference on Advanced Systems and Electric Technologies (IC-ASET)*, pages 192–195. IEEE, 2018. doi:10.1109/ASET.2018.8379857.
- [46] S Zhan and J Zhao. Advanced prognostic technique for improving the drilling performance of downhole tools. In *2015 First International Conference on Reliability Systems Engineering (ICRSE)*, pages 1–9. IEEE, 2015. doi:10.1109/ICRSE.2015.7366418.
- [47] S Zhan and I Ahmad. Real-time prognostic on downhole printed circuit board assembly of measurement-while-drilling/logging-while-drilling, 2015. US-Patent: 9,051,814.
- [48] H Reckmann, P Jogi, F Kpetehoto, S Chandrasekaran, and J Macpherson. MWD failure rates due to drilling dynamics. In *IADC/SPE Drilling Conference and Exhibition*. OnePetro, 2010. doi:10.2118/127413-MS.
- [49] A Kale, K Carter-Journet, T Falgout, L Heuermann-Kuehn, and D Zurcher. A probabilistic approach for reliability and life prediction of electronics in drilling and evaluation tools. In *Annual Conference of the PHM Society*. PHM Society, 2014. doi:10.36001/phmconf.2014.v6i1.2492.
- [50] A Kale, D Zhang, A David, L Heuermann-Kuehn, and O Fanini. Methodology for optimizing operational performance and life management of drilling systems using real time-data and predictive analytics. In *SPE Digital Energy Conference and Exhibition*. OnePetro, 2015.
- [51] A Mosallam, L Laval, F Youssef, J Fulton, and D Viassolo. Data-driven fault detection for neutron generator subsystem in multifunction logging-while-drilling service. In *PHM Society European Conference*, pages 1–4. PHM Society, 2018. doi:10.36001/phme.2018.v4i1.399.
- [52] K Carter-Journet, A Kale, D Zhang, E Pradeep, T Falgout, and L Heuermann-Kuehn. Estimating probability of failure for drilling tools with life prediction. In *SPE Asia Pacific Oil & Gas Conference and Exhibition*. OnePetro, 2014. doi:10.2118/171517-MS.
- [53] Q Yu, M Qi, S Wang, and G Zhai. Research on life prediction based on wavelet transform and ARMA model for space relay. In *4th IEEE Conference on Industrial Electronics and Applications*, pages 1275–1280. IEEE, 2009. doi:10.1109/ICIEA.2009.5138407.
- [54] Y Sun, Y Cao, Y Zhang, and C Xu. A novel life prediction method for railway safety relays using degradation parameters. *IEEE Intelligent Transportation Systems Magazine*, 10(3):48–56, 2018. doi:10.1109/MITS.2018.2842048.
- [55] W Rieder and T Strof. Relay life tests with contact resistance measurement after each operation. *IEEE transactions on components, hybrids, and manufacturing technology*, 14(1):109–112, 1991. doi:10.1109/33.76518.
- [56] M Hasegawa, N Kobayashi, and Y Kohno. Contact resistance characteristics of relays operated in silicone-vapor-containing and non-silicone atmospheres with different electrical load conditions. In *57th Holm Conference on Electrical Contacts (Holm)*, pages 1–7. IEEE, 2011. doi:10.1109/HOLM.2011.6034787.

- [57] X Zhou, L Zou, and R Briggs. Prognostic and diagnostic technology for DC actuated contactors and motor starters. *IEICE transactions on electronics*, 92(8):1045–1051, 2009. doi:10.1587/transele.E92.C.1045.
- [58] W Li, K Li, L Sun, S Zhao, and L Ji. The test data verifying & prediction model of electrical contact. In *50th IEEE Holm Conference on Electrical Contacts*, pages 429–436. IEEE, 2004. doi:10.1109/HOLM.2004.1353152.
- [59] L Huimin, Z Qingsen, and Z Guofu. Research and application of testing and analysing system for sealed relay’s time parameters. In *50th IEEE Holm Conference on Electrical Contacts*, pages 199–205. IEEE, 2004. doi:10.1109/HOLM.2004.1353118.
- [60] Y Xuerong, Y Qiong, and Z Guofu. Reliability assessment for electromagnetic relay based on time parameters degradation. In *11th International Conference on Electronic Packaging Technology & High Density Packaging*, pages 1269–1272. IEEE, 2010. doi:10.1109/ICEPT.2010.5582785.
- [61] Y Xuerong, M Yue, M Hang, and Z Guofu. Degradation failure model of electromagnetic relay. In *26th International Conference on Electrical Contacts (ICEC)*, pages 116–123. IET, 2012. doi:10.1049/cp.2012.0633.
- [62] Y Xuerong, R Fu, Y Wu, Y Lin, and G Zhai. Simulation and diagnosis of degradation failure on electromagnetic relay. In *IEEE Prognostics and System Health Management Conference (PHM-Harbin)*, pages 1–9. IEEE, 2017. doi:10.1109/PHM.2017.8079301.
- [63] J Guo, G Zhang, Y Bi, and Y Li. Life prediction of automotive electromagnetic relay based on wavelets neural network. *Chemical Engineering Transactions*, 62:1213–1218, 2017. doi:10.3303/CET1762203.
- [64] Z Wang, S Shang, X Huang, G Zhai, and W Chen. Research on degradation modeling of pick-up time for aerospace electromagnetic relay in long-term stockpile. In *61st IEEE Holm Conference on Electrical Contacts*, pages 182–186. IEEE, 2015. doi:10.1109/HOLM.2015.7355094.
- [65] Y Fang, L Zhigang, W Yachao, and Y Cuiping. Research on uncertainty of bounce time for electromagnetic relay and its application in operating reliability estimation. In *26th International Conference on Electrical Contacts (ICEC)*, pages 1–6. IET, 2012. doi:10.1049/cp.2012.0648.
- [66] X Pang, Z Li, M Tseng, K Liu, K Tan, and H Li. Electric vehicle relay lifetime prediction model using the improving fireworks algorithm–grey neural network model. *Applied Sciences*, 10(6):1940, 2020. doi:10.3390/app10061940.
- [67] Y Sun, Y Zhang, C Xu, and Y Cao. Failure mechanisms discrimination and life prediction of safety relay. *Journal of Traffic and Transportation Engineering*, 18:138–147, 2018.
- [68] P Slade. *Electrical contacts: principles and applications*. CRC press, 2017.
- [69] Y Zhao, E Zio, and G Fu. Remaining storage life prediction for an electromagnetic relay by a particle filtering-based method. *Microelectronics Reliability*, 79:221–230, 2017. doi:10.1016/j.microrel.2017.03.026.
- [70] L Kirschbaum, F Dinmohammadi, D Flynn, V Robu, and M Pecht. Failure analysis informing embedded health monitoring of electromagnetic relays. In *2018 3rd International Conference on System Reliability and Safety (ICSRS)*, pages 261–267. IEEE, 2018. doi:10.1109/ICSRS.2018.8688839.
- [71] A Najam, P Pieterse, and D Uhrlandt. Electrical modelling of switching arcs in a low voltage relay at low currents. *Energies*, 13(23):6377, 2020. doi:10.3390/en13236377.
- [72] G Yu, Y Chiu, X Zheng, Z Yuan, and Z Wang. Contact pressure of high-voltage dc power relay change and life prediction and structure optimization. *Advances in Mechanical Engineering*, 13(2):1–15, 2021. doi:10.1177/1687814021991666.
- [73] C Xinglei, Z Xue, C Mo, L Fubiao, and Z Zhefeng. Electrical lifespan prediction of high-voltage direct-current relay based on arc charge accumulation. In *62nd IEEE Holm Conference on Electrical Contacts*, pages 167–172. IEEE, 2016. doi:10.1109/HOLM.2016.7780027.
- [74] ICS. Aircraft - electromagnetic relays and contactors - part 1: General requirements. Standard, International Organization for Standardization, Geneva, CH, 1996.
- [75] K Li, F Yao, J Lu, and Z Li. Test and analysis of reliability for electromagnetic relay. In *46th IEEE Holm Conference on Electrical Contacts*, pages 79–82. IEEE, 2000. doi:10.1109/HOLM.2000.889915.
- [76] Y Fang, Z Li, W Li, and K Li. Concerning contact resistance prediction based on time sequence and distribution character. In *50th IEEE Holm Conference on Electrical Contacts*, pages 447–452. IEEE, 2004. doi:10.1109/HOLM.2004.1353155.

- [77] Y Fang, J Lu, J Zheng, and Z Huang. Research on the failure diagnostics parameters and the reliability prediction model of the electrical contacts. In *52nd IEEE Holm Conference on Electrical Contacts*, pages 69–72. IEEE, 2006. doi:10.1109/HOLM.2006.284067.
- [78] J Wan, K Xue, Y Chen, and M Tong. An acquisition and processing method for the key parameters of electromagnetic relays. In *Progress in Electromagnetics Research Symposium-Fall*, pages 1096–1100. IEEE, 2017. doi:10.1109/PIERS-FALL.2017.8293297.
- [79] G Zhai, S Wang, F Xu, and M Liu. Research on double-variable life forecasting based on model-building of super-path time and pick-up time for relays. *Chinese Society for Electrical Engineering*, 22(7):76–80, 2002.
- [80] X Cui, X Zhou, G Zhai, Z Fang, and F Liu. Electrical lifespan prediction of HVDC relay based on the accumulated arc erosion mass. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 8(3):356–363, 2017. doi:10.1109/TCPMT.2017.2768222.
- [81] L Li, Y Han, W Chen, C Lv, and D Sun. An improved wavelet packet-chaos model for life prediction of space relays based on volterra series. *PLOS One*, 11(6):e0158435, 2016. doi:10.1371/journal.pone.0158435.
- [82] Z Li, B Liu, M Yuan, F Zhang, and J Guo. Characterization of initial parameter information for lifetime prediction of electronic devices. *PLOS One*, 11(12):e0167429, 2016. doi:10.1371/journal.pone.0167429.
- [83] A Wileman and S Perinpanayagam. A prognostic framework for electromagnetic relay contacts. In *PHM Society European Conference*, volume 2, pages 1–7. PHM Society, 2014. doi:10.36001/phme.2014.v2i1.1531.
- [84] A Wileman and S Perinpanayagam. Integrated vehicle health management: An approach to dealing with lifetime prediction considerations on relays. *Microelectronics Reliability*, 55(9-10):2165–2171, 2015. doi:10.1016/j.microrel.2015.06.013.
- [85] Y Sun, Y Cao, M Zhou, T Wen, P Li, and C Roberts. A hybrid method for life prediction of railway relays based on multi-layer decomposition and RBFNN. *IEEE Access*, 7:44761–44770, 2019. doi:10.1109/ACCESS.2019.2906895.
- [86] J Liu, M Zhang, N Zhao, and A Chen. A reliability assessment method for high speed train electromagnetic relays. *Energies*, 11(3):652, 2018. doi:10.3390/en11030652.
- [87] S Sravanthi, R Dheenadhayalan, M Sakthivel, K Devan, and K Madhusoodanan. A method for online diagnostics of electromagnetic relays against contact welding for safety critical applications. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 5(12):1734–1739, 2015. doi:10.1109/TCPMT.2015.2498624.
- [88] H Sauer. *Modern Relay Technology*. Huethig, 1986.
- [89] S Sravanthi, R Dheenadhayalan, G Vinod, K Madhusoodanan, and K Devan. Reliability model of a relay output card with diagnostic circuitry for safety instrumented system. In *International Conference on System Reliability and Science (ICSRS)*, pages 130–136. IEEE, 2016. doi:10.1109/ICSRS.2016.7815851.
- [90] P Wadhawan, M Kumar, A Mayya, and P Marathe. Comparison and failure mode analysis of current signature based relay health monitoring schemes. In *6th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)*, pages 155–161. IEEE, 2017. doi:10.1109/ICRITO.2017.8342417.
- [91] Z Wang, G Zhai, W Ren, X Huang, and Q Yu. Research on accelerated storage degradation testing for aerospace electromagnetic relay. In *58th IEEE Holm Conference on Electrical Contacts*, pages 1–8. IEEE, 2012. doi:10.1109/HOLM.2012.6336597.
- [92] Z Wang, G Zhai, X Huang, and D Yi. Combination forecasting method for storage reliability parameters of aerospace relays based on grey-artificial neural networks. *International Journal of Innovative Computing, Information and Control*, 9(9):3807–3816, 2013.
- [93] Z Wang, S Shang, G Zhai, and W Ren. Research on storage degradation testing and life prediction based on ARMA and wavelet transform model for aerospace electromagnetic relay. In *60th IEEE Holm Conference on Electrical Contacts*, pages 1–8. IEEE, 2014. doi:10.1109/HOLM.2014.7031021.
- [94] Z Wang, S Fu, S Shang, and W Chen. New forecasting method of closing time for aerospace relay in storage accelerated degradation testing. In *11th International Conference on Reliability, Maintainability and Safety (ICRMS)*, pages 1–5. IEEE, 2016. doi:10.1109/ICRMS.2016.8050118.
- [95] Z Wang, Z Huang, J Wang, S Shang, and G Zhai. The failure mechanism of electromagnetic relay in accelerated storage degradation testing. In *63rd IEEE Holm Conference on Electrical Contacts*, pages 164–168. IEEE, 2017. doi:10.1109/HOLM.2017.8088080.

- [96] Z Wang, S Shang, J Wang, Z Huang, and F Sai. Accelerated storage degradation testing and failure mechanisms of aerospace electromagnetic relay. *Eksploatacja i Niezawodność - Maintenance and Reliability*, 19:530–541, 2017. doi:10.17531/EIN.2017.4.6.
- [97] W Guo, G Fu, B Wan, M Jiang, and Y Li. Storage PoF model of electromagnetic relays based on electrical contact theory. *Microelectronics Reliability*, 100:113381, 2019. doi:10.1016/j.microrel.2019.06.073.
- [98] Z Chen, H Mizukoshi, and K Sawa. Contact resistance characteristics of Ag material in breaking low-load DC arcs. *IEEE Transactions on Components, Packaging, and Manufacturing Technology: Part A*, 17(1):113–120, 1994. doi:10.1109/95.296376.
- [99] B Sun, X Jiang, K Yung, J Fan, and M Pecht. A review of prognostic techniques for high-power white LEDs. *IEEE Transactions on Power Electronics*, 32(8):6338–6362, 2016. doi:10.1109/TPEL.2016.2618422.

Chapter 5

Developed data-driven methods

Introduction

As argued in Chapter 1 this thesis proposes solutions which emphasise the role of data-driven techniques in order to enable informed maintenance decision making. Yet, to gain acceptance within an industrial setting such solutions need to sustain within a given data environment. Moreover, they should facilitate an expressive prognostic feedback under the application specific constraints. With these objectives in mind, the methods proposed here develop two distinctly different responses, both customised to the respective settings: (1) prediction of failure class probability of electronic assemblies at fleet-level (BHA-PCBA); (2) RUA estimation for the EMR on component-level (Electromagnetic-Relay-Remaining-Useful-Actuation (EMRUA)). Along these lines, Chapter 5.1 presents the developed BHA-PCBA pipeline, while Chapter 5.2 details the proposed strategy for the EMR. Here, special attention is paid to data-driven approaches that are challenged by high volumes of MVTD. Relevant DL terminology and key principles - in particular Convolutional Neural Networks and their adaption in PHM - are introduced in Chapter 5.2.1.

5.1 BHA-PCBA maintenance support framework

As elaborated in Chapter 3.1, the BHA is a complex system assembly. Within the upstream service industry, the individual BHA units resemble tool-fleets that are used interchangeably over multiple operations before being maintained according to their respective maintenance schedule. As detailed, the BHA holds a set of PCBAs. The electronics mounted on the PCBAs are liable to failure, since they are operated in harsh and often unpredictable operating conditions. The aim of the developed method is to determine a malfunctional behaviour of the BHA-PCBA and support the maintenance decision making within a potential PdM framework. Note, in practice a maintenance recommendation translates into a scrapping decision, i.e.

replacement of the entire board prior to the next run of the BHA. Replacing a whole PCBA is economically more efficient than first screening, then maintaining if necessary, and lastly qualifying the various components on the PCBA ¹.

5.1.1 Failure classification using machine learning

From a ML perspective, the problem of support for maintenance decision making is treated as a classification problem, where predicting the need for maintenance can be phrased as a two-class problem. One class represents the case when a PCBA needs replacement (failure $k=1$), while the other class indicates the PCBA is fit for a rerun (no-failure $k=0$). In addition to a hard decision boundary, the confidence of the algorithm in its prediction by means of a conditional probability output $p(y=k|x)$ is considered, where x denotes the model inputs and y the output corresponding to the predicted target class. Therefore, an algorithm is confident in a prediction if it outputs the probability of a sample belonging to one of the classes greater or equal to a user specified threshold. This threshold is denoted here as maintenance decision threshold TH_m . Translated into an engineering context, a recommendation to carry out replacement of the PCBA is recommended in all cases where the probability of failure is greater or equal to TH_m . Depending on operational circumstances TH_m can vary in accordance with the acceptable level of prediction uncertainty.

To facilitate BHA-PCBA failure prediction two different ML algorithms are considered, namely Random Forest and XGBoost. Independent of the chosen method, the hierarchical executed steps to derive a trained model (also referred to as offline-phase) are the following: (1) data processing, (2) data augmentation, (3) model training, and (4) performance evaluation.

5.1.2 Data sources

As prior discussed, data from various sources is collected during drilling containing information regarding the geological formations, tool health, and event logs. Therefore, the proposed approach utilises a combination of field data (surface data and telemetry data), post-run data (higher resolution memory data from the BHA), and subsequent reports (number of attempts and mission status) which are all aggregated into a single data set. The post-run data has been collected following each tool's arrival in a maintenance workshop from 208 missions over a six-month period. Recorded missions are of one specific type of BHA. Post-run data and field data records have been post-processed. Such processing of the post-run data provides the number of attempts of various BHA components to successfully or unsuccessfully establish an internal communication link. This internal communication network is

¹Parts of this chapter have already been published in my journal or conference article: L. Kirschbaum, et al., (2020).

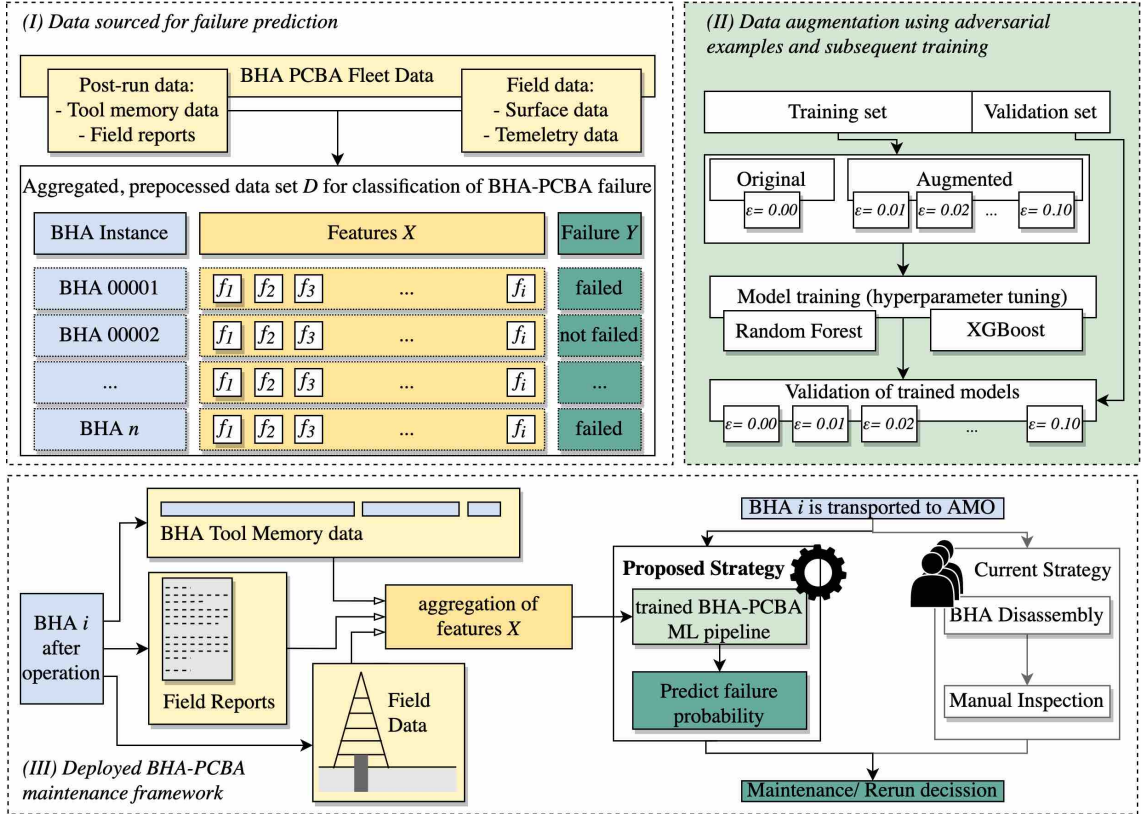


Figure 5.1: (I) A schematic display of the layout of the used data set. (II) The hierarchical training process for the RFC and the XGBoost Classifier using adversarial examples to augment the data set. (III) the deployed classification model predicting the failure probability for PCBA in a BHA i based on features X aggregated from multiple downhole tool data sources. This mitigates the need for labour intensive disassembly and manual inspection of the BHA-PCBA in order to make a informed maintenance/ rerun decision.

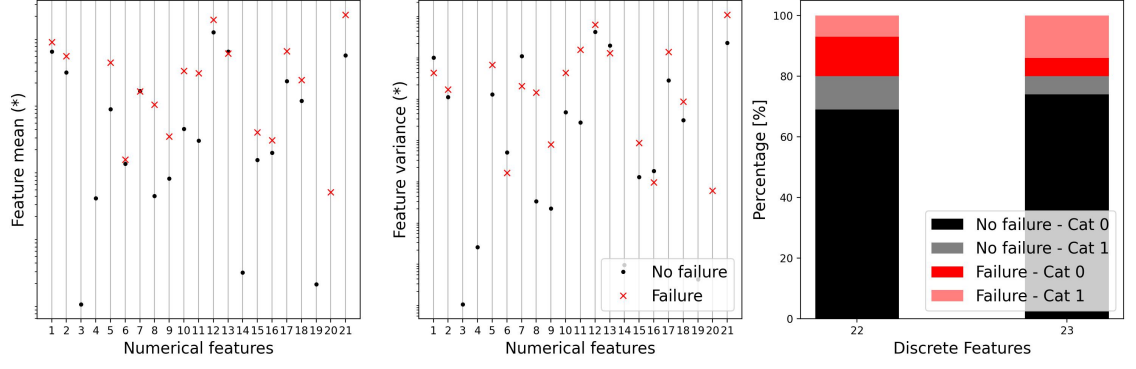


Figure 5.2: The numerical features (left and centre) and the discrete features (right) of the BHA-PCBA data set. (*) Due to data confidentiality the absolute values of the features can not be displayed.

resembled by various nodes. An accumulation of unsuccessful attempts or delays in communication at these nodes indicates faulty behaviour of a PCBA within the BHA. Concluding, as shown in Figure 5.1-(I), the data set is aggregated from BHA-PCBA post-run data and field data. It comprises out of $n = 208$ BHA instances, with $m = 23$ preprocessed and selected features. The j^{th} BHA instance holds the feature space vector $x_j^i = [f_j^1, \dots, f_j^m]$ and a status report classifying the BHA as healthy (80 % of the data) or failed (20 % of the data) using a label y_j . The data set can be represented as $D = (X_j^i, Y_j)$ holding the input matrix $X = [x_1, x_2, \dots, x_n]$ and the target class vector $Y = [y_1, y_2, \dots, y_n]$. Out of the m features, 21 are numerical and 2 are discrete. The mean and the variance of the numerical features of no-failure and failure BHA-PCBAs are displayed in Figure 5.2 on the (left) and in the (centre) respectively. Likewise the percentage of each category of the 2 used discrete features is displayed in Figure 5.2-(right). The discrete features only comprise two categories, e.g. *off* and *on*.

21 out of the 23 features are part of the BHA memory data. As detailed, they are either error flags or represent counts of unsuccessful communication attempts within the BHA-PCBA system. However, feature 1 and feature 2 represent two measures of time-till-failure that are derived using surface data. The reliability of the BHA-PCBAs can be modelled using a Weibull fit, cf. the survival function displayed in Figure 5.3. It is evident, that an increasing exposure to the drilling process decreases the reliability. In particular, feature 2 exhibits a steep initial decline of the reliability. This shows that the probability of failure increases as exposure time captured using feature 1 and feature 2 increases, cf. Chapter 2.1 .

As explained in Chapter 5.1.4, the employed ensemble algorithms do not require the features to be normalised. The remainder of this chapter details the subsequent data augmentation process, ML models, training, and performance evaluation.

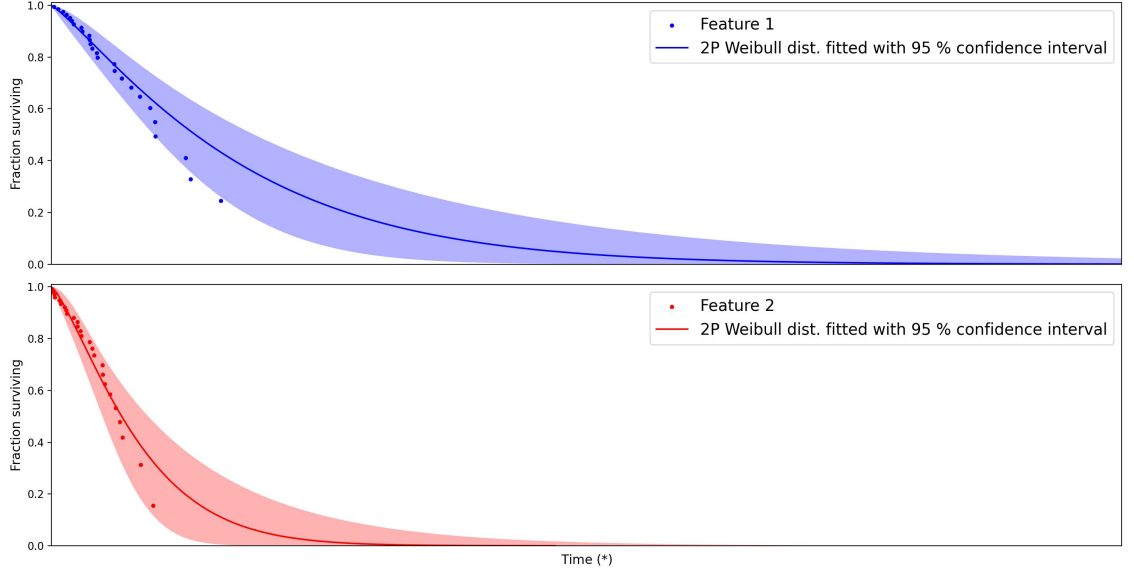


Figure 5.3: Survival plot of (top) feature 1 and (bottom) feature 2 incorporating a 2-Parameter Weibull fit. Samples that are labelled as failure are *fully observables* while no-failure samples are considered *right-censored*. (*) Due to data confidentiality the absolute values of the failure times and the parameters of the fit can not be displayed.

5.1.3 Data augmentation

As illustrated in Figure 5.1-(II), prior to model training the aggregated BHA-PCBA data set is split into a training and a validation set. The validation data is provided by the industrial partner independently of the training data and accounts for about one third of all available data, containing representative examples of failed and non-failed PCBAs. The training data is augmented, i.e. artificially created examples are added to the original data set, due to the limited extent of the data set in order to account for *out-of-distribution* examples and increase robustness towards outliers. Thus, the generalisation ability of the considered algorithms is improved. Note, there are no augmented examples in the validation set. Adversarial examples are only introduced to the training data set [1, 2]. These adversarial examples are carefully designed to be similar to a genuine training example but are misclassified by the algorithm. In particular, linear ML models are challenged by the introduction of adversarial examples. Misclassification of adversarial examples is not due to overfitting (e.g. a too complex decision boundary), but rather due to underfitting, since the model behaves too linear and is falsely extrapolating into regions where no examples have been observed. This is increasingly problematic in high dimensional data. To the naked eye, adversarial examples are almost imperceptible from the original input, but fool the model. While there are other methods, given an input x with target y , a similar approach to [1] is used to generate the augmented data set employing the fast gradient sign method to define an adversarial example as

$$x^* = x + \epsilon \operatorname{sign}(\nabla_x L(\theta, x, y)) \quad (5.1)$$

where ϵ is a small perturbation value, L is the loss function with respect to the model parameters θ , input x and target y . The loss function represents the error between the model's prediction and the expected output and is calculated after each training iteration. In order to obtain the gradient of the loss function, $\nabla_x L(\theta, x, y)$, the Ridge Classifier is employed [3]. A range of data sets containing adversarial examples with perturbation factors in the interval of $\epsilon = [0.00, 0.10]$ is created. The reason here is to investigate the effect of ϵ on the model performance. The newly generated data sets are further referred to as *adversarial training data sets*. The initial perturbation range has been determined by trial and error. For an elaborate discussion on possible perturbation values ϵ refer to [1].

5.1.4 Ensemble algorithms

While a DT model can be used for regression, the remainder of this section focuses on classification, i.e. predicting discrete values in a supervised learning context. A DT is a powerful ML method because of its invariance towards types of data and scales. It derives decision rules for categorical and continuous attributes alike; therefore, it mitigates distinct data preprocessing steps, e.g. normalisation [4]. Due to the rule-based nature, the DT offers interpretability in terms of how an estimate has been reached, e.g. highlighting the feature interdependence [5]. DT models form the basic building blocks of ensemble methods. Figure 5.4 provides the scheme of a DT which can be graphically abstracted as an inverted tree.

An efficient DT should be as small as possible. However, with an increasing feature space and data set size the computational complexity to derive the ideal tree rises exponentially [6]. Hence, a so called greedy heuristic approach is recursively deployed where on each node the next best local split is determined in order to maximise the class separability [7]. In general the *Gini Impurity* G_I , measuring the class variance for a feature as per Equation 5.2 is used.

$$G_I = 1 - \sum_{c=1}^C (p_c)^2 \quad (5.2)$$

Here C denotes the number of classes and p_c the respective class probability. Alternatively, the *Entropy* E for a subset of the data S is calculated as per Equation 5.3 - a measure of disorder in the data set - and the *information gain* which indicates how the entropy in the data set is reduced by further splitting the parent node feature according to a condition. The information gain is used to select a threshold to split continuous features.

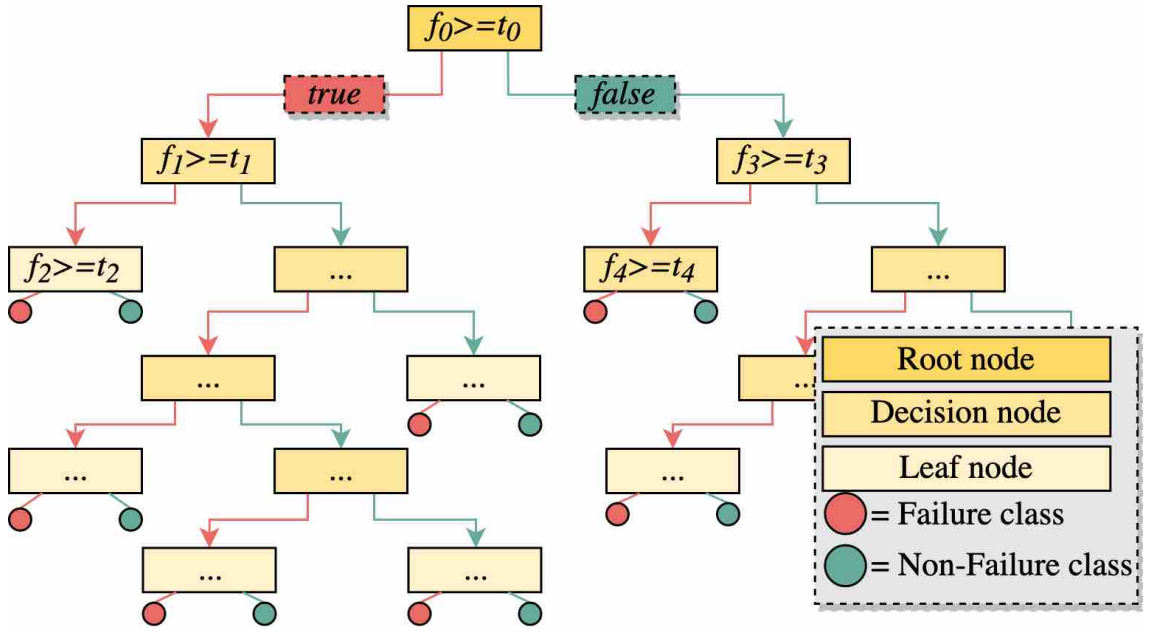


Figure 5.4: Schematic display of a DT, classifying an input based on f_n features into a *failure* and a *no-failure* class. The top node constitutes as *root node* from which the tree grows. A node contains a criterion relating to a feature which splits the attribute, e.g. $x \leq 3$. A binary split based on whether a condition/ threshold is *true* or *false* extends the DT. This might require continuous splits on the same feature or other features. If the split yields a set of features only relating to one class a *leaf node* terminates the tree at this branch. Contrary, if no clear class separation can be obtained a *decision node* is entered.

$$E(S) = \sum_{c=1}^C -p_c \log_2 p_c \quad (5.3)$$

During learning DT models have a tendency to overfit, e.g. because of the noise present in the training data. Further, DT models are sensitive to very small changes in the training data yielding highly varying results for each data set split. Therefore, different mechanisms have been established to reduce overfitting forcing the algorithm to build simple trees. For example, one can consider the following: the maximum tree depth can be limited; a minimum number of available samples is required at each node to allow a further split; pruning techniques remove redundant branches [8].

To reduce the susceptibility to variance in the training data set as well as overfitting from a very deep, single tree, the results of many individual, shallow DTs can be averaged to increase robustness. This popular ML algorithm is known as Random Forest and is one of the so called ensemble methods [9]. In general, ensemble methods are learning algorithms that first construct a set number of classifiers and then estimate a sample's class by casting a weighted vote of their predictions. It has long been observed that ensemble methods improve predictive performance over single based algorithms [10]. Different ensemble methods exist, cf. [11]. 1st the **bagging approach** is a randomisation-based technique where models are trained in parallel

without any interaction, e.g. Random Forest Classifier (RFC). Hence, the RFC holds a set of randomly generated DT models. The randomness is introduced via bootstrap-aggregation (known as bagging) where each DT is trained on a different subset of the training data, i.e. n random subsets $S = [S_1, S_2, \dots, S_n]$ are drawn from the training data while randomly subsampling the set of candidate splits at each node [12, 13, 14].

^{2nd} in contrast to bagging techniques in which DT models are grown to their maximum extent in parallel, boosting makes use of trees with fewer splits [15]. The **boosting approach** is a method where the ensemble members are trained sequentially, e.g. common implementations are AdaBoost or XGBoost, such that each subsequent tree aims to reduce the errors of the previous tree [12]. Hence, each DT learns from its predecessors and updates the residual errors [16]. The gradient descent is used to minimise a loss function, i.e. *Extreme Gradient Boosting Algorithms* implementations such as XGBoost, where models are sequentially added and optimised based on the residual of previous predictions, cf. [15, 17]. While ensemble learning can be implemented with any learning strategy, ensembles of DTs (e.g. RFC or XGBoost) are popular due to the computational efficiency of the DT model. They are known to achieve state-of-the-art performance on numerous supervised learning problems which has been demonstrated in [18].

The above-mentioned algorithms require careful hyperparameter tuning. Instead of the common grid-search, a random-search approach is adopted in order to determine the optimal set of hyperparameters. In this instance, a random uniform distribution is considered for each algorithm's hyperparameter space [19]. The choice for selecting such tuning technique is motivated by the fact that empirically and theoretically randomly chosen trials, no matter the distribution they are chosen from, are more efficient for hyperparameter optimisation than trials on a grid [20]. In addition, the randomised-search is coupled with a Stratified K-Fold Cross-Validation (SKF) technique in order to accommodate for the high imbalance in the two classes – a low number of failure instances compared to no-failure instances. The SKF incorporates folds of the data that preserve the percentage of samples for each class and thus allow the algorithm to train on balanced subsets in order to prevent overfitting on one of the two classes.

5.1.5 Performance evaluation

Each algorithm is evaluated based on specific classification scoring metrics. Some metrics are scoring qualitatively, others are quantifying the probabilistic outputs. Qualitative measures take the predicted class label as input depending on the decision threshold TH_m . Probabilistic scores, on the contrary, are calculated based on the output probability associated with its respective class prediction k . Since this is

	Model predicts: BHA PCBA is not faulty	Model predicts: BHA PCBA is faulty
Observed: BHA PCBA is faulty	True Negative (tn)	False Positive (fp)
Observed: BHA PCBA is not faulty	False Negative (fn)	True Positive (tp)

Figure 5.5: The confusion matrix for the proposed BHA-PCBA failure classification methodology using either RFC or XGBoost.

a binary classification task, discrete performance measures of the prediction of each algorithm can be derived from the confusion matrix as illustrated in Figure 5.5.

Traditionally, the *accuracy* denoted as acc in Equation 5.4, the *precision* in Equation 5.5, the *recall* in Equation 5.6, and the *F1-score* denoted as $F1$ in Equation 5.7 are used in this context. Accuracy can give misleading performance results on data sets. with imbalanced class distribution. Hence it is important to consider other metrics. Precision is selected, if only correctly classifying failures is more important than identifying all failures. On the contrary, if identifying all failures is paramount, whilst misclassifying some is acceptable, recall is chosen. The $F1$ score combines both precision and recall providing a unified metric. A comprehensive explanation of the metrics used in classification settings can be found in [21].

$$acc = \frac{tp + tn}{tp + tn + fp + fn} \quad (5.4)$$

$$precision = \frac{tp}{tp + fp} \quad (5.5)$$

$$recall = \frac{tp}{tp + fn} \quad (5.6)$$

$$F1 = \frac{2 * tp}{2 * tp + fp + fn} \quad (5.7)$$

The tp refers to all faulty BHA-PCBA instances that have been classified correctly, tn to all instances where no-failure has been correctly identified, fn to all instances which have been misclassified as no-failure, and fp to all instances where a failure has been predicted but no actual failure occurred. Further, the true positive rate $tpr = \frac{tp}{tp+fn}$ (equivalent to recall), i.e. the number of correctly classified failed BHA-PCBAs from all correctly classified instances, and the false positive rate $fpr = \frac{fp}{tn+fp}$ are considered. The fpr measures the ratio of all falsely predicted failures. Using

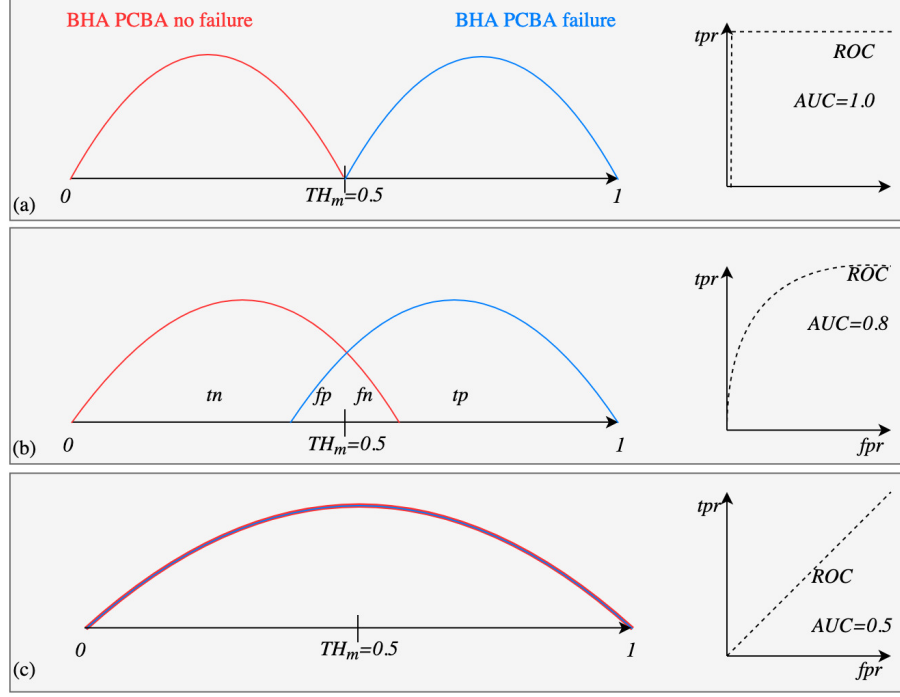


Figure 5.6: Left - BHA-PCBA probability distributions for no-failure class (red) and the failure class (blue). Right - ROC-AUC score curves. (a) A perfect model, (b) deviating model performance, (c) a model that can not differentiate between the classes.

the true positive rate and false positive rate the Receiver Operating Characteristic - Area Under the Curve (ROC-AUC) score can be used to assess how certain the algorithm is in its prediction [22]. It acts as a measure for class separability based on the predicted class probabilities. If the ROC-AUC increases, the class separability becomes more distinct. This class separability score is calculated as the Area Under the Curve (AUC). Consider the three edge cases illustrated in Figure 5.6 with a fixed decision threshold TH_m at 0.5. In Figure 5.6-(a) the prediction of a perfect model is shown. On the left the probability distribution curve for the no-failure class (red) and the failure class (blue) are illustrated. The ROC-AUC curve is displayed on the right plotting the fpr against the tpr . An $AUC = 1.0$ indicates full class separability, i.e. $fp = 0$ and $fn = 0$. This can also be seen on the left of 5.6-(a), as the distributions of each respective class are non-overlapping. However, if the model's performance deviates the distributions begin to overlap which is reflected in a reduced $AUC = 0.8$, consider 5.6-(b). If performance further decreases, the model is no longer able to differentiate between the classes, as can be seen in 5.6-(c).

5.2 EMR prognostics pipeline

This section presents the EMRUA-Pipeline and formulates the problem of Remaining Useful Actuation (RUA) prediction for Electromagnetic Relay (EMR) in Chapter 5.2.6. Research context is provided in Chapter 5.2.4 and Chapter 5.2.5 to underpin the decision for the selected DL method - the Temporal Convolutional Network

(TCN). However, first Chapter 5.2.1 introduces the basic principles of DL essential to the functionality of TCN. Subsequently chapter 5.2.2 presents examples of DL frameworks in the context of PHM.

5.2.1 CNN-RNN based methods for multivariate time series data

As discussed, data-driven prognostics methods are challenged by increasing amounts of data, i.e. BD. In particular, processing MVTD recorded at high rates over long periods of time is challenging for conventional ML models and demands considerable hardware resources. Autoregressive DL models such as the LSTM model have seen an increasing interest in a prognostic context, e.g. health monitoring and RUL prediction [23, 24] or automatic processing of written maintenance logs [25]. However, high dimensional inputs cause RNN-type models to be computational inefficient due to the required high numbers of trainable parameters. Further, valuable inter-sensor correlation might be lost [26]. Hence, CNN based architectures have been deployed in a supervised learning context to reduce the spatial dimensionality of the input by transforming signals into image-like representations that retain robust, lower-dimensional, non-linear feature-representation [27]. Consequently, research leverages the combination of CNN and RNN network topologies in supervised learning applications for PHM [26, 28, 29, 30]. The following section details CNN, CNN-AE, and RNN, some of which are the central DL building-blocks for TCN ².

5.2.1.1 Neural networks

NNs have been a topic of varying research interest for the last 70 years [31]. More recently they are being used in a wide variety of PHM applications summarised in [26, 32, 33, 34]. A NN's basic structure is made up of interconnected individual nodes or cells, often referred to as neuron, organised in layers through which the information flows in one direction (known as *feed-forward* NN). A node, cf. Figure 5.7, may have several connections to other nodes in the previous and subsequent layers. The neuron processes the received information from each connection by multiplying it with an assigned *weight* w_i and summing up these products, i.e. the weighted sum under consideration of the bias b . Activated by a non-linear function p , it sends this result to multiple neurons in the consecutive layer. Various activation functions have been considered including the *Sigmoid*-, the *Hyperbolic-Tangent* or the *Rectified Linear Unit (ReLU)*, cf. Table 5.1. [35] provides a comprehensive description of various other activation functions such as the parametric-, exponential-, or leaky-ReLU that are commonly employed. In general, the node is activated if the input

²Parts of this chapter have already been published in my journal or conference article: L. Kirschbaum, et al., (2021).

Table 5.1: Selected NN activation functions.

Name	Range	Abbreviation	Function $p(x)$
Sigmoid	$(0, 1)$	$\sigma(x)$	$p(x) = \frac{1}{1+e^{-x}}$
Hyperbolic tangent	$(-1, 1)$	$\tanh(x)$	$p(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
Rectified linear unit	$[0, \infty)$	$relu(x)$	$p(x) = \max(0, x)$

contains the learned feature. The *weights* and *biases* are initially set up random and then adjusted during training.

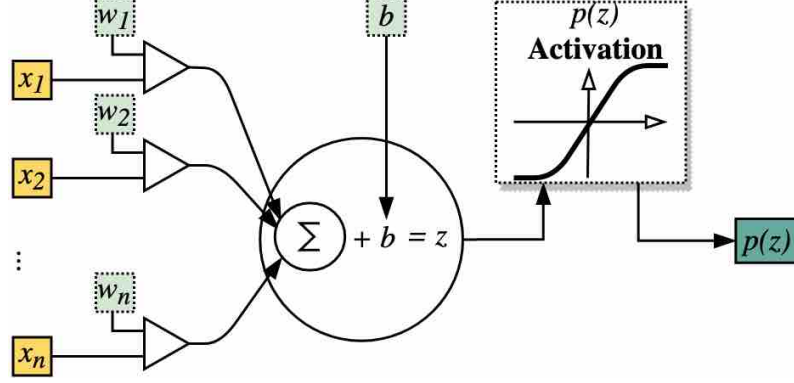


Figure 5.7: Single neuron schematic. Consider inputs \vec{x}_i , weight w_i for each input and a bias b using an activation function p .

The choice of the hyperparameters such as the activation function, the number of layers, the learning rate, or the initialisation of the weights and biases is decisive for the final performance of the NN. Thus, optimisation of these parameters and regularisation to prevent overfitting is a central objective when training NNs. In supervised applications the training process in order to adjust the weights and biases of the nodes in each layer comprises of three general steps. First, the NN will pass the input data \vec{x}_i received at the input layer forward through its architecture and outputs a prediction \hat{y}_i at the final layer. Second, the loss is calculated, using an error function E , e.g. for a regression task this may be the *mean squared error*. The prediction result \hat{y}_i is compared against the ground truth \vec{y}_i which yields an error. Third, the NN is trained using, e.g. stochastic gradient descent or, more commonly applied, the *Adam* optimiser [36]. Regularisation can be achieved by, e.g. the *L1-norm*, *L2-norm*, i.e. penalising the objective function, or *Early-Stopping*, i.e. terminating the training process if the validation error increases. A comprehensive explanation can be found in [35].

5.2.1.2 Convolutional neural networks

CNNs are a type of specialised DL architecture able to operate on various data input topologies (1D, 2D, or 3D). CNNs were first proposed more than two decades ago by

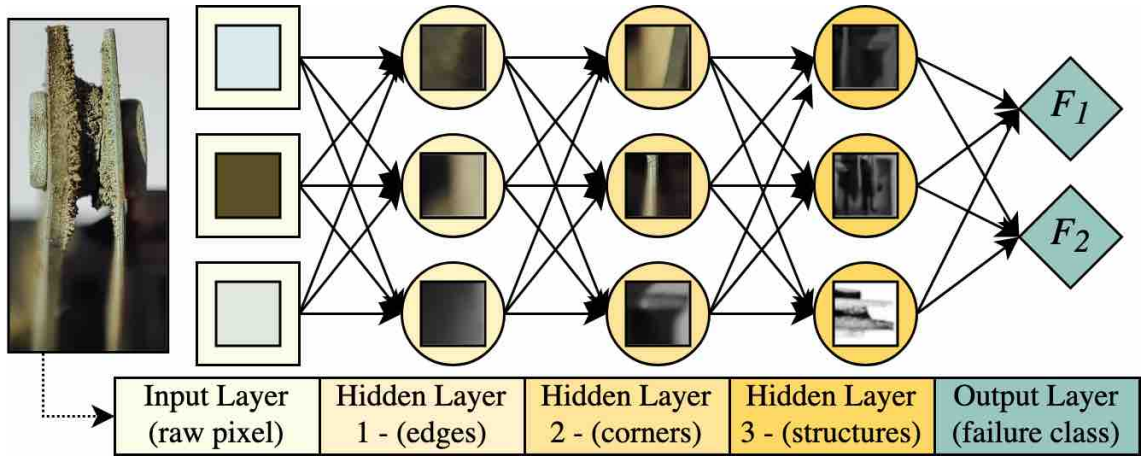


Figure 5.8: Concept of a CNN architecture for failure mode classification of EMR contacts.

[37, 38]. Relevant tasks are, i.e. pattern recognition, image classification, instance segmentation, or object detection [39, 40, 41]. In a PHM context CNN has been used to monitor infrastructure condition thriving on image data, e.g. cracks in rails, fault identification, and analysis of rotating machinery equipment by [42, 43, 44]. CNNs can automatically learn low-level and high-level abstractions from the input data. In order to illustrate the general working principle of CNN consider Figure 5.8. Here, the concept of a CNN for failure mode classification (class F_1 and F_2) of electrical contacts based on images of failed contacts is displayed. The network is trained with a data set of labelled images of failed EMR contacts (the label represents the failure mode). Note, learning the failure mode mapping directly from the raw sensory information is difficult, i.e the individual pixels of the image. However, using a CNN the mapping is divided into a series of sequential, simpler tasks. From the *visible* input layer the raw input is propagated through a series of *hidden* layers. These layers are learned, containing salient representations of the data. The complexity of the representations increases with the network depth. The representations of edges, contours, and structures in the 3rd hidden layer enables the trained NN to recognise the failure mode (class F_1 or F_2) depicted on the provided input image.

Just like the general NN, a CNN comprises a set of subsequent layers where the output of each layer forms the input to the consecutive layer [35]. Unlike the conventional NN, the CNN stacks a set of so called convolutional layers. The utilisation of these convolutional layers is unique to the CNN. A CNN can be made up out of many of subsequent convolutional layers, as can be seen in Figure 5.9-(top). Throughout the network the abstraction of the feature representation increases. Today's successfully deployed CNNs are very deep and contain millions of trainable parameters. Starting from the vanilla CNN architecture *LeNet*, a CNN referred to as *AlexNet* developed by [45] stacks many convolutional layers for image classification. Such architectures have been further refined by building even more complex, deeper network topologies, cf. [46]. The recent state-of-the-art CNN architecture for image

recognition, i.e. ResNet, contains more than 10 million trainable parameters, cf. [47].

If one considers an individual convolutional layer, a defined sequence of operations is performed, namely convolution, activation and pooling, cf. Figure 5.9-(bottom). The remainder of this section discusses the details of the operations performed on the input of each convolutional layer and the final Fully Connected Layers (FC)s layer. The details of these operations are schematically illustrated:

1. The convolution layer performs a linear operation. The convolution stage slides a set of learnable kernels (equivalent to a matrix of weights, cf. the concept of weights presented in Chapter 5.2.1.1) over the layers' input. The learned weights are then adjusted at each training iteration. These convolution operations are performed in parallel returning a set of output feature maps (one for each kernel). The convolution operation is displayed in Figure 5.10.
2. To increase non-linearity each feature map is activated using, e.g. the ReLU activation function. The activation operation is displayed in Figure 5.11-(left).
3. To improve invariance against small spatial changes in the input, i.e. translational invariance, pooling is applied. The pooling operation is displayed in Figure 5.11-(right).
4. The final layer is typically a FC layer which transforms the output of the last convolutional layer into a vector of class probabilities in the case of a supervised classification task or a single value in the case of a supervised regression task. The FC is displayed in Figure 5.12.

(1) Convolution Operation The first step is the actual convolution operation performed instead of matrix multiplication as it is the case for classical NNs. This is termed *convolution stage*. Note, the notation of this subsection refers to images since CNN is typically applied to images. The convolution is performed as the elementwise product given each value of an input of $I = H^{in} \times W^{in} \times C^{in}$ corresponding to the height, the width, and the depth (for an image this would be the number of colour channels) as well as a set of n kernels (during training the network aims to learn the set of kernels, i.e. the set of weights) $K = [k_1, k_2, \dots, k_n]$ (of size $k = H^k \times W^k \times C^{in}$, where H^k denotes the height and W^k the width of the kernel). In general, the kernel is significantly smaller than the spatial dimension of the input I [35]. The convolution operation results in an output A sometimes referred to as feature maps or activations. If $C^{in} \geq 1$ then the convolution operation is performed individually for each channel and summed up over all channels. Lastly, the bias is added. To extract various features at the same position, CNN architectures typically use multiple kernels. Hence, e.g. kernels used in CNNs in image classification applications can be of different type e.g. edge detection, sharpen, etc. and spatial extent, cf. [35] for an extensive discussion of typical kernels. The kernel parameters

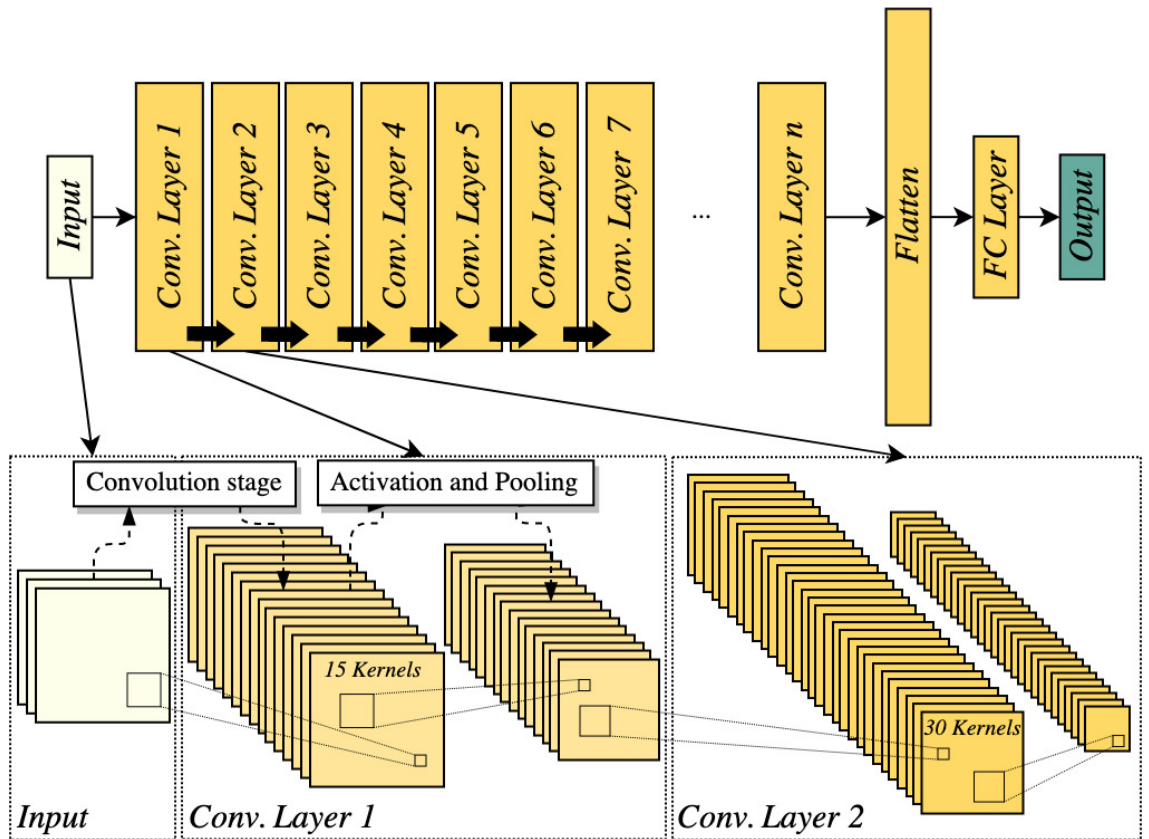


Figure 5.9: (Top) A CNN with n convolutional layers that each process the input from the prior layer and pass the results on to the consecutive layer. Typically, prior to the output layer, the results of the last convolutional layer are flattened into a single 1D vector which is then passed into a fully connected (FC) layer. (Bottom) Each convolution layer holds a set of feature maps which each correspond to a different kernel.

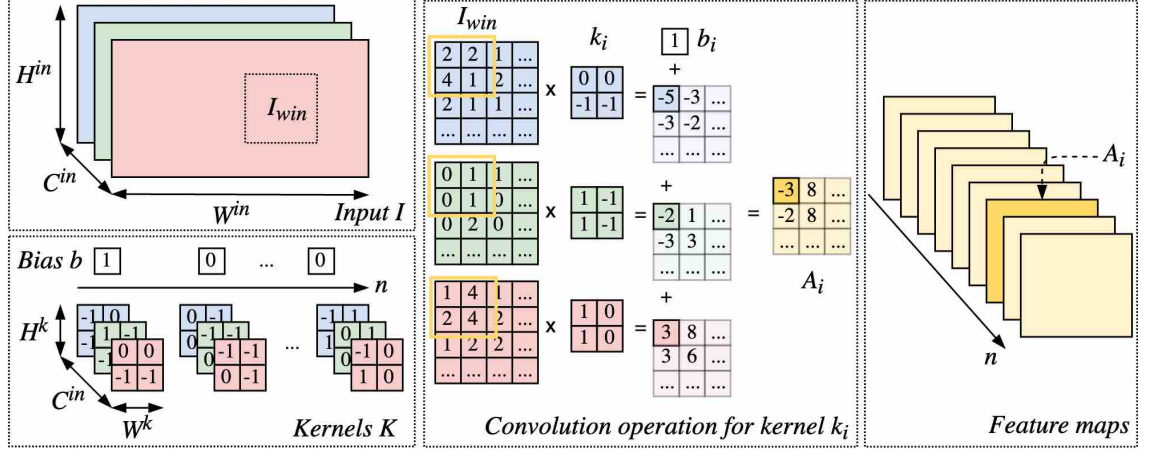


Figure 5.10: Convolution operation using n kernels K operated on an input I . Schematically illustrated for feature map A_i using a kernel k_i convoluted on all three channels for a representative section of the input I_{win} .

are learned during training. After the convolution operation n feature maps are derived as depicted in Figure 5.10. Besides the number and the spatial extent of the kernels, the output of the convolution operation depends on two more parameters. The **stride** s defines the step width at which the kernel is moving across the input. If $s = 1$, element-wise multiplication is performed on each value of the input. If $s \geq 1$, only every s^{th} of the input image is scanned by the kernel so that the convolution operation decreases the spatial extent of the feature maps A in comparison to I , cf. Equation 5.8. However, in order to maintain the original dimension of I in A **zero-padding** can be used. Thus, the input is padded with zeros around its outer boundaries.

$$A = ((H^{in} - H^k) + 1) \times ((W^{in} - W^k) + 1) \times n \quad (5.8)$$

Discrete convolution is motivated by the type of grid-structured data, e.g. images containing areas that locally correlate - irrespective of their absolute location [48]. Hence, detecting a pattern independent of its location in the input is achieved by sharing the weights across the entire input instead of being locally bound to one specific area of the input. This *sharing of the weights* reduces the number of trainable parameters significantly in comparison to traditional NN that employ matrix multiplication [35].

(2) Activation Operation Within the **detector stage**, the objective of the activation layer is to increase the non-linear behaviour of the network, i.e. the output is only activated for regions that exhibit the learned salient pattern, cf. Figure 5.11-(left). The non-linear ReLU function is commonly used in CNN. ReLU truncates the input for each element, according to 5.9.

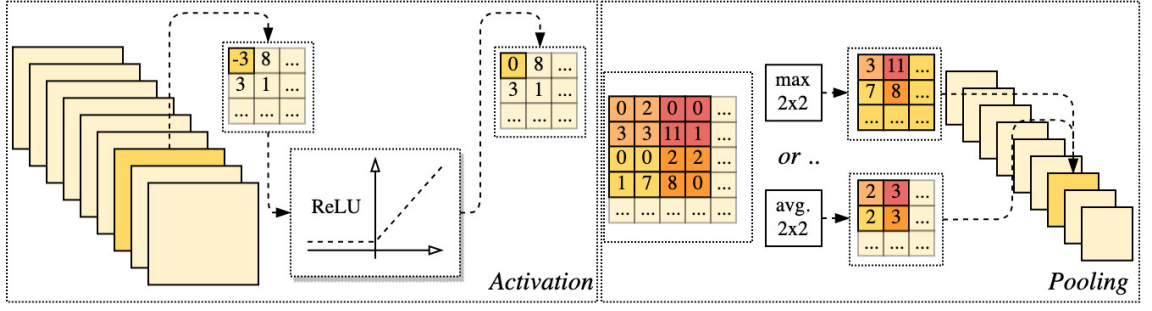


Figure 5.11: Activation as in the detector stage on the left, and pooling on the right (*max-pooling* only retrieves the maximum of the 2x2 sub-region; *avg.-pooling* takes the mean of the 2x2 sub-region).

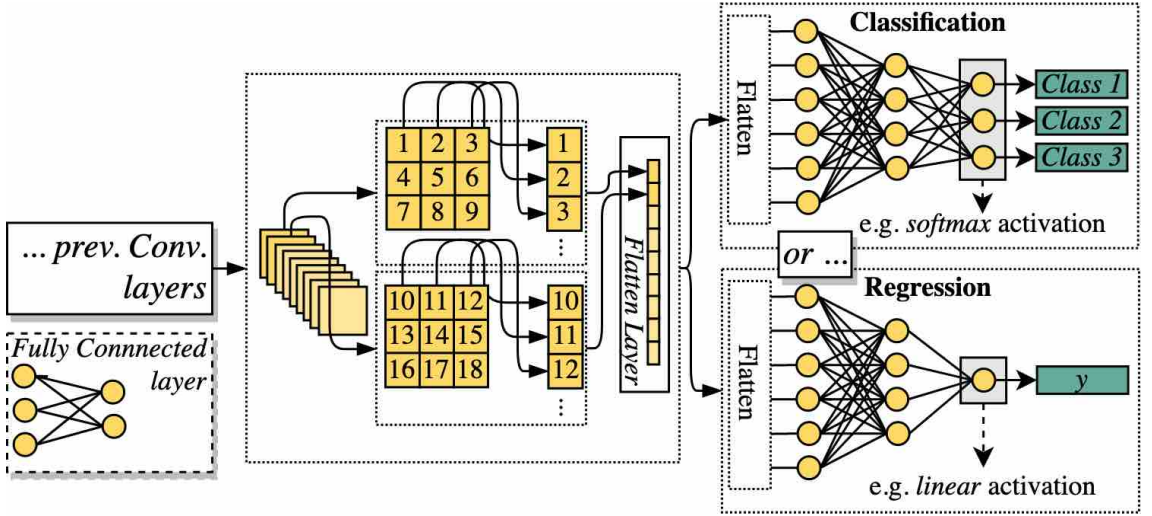


Figure 5.12: Fully connected (FC) layers in different configurations for supervised regression and classification tasks.

$$y(x) = \max(0, x) \quad (5.9)$$

(3) Pooling Operation Lastly, the **pooling** stage is applied, cf. Figure 5.11-(right). The aim of the pooling layer is twofold: 1^{st} to down-sample the input in order to reduce the required network parameters in the subsequent layers, therefore, improving computational efficiency [35]; 2^{nd} to reduce the effect of small local translations in the input, so called local translation invariance [48]. Pooling maps non-overlapping sub-regions of the input to a single numerical representation of this sub-region. Thereby, the majority of the pooled elements in the output remain constant, despite locally bound small variations. Two types of pooling exist: *max-pooling* returns only the maximum of the sub-region; *average-pooling* returns the mean of the sub-region.

Fully Connected Layer In practice, a set of Fully Connected Layers (FCs) are often used as the final network layers. Figure 5.12 illustrates how the final CNN layer needs to be complemented for classification and regression task respectively. Prior

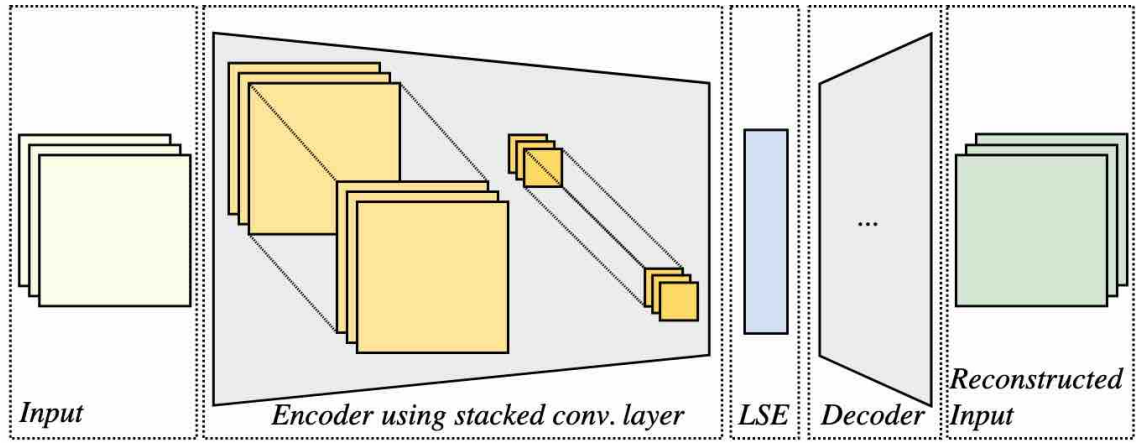


Figure 5.13: CAE operating on an 3D input using stacked convolutional layers to extract meaningful representations contained in a lower dimensional representation referred to as latent-space encoding (LSE). Reverse mirroring the structure in the encoder stage, the decoder reconstructs the original input solely from the latent-space encoding.

to the the FCs, the feature maps of the last pooled convolution layer are flattened into a single 1D vector. If a classification task is considered, the final fully connected layer is made up of as many nodes as there are classes using *softmax* activation to output class probabilities. In the case of regression one single node using linear activation returns a continuous value.

5.2.1.3 Convolutional Auto Encoder

A variation of the CNN is the CAE. AEs are NN structures that consist of two building blocks: the *encoder* and the *decoder*. The (general) aim of the AE is to replicate its input to the output. An under-complete AE constrains this copying task, i.e. learning the identity function, as it encodes the input to a lower dimensional representation and respectively reconstruct this input from the latent space, trying to minimise a loss function [49]. The lower dimensional representation acts as a bottleneck, enforcing the encoder to learn the prominent features of the input. If non-linear activation functions are employed during encoding, the AE is capable of learning expressive non-linear mappings of the input [35]. For a comprehensive study of AEs refer to [50]. CAE can be used for unsupervised automated feature extraction or denoising from the stacked raw multi-channel input data [49]. This DL architecture performs a set of subsequent operations on an input as depicted in Figure 5.13. The lower dimensional *latent-space encoding* - the spatial dimensionality of the large input is reduced through a set of subsequent convolutional layers - contains the learned expressive non-linear feature representations.

5.2.1.4 Recurrent neural networks

RNNs are a type of autoregressive NNs used for sequential tasks. They have been heavily deployed in natural language processing [51], but also improved performance

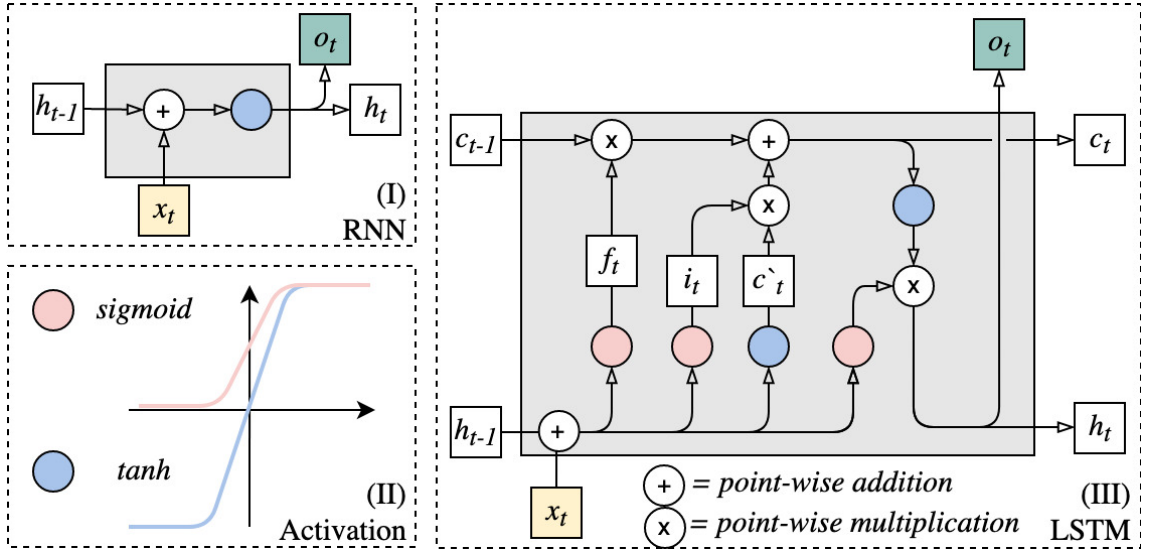


Figure 5.14: (I) Schematic of RNN cell; (II) the activation functions employed by RNN and LSTM; (III) Schematic of LSTM cell.

in tasks that relate to time series data $X_n = [x_1, x_2, \dots, x_{t-1}, x_t]$, e.g. predicting the next sequence step x_{t+1} . RNN extends the concept of the traditional feed forward NN using some kind of *sequential memory*, i.e. repeating nodes for each sequence step. A RNN transports the hidden state vector h_{t-1} from the previous cell to the current cell to derive an updated current state vector h_t in respect to the current sequence step x_t (point-wise addition of x_t and h_{t-1}), cf. Figure 5.14-(I). As detailed in Chapter 5.2.1.1, the non-linear behaviour of the RNN is achieved using the tanh activation function [35, 51], cf. Table 5.1.

RNN suffer from *short-term memory* due to a vanishing gradient during back-propagation [52] (the gradient of the errors at each training iteration inform the direction in which the layer weights are adjusted to reduce the error at the next training iteration, cf. Chapter 5.2.1.1 and [35]). However, the gradient not only shrinks from the networks output layer to the input layer, but also as it propagates backwards through each learned time-step damping the influence of long-term dependencies and earlier information. In addition, during training the vanishing gradient problem prevents the earlier layers from being updated. Hence, the RNN does not learn these long-term dependencies. This issue has been addressed in LSTMs aiming to improve short-term and long-term memory alike.

In 1997 [53] proposed the Long Short Term Memory Neural Network network, a modification of the RNN to address the above discussed short-term memory issue. Therefore, in addition to the hidden state vector of the previous sequence step h_{t-1} , a cell state vector c_{t-1} is introduced as illustrated in Figure 5.14-(III). As extensively discussed in [54] this modification improves the retention of long-term memory from previous cells. However, one should note, the design of LSTM reduces the vanishing-gradient problem but does not eliminate it. The cell state, the hidden state, and

the input of the current state x_t are combined using three *gates*. The forget gate f_t applies the σ activation function to the concatenated input of the previous hidden state h_{t-1} and the current sequence input x_t under consideration of the weights w_{f1} and w_{f2} . The bias of the forget gate is denoted as b_f . The forget gate controls the amount of information from the previous state that will contribute to the new state. A similar operation is performed on the input gate i_t . The same input as used by f_t and i_t is then regularised with the tanh activation to the range of $(-1, 1)$ returning the new preliminary cell state \hat{c}_t . i_t and \hat{c}_t are first combined using point-wise multiplication and then point-wise added to the product of f_t and c_{t-1} resulting in the current cell state c_t . Lastly, identical to the current hidden state h_t , the output gate o_t is derived by regularising c_t with tanh and then point-wise multiplying it with the input x_t and h_{t-1} which has been passed through the σ activation function. Like RNNs, LSTMs have found application in the domain of natural language processing [55], but also in PHM. Various modifications to the LSTM architecture have been proposed throughout literature [56]. For example, the Bidirectional LSTM [57], a LSTM coupled with an Attention Mechanism [58], or the so called Gated Recurrent Unit Neural Network (GRU) architecture [59].

5.2.2 Time series encoding enabling CNN-RNN topologies

A wide range of methods address the extraction of key insights from time series data, e.g. signal decomposition or forecasting. A detailed discussion of time series data and time series data analysis can be found in [60]. However, the focus of this section resides on the application of CNN to time series data. Motivated by the success in image representation learning, CNN's application has been extended from the analysis of condition monitoring image-like data to the application on time series data in the domain of PHM. A univariate time series $X_i = [x_1, x_2, \dots, x_t]$ consisting of one time-dependent variable can be encoded as a representation where $W = t$ and $H = 1$. This allows the usage of a 1D-CNN with a kernel of $H^k = 1$. Such an example for a PHM related application is presented by, e.g. [61]. Likewise, as in [62], a multivariate time series $X = [X_1, X_2, \dots, X_n]$ is encoded as 2D representations where $W = t$ and $H = n$ (n is the number of signals), cf. Figure 5.15-(top, centre). MVTD can also be encoded as 3D representations through windowing and stacking as visualised in Figure 5.15-(bottom). Further time series encoding methods to consider are Frequency Transformation, Inter-Sensor Correlation Maps [63, 64], Gramian Angular Field (GAF), Markov Transition Fields [65], or Recurrence Plots [66]. The findings of the authors suggest that prediction performance of CNN architectures can be improved if such 2D encodings are exploited in comparison to the raw 1D time series data. A comparative study confirming this assumption is presented in [27]. For example, the GAF can be seen as some form of data augmentation which allows the model to learn temporal dependencies as well as complex spatial

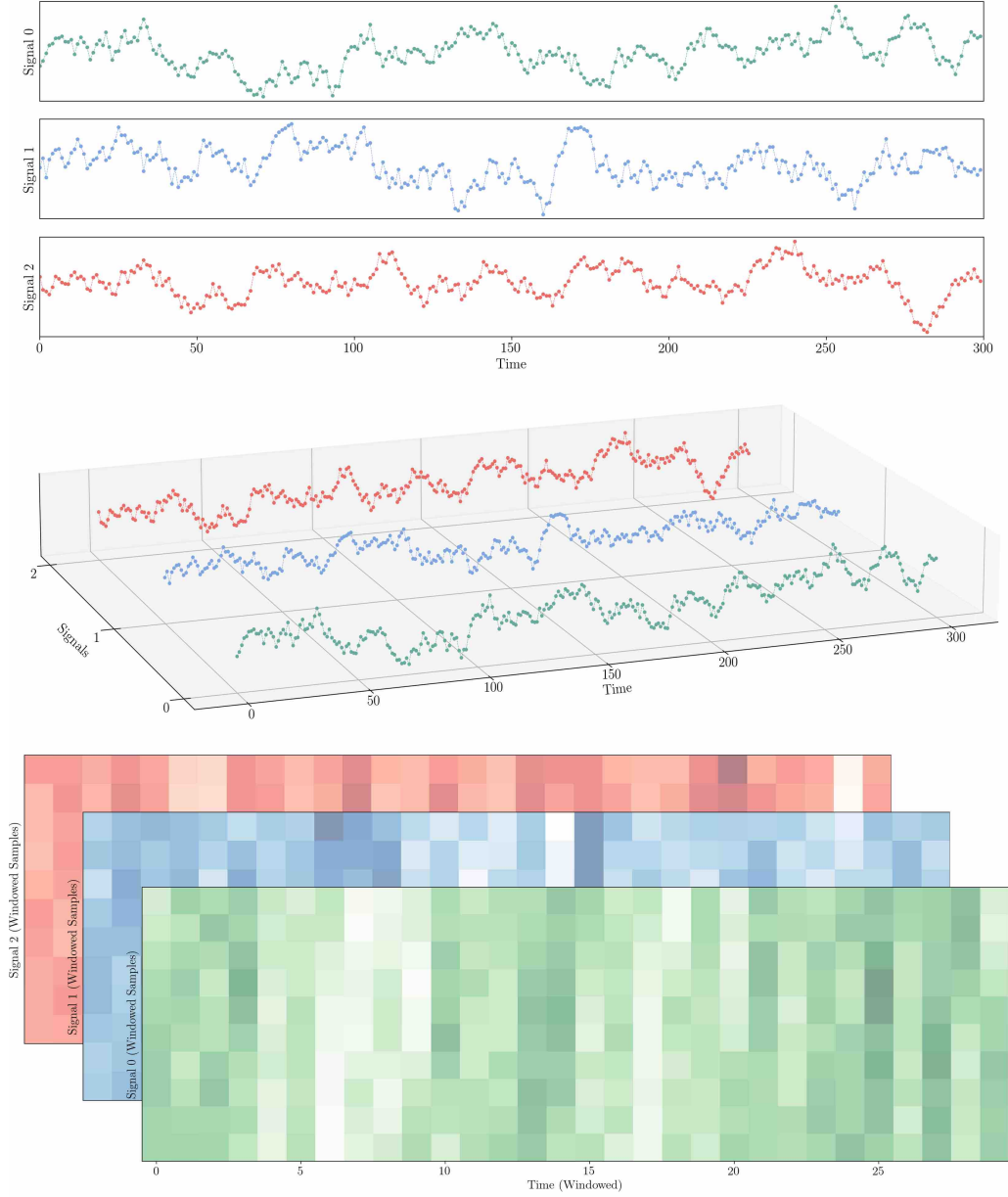


Figure 5.15: (Top) MVTD comprising three signals $n = 3$ of length t ($X = [X_{S1}, X_{S2}, X_{S3}]$). Each X_i can be represented as 1D Vector. (Centre) X can also be represented as a 2D matrix, imagine a 1-channel image where the width corresponds with the time $W = t$ (i.e. the number of samples), and the height with each individual signal, i.e. $H = n$. (Bottom) representation of X as 3D matrix. This can be achieved by windowing each signal and stacking these windows vertically. Hence the width is equal to $W = t/h$ where n is the size of the window, so that the height $H = h$ and the depth $C = n$.

dependencies of the signal. Its applicability in a PHM context for, e.g. anomaly detection is currently an active area of research [26].

CNN is indifferent towards sequential dependencies of the input, since it does not retain information from previous sequential steps to inform the current sequential step. On the other hand, a CNN combined with a LSTM forms a DL network that allows to derive sequential information from a spatial-temporal input. This has been shown to be useful for PHM [64, 67, 68]. Many different topological combinations of convolutional and autoregressive DL models can be tailored to a wide range of supervised learning tasks including anomaly detection, classification, and regression.

5.2.3 Review of topical literature

Not limited to diagnostics and prognostics purposes, CNN and RNN in various arrangements have been proposed for supervised tasks in order to cope with increasingly high volumes of time series data. Research shows how CNN can be used as standalone DL method or in combination with autoregressive DL methods to leverage the potential of data parallelism capabilities for dealing with large input feature maps, e.g. MVTD in high resolution. Table 5.2 provides an overview of relevant research not limited to PHM ³.

Table 5.2: Comparison of DL time series related methods (*UV* - Univariate time series data; *MV* - Multivariate time series data; *W* - Width of the input; *H* - Height of the input; *C* - Depth/Channels of the input; *t* - length of the time series window; *t_d* - length of down-sampled time series window; *t_s* - length of signal dependent time series window, i.e. it might be different for each signal depending on the sampling rate; *n* - number of individual signals; *n_f* - number of individual frequencies or frequency bands; * - data set not publicly available).

Ref.	Year	Type	Domain	Feature Map	Model
Anomaly Detection					
[69]	2016	<i>UV</i>	Time	1D - $W = t, H = 1, C = 1$	CNN based architecture; <i>DEAP Physiological</i> , 512 Hz, unlabelled [70]
[68]	2018	<i>UV</i>	Time	1D - $W = t_s, H = 1, C = 1$	DeepAnT; Yahoo time series anomaly benchmark
[64]	2018	<i>MV</i>	Time	3D - $W = t, H = t, C = n$	Multi-Scale Convolutional Recurrent Encoder-Decoder (MSCRED); Power Plant*, 5 anomaly classes, labelled
[67]	2019	<i>MV</i>	Time	1D per signal - $W = t, H = 1, C = 1$	Multi-head CNN-RNN; Elevator*, 3 anomaly classes, labelled.
[71]	2019	<i>UV</i>	Freq.	1D using spectral residual transform - $W = t, H = 1, C = 1$	Salience Residual CNN (SR-CNN); Yahoo time series anomaly benchmark
Continued on next page					

³Parts of this chapter have already been published in my journal or conference article: L. Kirschbaum, et al., (2021) and L. Kirschbaum, et al., (2022).

Table 5.2 – continued from previous page

Ref.	Year	Type	Domain	Feature Map	Model
[72]	2019	MV	Time	2D - $W = t, H = 1, C = n$	time series segmentation CNN based on U-Net architecture; Dodgers Loop [73] and Gasoil Plant Heating
[74]	2020	UV	Time	1D - $W = t, H = 1, C = 1$	Robust time series anomaly detection (RobustTAD) based on U-Net; Yahoo time series anomaly benchmark
Classification					
[75, 76]	2014	MV	Time	1D per signal - $W = t_s, H = 1, C = 1$	Multi Channel Depp CNN (MC-DCNN); PAMAP2, 12 classes [77].
[78]	2015	MV	Time	2D - $W = t, H = n, C = 1$	CNN based architecture; Opportunity Activity Recognition, 18 classes, 113 signals, 30Hz [79]
[65]	2015	UV	GAF	2D $W = t, H = t, C = 1$	Tiled CNN; GunPoint, 2 classes [80]
[81]	2016	MV	Time, Freq.	2D per signal - $W = t, H = n_f, C = 1$	Multi Scale CNN (MS-CNN); UCR time series classification [80]
[82]	2017	UV	Time	1D - $W = t, H = 1, C = 1$	CNN; UCR time series classification [80]
[83]	2017	UV	Time	1D - $W = t, H = 1, C = 1$	LSTM-CNN; UCR time series classification [80]
[62]	2017	MV	Time, Freq.	1D - $W = t, H = 1, C = 1$	Adaptive Multi-Sensor data Fusion Deep CNN; Planetary Gearbox*, 7 classes, 4 signals, 20kHz
[84]	2017	UV	Freq.	2D - $W = t, H = n_f, C = 1$	Wavelet Transform CNN (WT-CNN); Gearbox*, 10 classes, 24kHz
[85]	2017	UV	Freq.	2D - $W = t, H = n_f, C = 1$	Short-Fourier and Wavelet-Transform based CNN; MFPT Bearing Fault, multiple fault classes [86]
[87]	2017	UV	Freq.	2D - $W = t, H = n_f, C = 1$	Energy Fluctuated Multi-Scale Feature CNN; CWRU, 2 fault classes
[88]	2017	UV	Time	2D - $W = \sqrt{t}, H = \sqrt{t}, C = 1$	Temporal CNN; CWRU, 2 fault classes
[89]	2017	UV	Time	1D - $W = t, H = 1, C = 1$	CNN; Electrocardiogram*, 12 classes, 200Hz
[66]	2018	UV	Time	1D - $W = t, H = 1, C = 1$	CNN; UCR time series classification [80]
[90]	2019	MV	Time	3D - $W = t, H = 2, C = 2$	Multivariate CNN (MVCNN); PHM 2015 Data Challenge
[91]	2019	MV	Time	2D - $W = n, H = t, C = 1$	Multi-Task Attention CNN; Air Quality*
Regression					
[61]	2016	MV	Time	2D - $W = t, H = n, C = 1$	CNN; C-MAPSS aircraft engine, PHM 2018 Data Challenge
[92]	2018	MV	Time	2D - $W = n, H = t, C = 1$	CNN-LSTM; PEMS Traffic, NREL Solar Energy

Continued on next page

Table 5.2 – continued from previous page

Ref.	Year	Type	Domain	Feature Map	Model
[28]	2019	MV	Time	2D - $W = t, H = n, C = 1$	CNN-Bi-LSTM; IHEPC Electricity
[30]	2020	MV	Time	2D - $W = t, H = n, C = 1$ and a separate 1D statistical feature vector	Multi-Channel and Multi-Scale CNN-LSTM; AEP, COMED, DAYTON.

To enable modern maintenance by creating actionable insights from data, e.g. data-driven CBM and PdM, researchers are confronted with high volumes of MVTD. Thus, deriving meaningful insights from high volumes of time series data in a data-driven context may be challenging:

- dealing with high dimensional inputs (of different scale) in an efficient manner [91];
- extracting relevant features may require domain expertise;
- improving the utilisation of information from spatial, temporal and inter-sensor correlations [64, 93];
- applicable to different types of MVTD, long term dependencies and slow trends, often masked by noise - rendering many traditional time series classification and forecasting methods unsuitable [71, 91, 94].

With the advent of AI research is addressing this challenge using ML and increasingly DL also for RUL prediction of electronics [95, 96, 97, 98]. As evidenced, DL is a promising candidate, since it supplies a scalable architecture for automatic feature extraction from raw data [69] and the ability to learn complex time dependent degradation processes [99]. Inter alia, solutions combining DL techniques, capable of operating directly on large inputs while preserving sequential aspects have become a popular choice for classification and regression tasks. Standalone CNN architectures or in combination with RNN that employ techniques such as multivariate-time series imaging resonate with high-volumes of MVTD as recent publication have demonstrated [26, 27, 61, 75, 76, 78, 85, 87, 91]. CNN can extract local and global, spatial features that show good generalisation qualities due to the sharing of parameters [68]; or are capable to operate directly on raw multivariate input data [61]. In particular, approaches utilising CNN in the form of CAE for automated feature extraction while drawing on the autoregressive power from RNN-based architectures are accepted combinations [29, 28]. Such approaches have been further refined to improve performance of CNN-LSTM architectures, e.g. multi-input NN topologies which consider statistical feature vectors derived from windowed time series data in parallel to the CNN-LSTM [30].

Although accurate for auto-regressive tasks, one major shortcoming of the above reviewed approaches is their computational efficiency, since RNNs somewhat limit the

efficiency of CNN. They require considerable computational resources particularly when dealing with long inputs and evince deficits in retaining extended long-term dependencies [94]. A recent, novel DL architecture addressing above challenges is referred to as Temporal Convolutional Network (TCN). In contrast to, e.g. CNN-LSTM, TCN extends CNN's traditional applicability to sequence classification or forecasting problems without the need to integrate a RNN architecture. It exhibits improvements in training time as well as performance [100]. Unlike, e.g. LSTM, TCN is able to be trained on input sequences, irrespective of the length of the input, since the number of trainable parameters per layer only depends on the number of input features, filters and the kernel-size. The working principle, its advantages over, e.g. CNN-LSTM and recent applications implementing TCN for State of Health (SOH) and RUL are detailed in the following Chapter 5.2.4.

5.2.4 Principles of TCN

Proposed by [100], Temporal Convolutional Network (TCN) is a novel, autoregressive DL architecture based on 1D-CNN. TCN extends the functionality of CNN used for, e.g. image classification tasks towards sequence classification and forecasting [101]⁴.

TCN compensates the shortcomings of classical autoregressive DL models when it comes to learning very long sequences [102]. Advantages are the mitigation of the vanishing or exploding gradient problem when back-propagating through time as often encountered with LSTM, reduction of memory usage as well as training and inference time over traditional RNN architectures [103]. Compared to LSTM, TCN requires less trainable parameters to store intermediate results [100]. To elaborate, 1D-convolution adopted in TCN shares the learned filters across the entire input feature map of length l per input channel c . This can be attributed to the parallelism of the convolution operation. Given a sequence $x_l = [x_1, x_2, \dots, x_{i-1}, x_i]$, retrieving a result for x_i using RNNs depends on the prediction of x_{i-1} and all previous time steps. However, convolution can operate in parallel on the entire sequence x_l as the same kernel k is shared across the entire layer. Lastly, controlling the size of the receptive field r can be accomplished by different means providing greater flexibility in the design of the architecture [100].

TCN shares the ability to map an arbitrary-length input sequence x_l to an output sequence y_l of the same length using 1D-convolution. However, in sequence modelling it is important that an output y_i only depends on the current and previous inputs $[x_1, x_2, \dots, x_{i-1}, x_i]$. Figure 5.16-(I) displays the concept of causal convolution. Consider a multivariate input sequence with $c = 2$ so that

⁴Parts of this chapter have already been published in my journal or conference article: L. Kirschbaum, et al., (2022).

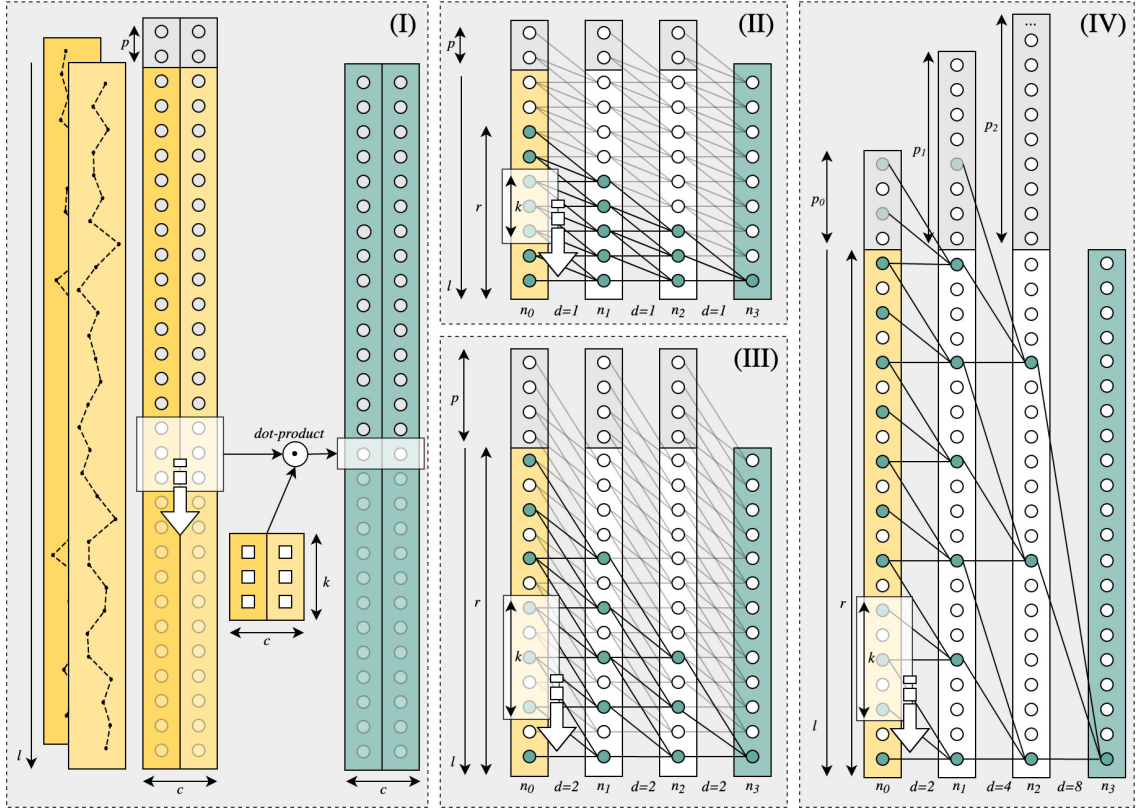


Figure 5.16: (I) Causal 1D casual convolution with multi-channel input $c = 2$ for a sequence of length l and a kernel of size $k = 3$, zero-padding $p = 2$; (II) 1D causal convolution with $k = 3$, $d = 1$, $p = 2$, no full history coverage as $r < l$; (III) 1D causal convolution with $k = 3$, constant dilation $d = 2$, $p = 4$, full history coverage as $r = l$; (IV) 1D causal convolution with $k = 3$, exponential dilation with $d_b = 2$ and $d = [2, 4, 8]$, full history coverage as $r = l$.

$X = [X_1, X_2]$, cf. Figure 5.16-(I)-(left). This input feature map is convolved moving sets of learnable kernels of size $k = 3$, unique for each channel c , in one direction along one axis only across the multivariate input, cf. Figure 5.16-(I)-(centre). This outputs a 2D tensor \hat{X} of the same length l and width c , cf. Figure 5.16-(I)-(right). Hence, at each training iteration in each convolution layer, the learnable kernel of size k is shifted across the input with a step-width of $s = 1$ utilising the same kernel weights across the entire input channel. In practice, if $c > 1$, the 1D-convolution can be imagined like a 2D-CNN, where the filters are restricted to the channel. The number of weights used in the model depends on the kernel size k , the number of filters f and the network depth n . As one can see in 5.16-(I), in order to retain the same sequence length for the output y_l , zero padding at the beginning of the sequence is required. In the case of simple 1D causal convolution the padding length is $p = k - 1$.

The receptive field of size r , i.e. the number of elements in the input x_l which relate to an output y_i , is important to consider. In the case of TCN, the width of r defines how far back the model's horizon reaches. If r covers the entire input length it is termed *full history coverage*. As one can see in Figure 5.16-(II), the receptive field r grows linearly with the network depth n as $r = 1 + n(k - 1)$ if k is constant throughout the entire network. Therefore, increasing r can be achieved by either increasing the depth of the network n or the kernel size k . Hence, due to this linear relationship between network depth n and receptive field r , achieving full history coverage for sequences where l is large would require very deep networks. In turn, this may cause problems with the vanishing gradient and undo the advantages of TCN over RNN based architectures.

To circumvent this problem *dilation* is introduced. Dilation, somewhat similar to the step-width used in classical CNNs, spreads out the kernel across the input skipping certain elements depending on the dilation step-width d . A kernel of size $k = 3$ and dilation $d = 1$ would convolve over an input of $l = 3$. Contrary, if $d = 2$, the same kernel would cover an input of $l = 5$ with holes at the 2^{nd} and 4^{th} elements. This concept is introduced in Figure 5.16-(III). The receptive field grows as $r = n(1 + d(k - 1))$, depending not only on n and k , but also on d . However, if d is a constant r still grows linearly. Hence, to more effectively increase r along the network depth, d should grow exponentially as illustrated in Figure 5.16-(IV). This yields a dilation of $d_i = d_b^{n_i-1}$ and r as per Equation 5.10.

$$r = 1 + \sum_{i=0}^{n-1} (k - 1)d_b^i = 1 + (k - 1)\frac{d_b^n - 1}{d_b - 1} \quad (5.10)$$

Typically, the dilation is increased with the base of $d_b = 2$. To achieve full history coverage, i.e. $r \geq l$, the minimum number of required layers is

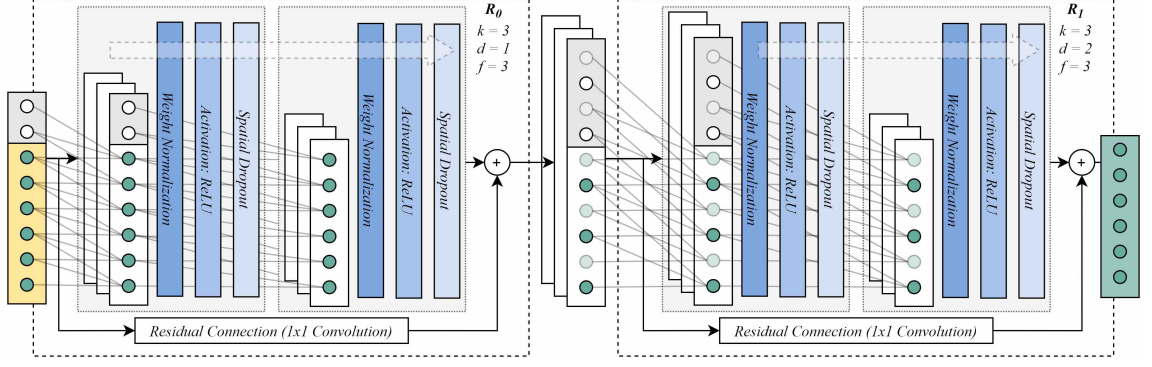


Figure 5.17: A TCN consisting of two residual blocks $R = [R_0, R_1]$ with filters $f = 3$ and a kernel size of $k = 3$. The first residual block R_0 employs two casual 1D convolution layers with dilation $d = 1$ and padding $p = 2$. The second residual block R_1 employs 2 causal 1D convolution layers with a dilation of $d = 2$ and padding $p = 4$. After each convolution layer, sequentially weight normalisation, activation using ReLU (for non-linearity), and spatial dropout for regularisation are employed. A residual connection is used to stabilise the network during training.

$$n_l = \log_{d_b} \left(\frac{(1-l)(d_b-1)}{(k-1)} + 1 \right) \quad (5.11)$$

where the padding for each layer is $p_i = d_b^{n_{i-1}}(k-1)$. To avoid gridding, i.e. an incomplete coverage of the elements in the input x_l within r , the kernel should be chosen as $k \geq d_b$ [104].

Adapted from [105], the authors in [100] utilise a structural element referred to as *residual block* replacing the simple 1D convolutional layer. TCN encapsulates this structural element to improve the stability of the architecture as the model learns a modification of the input feature map [100]. A network using residual blocks R is displayed in Figure 5.17 with $d = [1, 2]$, $f = 3$, and $k = 3$ resulting in a network with $R = [R_0, R_1]$ blocks.

The proposed structure alters the typical CNN building block, comprising $h = 2$ 1D-convolution layers using the same k and d . The receptive field r which allows full history coverage for this architecture can be expressed as

$$r = 1 + \sum_{i=0}^{n-1} h(k-1)d_b^i = 1 + h(k-1) \frac{d_b^n - 1}{d_b - 1} \quad (5.12)$$

$$n_l = \log_{d_b} \left(\frac{(1-l)(d_b-1)}{(k-1)h} + 1 \right) \quad (5.13)$$

Figure 5.18 visualises the growth of the receptive field after each TCN convolution layer. One can see that dilation only increases the size of the receptive field effectively if the dilation is increased alongside the network depth (cf. *constant dilation* and

dilation with d_b). As becomes apparent, the introduction of residual blocks nearly doubles the size receptive field.

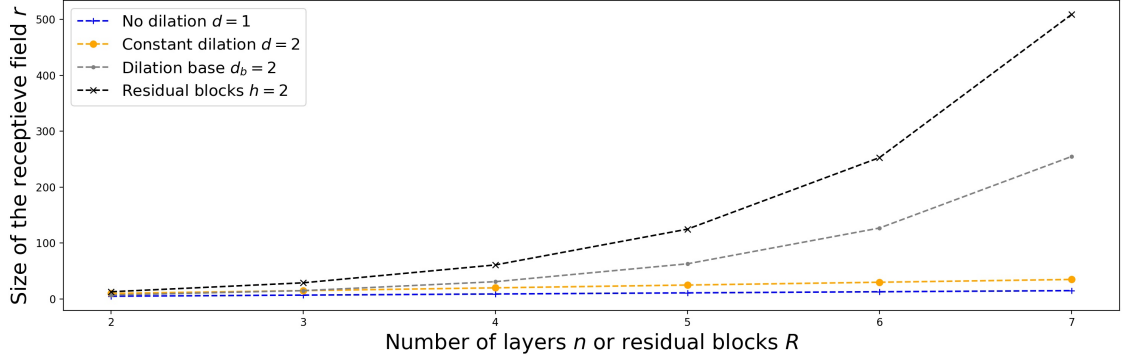


Figure 5.18: The size of the receptive field r depends on the number of residual blocks, the dilation base and the kernel size (here $k = 3$). Four cases are considered: (blue) - no dilation $d = 1$, $k = 3$; (orange) - constant dilation $d = 2$, $k = 3$; (grey) - dilation with $d_b = 2$, $k = 3$; (black) - usage of residual blocks architecture $h = 2$.

After each convolution layer, *weight-normalisation* to normalise the convolution outputs and reduce effects of an exploding gradient, activation for non-linearity using ReLU and regularisation using spatial dropout to minimise overfitting are employed. Contrary to dropout (randomly dropping out some elements in the feature maps) spatial dropout randomly drops out entire feature maps. The number of feature maps to be dropped out is controlled by the dropout rate. To retain the input sequence length l between residual blocks, a residual connection operates a 1×1 convolution directly on the blocks input and is then element-wise added to the blocks output. This stabilises the network and counteracts the vanishing gradient problem encountered when back-propagating the errors through deep networks [100].

5.2.5 Applications of TCN in a prognostics context

In [106], the authors propose the use of TCN for RUL estimation as an alternative to existing RNN-DL strategies. The method improves forecasting accuracy compared to, e.g. LSTM. K-Means Clustering is employed to distinguish between different component operating states. Evaluation is performed using a piece-wise linear RUL curve on the turbofan-engine degradation *NASA C-MAPPS* data set [107]. Likewise, verified on the *NASA C-MAPPS* data set, in [108] a refined approach utilising TCN for piece-wise linear RUL is presented. Instead of passing the multivariate input sequence directly into TCN, as in [106], a distributed hierarchical attention network is employed. The feature map weighs each channel, i.e. each sensor, in regard to its contribution to RUL. Subsequently, estimation of a RUL sequence using TCN, an attention network determines the contribution of each time-step to the final RUL. RUL prediction combining TCN and a sliding-window, fixed-size multi-channel feature map is evaluated on the *NASA C-MAPPS* data set by [109]. The proposed

method improves performance on some samples compared to [106, 108]. Moreover, it is demonstrated that an increase of the TCN’s receptive field through either the kernel-size or the dilation can both have a positive impact, though the number of trainable parameters increases.

Research on the performance of linear RUL estimation for bearings using TCN is presented by [110, 111]. In [110], the authors present a DL architecture operating on statistical features for windowed segments of MVTD in the time- and frequency-domain. The proposed model exhibits an overall improved performance compared to LSTM or GRU and reduced model training time. The approach is evaluated on the *PRONOSTIA* bearing data set [112]. Evaluated on the same data set, in [111] the authors replace the feature extraction process with Empirical Mode Decomposition that creates an input feature map with a distinct number of intrinsic mode functions revealing different characteristics of the input in the time domain. The authors then demonstrate how TCN performance, generalisability, and robustness of the RUL forecast can be further tuned using dilation to increase the size of the receptive field. However, compared to the approach proposed by [110] the improvements are marginal.

As in [111], in [104] the authors apply Empirical Mode Decomposition in combination with TCN for SOH and RUL prediction of lithium-ion batteries. In this context SOH is defined as the capacity at t over the initial capacity. The RUL refers to the number of cycles remaining till EOL, where the EOL is defined as a threshold capacity. For both tasks the proposed architecture exhibits improved performance compared to RNN architectures. The authors attribute this to the increased receptive field of the TCN. The proposed approach is evaluated on the *NASA lithium-ion battery* data set [113].

In [114] the TCN model serves as a forecasting model for piece-wise linear RUL estimation in critical nuclear power plant infrastructure, namely some electric valves. The authors propose to use CAE for robust feature extraction prior to RUL estimation by TCN. It is demonstrated that such feature extraction improves robustness against sensor noise over the direct use of the raw data. As emphasised before, the authors underpin the importance of hyperparameter selection for TCN, i.e. the number of causal layers, the size of the receptive field, and the dilation. It is concluded that under some conditions increasing the number of layers and the size of the receptive field may deteriorate the performance. Therefore, the choice of TCN hyperparameters is not trivial and appears to be highly application dependent.

While TCN has shown promising results in various engineering disciplines, a comprehensive review of the impact of TCN hyperparameters and possible sub-sampling strategies have not yet been subject of extensive research. Moreover, none of the reviewed diagnostic and prognostic methods that use TCN integrate a measure of

uncertainty with their forecast. This is why the following section addresses these very challenges, inter alia focusing on a computationally efficient uncertainty estimation for TCN during inference.

5.2.6 EMRUA pipeline

In this section the Electromagnetic-Relay-Remaining-Useful-Actuation (EMRUA) pipeline is introduced. The pipeline performs a set of sequential steps to estimate the EMR's Remaining Useful Actuation (RUA) at any point during the components life. For the sake of clarity - rather than measuring the remaining time to failure as RUL [115, 116] - this work is concerned with the estimation of the number of RUA which refers to the number of EMR making and breaking actuation left until an EMR has failed. A failure can occur due to one of the multiple reasons discussed in Chapter 3.2.2 ⁵.

Figure 5.19 provides an overview of EMRUA. In general, two major stages are distinguished:

1. **During training and testing:** Suitable input feature representations extracted from the raw data are determined; a range of sub-sequence selection strategies to sample from MVTD are considered; TCN-model combinations for RUA predictions are tuned (the aim is to minimise the loss between target RUA-sequence and estimated RUA-sequence).
2. **During inference:** Once the best model has been determined, RUA estimations under consideration of uncertainty can be performed.

5.2.6.1 Data extraction

MVTD snippets are recorded during EMR switching. Together, each contact making and the consecutive contact breaking comprises one single actuation A . For each actuation A_j from c sensors, a set of sensor signals $X = [X_1, X_2, \dots, X_c]$ is recorded. The i^{th} signal is $X_i = [x_0, x_1, x_2, \dots, x_s]$ of length s , i.e. samples for the contact making and breaking respectively. Figure 5.19-(I) illustrates the process of data extraction for the CV and CI within EMRUA from recorded EMR making and breaking actuations. While EMRUA utilises only CV and CI waveforms, the principal architecture is not limited to these two input signals. The properties of CV and CI waveforms for normally open EMR contacts have been discussed in Chapter 3.2.5.2 and Chapter 3.2.5.1.

⁵Parts of this chapter have already been published in my journal or conference article: L. Kirschbaum, et al., (2022).

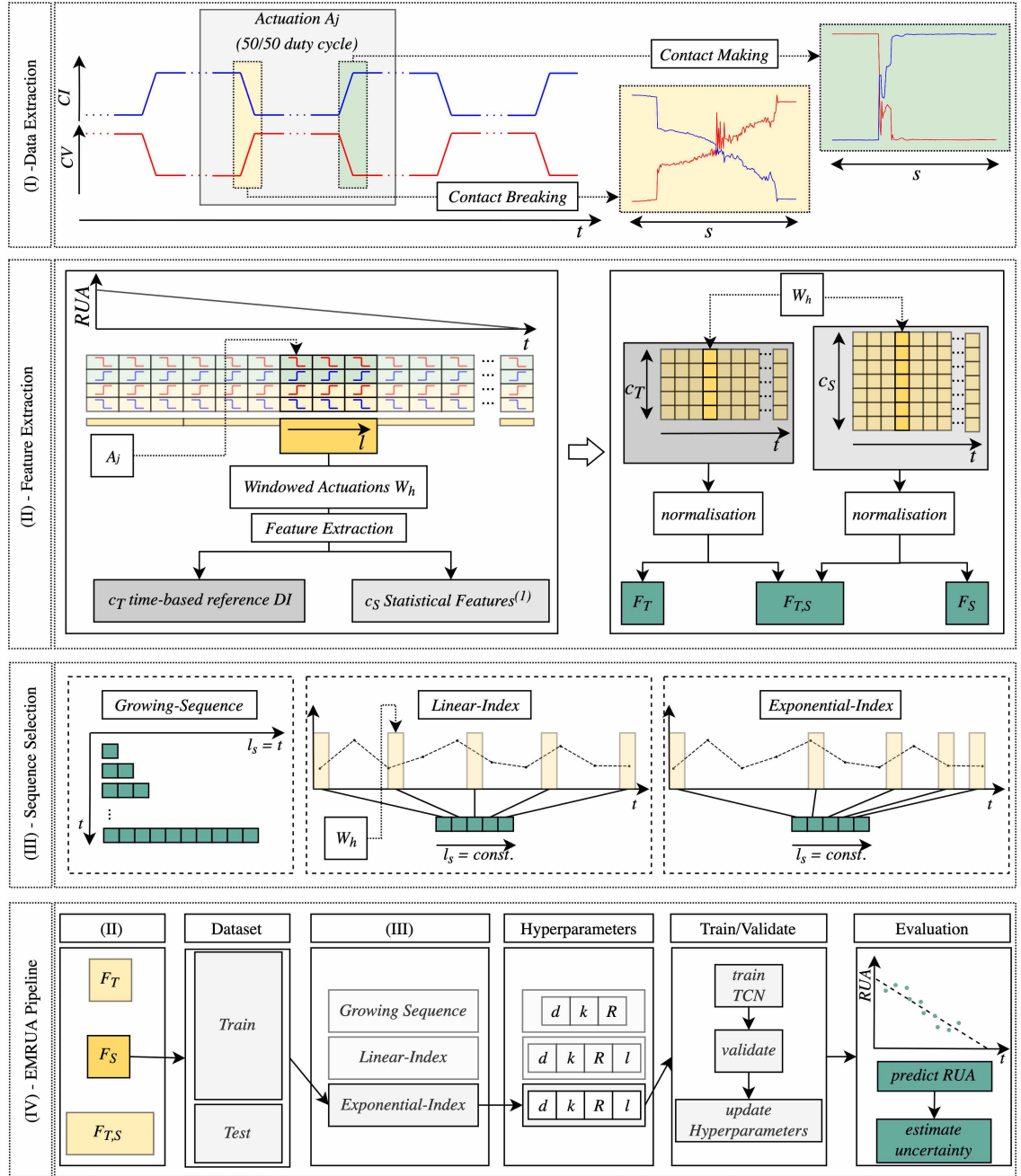


Figure 5.19: (I) The sampling strategy for CV and CI from the EMR opening and closing waveform during switching; (II) the process of feature extraction and preparation as TCN input feature map; (III) schematic display of the three sequence selection strategies GI, LI, and EI; (IV) the EMRUA pipeline to learn the mapping from CV and CI waveforms to the RUA target, cf. Chapter 5.2.6.3 and Chapter 5.2.6.4. ⁽¹⁾ - the extracted statistical features are detailed in Chapter 7.3.

5.2.6.2 Feature extraction

The feature extraction process follows a set of consecutive steps, schematically illustrated in Figure 5.19-(II). The aim is to derive features which depict the underlying degradation process that is common among the population of sampled EMRs. A set of features is derived for l actuations, where each actuations comprises the making CV, the making CI, the breaking CV, and the breaking CI. The resulting features are somewhat similar to down-sampled representations of the original waveforms A_j . Subsequently, each feature in the group A_j is averaged providing the feature vector W_h . Two sets of derived features are distinguished being time-based reference DIs denoted as feature-set F_T consisting of c_T features, e.g. the BT, AT, RT, etc.; and a feature-set F_S composed of c_S features representing a combination of amplitude-based reference DIs and statistical features such as the *mean*, *max*, *min*, etc. extracted from CV and CI waveforms. Normalisation of the derived feature sets is performed to make them suitable inputs for the TCN. Therefore, each feature is scaled to a range of $[0, 1]$, as per Equation 5.14,

$$F_i^{norm} = \frac{F_i - \min(F_i)}{\max(F_i) - \min(F_i)} \quad (5.14)$$

where F_i denotes the i^{th} feature of the feature set F . In addition to F_T and F_S a combination being $F_{T,S} = [F_T, F_S]$ is also considered as a data set. The changes in the waveforms due to deterioration of the EMR's contacts from switching under load are extensively discussed in the Chapter 7.2. The respective sets of features are introduced in greater detail in Chapter 7.3.

5.2.6.3 Sequence selection

As elaborated, TCN performs causal-convolution over the input feature map F , to estimate a RUA target sequence of the same length. As the kernel is shared across the entire input sequence, the number of trainable parameters in the model is independent of the sequence length l_s . It only depends on the number of considered input features c , the number of filters f , the kernel-size k , and the number of residual blocks R . However, different strategies can be employed to select a subset of representative windows W from the interval $T = [0, t]$, i.e. in order to sub-sample the waveform. The different processes of considered sequence selection strategies are pictured in Figure 5.19-(III).

Growing sequence indexing From an input sequence of length $l_s = t$, Growing-Sequence Indexing (GI) considers every single making and breaking actuation for the entire interval $T = [0, t]$. Hence, the sequence will grow over the course of the life of the EMR until the EOL. In practice this poses a challenge as the extension

of the sequence length needs to be considered during the model training stage. As input to the model serves a randomly selected batch B containing b examples which may vary in their length l_s , but remain constant in regard to the number of features c , cf. [117]. However, all sequences within one training batch need to have the same length. Hence, all b randomly selected sequences are padded to the maximum sequence length $l_{s(max)}$ encountered in the batch B , as displayed in Figure 5.20. A sequence ranges always from the first switching actuation to an actuation at time t relative to the EOL of the respective sample.

Linear sequence indexing In the case of Linear Indexing (LI), the entire degradation sequence is equally considered. $l_s \in \mathbb{Z}^+$ actuations are selected, evenly spaced in the interval of $T = [0, t]$. Hence, even as t increases, l_s remains constant. The selection of l_s actuations can be expressed as

$$A = \{A_0, A_{\frac{t}{l_s-1}}, \dots, A_t\} \quad (5.15)$$

Exponential sequence indexing Contrary to Section 5.2.6.3, this Exponential Indexing (EI) selection strategy favours recent actuations because l_s actuations are selected in the interval $T = [0, t]$, expressed as

$$A = \{A_0, A_{t-(\frac{t^2}{l_s-1})^2}, \dots, A_t\} \quad (5.16)$$

5.2.6.4 RUA estimation and uncertainty

RUA behaves linear - consider the actuations which have *passed* at any t against the actuations still *left*. Equation 5.17 defines RUA for the i^{th} actuation a_i ,

$$RUA(a_i) = a_{EOL} - a_i \quad \text{where} \quad a_i \leq a_{EOL}, a_i \in \mathbb{Z}^+ \quad (5.17)$$

However, since TCN is capable of mapping each input $X_{[1,i]}$ to an output $Y_{[1,i]}$, the problem at hand is concerned with a *sequence to sequence* mapping task. Thus, RUA can be imagined as a vector, cf. Equation 5.18.

$$RUA(a_{[1,i]}) = a_{EOL} - \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_i \end{bmatrix} \quad (5.18)$$

As evidenced in Chapter 5.2.5, PHM methods employing TCN are yet to address uncertainty. The remainder of this section is, therefore, concerned with the integra-

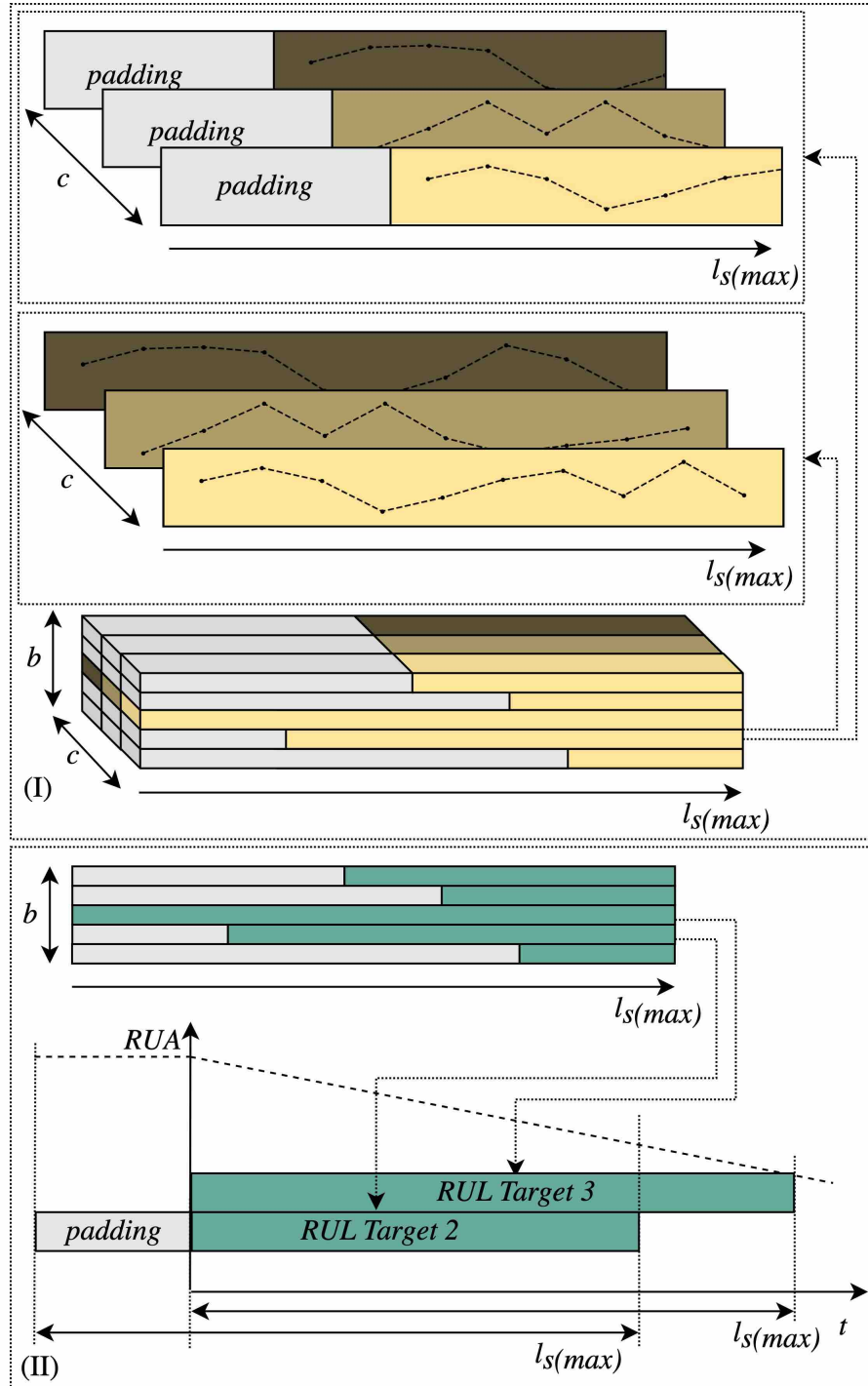


Figure 5.20: (I) The input batch B of batch-size $b = 5$ with a maximum sequence length $l_s(max)$ and features $c = 3$; (II) the RUA target sequence for each instance in the input batch.

tion of a method to estimate uncertainty into a TCN-based prognostics architecture that can be used during the inference phase. Monte Carlo Dropout (MCD) can be applied in order to determine uncertainty during inference in DL. MCD has been proposed by [118]. It approximates the Bayesian Gaussian Process providing a highly computational efficient solution for DL to estimate the posterior distribution [119]. Note, the effect of the model's input dimension on the computational complexity is an important aspect to consider [116]. The number of used features during inference has a negligible effect, however, the length of the input time series might be of concern in terms of computational efficiency. MCD achieves uncertainty estimation by utilising dropout during inference on the trained NN model f which results in a different prediction of $Y_t = [y_0, y_1, \dots, y_t]$ for an input sequence $X_t = [x_1, x_2, \dots, x_t]$ at each forward pass through f , since the dropout-mask δ^i is selected at random. This yields

$$Y_t^i = f(X_t|\delta^i) + \epsilon \quad (5.19)$$

where $\epsilon \sim \mathcal{N}(\mu, \sigma^2)$ represents the Gaussian distributed process noise stemming from measurement errors. The distribution derived from averaging over N forward passes is somewhat similar to an ensemble of N trained models. Consequently, one can compare this to estimating a distribution of the learned weights per layer that can be approximated using a relatively small number of forward passes, i.e. $N \leq 1000$ [120]. To facilitate MCD, [121] suggests to employ dropout after each layer. [119] point out it is important that the dropout rate is kept constant and not tuned along other hyperparameters during training.

Making use of a batch B with the size b during inference, the RUA for N forward passes can be efficiently estimated in parallel so that $b = N$, where each feature map X_t in B represent the same input. This yields the total input $B_t = [X_t^0, X_t^1, \dots, X_t^b]$. The posterior distribution of Y_t is then estimated in parallel, since the dropout mask $\Delta_N = [\delta_0, \delta_1, \dots, \delta_b]$ for each X_t^i is chosen at random across the input B_t . Therefore, Equation 5.19 can be amended to

$$Y_t^N = f(B_t|\Delta_N) + \epsilon \quad (5.20)$$

The process is illustrated in Figure 5.21. It is evident that DL methods deploying convolution are especially well suited NN architectures to integrate MCD in order to quantify uncertainty due to the parallel computing capabilities exhibited by those architecture. However, one should be careful with a combination of very long sequences and a large N , since this requires significant memory overhead because of the size of the resulting input feature map.

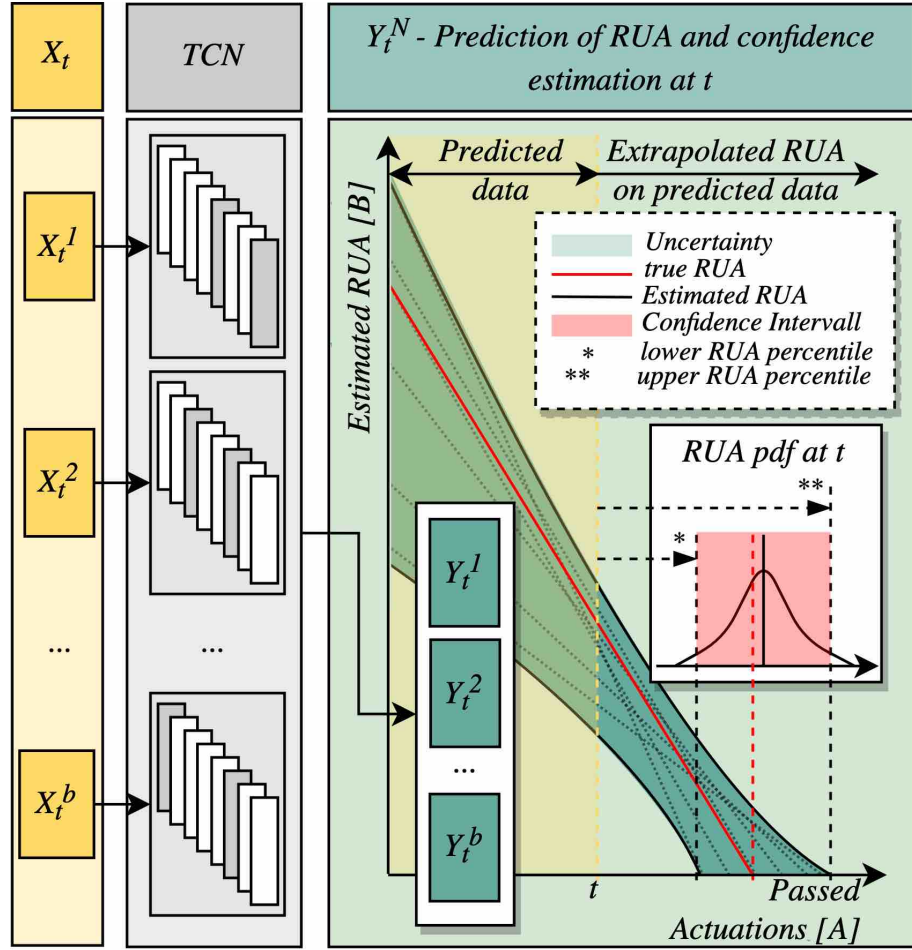


Figure 5.21: Parallel sampling of RUA from an input X_t using a batch B with b different dropout masks during inference, predicting Y_t^N posterior RUA sequences.

The RUA is estimated from Y_t^N , as summarised in Figure 5.21. The best linear fit for each predicted RUA sequence Y_t^i is determined reducing the residual sum of squares. The *mean* as in Equation 5.21 and the *variance* as in Equation 5.22 for each a_i is then calculated from the N extrapolated linear RUA trajectories (ranging from a_0 till the a_{EOL}) to derive a confidence interval. Within this interval the model is 95% confident that the true mean of the population is contained, i.e. the true RUA.

$$\mu(a_i) = \frac{1}{N} \sum_{j=0}^N m_j(a_i) + b_j \quad (5.21)$$

$$\sigma(a_i)^2 = \frac{\sum_{j=0}^N ((m_j(a_i) + b_j) - \mu(a_i))^2}{(N - 1)} \quad (5.22)$$

5.2.7 Model configuration and scoring

An overview of the model hyperparameters is provided in Table 5.3. The model consists of a number of stacked residual blocks, depending on the final width of the receptive field r ; followed by a final FC layer using linear activation. Dropout is employed within each residual block. Model performance is evaluated using Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE) where RUA_i denotes the target and RUA_i^* the estimated RUA.

$$MAE(RUA_i^*, RUA_i) = \frac{1}{N} \sum_{i=0}^N |RUA_i^* - RUA_i| \quad (5.23)$$

$$MAPE(RUA_i^*, RUA_i) = \frac{100\%}{N} \sum_{i=0}^N \frac{|RUA_i^* - RUA_i|}{RUA_i} \quad (5.24)$$

$$RMSE(RUA_i^*, RUA_i) = \sqrt{\frac{1}{N} \sum_{i=0}^N (RUA_i^* - RUA_i)^2} \quad (5.25)$$

In addition to the RMSE, a scoring metric proposed by [122] is adopted to evaluate the performance under consideration of the model's estimated uncertainty, cf. for a prognostics use case [97, 123]. An accuracy zone α is defined that provides bounds of allowed deviation from the targeted RUA. This zone is limited by an upper threshold α^+ and a lower threshold α^- . α is selected according to the needs of the specific application. Here, $\alpha = 0.2$ is chosen for evaluating the EMRUA. The α_{sc} is calculated by counting the frequency of RUA estimates within the α^\pm bounds, cf. [122, 124].

Table 5.3: Evaluated TCN model hyperparameters.

Hyperparameter		Value
Filters	f	64
Kernel	k	[3, 5, 7]
Dilation base	d_b	[2, 3, 4]
Residual blocks	R_n	[6, 7, 8]
Dropout rate	-	0.5
Activation	-	ReLU
Optimisation	-	Adam
Learning Rate	-	0.0001
Batch size	b	[400, 200, 100]
FC-Layer Units	-	1
FC-Layer Activation	-	Linear

Summary

This chapter has detailed the developed methodologies for the two case studies this dissertation is concerned with, namely the BHA-PCBA maintenance support framework and the EMR prognostics pipeline. Both approaches rely on AI, but differ in terms of the selected method, their prognostics output, and the depth of implementation. For the BHA-PCBA the choice of method, i.e. RFC or XGBoost Classifier, is justified by the complexity of the monitored systems, unknown sensor placement, failure modes, and a low resolution as well as availability of monitoring data. For the second use case, i.e. the EMR-RUA estimation, the high volume of MVTD that needs to be processed (high sampling rates over extended periods of time) is the determining selection criterion that motivates the use of an appropriate DL model, i.e. TCN combined with an efficient implementation of MCD. The following Chapter 6 discusses the ALT-EMR design. Hereinafter, the results of the respective case studies are presented in Chapter 7.

References

- [1] I Goodfellow, J Shlens, and C Szegedy. Explaining and harnessing adversarial examples. *arXiv preprint*, 2014. arXiv:1412.6572.
- [2] C Szegedy, W Zaremba, I Sutskever, J Bruna, D Erhan, I Goodfellow, and R Fergus. Intriguing properties of neural networks. *arXiv preprint*, 2013. arXiv:1312.6199.
- [3] A Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.
- [4] L Rokach and O Maimon. Decision trees. In *Data mining and knowledge discovery handbook*, pages 165–192. Springer, 2005. doi:10.1007/0-387-25465-X_9.
- [5] A Myles, R Feudale, Y Liu, N Woody, and S Brown. An introduction to decision tree modeling. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 18(6):275–285, 2004. doi:10.1002/cem.873.
- [6] H Laurent and R Rivest. Constructing optimal binary decision trees is NP-complete. *Information processing letters*, 5(1):15–17, 1976. doi:10.1016/0020-0190(76)90095-8.

- [7] Y Song and L Ying. Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry*, 27(2):130, 2015. doi:10.11919/j.issn.1002-0829.21504.
- [8] W Mohamed, M Salleh, and A Omar. A comparative study of reduced error pruning method in decision tree algorithms. In *2012 IEEE International conference on control system, computing and engineering*, pages 392–397. IEEE, 2012. doi:10.1109/ICCSC.2012.6487177.
- [9] L Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001. doi:10.1023/A:1010933404324.
- [10] T Dietterich. Ensemble methods in machine learning. In *International workshop on multiple classifier systems*, pages 1–15. Springer, 2000. doi:10.1007/3-540-45014-9_1.
- [11] B Mahesh. Machine learning algorithms - a review. *International Journal of Science and Research (IJSR)*, 9:381–386, 2020. doi:10.21275/ART20203995.
- [12] L Breiman. Random forest. *Machine Learning*, 45:5–32, 2001. doi:10.1023/A:1010933404324.
- [13] L Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996. doi:10.1007/BF00058655.
- [14] G Biau and E Scornet. A random forest guided tour. *Test*, 25(2):197–227, 2016. doi:10.1007/s11749-016-0481-7.
- [15] T Chen and C Guestrin. Xgboost: A scalable tree boosting system. In *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. ACM, 2016. doi:10.1145/2939672.2939785.
- [16] Y Freund and R Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997. doi:10.1006/jcss.1997.1504.
- [17] T Chen, T He, M Benesty, V Khotilovich, Y Tang, and H Cho. Xgboost: extreme gradient boosting. *R package version 0.4-2*, 1(4):1–4, 2015.
- [18] R Caruana and A Niculescu-Mizil. An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd International Conference on Machine learning*, pages 161–168. ACM, 2006. doi:10.1145/1143844.1143865.
- [19] P Probst, M Wright, and A Boulesteix. Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(3):e1301, 2019. doi:10.1002/widm.1301.
- [20] J Bergstra and Y Bengio. Random search for hyper-parameter optimization. *Journal of machine learning research*, 13(2), 2012.
- [21] C Ferri, J Hernández-Orallo, and R Modroiu. An experimental comparison of performance measures for classification. *Pattern Recognition Letters*, 30(1):27–38, 2009. doi:10.1016/j.patrec.2008.08.010.
- [22] Jin Huang and Charles X Ling. Using auc and accuracy in evaluating learning algorithms. *IEEE Transactions on knowledge and Data Engineering*, 17(3):299–310, 2005. doi:10.1109/TKDE.2005.50.
- [23] Y Wu, M Yuan, S Dong, L Lin, and Y Liu. Remaining useful life estimation of engineered systems using vanilla LSTM neural networks. *Neurocomputing*, 275:167–179, 2018. doi:10.1016/j.neucom.2017.05.063.
- [24] R Zhao, R Yan, J Wang, and K Mao. Learning to monitor machine health with convolutional bi-directional LSTM networks. *Sensors*, 17(2):273, 2017. doi:10.3390/s17020273.
- [25] C Su, L Tsai, S Huang, and Y Li. Deep learning-based real-time failure detection of storage devices. In *International Conference on Applied Human Factors and Ergonomics*, pages 160–168. Springer, 2019. doi:10.1007/978-3-030-20454-9_16.
- [26] O Fink, Q Wang, M Svensen, P Dersin, W Lee, and M Ducoffe. Potential, challenges and future directions for deep learning in prognostics and health management applications. *Engineering Applications of Artificial Intelligence*, 92:103678, 2020. doi:10.1016/j.engappai.2020.103678.
- [27] G Garcia, G Michau, M Ducoffe, J Gupta, and O Fink. Temporal signals to images: Monitoring the condition of industrial assets with deep learning image processing algorithms. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 2021. doi:10.1177/1748006X21994446.
- [28] T Le, M Vo, B Vo, E Hwang, S Rho, and S Baik. Improving electric energy consumption prediction using CNN and Bi-LSTM. *Applied Sciences*, 9(20):4237, 2019. doi:10.3390/app9204237.

- [29] Z Khan, T Hussain, A Ullah, S Rho, M Lee, and S Baik. Towards efficient electricity forecasting in residential and commercial buildings: A novel hybrid CNN with a LSTM-AE based framework. *Sensors*, 20(5):1399, 2020. doi:10.3390/s20051399.
- [30] X Shao, C Kim, and P Sontakke. Accurate deep model for electricity consumption forecasting using multi-channel and multi-scale feature fusion CNN-LSTM. *Energies*, 13(8):1881, 2020. doi:10.3390/en13081881.
- [31] L Deng and D Yu. Deep learning: methods and applications. *Foundations and trends in signal processing*, 7(3–4):197–387, 2014. doi:10.1561/20000000039.
- [32] R Zhao, R Yan, Z Chen, K Mao, P Wang, and R Gao. Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing*, 115:213–237, 2019. doi:10.1016/j.ymssp.2018.05.050.
- [33] S Khan and T Yairi. A review on the application of deep learning in system health management. *Mechanical Systems and Signal Processing*, 107:241–265, 2018. doi:10.1016/j.ymssp.2017.11.024.
- [34] G Biscardini, D Branson, C Del Maestro Adrian, K Hawsey, N Molavi, and R Morrison. Oil and gas trends 2019-building growth strategies on shifting sands. *Proc. 22nd Annu. Global CEO Surv. Trends*, pages 1–22, 2019.
- [35] I Goodfellow, Y Bengio, and A Courville. *Deep learning*, volume 1. MIT press Cambridge, 2016.
- [36] D Kingma and J Ba. Adam: A method for stochastic optimization. *arXiv preprint*, 2014. arXiv:1412.6980.
- [37] Y LeCun, B Boser, J Denker, D Henderson, R Howard, W Hubbard, and L Jackel. Hand-written digit recognition with a back-propagation network. *Advances in neural information processing systems*, 2, 1989. NIPS.
- [38] Y LeCun, L Bottou, Y Bengio, and P Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. doi:0.1109/5.726791.
- [39] K He, G Gkioxari, P Dollár, and R Girshick. Mask R-CNN. *arXiv preprint*, 2017. arXiv:1703.06870.
- [40] A Dhillon and G Verma. Convolutional neural network: a review of models, methodologies and applications to object detection. *Progress in Artificial Intelligence*, 9(2):85–112, 2020. doi:10.1007/s13748-019-00203-0.
- [41] A Odo, S McKenna, D Flynn, and J Vorstius. Towards the automatic visual monitoring of electricity pylons from aerial images. In *15th International Conference on Computer Vision Theory and Applications*, pages 566–573. SciTePress, 2020.
- [42] X Gibert, V Patel, and R Chellappa. Deep multitask learning for railway track inspection. *IEEE transactions on intelligent transportation systems*, 18(1):153–164, 2016. doi:10.1109/TITS.2016.2568758.
- [43] O Janssens, V Slavkovikj, B Vervisch, K Stockman, M Loccufer, S Verstockt, R van de Walle, and S van Hoecke. Convolutional neural network based fault detection for rotating machinery. *Journal of Sound and Vibration*, 377:331–345, 2016. doi:10.1016/j.jsv.2016.05.027.
- [44] X Guo, L Chen, and C Shen. Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. *Measurement*, 93:490–502, 2016. doi:10.1016/j.measurement.2016.07.054.
- [45] A Krizhevsky, I Sutskever, and G Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012. doi:10.1145/3065386.
- [46] C Szegedy, W Liu, Y Jia, P Sermanet, S Reed, D Anguelov, D Erhan, V Vanhoucke, and A Rabinovich. Going deeper with convolutions. In *IEEE conference on computer vision and pattern recognition*, pages 1–9. IEEE, 2015. doi:10.1109/CVPR.2015.7298594.
- [47] K He, K Zhang, S Ren, and J Sun. Deep residual learning for image recognition. *arXiv preprint*, 2015. arXiv:1512.03385.
- [48] Y LeCun and G Bengio, Yand Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015. doi:10.1038/nature14539.
- [49] J Masci, U Meier, D Cireşan, and J Schmidhuber. Stacked convolutional auto-encoders for hierarchical feature extraction. In *International conference on artificial neural networks*, pages 52–59. Springer, 2011. doi:10.1007/978-3-642-21735-7_7.
- [50] D Bank, N Koenigstein, and R Giryes. Autoencoders. *arXiv preprint*, 2020. arXiv:2003.05991.
- [51] A Graves, A Mohamed, and G. Hinton. Speech recognition with deep recurrent neural networks. *arXiv preprint*, 2013. arXiv:1303.5778.

- [52] A Sherstinsky. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena*, 404:132306, 2020. doi:10.1016/j.physd.2019.132306.
- [53] S Hochreiter and J Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997. doi:10.1162/neco.1997.9.8.1735.
- [54] Alex Sherstinsky. Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network. *Physica D: Nonlinear Phenomena*, 404:132306, 2020. doi:10.1016/j.physd.2019.132306.
- [55] S Bowman, G Angeli, Cr Potts, and C Manning. A large annotated corpus for learning natural language inference. *arXiv preprint*, 2015. arXiv:1508.05326.
- [56] F Gers, J Schmidhuber, and F Cummins. Learning to forget: Continual prediction with LSTM. *Neural computation*, 12(10):2451–2471, 2000. doi:10.1162/089976600300015015.
- [57] A Graves and J Schmidhuber. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural networks*, 18(5-6):602–610, 2005. doi:10.1016/j.neunet.2005.06.042.
- [58] Y Wang, M Huang, X Zhu, and L Zhao. Attention-based LSTM for aspect-level sentiment classification. In *Conference on empirical methods in natural language processing*, pages 606–615. ACL Anthology, 2016. doi:10.18653/v1/D16-1058.
- [59] K Cho, B Van Merriënboer, C Gulcehre, D Bahdanau, F Bougares, H Schwenk, and Y Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint*, 2014. arXiv::1406.1078.
- [60] James Douglas Hamilton. *Time series analysis*. Princeton university press, 2020.
- [61] G Babu, P Zhao, and X Li. Deep convolutional neural network based regression approach for estimation of remaining useful life. In *International conference on database systems for advanced applications*, pages 214–228. Springer, 2016. doi:10.1007/978-3-319-32025-0_14.
- [62] L Jing, T Wang, M Zhao, and P Wang. An adaptive multi-sensor data fusion method based on deep convolutional neural networks for fault diagnosis of planetary gearbox. *Sensors*, 17(2):414, 2017. doi:10.3390/s17020414.
- [63] D Song, N Xia, W Cheng, H Chen, and D Tao. Deep r-th root of rank supervised joint binary embedding for multivariate time series retrieval. In *24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, page 2229–2238. ACM, 2018. doi:10.1145/3219819.3220108.
- [64] C Zhang, D Song, Y Chen, X Feng, C Lumenazu, W Cheng, J Ni, B Zong, H Chen, and N. Chawla. A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data. *arXiv preprint*, 2018. arXiv:1811.08055.
- [65] Z Wang and T Oates. Imaging time-series to improve classification and imputation. *arXiv preprint*, 2015. arXiv:1506.00327.
- [66] N Hatami, Y Gavet, and J Debayle. Classification of time-series images using deep convolutional neural networks. In *Tenth international conference on machine vision (ICMV)*, volume 10696, page 106960Y. International Society for Optics and Photonics, 2018. doi:10.1117/12.2309486.
- [67] M Canizo, I Triguero, A Conde, and E Onieva. Multi-head CNN-RNN for multi-time series anomaly detection: An industrial case study. *Neurocomputing*, 363:246 – 260, 2019. doi:10.1016/j.neucom.2019.07.034.
- [68] M Munir, S Siddiqui, A Dengel, and S Ahmed. DeepAnT: A deep learning approach for unsupervised anomaly detection in time series. *IEEE Access*, 2018. doi:10.1109/ACCESS.2018.2886457.
- [69] K Wang, Y Zhao, Q Xiong, M Fan, G Sun, L Ma, and T Liu. Research on healthy anomaly detection model based on deep learning from multiple time-series physiological signals. *Scientific Programming*, 2016, 2016. doi:10.1155/2016/5642856.
- [70] S Koelstra, C Muhl, M Soleymani, J Lee, A Yazdani, T Ebrahimi, T Pun, A Nijholt, and I Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 3(1):18–31, 2011. doi:10.1109/T-AFFC.2011.15.
- [71] H Ren, B Xu, Y Wang, C Yi, C Huang, X Kou, T Xing, M Yang, J Tong, and Q Zhang. Time-series anomaly detection service at Microsoft. *arXiv preprint*, 2019. arXiv:1906.03821.
- [72] T Wen and R Keyes. Time series anomaly detection using convolutional neural networks and transfer learning. *arXiv preprint*, 2019. arXiv:1905.13628.
- [73] J Hutchins. Dodgers loop sensor data set. *UCI Machine Learning Repository*, 2006.

- [74] J Gao, X Song, Q Wen, P Wang, L Sun, and H Xu. Robusttad: Robust time series anomaly detection via decomposition and convolutional neural networks. *arXiv preprint*, 2020. arXiv:2002.09545.
- [75] Y Zheng, Q Liu, E Chen, Y Ge, and J Zhao. Time series classification using multi-channels deep convolutional neural networks. In *International conference on web-age information management*, pages 298–310. Springer, 2014. doi:10.1007/978-3-319-08010-9_33.
- [76] Y Zheng, Q Liu, E Chen, Y Ge, and J Zhao. Exploiting multi-channels deep convolutional neural networks for multivariate time series classification. *Frontiers of Computer Science*, 10(1):96–112, 2016. doi:10.1007/s11704-015-4478-2.
- [77] A Reiss. UCI machine learning repository: PAMAP2 physical activity monitoring data set, 2016.
- [78] J Yang, M Nguyen, P San, X Li, and S Krishnaswamy. Deep convolutional neural networks on multichannel time series for human activity recognition. In *Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI’15*, page 3995–4001. AAAI Press, 2015. doi:10.5555/2832747.2832806.
- [79] D Roggen, A Calatroni, M Rossi, T Holleczeck, K Förster, G Tröster, P Lukowicz, D Bannach, G Pirkel, and A Ferscha. Collecting complex activity datasets in highly rich networked sensor environments. In *2010 Seventh international conference on networked sensing systems (INSS)*, pages 233–240. IEEE, 2010. doi:10.1109/INSS.2010.5573462.
- [80] H Dau, A Bagnall, K Kamgar, C Yeh, Y Zhu, S Gharghabi, C Ratanamahatana, and E Keogh. The UCR time series archive. *IEEE/CAA Journal of Automatica Sinica*, 6(6):1293–1305, 2019. doi:10.1109/JAS.2019.1911747.
- [81] Z Cui, W Chen, and Y Chen. Multi-scale convolutional neural networks for time series classification. *arXiv preprint*, 2016. arXiv:1603.06995.
- [82] Z Wang, W Yan, and T Oates. Time series classification from scratch with deep neural networks: A strong baseline. In *International joint conference on neural networks (IJCNN)*, pages 1578–1585. IEEE, 2017. doi:10.1109/IJCNN.2017.7966039.
- [83] F Karim, S Majumdar, H Darabi, and S Chen. LSTM fully convolutional networks for time series classification. *IEEE Access*, 6:1662–1669, 2017. doi:10.1109/ACCESS.2017.2779939.
- [84] Y Liao, X Zeng, and W Li. Wavelet transform based convolutional neural network for gearbox fault classification. In *IEEE Prognostics and System Health Management Conference (PHM-Harbin)*, pages 1–6. IEEE, 2017. doi:10.1109/PHM.2017.8079274.
- [85] D Verstraete, A Ferrada, E Droguett, V Meruane, and M Modarres. Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings. *Shock and Vibration*, 2017:1–17, 2017. doi:10.1155/2017/5067651.
- [86] E Bechhoefer. A quick introduction to bearing envelope analysis. *Green Power Monit. Syst*, 2016.
- [87] X Ding and Q He. Energy-fluctuated multiscale feature learning with Deep ConvNet for intelligent spindle bearing fault diagnosis. *IEEE Transactions on Instrumentation and Measurement*, 66:1926–1935, 2017. doi:10.1109/TIM.2017.2674738.
- [88] W Zhang, G Peng, and C Li. Bearings fault diagnosis based on convolutional neural networks with 2-d representation of vibration signals as input. In *MATEC web of conferences*, volume 95, page 13001. EDP Sciences, 2017. doi:10.1051/mateconf/20179513001.
- [89] P Rajpurkar, A Hannun, M Haghpanahi, C Bourn, and A Ng. Cardiologist-level arrhythmia detection with convolutional neural networks. *arXiv preprint*, 2017. arXiv:1707.01836.
- [90] C Liu, W Hsaio, and Y Tu. Time series classification with multivariate convolutional neural network. *IEEE Transactions on Industrial Electronics*, 66(6):4788–4797, 2018. doi:10.1109/TIE.2018.2864702.
- [91] Y Chen and H Rangwala. Attention-based multi-task learning for sensor analytics. In *IEEE International Conference on Big Data (Big Data)*, pages 2187–2196. IEEE, 2019. doi:10.1109/BigData47090.2019.9006207.
- [92] G Lai, W Chang, Y Yang, and H Liu. Modeling long-and short-term temporal patterns with deep neural networks. In *41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 95–104. ACM, 2018.
- [93] J Cheng, K Huang, and Z Zheng. Towards better forecasting by fusing near and distant future visions. In *AAAI Conference on Artificial Intelligence*, volume 34, pages 3593–3600. AAAI Press, 2020.
- [94] R Wan, S Mei, J Wang, M Liu, and F Yang. Multivariate temporal convolutional network: A deep neural networks approach for multivariate time series forecasting. *Electronics*, 8(8):876, 2019. doi:10.3390/electronics8080876.

- [95] L Kirschbaum, D Roman, G Singh, J Bruns, V Robu, and D Flynn. AI-driven maintenance support for downhole tools and electronics operated in dynamic drilling environments. *IEEE Access*, 8:78683–78701, 2020. doi:10.1109/ACCESS.2020.2990152.
- [96] M Pecht and M Kang. *Prognostics and health management of electronics: Fundamentals, machine learning, and the internet of things*. John Wiley and Sons, Ltd., 2018. doi:10.1002/9781119515326.
- [97] D Roman, S Saxena, V Robu, M Pecht, and D Flynn. Machine learning pipeline for battery state-of-health estimation. *Nature Machine Intelligence*, 3(5):447–456, 2021. doi:10.1038/s42256-021-00312-3.
- [98] D Roman, S Saxena, J Bruns, R Valentin, M Pecht, and D Flynn. A machine learning degradation model for electrochemical capacitors operated at high temperature. *IEEE Access*, 9:25544–25553, 2021. doi:10.1109/ACCESS.2021.3057959.
- [99] M Xia, T Li, T Shu, J Wan, C De Silva, and Z Wang. A two-stage approach for the remaining useful life prediction of bearings using deep neural networks. *IEEE Transactions on Industrial Informatics*, 15(6):3703–3711, 2018. doi:10.1109/TII.2018.2868687.
- [100] S Bai, J Kolter, and V Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint*, 2018. arxiv:1803.01271.
- [101] A van den Oord, S Dieleman, H Zen, K Simonyan, O Vinyals, A Graves, N Kalchbrenner, A Senior, and K Kavukcuoglu. Wavenet: A generative model for raw audio. *arXiv preprint*, 2016. arXiv:1609.03499.
- [102] C Pelletier, G Webb, and F Petitjean. Temporal convolutional neural network for the classification of satellite image time series. *Remote Sensing*, 11(5):523, 2019. doi:10.3390/rs11050523.
- [103] M Nan, M Trăscău, A Florea, and C Iacob. Comparison between recurrent networks and temporal convolutional networks approaches for skeleton-based action recognition. *Sensors*, 21(6):2051, 2021. doi:10.3390/s21062051.
- [104] D Zhou, Z Li, J Zhu, H Zhang, and L Hou. State of health monitoring and remaining useful life prediction of lithium-ion batteries based on temporal convolutional network. *IEEE Access*, 8:53307–53320, 2020. doi:10.1109/ACCESS.2020.2981261.
- [105] K He, X Zhang, S Ren, and J Sun. Deep residual learning for image recognition. In *IEEE conference on computer vision and pattern recognition*, pages 770–778. IEEE, 2016. doi:10.1109/CVPR.2016.90.
- [106] J Wenqiang, C Jian, and C Yi. Remaining useful life prediction for mechanical equipment based on temporal convolutional network. In *14th IEEE International Conference on Electronic Measurement & Instruments (ICEMI)*, pages 1192–1199. IEEE, 2019. doi:10.1109/ICEMI46757.2019.9101706.
- [107] A Saxena and K Goebel. Turbofan engine degradation simulation data set. *NASA Ames Prognostics Data Repository*, pages 1551–3203, 2008.
- [108] Y Song, S Gao, Y Li, L Jia, Q Li, and F Pang. Distributed attention-based temporal convolutional network for remaining useful life prediction. *IEEE Internet of Things Journal*, 2020. doi:10.1109/JIOT.2020.3004452.
- [109] J Chen, D Chen, and G Liu. Using temporal convolution network for remaining useful lifetime prediction. *Engineering Reports*, 3(3):e12305, 2021. doi:10.1002/eng2.12305.
- [110] C Liu, L Zhang, and C Wu. Direct remaining useful life prediction for rolling bearing using temporal convolutional networks. In *IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 2965–2971. IEEE, 2019. doi:10.1109/SSCI44817.2019.9003163.
- [111] W Yang, Q Yao, K Ye, and C Xu. Empirical mode decomposition and temporal convolutional networks for remaining useful life estimation. *International Journal of Parallel Programming*, 48(1):61–79, 2020. doi:10.1007/s10766-019-00650-1.
- [112] P Nectoux, R Gouriveau, K Medjaher, E Ramasso, B Chebel-Morello, N Zerhouni, and C Varnier. PRONOSTIA: An experimental platform for bearings accelerated degradation tests. In *IEEE International Conference on Prognostics and Health Management*, pages 1–8. IEEE, 2012.
- [113] B Saha, K Goebel, S Poll, and J Christophersen. Prognostics methods for battery health monitoring using a Bayesian framework. *IEEE Transactions on instrumentation and measurement*, 58(2):291–296, 2008. doi:10.1109/TIM.2008.2005965.
- [114] H Wang, M Peng, R Xu, A Ayodeji, and H Xia. Remaining useful life prediction based on improved temporal convolutional network for nuclear power plant valves. *Frontiers in Energy Research*, 8:296, 2020. doi:10.3389/fenrg.2020.584463.

- [115] M Pecht. IEEE standard framework for prognostics and health management of electronic systems. Standard, IEEE Standards Association, NJ, USA, 2017. doi:10.1109/IEEESTD.2017.8227036.
- [116] K Goebel, M Daigle, A Saxena, I Roychoudhury, S Sankararaman, and J Celaya. *Prognostics: The science of making predictions*. CreateSpace Independent Publishing Platform, 2017.
- [117] S Smith, P Kindermans, C Ying, and Q Le. Don’t decay the learning rate, increase the batch size. *arXiv preprint*, 2017. arXiv:1711.00489.
- [118] Y Gal and Z Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *International conference on machine learning*, pages 1050–1059. MLR press, 2016.
- [119] B Lakshminarayanan, A Pritzel, and C Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *arXiv preprint*, 2016. arXiv:1612.01474.
- [120] H Nguyen, J Liu, and E Zio. A long-term prediction approach based on long short-term memory neural networks with automatic parameter optimization by tree-structured Parzen estimator and applied to time-series data of NPP steam generators. *Applied Soft Computing*, 89:106116, 2020. doi:10.1016/j.asoc.2020.106116.
- [121] R Seoh. Qualitative analysis of Monte Carlo Dropout. *arXiv preprint*, 2020. arXiv:2007.01720.
- [122] A Saxena, J Celaya, B Saha, S Saha, and K Goebel. Metrics for offline evaluation of prognostic performance. *International Journal of Prognostics and health management*, 1(1):4–23, 2010. doi:10.36001/ijphm.2010.v1i1.1336.
- [123] Y Lei, N Li, L Guo, N Li, T Yan, and J Lin. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical systems and signal processing*, 104:799–834, 2018. doi:10.1016/j.ymssp.2017.11.016.
- [124] V Kuleshov, N Fenner, and S Ermon. Accurate uncertainties for deep learning using calibrated regression. In *International Conference on Machine Learning*, pages 2796–2804. MLR press, 2018.

Chapter 6

ALT-EMR design

Introduction

To demonstrate the merits of an EMR prognostics solution the lack of relevant life-cycle data constitutes a major obstacle, as has been pointed out before in Chapter 4.2.3. To address this shortcoming, as part of my dissertational project an ALT-EMR platform has been developed. Its objective is the aggregation of high resolution life-cycle data from different sensing sources. The analysis of these data allows tracing of the EMR deterioration processes. Throughout the development and testing phase of the ALT some alterations had to be realised to accommodate challenges encountered with the initial design. For example, in order to reliably capture life-cycle data at very high rates continuously over a long duration, considerable changes in the selection, design, and operation of hardware as well as software were made.

This chapter opens with a review of established methods for ALT-EMR in Chapter 6.1, since such considerations have informed the initial design of the life-cycle platform. The characteristics of the tested EMR are presented in Chapter 6.2. A description of the experimental platform is given in Chapter 6.3. Here, the design and selection of the testing hardware, the implementation of the software control, and data acquisition procedures are discussed in detail. The chapter concludes with an overview of the experiments' specifics and a summary of the aggregated data ^{1 2}.

6.1 Review of testing methodologies

As previously stated, the contacts of EMRs are most liable to failure, since the mechanical life of the EMR is significantly higher [1]. Hence, the EMR life is heav-

¹*Parts of this chapter have already been published in my journal or conference article: L. Kirschbaum, et al., (2021) and L. Kirschbaum, et al., (2022).*

²At the time of completion of this PhD thesis, the life cycle data is not published. However, publication is planned and will be announced on <https://smartsystems.hw.ac.uk/>.

ily impacted by the contact life - to be more precise by the electrical life of the contacts. Therefore, the focus of the proposed life-cycle experiments are contact related failure modes. In particular, when testing EMR it is important not to induce uncharacteristic failure modes through poor choice of test-parameters. Hence, a common approach to test only the contacts is a *model switch*, i.e. a device which allows to precisely control the switching parameters [2, 3]. However, whilst yielding results applicable to the contact material, carefully tuning such replica to match the properties of an EMR under test is challenging [1]. Factors such as closing velocity of the contacts or contact bounce need to be considered.

Despite the fact that alterations in CR are not necessarily reliable indicators for EMR wear, cf. Chapter 4.2.4, it remains a common measure to judge EMR performance. [4, 5] specify dry-contact measurements preferable at low test currents and voltages of less than 80 mV. Dry-contact measurements might not return accurate CR measurements if the contacts are under load. Fritting, cf. Chapter 3.2.4, can be the cause for significantly higher CR readings. In regard to contact testing [1] points out, although it is industry standard to test at maximum rated specifications, no assurance can be given that the degradation behaviour of electrical contacts at lower loads will be similar. For example, testing using high loads will initially circumvent problems like contact film contamination encountered only at lower loads.

In order to accelerate the general degradation or one particular degradation type, one can explore different means. The most apparent approach is an increase in switching frequency under the assumption that this does not alter the overall degradation behaviour of the contacts and the EMR. Alternatively, an increase of the contacts stress through elevated current levels can be considered. Testing under elevated ambient temperature might promote degradation effects like contact corrosion or the possibility of coil failure [6]. Multiple factors influence the EMR degradation. As pointed out in Chapter 3.2.4, the severity of electrical arcing and, thus, the contact degradation not only depends on the EMR design. Other factors have to be considered, e.g. the electric circuit in which the EMR is placed. Likewise, external factors such as ambient temperature and atmosphere should be taken into account. While many more aspects could be discussed [1], Table 6.1 focuses on predominant considerations when designing an EMR-ALT experiment.

Table 6.1: Aspects to consider when testing EMR.

Factor	Description
Power	
AC	In AC application the contacts constantly change polarity. Hence, the anodic and subsequent cathodic erosion lead to a comparable net erosion on both contacts.

Continued on next page

Table 6.1 – continued from previous page

Factor	Description
DC	Contrary, the polarity of the contacts in DC applications does not change. Hence, typically the majority of erosion is encountered on the anode (craters). The cathode gains material (pips). Refer to Chapter 3.2.4.1.
Load	
Lamp (filament)	A high inrush current and variations of contact bounce accelerate contact erosion.
Inductive	In DC circuits inductive loads can cause a current lag of up to $7ms$ [7]. This leads to longer lasting arcs during <i>break</i> , while no arcing and therefore no erosion takes place during <i>make</i> due to the current lag. Effects of bouncing might become irrelevant, refer to Chapter 3.2.4.2.
Resistive	Purely resistive loads have the smallest effect on contact erosion as inductive effects are mitigated, though arcing during bouncing has to be considered.
Contact Bounce	Contact bounce can lead to contact welding. The frequency of short unstable welds typically increases towards the EOL as the contact force and the over-travel are being reduced [8]. Notice, in DC circuits the contribution of the one-sided material loss is problematic. The frequency of the re-bounce effects the rate of erosion. High-frequent, short-gap bouncing accounts for the highest material transfer rates [1].
Arcing Suppression	
Protective Gas	Testing contacts in inert gases or nitrogen will increase the contact life and mitigate contact contamination effects.
Blowout Magnet	For DC a permanent magnetic field next to the contacts can be introduced into the EMR design to reduce effects of arcing when switching an inductive load as the arc is drawn away from the contact dispersing heat [9, 7]. Hence, such contacts are subjected to reduced rates of electrical erosion.
Sealing	
Enclosed	The contact life might be considerably longer if contacts are tested in a plastic enclosure typically found as a sealing enclosure in many EMR applications[10]. The out-gassing of the plastic leads to a contamination of the contact surfaces through deposited particles and the arc moves across the contact surface at each operation. This in turn leads to an even degradation of the entire surface. Contact life can be increased through adsorption effects from <i>getter-materials</i> used to reduce contact degradation through film-formation adsorbing harmful molecules [7].
Unsealed	The arc will more likely develop a static anode and cathode fall region, accelerating the degradation process at these spots and leading to earlier failures. Effects of contact activation and reaction of the contact material with the ambient air should be taken into account.
Switching Frequency	The switching frequency depends on the operate and release properties of the EMR, subject to the contact mass and velocity. In the presence of arcing, heat dissipation from the contacts should be considered, in order to not falsify the failure mode.

In DC testing, the effects of switching the anode and cathode among the contacts in the test setup may alter the degradation pattern, since cost-efficient consumer-EMRs tend to have a thicker stationary contact carrier and a thinner movable contact

Table 6.2: Specifications of the selected EMR under test.

Parameter	Specification
Coil Rating	5 VDC, 105 mA
Operate/ Release V	at max. 70 %/ min. 15 %
Operate/ Release t	15 ms/ 5 ms
Contacts Rating	30 VDC, 10 A (resistive load)
Contact Resistance	30 m Ω
Contact Type	Plated Copper Rivets
Contact Plating	$AgSnO_2In_2O_3$
Max. op. Temp.	70°C
Casing	Flux-Protection/ Sealing

carrier, i.e. realised as an integrated spring. Therefore, the moving contact carrier may experience severe heating during operation as it is much thinner. This can accelerate the deterioration in an unintended way. Ultimately, the thinner, moving spring contact carrier might melt because of the increased heat build-up. It should be stressed that such failure mode needs to be carefully examined, since its occurrence greatly depends on the design and dimension of the contact-carriers. To reduce the heat build-up while testing in a DC setting, the experimental design can specify the static contact carrier side as anode [1].

6.2 Selected test component

While EMRs are available in many different designs and contact arrangements, the choice for the EMR under test was greatly motivated by the EMRs availability as well as its simplicity in order to be able to facilitate a subsequent failure analysis. Hence, a general-purpose Single Pole Single Throw (SPST)-Normally-Open (NO) EMR was selected (an unsealed and sealed variant with the same specification), cf. Table 6.2.

The selected EMR comprises a coil, an armature which is pulled down once the coil is energised and snaps back once de-energised, and two contact carriers of which one is *static* and one is *moving*. The latter one is designed as contact spring retracting the contact when the coil is de-energised, cf. Figure 6.1. The maximum operating conditions depend on the type of load and whether AC or DC is switched. For the developed test stand the maximum specified operating conditions of a resistive DC load are considered for the selection of an appropriate test load, cf. Figure 6.2.

6.3 Life-cycle platform

This section introduces the ALT-EMR platform. The experiment was designed from the outset as an iterative process and has yielded two distinctly different data sets. Data obtained during the first run is termed Data Set 1 (DA1); the second

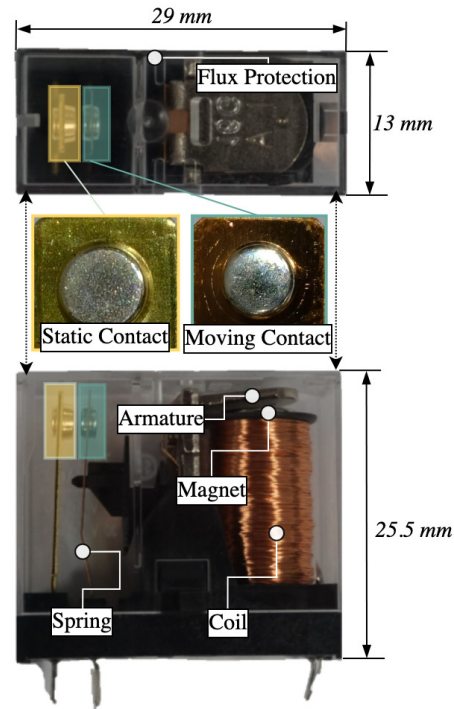
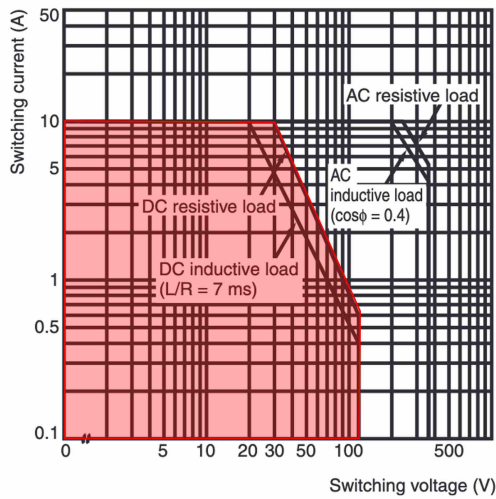
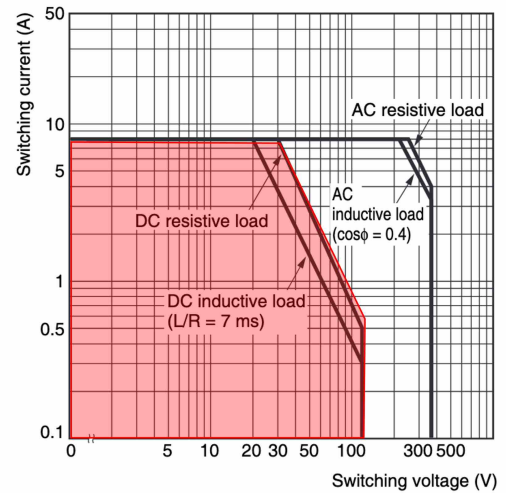


Figure 6.1: The selected test EMR (unsealed variant), top- and side-view of coil and moving armature.



(a) Operational threshold of the unsealed test EMR, adopted from data sheet [11]; the red-shaded area corresponds with the allowed DC resistive load.



(b) Operational threshold of the sealed test EMR, adopted from data sheet [11]; the red-shaded area corresponds with the allowed DC resistive load.

Figure 6.2: EMR test thresholds.

run produced Data Set 2 (DA2). Table 6.3 provides a detailed description of the experimental setup.

- **DA1:**
 - Purpose: Full coverage of EMR life-cycle data
 - EMR Type: 16 unsealed EMRs
 - ALT-Mode: High temperature ($60\text{ }^{\circ}\text{C}$), low switching frequency (0.25 Hz)
 - Sampling: Each actuation at 25 kHz
 - Measurements: CV, CI
- **DA2:**
 - Purpose: Analysis of EMR failure modes and mechanisms
 - EMR type: 4 unsealed EMRs and 10 sealed EMRs
 - ALT-Mode: Low temperature ($30\text{ }^{\circ}\text{C}$), high switching frequency (20 Hz)
 - Sampling: Every 1000th actuation at 150 kHz
 - Measurements: CV, CI, CC, CR

The remainder of this chapter details the experimental process and the collected data sets.

6.3.1 Design and hardware selection



Figure 6.3: The lab and the life-cycle test platform. On the left, the oven can be seen in which the EMRs have been tested. A port on the side of the oven allows to connect the test equipment and sensors with the test stand and the DAQ, shown on the right side.

Various sources of hardware are required, i.e. recording and control equipment, sensors, electrical loads, power supplies, and an oven. The laboratory setup of the

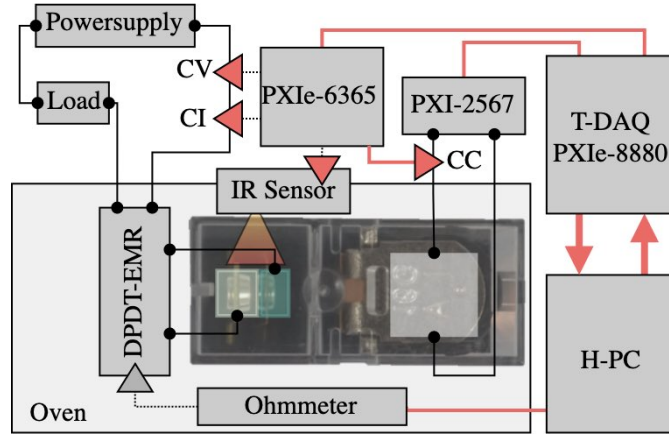


Figure 6.4: Simplified overview of the experimental platform.

experiment is shown in Figure 6.3. The central component of the experimental setup is the *deterministic* National Instruments (NI) PXIe-8880 controller³. As detailed in Figure 6.4, this module, termed Data Acquisition Unit (DAQ), controls the data acquisition and the switching of the tested EMR. LabVIEW-RealTime is deployed as operating system. In order to log data and monitor the experiment the DAQ is interfaced with an independent *host* Control-PC, termed Host PC Unit (H-PC), via a network Ethernet connection. In combination with the DAQ, a NI PXI-2567 module is used to precisely trigger switching of the test EMR according to a defined switching frequency and duty-cycle. Experimental data is collected using the multi-input NI PXIe-6365 module. Depending on the data set two or three measurements are taken at high resolution: Contact Voltage (CV), Contact Current (CI) and Coil Current (CC). CC is directly recorded whilst the amplitude of the CV signal is reduced using a voltage divider to scale the signal into an interval of ± 10 V due to the limited input range of the measurement hardware. As per specification for NI PXIe-6365 for the full ± 10 V range the accuracy of the measurements is ± 0.0166 %. Measurements are set up as *differential-referenced*, i.e. connected to the DAQ ground using two bias resistors for each channel. CI is recorded with a hall-effect current sensor (LEM HX 03-P/SP2) outputting a voltage equivalent of $[2.5, 5]$ V for the range of $[0, 9]$ A. An external power supply (AimTTI CPX400DP) and two parallel, variable resistors set up with opposing winding to reduce effects of load inductance are used to simulate a resistive test-load. The load is continuously cooled using two fans. The power supply is connected via serial RS-232 to the experimental setup in order to control voltage and current while providing the option of a safe, automated shutdown.

A PCB has been developed to house the necessary sensors and components and to facilitate the exchange of the EMR under test, once it has failed, cf. Appendix A.1.

³In this case, *determinism* refers to software determinism, i.e. hardware that can execute a software loop at very low rates of jitter.

The PCB is placed in an oven in order to control the ambient temperature during the experiment. For reference the ambient temperature in the oven is monitored using a type-K thermocouple. In addition to the ambient temperature, an infrared temperature sensor (CTL-Fast) is placed behind the EMR's static contact rivet pointing at the EMR enclosure. The temperature of the enclosure is measured on a 0.7 mm wide black surface spot. To facilitate CR-measurements, a secondary circuit can be switched in using a Double Pole Double Throw (DPDT)-EMR positioned as close as possible to the EMR under test. This circuit is connected to the 4-wire measurement setup of a Milli-Ohmmeter (GOM-805) which is in turn directly linked to the H-PC via a serial RS-232 connection. As increases of CR reside in the m Ω range, a measurement range resolution of 500 m Ω is selected with an accuracy of ± 0.05 %. The selected Ohmmeter facilitates different CR measurement modes.

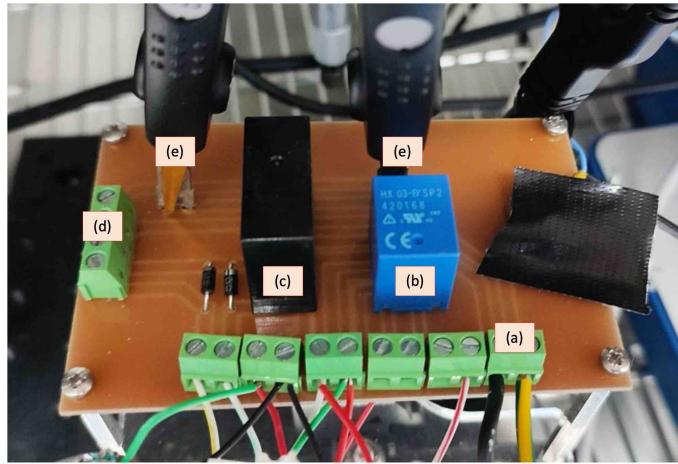


Figure 6.5: The PCB (length 100 mm, width 60 mm) is situated in the oven and facilitates the exchange of the EMR test samples. Further sensors, terminals, and connection points to the Ohmmeter *4-wire* measurement setup are integrated. The infrared temperature sensor can be adjusted to point directly at the EMR enclosure. (a) connection to the test-load; (b) current sensor; (c) DPDT-EMR to switch in the CR circuit; (d) connection to the test EMR adaptor board on which the test sample is soldered onto; (e) terminals and clamps of the *4-wire* setup.

6.3.2 Control and data acquisition

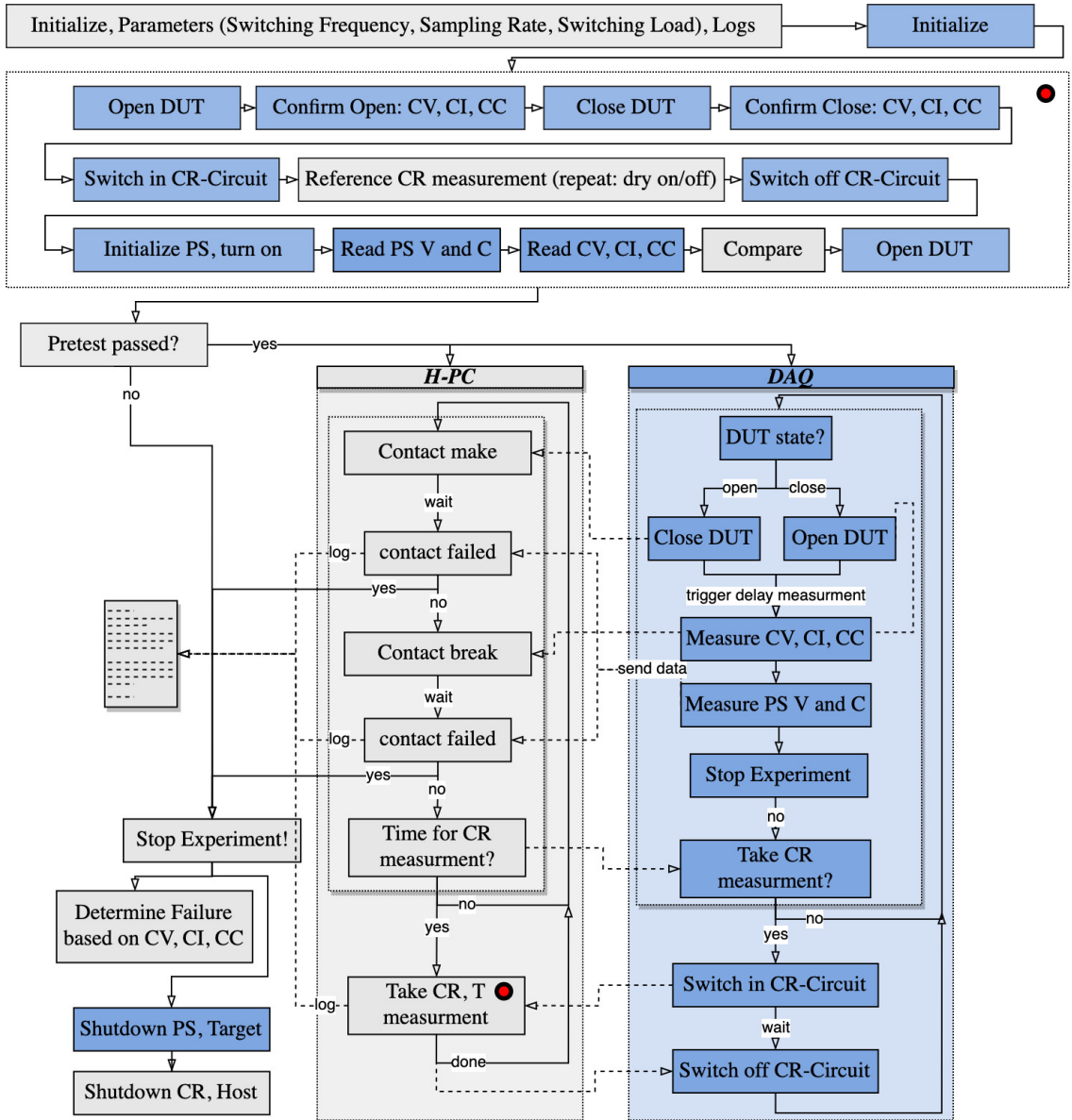


Figure 6.6: Simplified schematic of the experiment (test EMR - Device Under Test (DUT); Power Supply (PS)). The H-PC is used to initialise all parameters and forward them to the DAQ. The EMR switching and data acquisition is triggered by the NI PXIe-8880 and its sub-modules. Data is send for logging and analysis to the H-PC. Further the H-PC controls the switch-in of the CR measurement circuit.

The test platform and all connected hardware is monitored via NI LabVIEW deployed on the H-PC. Life-cycle experiments are launched through the User-Interface, cf. Appendix A.2. Each experiment is initialised based on a set of pre-configured test parameters. First, an automatic pre-test of all components is conducted; if successful, the subsequent EMR switching and sampling software-loops on the H-PC and the DAQ are initiated, cf. Appendix A.3 and Appendix A.4. Sampling starts prior to and ends after each making and breaking actuation respectively. The interval in-between actuations is not recorded. The EMR actuation itself is very fast (< 1 ms), with significant delay between the switching trigger and PT during mak-

ing and RT during breaking. Control commands are broadcasted from the H-PC to the DAQ; experimental data is streamed from the DAQ using a buffered queue. The DAQ incorporates all switching functionality, measurements and control of external components such as the power supply. Acquired data is logged in the *.hdf5* file format, cf. the simplified experiment workflow in Figure 6.6.

In the case of DA1, CV and CI are recorded at high resolution for each actuation. Likewise, supplementary measurements of the ambient temperature are taken at lower resolution. In the case of DA2, CV, CI, and CC high resolution measurements are recorded in combination with CR, ambient temperature, and EMR enclosure temperature measurements at regular intervals. To facilitate CR measurements switching of the contacts is paused, the EMR under test is then closed. The DPDT-EMR - controlled via a secondary channel on the NI PXI-2567 - switches from the test-circuit to the CR-measurement-circuit. The Ohmmeter then takes several readings. Simultaneously, using the infrared temperature sensor the EMR casing temperature is recorded in close proximity to the EMR contacts.

6.3.3 Specifics of the experiments conducted

Table 6.3: Comparison of the two sets of life-cycle experiments performed. (*) Logged waveforms every 10^{th} actuation just prior to failure.

Experiment	DA1	DA2
Parameters		
Tested EMR	Unsealed	Sealed & Unsealed
Sample size	16	14
Mode of ALT	Temperature	Switching frequency
Temperature	60 °C	30 °C
Switching frequency	0.25 Hz	20 Hz
Duty Cycle	50/50 %	50/50 %
Load	30 VDC/ 6 A	35 VDC/ 6 A
Load type	resistive (cooled)	resistive (cooled)
Measurements		
Trigger Signal	Analog voltage read back	Digital trigger (timestamped)
Coil Current (CC)	-	every 1000 th actuation at 150 kHz (*); resistor
Contact Voltage (CV)	every actuation at 25 kHz; voltage divider	every 1000 th actuation at 150 kHz (*); voltage divider
Contact Current (CI)	every actuation at 25 kHz; current sensor	every 1000 th actuation at 150 kHz (*); current sensor
CR measurements	-	Dry-/Full-mode every 10000 th actuations
Contact Temperature	-	Infrared every 10000 th actuations
Ambient temperature	Thermocouple K-Type	Thermocouple K-Type
Additional Sensors	Acoustic Emission, Vibration	-
Logging		
Continued on next page		

Table 6.3 – continued from previous page

Experiment	DA1	DA2
Data format	raw <i>.txt</i> files converted to <i>.hdf5</i>	<i>.hdf5</i> , supplementary measurements of CR as <i>.csv</i>
Data structure	File-Keys (Actuations)- Keys (Open/Closed)- Keys (data/t0)- data ([Samples,Sensors]), t0 (timestamp)	Files (breaking/making)- Keys (Actuations)- data ([Samples,Sensors]), attributes (timestamp,dt)

The major differences and similarities regarding the two sets of experiments conducted are highlighted in Table 6.3. Foremost, in DA1 only unsealed EMRs have tested, whereas DA2 comprised sealed and unsealed components. In DA1 acceleration of the EMR deterioration has been achieved through testing at an elevated ambient temperature. However, upon conclusion of the experiment DA1 it was found that the rate of contact erosion was predominantly affected by the number of conducted switching operations until failure. Therefore, to stimulate a realistic degradation pattern, the mode of acceleration in DA2 was altered from elevated ambient temperature to an increased switching frequency. Regarding the type of measurements taken, CV, CI, and CC measurements were recorded at higher resolution in DA2 compared to DA1, but at lower rate. In addition, DA2 aggregated CR measurements in order to allow for a direct comparison with results reported in the literature.

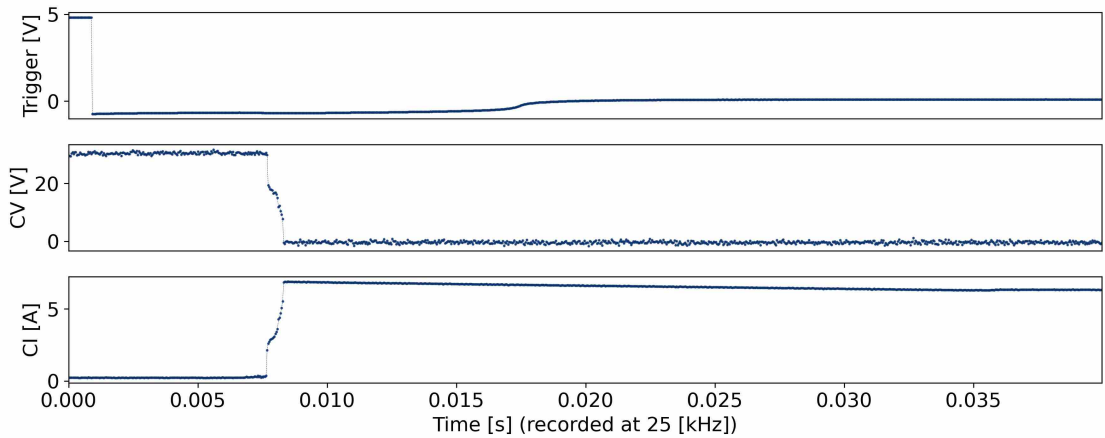


Figure 6.7: DA1 - Trigger Signal, CV, and CI measurements of the making actuation sampled over an interval of 0.04 s. The contacts start closing after a delay of 8 ms with an actual closing duration of 1 ms.

In the case of DA1, the trigger signal which actuates the EMR by energising the coil, the CV, and the CI have been sampled at 25 kHz for a window of 0.04 s yielding 1000 measurement points for each making and breaking operation respectively, cf. Figure 6.7 and 6.9. Due to the high resolution of measurements in DA2, sampling

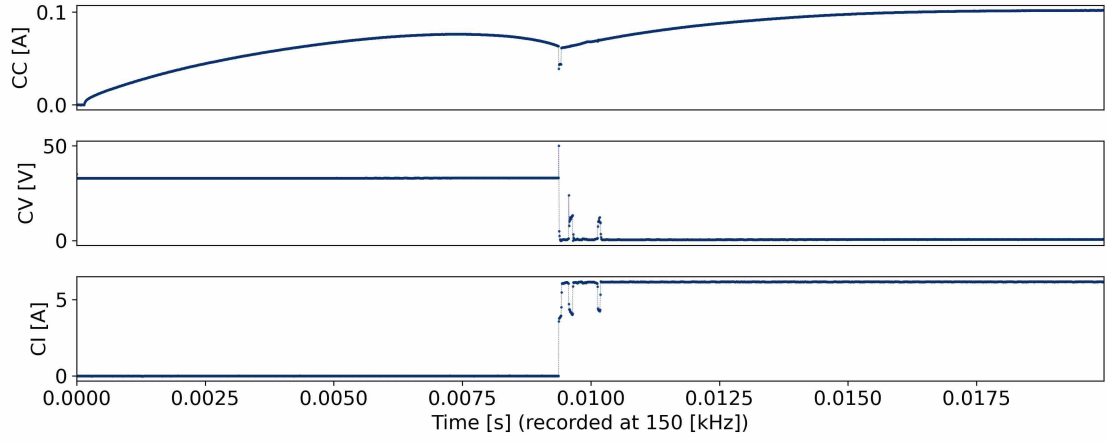


Figure 6.8: DA2 - CC, CV, and CI. As the coil is being energised the contact armature begins to move and closes completely at 0.009 s which is indicated by the inductive nick of the CC waveform. The higher sampling rate in comparison to DA1 provides a more detailed picture of the closing process showing two distinct bounces just after the initial *make* at 0.009 s and 0.011 s.

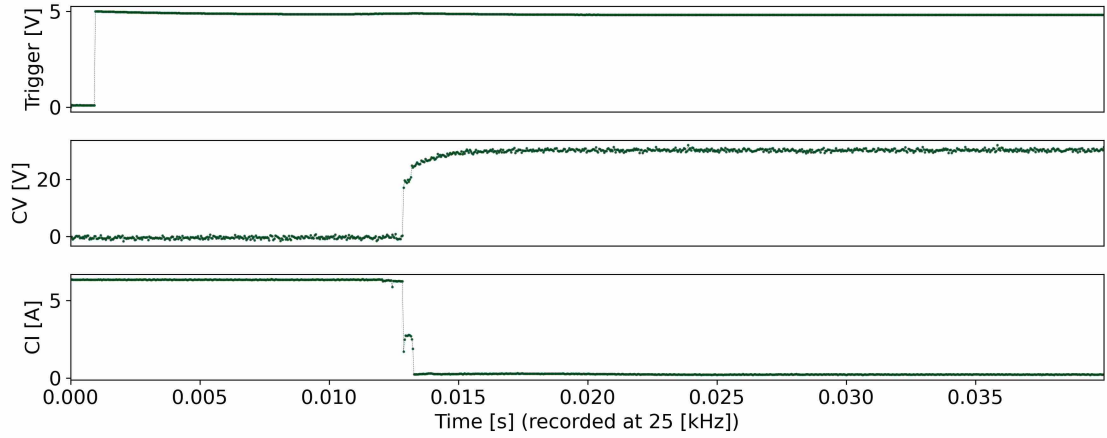


Figure 6.9: DA1 - Trigger Signal, CV, and CI measurement of the breaking actuation. The delay between the initial coil de-energisation and the opening of the contacts is ± 10 ms.

did not start with the actual trigger of the EMR actuation because the initialisation of any recording channel is affiliated with a certain degree of jitter at this rate. Thus, recording commenced with the actual energisation of the coil, cf. Figure 6.8 and 6.10. The sampling rate is set to 150 kHz, sampled over a window of 0.02 ms which yields a higher resolution of the signals CC, CV, and CI with 3000 measurement points each for the making and the breaking actuation operation respectively. However, only every 1000th making and breaking actuation is logged. In addition, the closed-circuit voltage v_{cc} and open-circuit voltage v_{oc} as further detailed in Chapter 7.2 have been logged for every actuation only in DA2.

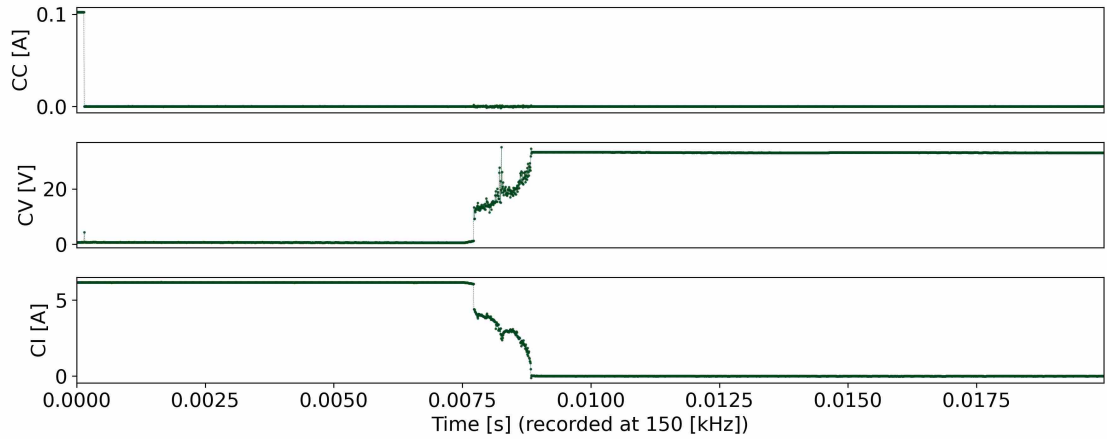


Figure 6.10: DA2 - CC, CV, and CI providing a higher resolution signal of the contact breaking lasting from 0.0075 s to 0.009 s.

6.3.4 Data set DA1

As for DA1, life-cycle experiments have been conducted at 30 VDC, 6 A and elevated ambient temperature (60 °C). Overall, 16 unsealed EMR samples have been tested, cf. Table 6.4 and Appendix A.5. All tested EMRs failed stuck-closed.

Table 6.4: Overview of the recorded data in DA1.

Sample	Type	Actuations	Time [hr]	Failure
DA1-01	Unsealed	94190	104.7	Stuck-Closed
DA1-02	Unsealed	120739	134.5	Stuck-Closed
DA1-03	Unsealed	100868	112.1	Stuck-Closed
DA1-04	Unsealed	109970	122.8	Stuck-Closed
DA1-05	Unsealed	98968	110.0	Stuck-Closed
DA1-06	Unsealed	174264	193.6	Stuck-Closed
DA1-07	Unsealed	133251	148.1	Stuck-Closed
DA1-08	Unsealed	162220	180.3	Stuck-Closed
DA1-09	Unsealed	79826	88.7	Stuck-Closed
DA1-10	Unsealed	77438	86.0	Stuck-Closed
DA1-11	Unsealed	91543	101.8	Stuck-Closed
DA1-12	Unsealed	102729	114.1	Stuck-Closed
DA1-13	Unsealed	80014	88.9	Stuck-Closed
DA1-14	Unsealed	198747	220.8	Stuck-Closed
DA1-15	Unsealed	78968	87.7	Stuck-Closed
DA1-16	Unsealed	113246	125.8	Stuck-Closed

6.3.5 Data set DA2

Ten sealed and four unsealed EMRs were tested for DA2, cf. Table 6.5 and Appendix and A.6. Unlike DA1, the tests were conducted at higher switching frequency but lower ambient temperature (30 °C). This significantly affected the EMR life. The mean time to failure increased to approximately four million actuations. This is equivalent to a test duration of 1 – 2 weeks per EMR. Because of excessive contact erosion, nine out of ten of the EMRs failed stuck-open. The unsealed EMRs

predominantly failed stuck-closed due to contact welding.

Table 6.5: Overview of the recorded data in DA2. The time to failure is not provided because the CR measurements have been automatically taken every 10000th actuation which took several seconds at which the experiment was halted.

Sample	Type	Actuations	Failure
DA2-01S	Sealed	2550135	Stuck-Open
DA2-02S	Sealed	3494924	Stuck-Open
DA2-03S	Sealed	2836161	Stuck-Open
DA2-04S	Sealed	3683048	Stuck-Open
DA2-05U	Unsealed	3417789	Stuck-Open
DA2-06S	Sealed	4350630	Stuck-Open
DA2-07S	Sealed	3556590	Stuck-Open
DA2-08S	Sealed	4283131	Stuck-Open
DA2-09U	Unsealed	6194520	Stuck-Closed
DA2-10S	Sealed	3075211	Stuck-Closed
DA2-11S	Sealed	3483071	Stuck-Open
DA2-12S	Sealed	2455181	Stuck-Open
DA2-13U	Unsealed	5574044	Stuck-Closed
DA2-14U	Unsealed	7460384	Stuck-Closed

Summary

The central objective of this chapter has been the summary of the experimental workflow that was used to generate ALT-EMR data sets. The specification of the designed test rig, including the hardware and software components were discussed. The experiments yielded two sets of data that are suitable for different purposes. In Chapter 7.2 the analysis of the failure mechanisms acting on the EMR contacts is presented utilising the high resolution CV and CI waveform recordings obtained at 150 kHz at each 1000th actuation, i.e. the data set DA2. Chapter 7.3 makes use of data set DA1 to evaluate the performance of the EMRUA pipeline previously introduced in Chapter 5.2.

References

- [1] P Slade. *Electrical contacts: principles and applications*. CRC press, 2017.
- [2] Z Chen and G Witter. A study of dynamic welding of electrical contacts with emphasis on the effects of oxide content for silver tin indium oxide contacts. In *56th IEEE Holm Conference on Electrical Contacts*, pages 1–6. IEEE, 2010. doi:10.1109/HOLM.2010.5619552.
- [3] T Mutzel, M Bender, and R Niederreuther. The effect of material composition on dynamic welding of electrical contacts. In *59th IEEE Holm Conference on Electrical Contacts*, pages 1–7. IEEE, 2013. doi:10.1109/HOLM.2013.6651432.
- [4] ICS. Aircraft — electromagnetic relays and contactors. Standard, International Organization for Standardization, Geneva, CH, 1996.
- [5] IEC. Electromechanical elementary relays - part 1: General and safety requirements. Standard, International Electrotechnical Commission, Geneva, CH, 2015.
- [6] J Liu, M Zhang, N Zhao, and A Chen. A reliability assessment method for high speed train electromagnetic relays. *Energies*, 11(3):652, 2018. doi:10.3390/en11030652.
- [7] H Sauer. *Modern Relay Technology*. Huethig, 1986.

- [8] Z Chen and K Sawa. Effect of arc behavior on material transfer: A review. In *42nd IEEE Holm Conference on Electrical Contacts*, pages 238–251. IEEE, 1996. doi:10.1109/HOLM.1996.557203.
- [9] W Godsey. Arc suppression for relay contacts in dc service. *IRE Transactions on Component Parts*, 4(2):36–42, 1957. doi:10.1109/TCP.1957.1135909.
- [10] G Witter and L Polevoy. Contact erosion and material transfer for contacts in automotive relays. In *42nd IEEE Holm Conference on Electrical Contacts*, pages 223–228. IEEE, 1996. doi:10.1109/HOLM.1996.557201.
- [11] Omron G2R PCB Power Relay. https://omronfs.omron.com/en_US/ecb/products/pdf/en-g2r.pdf, 2022. Accessed: 2022-15-01.

Chapter 7

Results and analysis

Introduction

This chapter presents the results for the BHA-PCBA case study in Chapter 7.1, analyses the observed EMR failure modes and mechanisms based upon DA2 in Chapter 7.2, and details the results obtained when deploying EMRUA on DA1 in Chapter 7.3. With respect to the two case studies set up in this dissertation, cf. Chapter 5, the results presented in this Chapter exhibit prominent differences based on the data environment and the prognostic response returned.

7.1 BHA-PCBA maintenance support framework

The following section describes and examines the results based on a BHA-PCBA data set provided for this research, cf. Chapter 5.1.2 ¹.

7.1.1 Model accuracy

As elaborated in Chapter 5.1, the training data set is augmented using adversarial examples. Various perturbation factors have been studied, namely the original training data $\epsilon = 0.00$ and the following augmented data sets $\epsilon = 0.02$, $\epsilon = 0.05$, $\epsilon = 0.1$. A visualisation of the effects on the feature space X_n through such augmentation is depicted using the correlation heatmap in Figure 7.1. The RFC ² and

¹*Parts of this chapter have already been published in my journal or conference article: L. Kirschbaum, et al., (2020).*

²The `python sci-kit learn` implementation is used for the RFC, i.e. `sklearn.ensemble.RandomForestClassifier` (version 0.22.2) [1]. The model `rfc` is configured as follows: the Gini Impurity is used to determine the best splits `rfc.criterion = 'gini'`; `rfc.bootstrap = True`; `rfc.n_estimators` search-space `[10, 300]`; `rfc.max_depth` search-space `[2, 110]`; `rfc.max_features` search space `['log2', 'sqrt']`; `rfc.min_samples_split` search-space `[2, 5, 10]`; `rfc.min_samples_leaf` search-space `[1, 2, 4]`; all other parameters *default*.

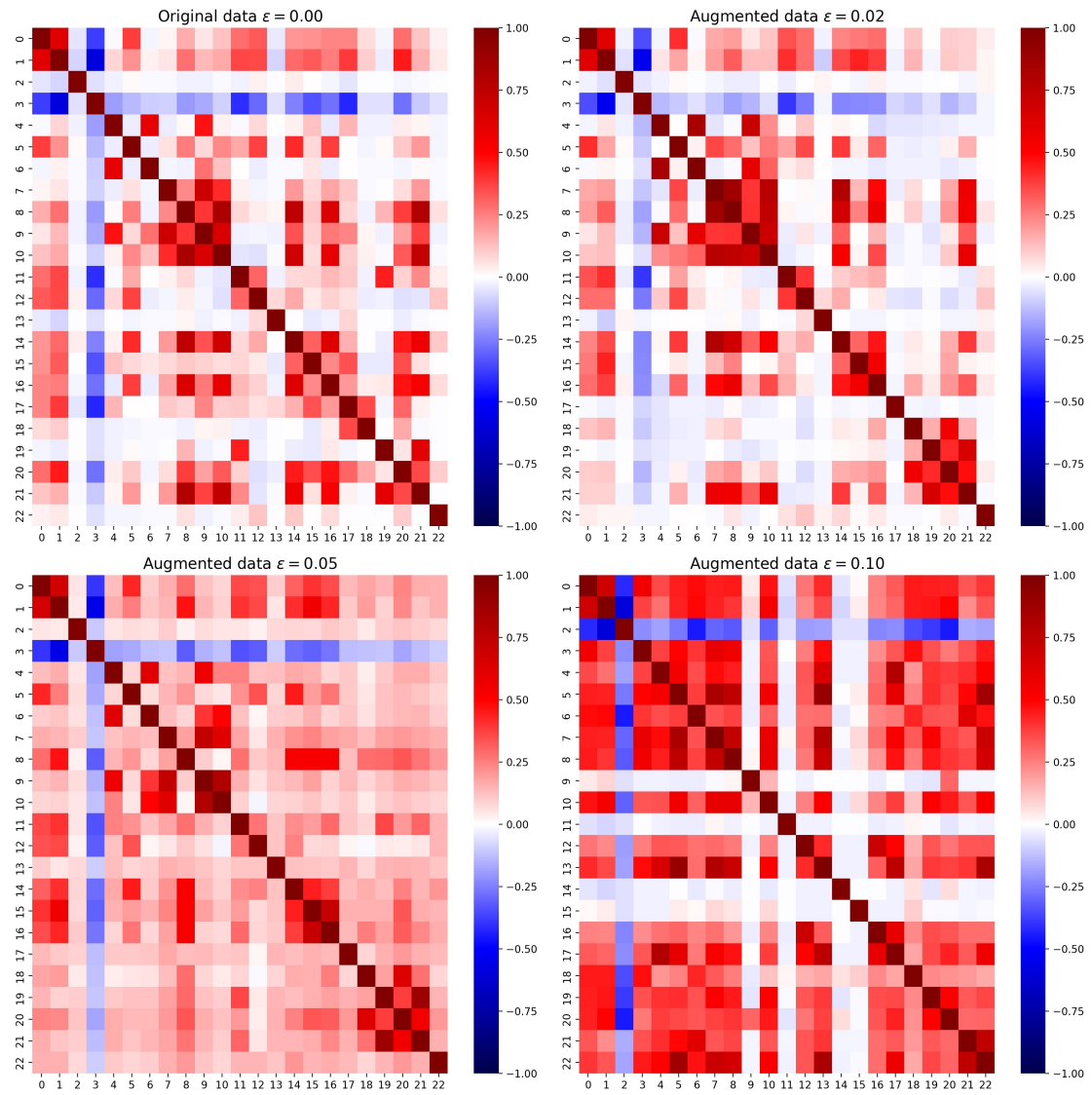


Figure 7.1: The correlation heatmap shows the correlation of each feature pair (numbered 0–22). The original data set $\epsilon = 0.00$ (top-left) and the augmented data sets are depicted.

Table 7.1: Performance comparison of RFC and XGBoost Classifier using different scoring functions.

Classifier	Data set	Accuracy	F1	ROC-AUC
RFC	$\epsilon = 0.00$	0.55	0.47	0.59
	$\epsilon = 0.02$	0.85	0.82	0.96
	$\epsilon = 0.05$	0.85	0.83	0.94
	$\epsilon = 0.10$	0.60	0.45	0.62
XGBoost	$\epsilon = 0.00$	0.53	0.39	0.61
	$\epsilon = 0.02$	0.90	0.89	0.96
	$\epsilon = 0.05$	0.90	0.89	0.93
	$\epsilon = 0.10$	0.60	0.37	0.65

XGBoost Classifier ³ are trained in turn on each of the four training data sets. For example, the randomised grid search based hyperparameter selection for the RFC trained on the augmented data set $\epsilon = 0.02$ yields 200 parallel trees (the number of estimators) with a maximum depth of 50 as the best hyperparameter set. Trained on the same data set, the XGBoost Classifier uses 100 sequential trees with a maximum depth of 5 per tree. The results for the standard decision threshold $TH_m = 0.5$ are depicted in Table 7.1. As the results indicate, the XGBoost Classifier improves the performance over the RFC. In fact, using perturbation factors of $\epsilon = 0.02$ and $\epsilon = 0.05$ an accuracy of 90 % has been obtained. This demonstrates that the introduction of adversarial examples as additional learning samples to the training data improve model generalisation and reduce overfitting on the training data. However, as one can see in Table 7.1 the performance of the RFC and the XGBoost Classifier starts deviating if the perturbation continues to increase, i.e. a training data set is augmented with adversarial examples using a perturbation of $\epsilon = 0.10$. Consider the correlation heatmap for the original data $\epsilon = 0.00$ in Figure 7.1-(top-left) and the data augmented with a perturbation factor $\epsilon = 0.10$ in Figure 7.1-(bottom-right). The correlation between the features of the augmented data set $\epsilon = 0.10$ has increased considerably among the majority of considered features in comparison to the original data set $\epsilon = 0.00$. As observed by [3], the obtained results exemplify that high perturbation factors introduce spurious correlations between features which in turn leads to a correlation bias causing instabilities in the DT based algorithms.

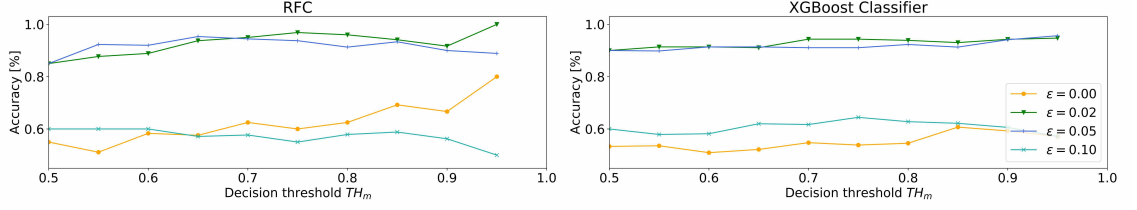


Figure 7.2: Accuracy score depending on decision threshold TH_m for RFC and XGBoost Classifier.

7.1.2 Maintenance threshold interpretation

Since downhole tool failures can be very costly, the operational and maintenance support provided by the BHA-PCBA framework needs to be adjusted according to the risk exposure of the operation. Hence, a more cautious prediction shifts the priority towards the mitigation of all BHA-PCBA failures during drilling which comes at the cost of over-maintenance. Therefore, if the maintenance decision threshold TH_m is increased, the algorithms predictions are treated more conservatively. The predictions for both classes are filtered according to $p(y = k|x) \geq TH_m$, corresponding to the confidence threshold in the interval $0.5 \leq TH_m \leq 1$. All test instances are rejected if the predicted probability is below TH_m . Subsequently such a BHA is treated as suspended, thus, maintenance is required. The results in Figure 7.2 are obtained by plotting the accuracy of both classes while increasing TH_m . As TH_m increases, it can be observed, in particular for RFC, how the accuracy increases. Likewise, the rate of BHAs classified as possibly failed, i.e. fp , increases. The augmented data sets $\epsilon = 0.02$ and $\epsilon = 0.05$ yield stable results, whereas the performance of the XGBoost Classifier is only marginally affected by TH_m .

Because the forecast certainty depends on the decision threshold TH_m and, in turn, on the separability of the predicted probabilities for the *failure* and *no-failure* class, the ROC-AUC score introduced in Chapter 5.1 is used as the guiding metric to assess the model performance. Consider the ROC-AUC score displayed in Table 7.1; the RFC or the XGBoost Classifier exhibit the best performance if trained on the augmented data set $\epsilon = 0.02$. Therefore, the consecutive analysis is performed only on the models trained on the augmented data set $\epsilon = 0.02$. Exemplary, Table 7.2 presents distinct results of the trained RFC model during inference, i.e. predicting the *no-failure*/ *failure* probability for BHA-PCBAs. The results for a BHA fleet under consideration of three operational risk scenarios which would require an adjustment of TH_m are shown. In a high-risk scenario the operator might demand a higher confidence in the accuracy of tn and tp predictions due to the increased

³The *XGBoost* python package provides the XGBoost Classifier using `xgboost.sklearn.XGBClassifier` (version 0.7) [2]. The model `xgbc` is configured as follows: `xgbc.learning_rate=0.1`; `xgbc.n_estimators` search-space [10,300]; `xgbc.max_depth` search-space [2,110]; `xgbc.grow_policy = 0`; `xgbc.objective = 'binary:logistic'`; `xgbc.colsample_bytree = 0.8`; all other parameters *default*.

Table 7.2: Maintenance support for BHA units considering the *no-failure* probability for three risk scenarios by adjusting TH_m ; results of RFC, data set augmentation $\epsilon = 0.02$; *accept* indicates a BHA re-run recommendation without prior maintenance of the electronics, *reject* indicates that maintenance of BHA-PCBA is recommended prior to a re-run of the unit as the probability of *no-failure* throughout the operation is too low.

Operational Risk BHA Unit	no-failure (probability)	Low $TH_m = 0.5$	Medium $TH_m = 0.6$	High $TH_m = 0.7$
BHA 00001	0.82	<i>accept</i>	<i>accept</i>	<i>accept</i>
BHA 00002	0.68	<i>accept</i>	<i>accept</i>	<i>reject</i>
BHA 00003	0.95	<i>accept</i>	<i>accept</i>	<i>accept</i>
BHA 00004	0.59	<i>accept</i>	<i>reject</i>	<i>reject</i>
...				
BHA j	0.64	<i>accept</i>	<i>accept</i>	<i>reject</i>

severity of a BHA failure during drilling. Under such circumstances an operator considers a BHA-PCBA only fit for re-run if the algorithm predicts *no-failure* with a probability higher than $TH_m = 0.7$. If the model's predictions fall below the specified threshold TH_m , then the BHA-PCBA is suspended and maintenance is recommended. On the contrary, in a low risk application, e.g. $TH_m = 0.5$, the above predictions would have met the decision threshold criterium and the BHA-PCBA would be considered fit for a re-run.

In order to better understand the interpretation of TH_m the relevance of the proposed approach is explored through an example of a maintenance business case. Maintenance expenditures are inevitably linked to tool design and its reliability, spare-parts costs, labour costs, the employed maintenance strategy, and personnel competency. In this example, only tool design and reliability are considered for a BHA fleet. Personal competency is not considered. Thus, a simplified cost model is introduced. This cost model accounts for spare-parts costs (average PCBA replacement cost), labour costs (average cost of maintaining a tool without parts), and infield failure costs. The total costs of failure are determined through Equation 7.1

$$C_f = \overline{C}_f F_{fn} \quad (7.1)$$

The total infield failure costs C_f are calculated by the average failure cost \overline{C}_f , depending on the severity of the failure, and the total number of failures F_{fn} which have not been correctly identified by the model, i.e the false negative (fn) classifications.

$$C_m = (\overline{C}_p + \overline{C}_l)(F_{tp} + F_{fp} + BHA_{nc}) \quad (7.2)$$

The total maintenance costs C_m are calculated according to Equation 7.2. \overline{C}_p represent the average part cost and \overline{C}_l the average labour cost; F_{tp} the total number

of correctly classified failures, i.e. the true positive (tp) classifications and F_{fp} the total number of falsely classified failures, i.e. the number of false positive (fp) classifications; BHA_{nc} is the total number of BHA units which have not been considered. Such instances are discarded because the prediction for a failure or no-failure lies below the decision threshold TH_m . Subsequently the overall costs C_t are calculated as per Equation 7.3.

$$C_t = C_m + C_f \quad (7.3)$$

Twelve scenarios to evaluate the efficiency of the proposed approach are considered. The scenarios are created by assuming three average failure cost factors of $\overline{C}_f = 1$, $\overline{C}_f = 3$, and $\overline{C}_f = 10$. Here $\overline{C}_f = 1$ – a low impact failure – represents a scenario in which a failure does not impact the business significantly and $\overline{C}_f = 10$ – a high impact failure – relates to a scenario where a failure can lead to severe loss of business. Fleet A , B , C , and D can be assumed to represent various reliability levels of the BHA fleet. An outdated fleet can have a low level of reliability, i.e. referred to as fleet A . Contrary, a lightly used, proactively maintained, younger fleet can have a high level of reliability, i.e. fleet D . Fleet reliability, in this example, is modelled as the percentage of *no-failure* and *failure*, e.g. 90 % fleet reliability equates to 90 % *no-failures* and 10 % *failures* in the scrutinised test data set. The reliability levels chosen here are not necessarily representative for a real fleet, but rather serve as a mathematical exercise to select the optimal threshold TH_m in the context of the proposed BHA-PCBA maintenance support framework. Figure 7.3 visualises the results of the considered business case for each combination of fleet reliability (A , B , C , and D) and average failure costs. The results for the RFC classifier are represented by a solid line and for the XGBoost Classifier by a dashed line. The first column displays the distribution of the predicted *no-failure* probabilities derived on the test set. A right skewed distribution indicates better class separation, i.e. the models capability to distinguish between the *no-failure* and *failure* instances. Each of the graphs shown in the three columns of the average failure cost scenarios ($\overline{C}_f = 1$, $\overline{C}_f = 3$, and $\overline{C}_f = 10$) plots the decision threshold TH_m ranging from 0.5 to 0.8 against the not considered PCBAs BHA_{nc} (right y-axis, blue lines) and the resulting costs of infield failure C_f (left y-axis, green lines).

The results displayed in this Figure 7.3 indicate that it is not necessary to have one superior approach for all scenarios. A variety of solutions should be deployed. Consequently, a choice is made based on the fleet reliability, available data, available maintenance infrastructure, availability of spare parts, and the business case. Various conclusions can be drawn from the analysis of the above stated business case. In general, the XGBoost Classifier allows a better distinction between *no-failure* and *failure* classes compared to RFC (consider the class probability distribution in

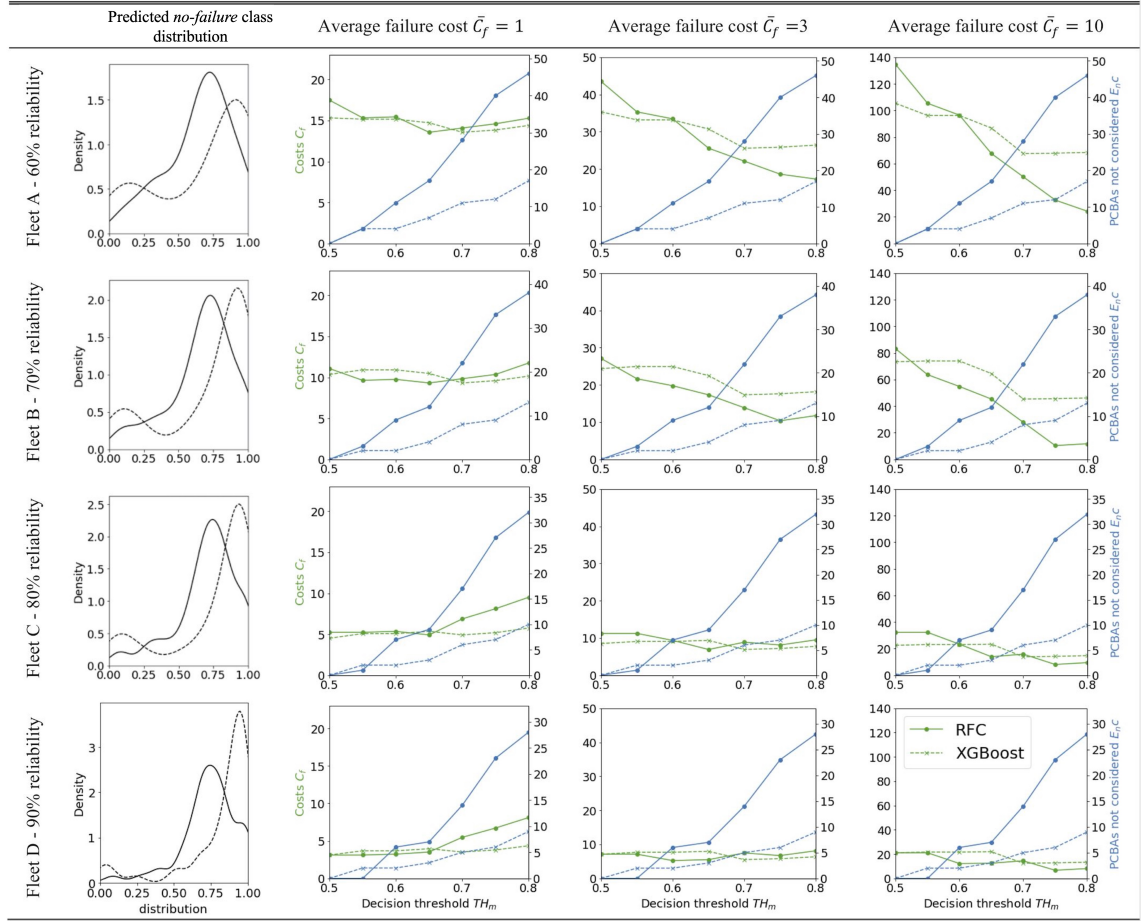


Figure 7.3: Business case – predictions based on models trained on the augmented data set $\epsilon = 0.02$; Models: solid line - RFC, dashed line - XGBoost Classifier; BHA_{nc} - blue lines; C_f - green lines.

7.3-(left)). However, for a high failure cost scenario, over-maintenance will further mitigate costs as the results for RFC indicate. If the failure costs increase relative to the maintenance costs, selecting a higher threshold TH_m can considerably reduce costs. The impact of a higher TH_m is heavily depended on the fleet reliability and decreases when the fleet reliability increases. Two root-causes are identified for lower expenses at higher TH_m . Firstly, as the number of BHA units increases for which predictions are not considered due to a non-meaningful forecast, the amount of misclassified entities declines (the *no-failure* probability is below TH_m). Simultaneously, the accuracy increases. Secondly, a misclassified failure *fp* tends to have a lower *no-failure* probability and will rather be considered as a BHA unit requiring maintenance when the threshold TH_m increases. Thus, not identified in-field failures are mitigated. This trend gains significance as the fleet reliability decreases and failures become more costly. However, predictions for fleets with a high reliability and low impact failures indicate the best threshold being $TH_m = 0.5$.

```

=====
Version PCBA_001, 01st April 2020
Review readme.txt - You might want to retrain the model if new data is available
=====
Retrain (y/n):y (1)
=====
RESULTS
Accuracy Score 0.87
F1 Score: 0.85
ROCAUC Score: 0.94
----> XGBoost Retrained (a)
=====
Enter data path to predict (e.g. <example.csv>, no LABEL!):example.csv (2)
=====
#####
RESULTS (0 == No Failure; 1 = Failure)
The predicted class:
[1 1 1 1 1 0 1 1 1 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0]
The predicted probabilities for a failure are:
[0.83 0.87 0.89 0.58 0.77 0.02 0.93 0.77 0.91 0.72 0.59 0.12 0.82 0.77
0.35 0.01 0.01 0.03 0.16 0.1 0.02 0.01 0.29 0.02 0.16 0.03 0.03 0.06
0.11 0.03]
#####
Enter export path (e.g. <results.csv>,:results.csv (3)
Results exported as: results.csv

```

Figure 7.4: The developed API using *python*. User interaction is supported through: (1) the option to retrain the selected model with new data; (a) returns the scoring of the retrained model on the validation data set; (2) selection of a data set containing BHA units for which *failure* and *no-failure* estimations are required; (b) the returned estimates showing the absolute class estimations at $TH_m = 0.5$ and the distinct probabilities in case a higher risk business case is considered; (3) provides the option to export the results.

7.1.3 Deployed API

A key deliverable for the industrial usage of the proposed framework is the provision of a custom Application Programming Interface (API) that allows to retrain the pool of developed ML models once new data is available. Further, it provides predictions for any number of BHA units and facilitates the export of these results. The interface of the developed API is presented in Figure 7.4.

7.1.4 Discussion of PdM integration and data constraints

The proposed fleet-level method provides the means to predict a failure of PCBAs within the BHA without the need for tedious disassembly and manual inspection of the built-in electronics. This is achieved by using an aggregated feature set from field data and diagnostic tool memory data. In order to integrate such an approach into an data-driven PdM framework two distinct phases should be considered, cf. [4]. First, the collected data serves as training data for the offline training phase. Second, once deployed, newly collected data of the same format acts as an input during the online phase to predict imminent PCBA failures. In order to extend the capabilities of the presented method in a PdM framework, additional data sources should be consulted and the model regularly retrained to account for changes in the overall BHA-PCBA fleet reliability and newly acquired data.

The complex nature of the BHA (a multi-component system) with different interacting failure modes and mechanisms poses a challenge. Thus, at the moment, a

circuit-level or even component-level resolution owning a diagnostics or prognostics character is not achievable given the low availability and resolution of the data as well as missing failure diagnosis and sensor localisation. Nonetheless, in order to reduce sources of uncertainty for the proposed fleet-level approach and to improve the models maintenance support capabilities, some form of time series data would be beneficial, rather than only aggregating data in form of sparse snapshots for each BHA unit. The degradation process and its propagation could then be adequately mapped, e.g. a continuous recording of tool memory data. To facilitate such an approach, detailed specifications of the environmental influences through extensive and continuous measurements of the experienced dynamic operating conditions and the sensor-placements within the BHA are required in order to derive meaningful relations between such data and the diagnostic tool memory data.

7.2 EMR failure analysis of silver-plated contacts

Silver-based contacts (designed as pure silver, silver-alloy, or silver-metal oxide contacts) are widely used as contact material for EMRs. They can be either welded directly onto the contact carrier or on a copper-based contact-rivet that is then placed on the contact carrier strip. Besides economic considerations, silver-based contacts possess a low CR. This is due to comparably lower oxidation rates, despite the effective temperature increase on the contact's surfaces initiated through electrical arcing [5]. In addition, the oxides formed from silver are unstable at higher temperatures. However, since pure silver has a high tendency for contact welding, silver-composite materials are used instead. Internal-oxidation of these materials increases the material hardness improving its resistance against arcing while it reduces the weld strength, contact sticking, and material loss [6, 7]. On the downside, in comparison to raw silver, silver-composite materials exhibit an operating temperature increase. This relates to the decreasing electrical conductivity that is caused by increased rates of internal-oxidation [8]. Manufactured through such internal-oxidisation processes, EMR applications utilise silver-metal oxides in low power applications of up to 20 A [7]. Up to 15 % oxides are common among these types of electrical contacts [9]. Still, predominant failure mechanisms are connected to arcing leading to excessive contact erosion. In general, the plated surface material is subject to continuous decomposition through evaporation, splattering, and welding as the silver and the oxidised metal dissociate [10]. The improvement of contact materials is an active field of research. For example, [11, 12] report results of contact erosion rates for silver-oxide contacts under the influence of elevated load currents. Toxic Silver-Cadmium-Oxide $AgCdO$ contacts exhibit lower erosion rates than Silver-Tin-Oxide ($AgSnO_2$) contacts. Oxidised Silver-Tin-Indium $AgSnO_2In_2O_3$ contacts show even

further reduced rates of erosion in comparison to, e.g. *AgCdO* contacts [8] ⁴.

Certainly, silver-based contact materials improve the performance of electrical contacts. And yet, in terms of monitoring the health of contacts, such materials present a challenge because the environment in which they operate significantly alters the prevailing failure mechanisms and its propagation. In order to understand the rationale behind this, it is helpful to investigate the formation of contact faults due to contact erosion and contact welding. Consider the case of silver-based *AgSnO₂In₂O₃* contacts found in the tested unsealed and sealed EMRs that, as a matter of a fact, exhibit these very diverging degradation patterns. So as to identify the underlying root-causes, the closed-circuit voltage v_{cc} , open-circuit voltage v_{oc} , CR, and contact temperature measurements are aggregated alongside high resolution CV, CI, and CC recordings. The analysis of the evolution and manifestation of experienced failures is supported by an examination of the contact rivets by the means of Computer Tomography Scan (CTS). Finally, the results are evaluated with regard to the applicability of CR as a DI for silver-based EMR contacts. Possible alternatives are then presented. The following examination places the focus on DA2.

7.2.1 Observed failure modes and mechanisms

Measurements of v_{cc} after contact making are taken at t_m (i.e. PT+BT; measured from first coil voltage increase trigger) just past the *maximum allowed contact making time* $t_{m,max} = 15$ ms. Likewise, measurements of v_{oc} after contact breaking are recorded at t_b (i.e. RT+AT; measured from first coil voltage drop trigger) just past the *maximum allowed contact breaking time* $t_{b,max} = 5$ ms. Displayed in Figure 7.5-(top), if EMR contacts are operating nominally, the measured $v_{cc}(t_m)$ settles at a relatively low value; the measured $v_{oc}(t_b)$ retains a high value. However, if the EMR degrades, contact making and contact breaking are liable to varying degrees of drift. As a rule, such changes of the waveform become more apparent towards the EOL, since various failure mechanisms act on the EMR components, predominantly the contact's surface. Hence, two cases of non-nominal actuations can be distinguished, cf. 7.5-(bottom). First, during contact making, contacts might not have fully settled when the reference v_{cc} measurement is taken at t_m yielding an elevated v_{cc} reading beyond the specified time $t_{m,max}$ threshold, i.e. in the case of DA2 $v_{cc} > 0$ V. Second, contact breaking may take longer and the sensed v_{oc} taken at (t_b) is lower than the required open-circuit voltage required at $t_{b,max}$, i.e. $v_{cc} < 35$ V.

⁴*Parts of this chapter have already been published in my journal or conference article: L. Kirschbaum, et al., (2022) and L. Kirschbaum, et al., (2018).*

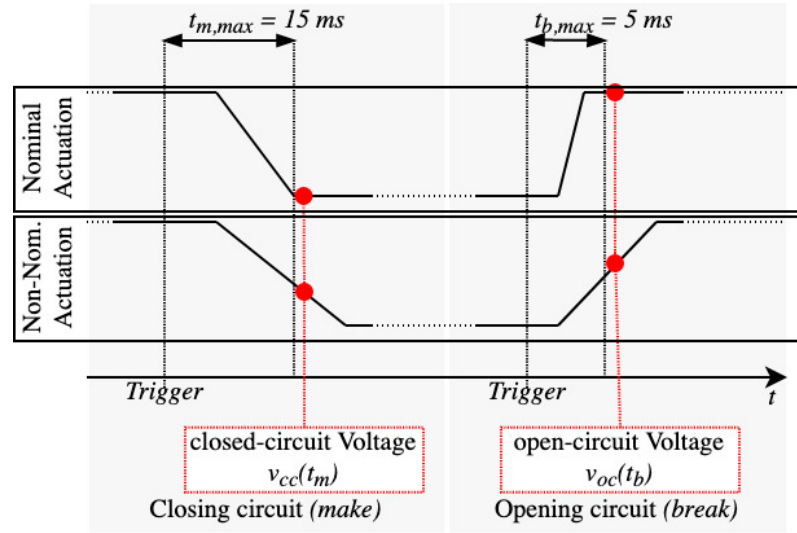


Figure 7.5: Schematic display of measurements of the v_{cc} and v_{oc} of nominal (top) and non-nominal actuation (bottom).

7.2.1.1 Sealed EMR

First, consider the case of sealed EMRs. As for contact making, depicted in Figure 7.6, the measured v_{cc} resides nominally around 0.1 V due to the inherent CR. The majority of non-nominal closing actuations with prolonged making time ($v_{cc}(t_{m,max}) > 0.1\text{ V}$) predominantly occur within the last 10 % of the EMR's life. This indicates that contact is not sufficiently established within 15 ms past the initial coil voltage increase. In such cases, the measured v_{cc} settles around $\pm 10\text{ V}$, but in general does not reach the v_{oc} prior to the actual stuck-open failure. In many instances the non-nominal actuations appear in batches after which the actuations settle again to the nominal behaviour over the extent of certain intervals.

Multiple reasons can be identified that contribute to the growing number of non-nominal operations towards the EOL. First, consider reference-times derived for EMR DA2-03S, cf. displayed in Figure 7.7. Both, the PT and BT, determined according to the criteria stated in Table 7.3, exhibit a subtle upwards trend throughout the EMR's life. Likewise, deviations from the nominal bounce duration become more frequent. Directly related to the BT and, therefore, the duration of the individual bounces, is the AT during bouncing. The increase in AT suggests that the cumulative time during bouncing at which arcing is present rises towards the EOL. The increase in BT and AT can be attributed to accelerated contact erosion. This reduces the contact force because the contact over travel diminishes due to the reduction in contact thickness. Furthermore, a reduced contact force might lead to a higher number of individual bouncing events. As can be seen in Appendix B.1, the CC waveform is also subjected to change over the course of the EMR life, whereas the shape of the salient inductive nick changes. This mirrors the observed PT increase. Fluctuations due to an increased coil resistance over the course of the

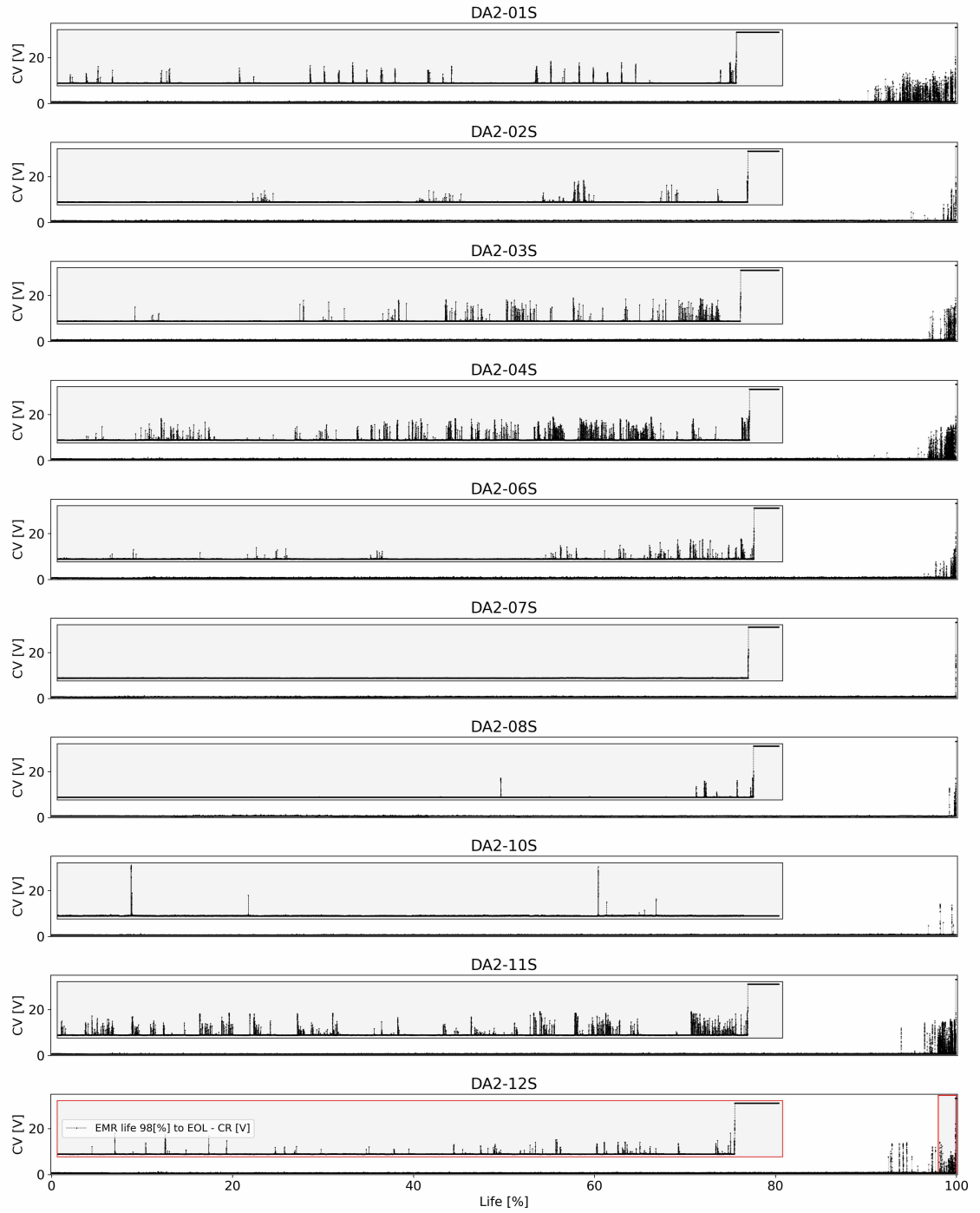


Figure 7.6: DA2 - Sealed EMR v_{cc} . The magnified section displays the voltage of the last 2 % of the EMR life.

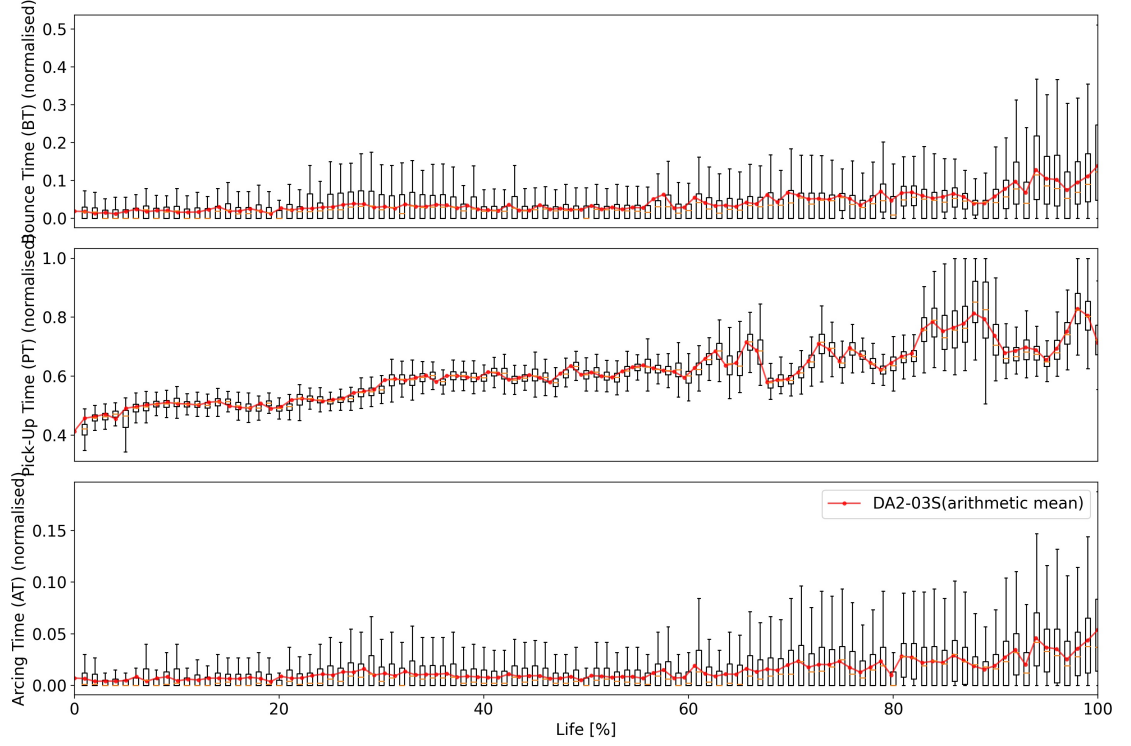


Figure 7.7: DA2-03S contact making actuations; actuations are averaged over each 1% bin of the EMR life; (top) normalised BT, (centre) normalised PT, (bottom) normalised AT (cf. Table 7.3).

EMR life relate to the travel of the armature. In such cases, the closing velocity of the moving contact carrier is reduced if the coil resistance increases. Lastly, higher contact force to establish current carrying paths is required. Here, the build up of non-conducting layers in combination with increasing contact surface roughness and the contact force needed to break such contaminating layers has to be considered, cf. the effects of fretting detailed in Chapter 3.2.4.

So far, only detrimental effects relevant to the deterioration of the EMR contacts during contact making have been discussed. Thus, in Figure 7.8 the v_{oc} is displayed, nominally residing at ± 35 V. Voltages lower than this nominal v_{oc} indicate extended arcing beyond 5 ms. If $v_{oc} = v_{cc}$, the contacts have failed to break sufficiently. Unlike non-nominal making operations, non-nominal breaking operations appear throughout the entire EMR life. However, most of these non-nominal breaking actuations are only of extended arcing duration, since they do not reach v_{cc} . No coherent trend among the tested sealed EMRs can be recognised that would suggest any correlation between the EMR life and the occurrence of non-nominal breaking events.

Similar to the analysis carried out for the making operation, consider reference-

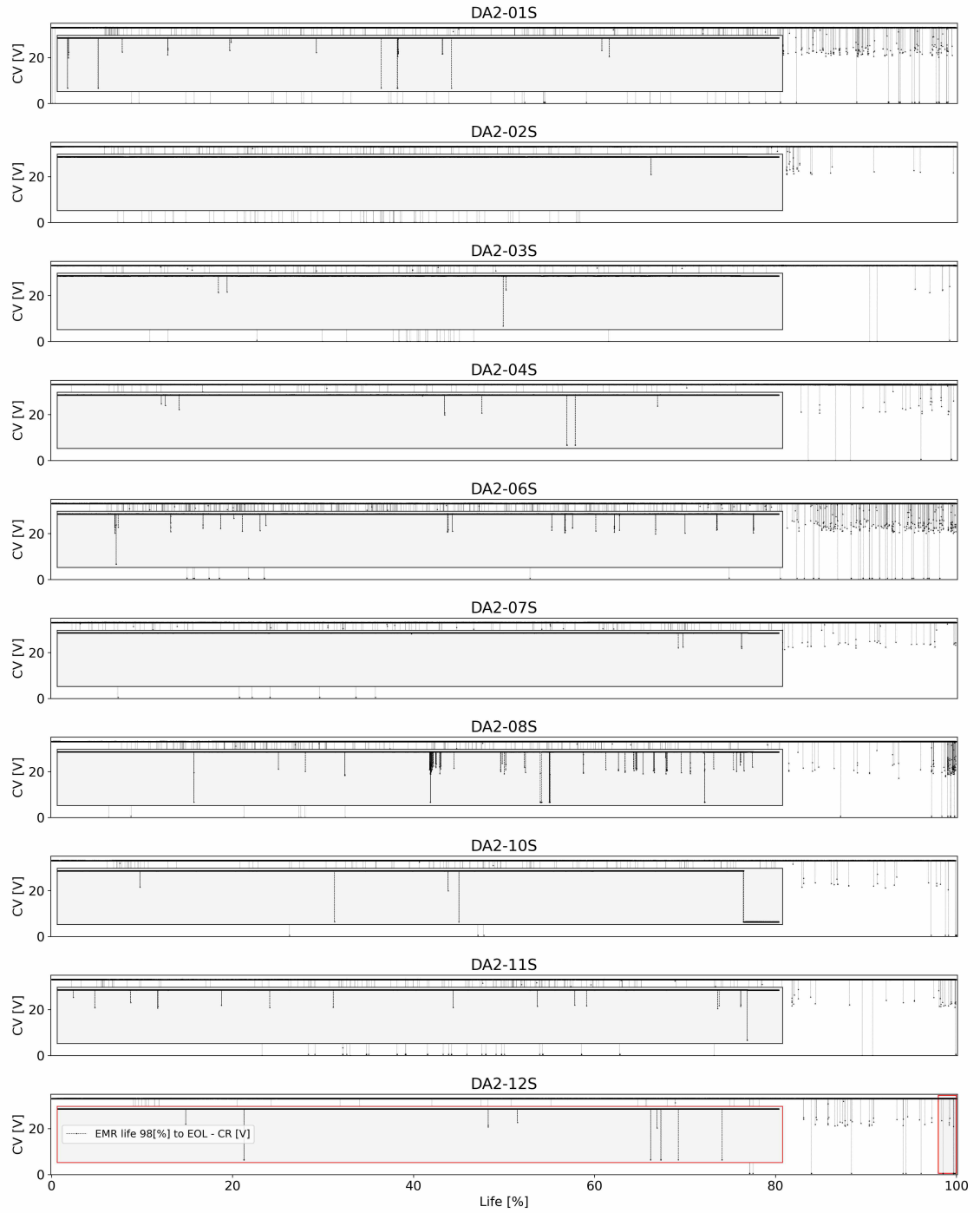


Figure 7.8: DA2 - Sealed EMR v_{oc} . The magnified section displays the CV of the last 2 % of the EMR life.

Table 7.3: The table defines how the time-based reference DIs used in this thesis have been calculated; PT, BT, and AT during contact making and RT, and AT during contact breaking. These have been introduced in general in Chapter 4.2.2 and Figure 4.4.

Time-based DI	Definition
Making	
PT (Pick-Up Time)	Determine the first occurrence of the voltage below the v_{cc} and the $CI > 0$. This is the first falling edge of the CV making waveform and the first rising edge of the CI making waveform.
BT (Bounce Time)	Detect the last falling edge of the CV making waveform where the voltage has been above the minimum arcing voltage (13 V) and settles below v_{cc} . BT is defined as the time passed between PT and the time at which the voltage falls below the v_{cc} in regards to this last falling edge of the CV.
AT (Arcing Time)	Determine the duration of the signal where the CV making waveform is below v_{oc} but above the minimum arcing voltage (13 V) in-between PT and PT+BT.
Breaking	
RT (Release Time)	Determine the first occurrence on the CV breaking waveform of the voltage rising above the minimum arcing voltage (13 V) and the current of the CI breaking waveform dropping below 6 A.
AT (Arcing Time)	Determine the interval between PT and the CV reaching 90 % of the v_{oc} .

times derived for the breaking actuations, cf. DA2-03S displayed in Figure 7.9. An increasing occurrence of micro-welding causes the contacts to stick more frequently towards the EOL as the subtle increase of RT documents, cf. Table 7.3. Moreover, the spring force relaxation due to material ageing is likely to contribute to a reduction in contact retraction force and, consequently, velocity. To illustrate this phenomenon, consider the windowed high resolution CV and CI breaking waveforms in Appendix B.2. Though, no increase in AT can be observed which implies that degradation is not accelerated by growing arcing duration during breaking. These results are consistent with those shown in Figure 7.8 with respect to v_{oc} , since the EMR does not exhibit excessive arcing beyond $t_{b, max}$ for all sealed EMRs.

Apart from DA2-10S, all sealed EMRs failed stuck-open. As mentioned before, this failure mode can be identified through the immediate increase of the v_{cc} to ± 35 V at the EOL, cf. Figure 7.6. At this point, the erosion of contact material, the contamination of the contact surface (through a reduction of the contact force fostered by the widening contact gap), and the spring force relaxation have progressed too far to establish any current carrying path among the contact interfaces.

The high resolution CV measurements taken over the last couple of actuations prior to the occurrence of the stuck-open failure provide further details regarding the manifestation of the failure. Measurements relating to EMR DA2-03S are presented in Figure 7.10. During contact making, at actuation 2833140, continuous bouncing of the EMR contacts can be observed. However, at the consecutive break operation 25 ms later the contacts have settled and good contact has been established, since the measured CV at the start of the recording is equivalent to v_{cc} . At the

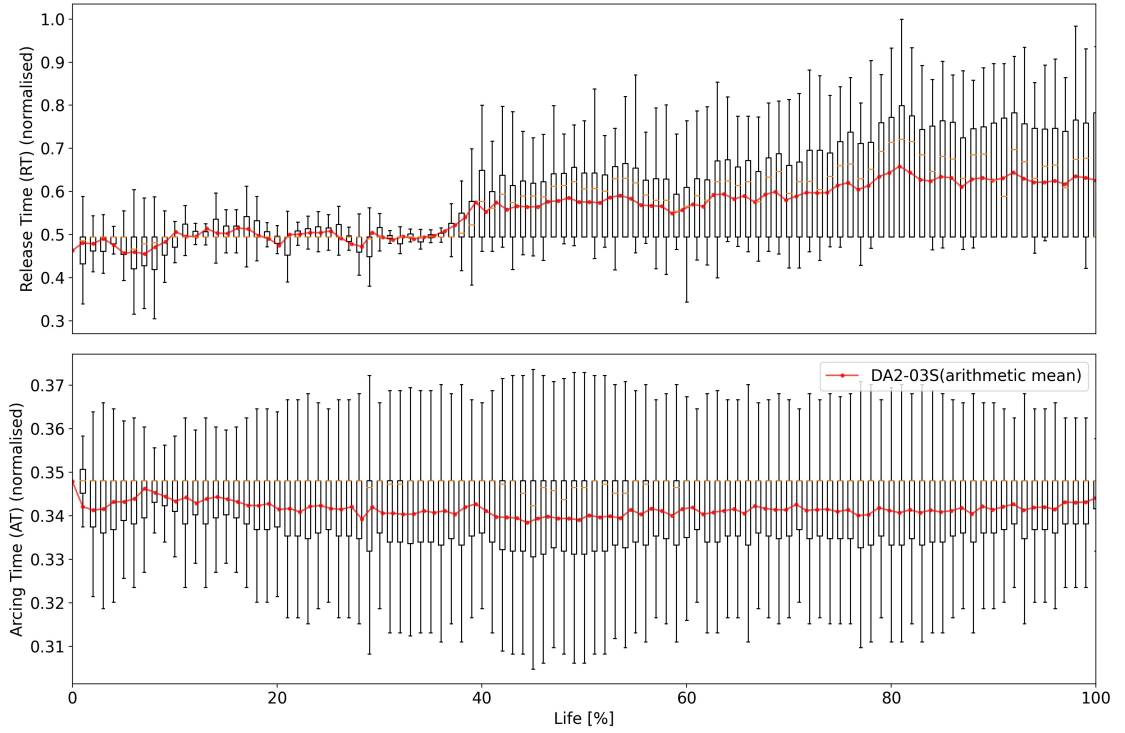


Figure 7.9: DA2-03S contact breaking actuations; actuations are averaged over each 1% bin of the EMR life; (top) normalised RT; (bottom) normalised AT.

making operation 10 actuations later, i.e. actuation 2833150, the contacts continue to bounce beyond $t_{m,max}$ for almost the entire 25 ms period in-between switching, cf. the break operation 2833150. This continuous bouncing occurs for all following actuations until the making actuation 2833160. Now, followed by this period of extended bouncing, the contacts touch only one time. The measured CV instantly settles around the minimum arc voltage of the contacts; no consecutive bouncing nor acceptable contact making can be observed from this point onward. On the contrary, a continuous burning arc establishes between the closely spaced non-touching contact surfaces. It lasts until the consecutive break operation 2833160. Here, the arc is extinct when the armature and, therefore, the moving contact carrier are fully retracted. No contact making can be seen during the following making actuation, cf. make 2833170, despite the armature does indeed travel forward pushing the moving contact carrier which is indicated by the inductive nick of the CC (also compare the CV waveform settling around ± 35 V).

The data considered so far show a distinct pattern of how the failures evolve in time. Despite a continuously increasing PT, AT during making, and BT throughout the course of any sealed EMR's life, a severe manifestation of failure becomes evident only in the last several 100 actuations. The first indicator of imminent failure is excessive bouncing lasting for multiple *ms* over the course of many actuations. This, in turn, causes significantly higher rates of contact erosion, severe material

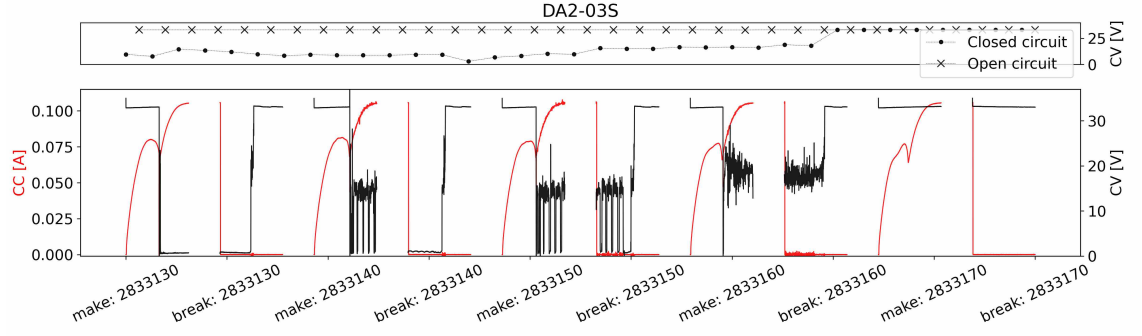


Figure 7.10: DA2-03S sealed EMR failed stuck-open; (top) v_{oc} and v_{cc} measurements; (bottom) CV and CC of the actuations sampled at 150 kHz for 2 ms every 10th making and breaking actuation.

ejection, dispersion, and redistribution, e.g. from micro welding. As mentioned before, eventually the over-travel is too low and the contact force too small to make good contact, in particular, when acting against blow-off forces of arcing contact interfaces. Prior to the actual failure, during contact making a continuous ambient arc establishes burning the entire period in-between contact making and contact breaking (i.e. ± 25 ms). This arc is comparably more powerful than the short interval arcs during bouncing or during contact breaking, in particular, whenever the contacts are pushed very close together by the armature but do not establish any contact. Its impact initiates the final deterioration of the contact surface leaving too little material on the contact rivets so that the contacts fail stuck-open.

Further, consider a set of sealed EMR contacts subjected to a stuck-open failure. On the anode, the majority of the original contact surface has eroded leaving a concave contact surface in the rivet body. Many small craters that are scattered over the entire contact surface can be recognised, cf. Figure 7.11-(a). The rendering of the CTS, in Figure 7.12-(a) highlights this, indicating that only the edge of the original contact plating material is left. Contrary, in Figure 7.11-(b) the moving contact forms a convex shape studded with pips, cf. Figure 7.12-(b). Both contact carriers reveal, in the vicinity of the contact rivet, sputtered black material on the carrier surfaces. This material consists of silver-oxides and carbons as EDX analysis suggests, cf. Appendix B.5.

The examination of the worn contact rivets with the help of CTS documents an unidirectional material transfer and the erosion of electrical contacts that is to be expected in DC switching circuits. The predominant material loss is found to be at the anode (formation of craters). In contrast, a material gain can be observed at the cathode (formation of pips). Since the operation of the arc in ambient air in DC circuits is relatively short, metallic arcing from the static contact carrier rivet (anode) to the moving contact carrier rivet (cathode) is responsible for this one-sided re-distribution of contact material and the concurrent net-loss of material

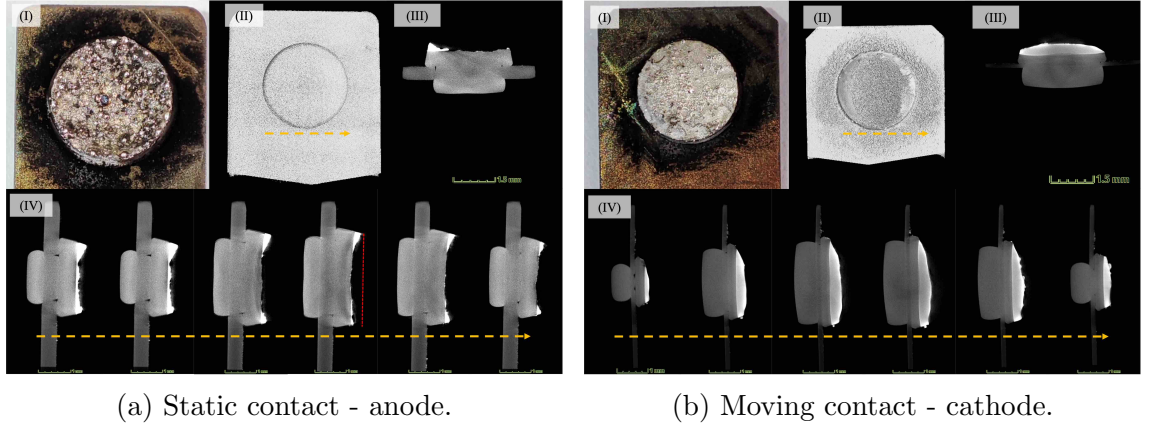


Figure 7.11: DA2-02S: CTS of sealed EMR; (I) photograph of the static contact carrier (anode) after failure; (II),(III) and (IV) cross-sections, the brighter grey area on the contact rivet can be identified as oxidised silver and the silver contact plating material, while the darker grey area is the copper contact rivet body.

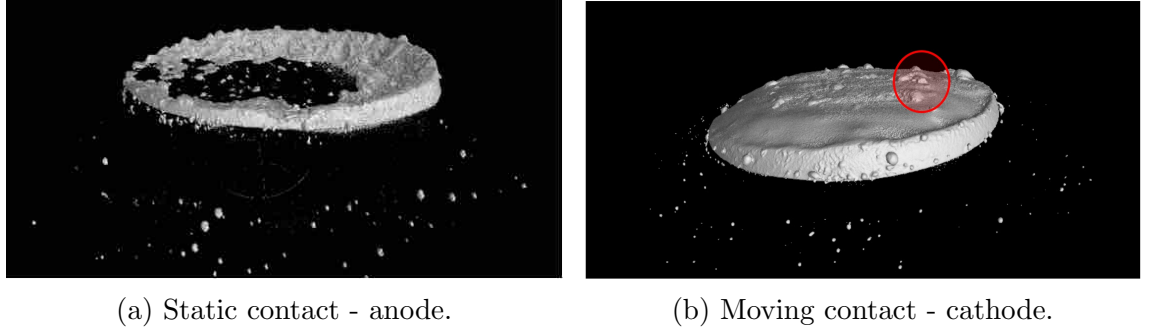


Figure 7.12: DA2-02S: Rendering of the CTS. The red circle indicates contact pips.

during switching. The governing failure mechanism acting on the contacts of the sealed EMRs are excessive bouncing and electrical arcing between the EMR contacts during contact making.

7.2.1.2 Unsealed EMR

A similar analysis of the degradation process can be performed for the unsealed EMRs which, unlike the sealed EMRs, predominantly failed stuck-closed. First, consider the v_{cc} measurements that are shown in Figure 7.13. No accumulation of non-nominal making actuations towards the EOL can be observed as it has been documented for the sealed EMRs. Likewise, increases of PT, BT, and AT are less apparent, consider the example of the unsealed EMR DA2-09U presented in Figure 7.14. In particular, this is the case for PT. One finds that the changes of the CC waveform over the course of the EMR life are comparably smaller than the ones reported for the sealed EMR, cf. Appendix B.3. In addition, less variance of the shape of the CV and CI waveforms, i.e. shift and bouncing, can be observed compared to the sealed waveform characteristics, cf. Appendix B.1. Hence, despite a deteriorating making actuation, the contact making process remains within the specified $t_{m,max} = 15$ ms threshold.

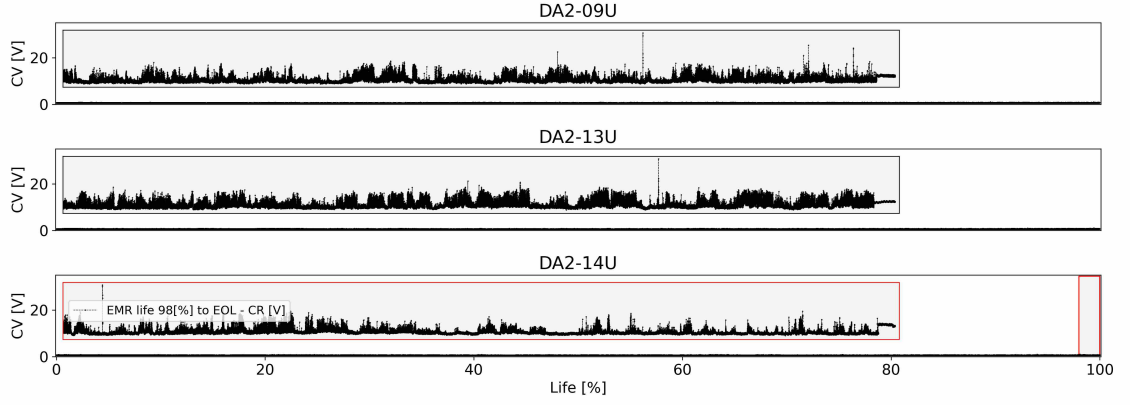


Figure 7.13: DA2 - Unsealed EMR v_{cc} . The magnified section displays the CV of the last 2 % of the EMR life.

While the sealed EMRs seemed to be predominantly affected by effects appearing during the making actuation, the unsealed EMRs do not exhibit such behaviour. Consider instead the v_{oc} measurements for the unsealed EMR DA2-09U displayed in Figure 7.15. These recordings reveal some events of extended contact breaking that seem to become increasingly frequent towards EOL, cf. Appendix B.4. The changes of the shape as well as a shift of the CV and CI waveforms signal an increase in RT, though this behaviour is not as distinct as it has been observed for the sealed EMRs.

In order to understand the development of stuck-closed failure, the focus resides on the contact sticking actuations during contact breaking that occur throughout the EMR life. To illustrate the formation of the stuck-open failure we present the case of the unsealed EMR DS2-09U displayed in Figure 7.16. The contacts stick together over the course of 80 actuations (6184190 to 6184270). The changed CC waveform for this period does not display the typical inductive nick which indicates the change in coil inductance when the armature closes. As detailed in Chapter 3.2.4, the contacts have welded together during contact breaking or more likely during contact making. The strength of the weld is greater than the spring force retracting the moving contact carrier during the break actuation. However, the repetitive stress of the making actuations (trying to push the contacts further together) and the breaking actuations (trying to pull the moving contact carrier back) excites the weld. The contact bridge eventually ruptures around actuation 6184270, allowing the contacts to operate nominally. In the end, the EMRs are considered failed (stuck-closed), whenever the contacts exhibit such a behaviour over a period that is deemed unacceptable (application dependent).

We may conclude then, that an unsealed EMR similar in design and contact material as its sealed cognate exhibits less severe damage on the anode contact surface. This

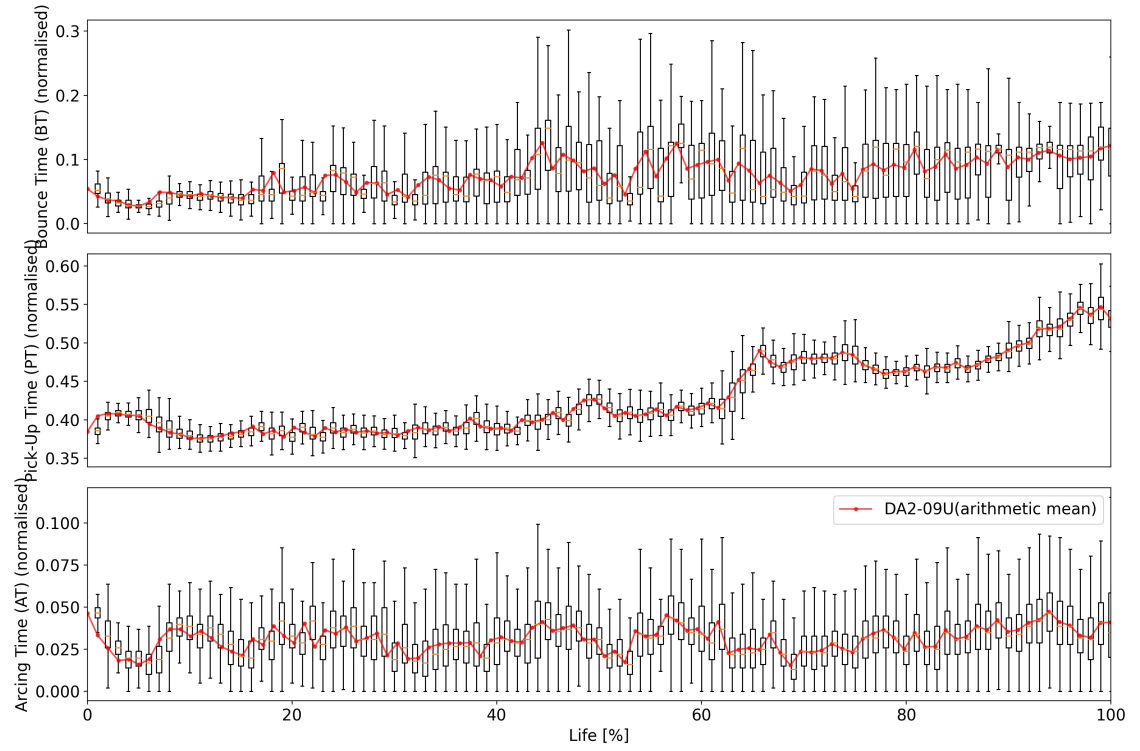


Figure 7.14: DA2-09U contact making actuations; actuations are averaged over each 1% bin of the EMR life; normalised BT, (centre) normalised PT, (bottom) normalised AT.

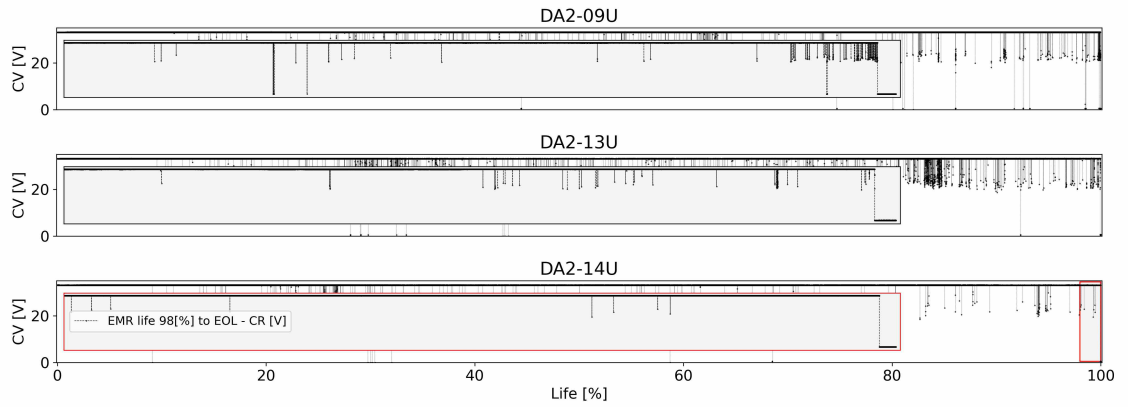


Figure 7.15: DA2 - Unsealed EMR v_{oc} . The magnified section displays the CV of the last 2 % of the EMR life.

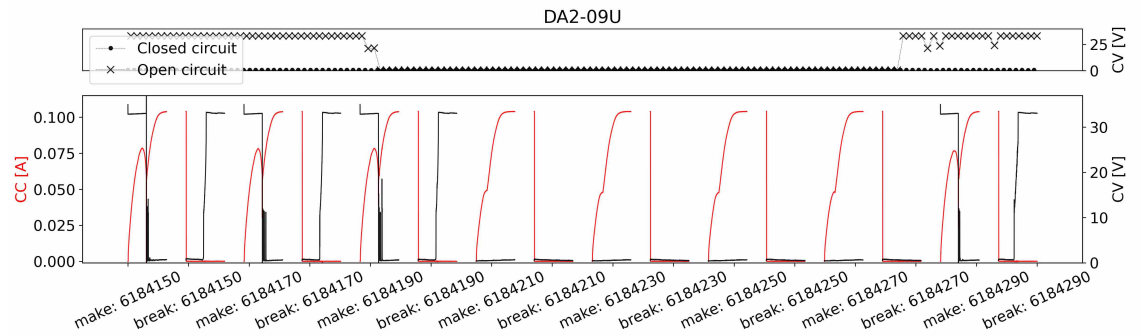


Figure 7.16: DA2-09U unsealed EMR failed stuck-closed; (top) v_{oc} and v_{cc} measurements; (bottom) CV and CC of the actuations sampled at 150 kHz for 2 ms every 20th making and breaking actuation.

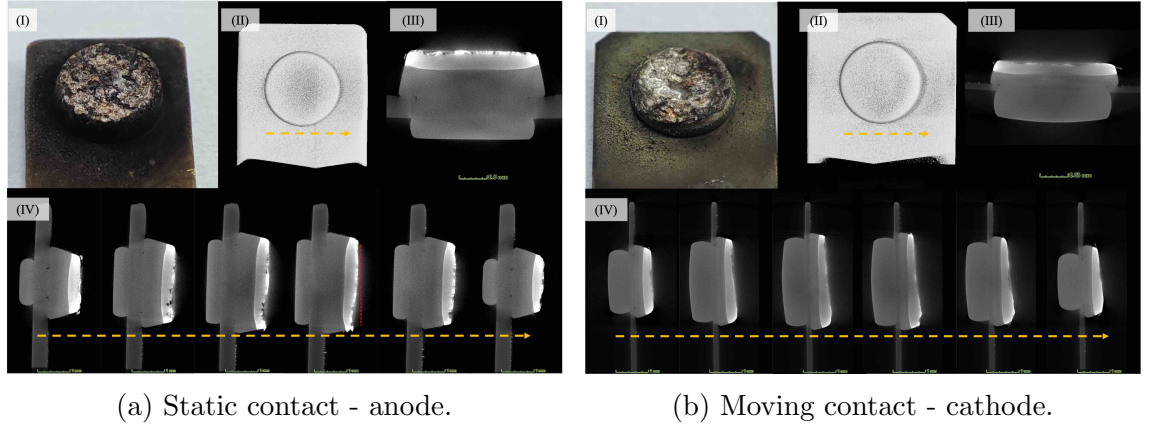


Figure 7.17: DA2-09U: CTS unsealed EMR; (I) photograph of the static contact carrier (anode) after failure; (II),(III) and (IV) cross-sections, the brighter grey area on the contact rivet can be identified as oxidised silver and the silver contact plating material, while the darker grey area is the copper contact rivet body.

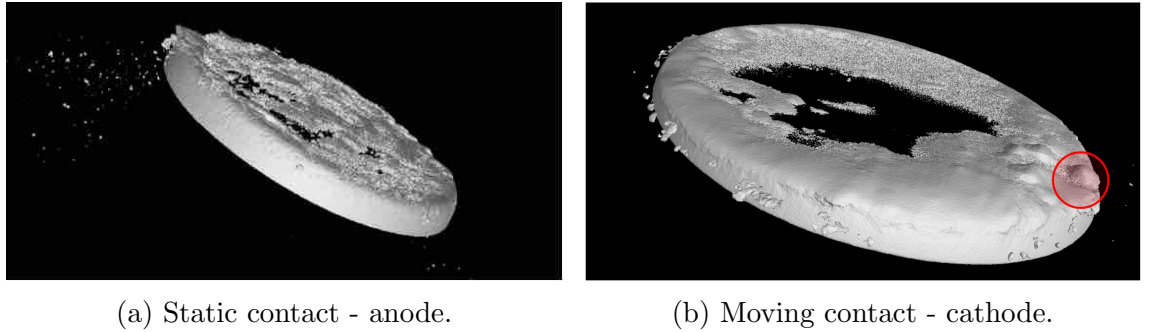


Figure 7.18: DA2-09U: Rendering of the CTS. The red circle indicates contact pips.

can be seen using CTS in Figure 7.17-(a) and Figure 7.18-(a). Unlike the sealed EMRs, the unsealed EMRs did not experience excessive bouncing during contact making beyond the maximum allowed $t_{m,max}$. Therefore, it can be argued that the gross of electrical erosion in unsealed EMRs must have occurred during contact breaking. Since erosion from electrical arcing does not seem to be as evenly distributed over the entire contact surface, it is likely that the arc developed a static anode-fall and cathode-fall region. The distribution of the contact asperities underpins this hypothesis. The predominant loss that led to the formation of crater structures on the anode appears to be most severe in the bottom section of the contact's surface, cf. Figure 7.17-(a)-(IV). Also note, the sputtered contact material that is distributed around the contact rivet, partially consisting of oxidised contact plating material and carbons, cf. Appendix B.5. Correspondingly, the cathode's surface gained material at the bottom section. Although, the unsealed EMRs experience comparably less material loss, they also exhibit the typical pip structure on the cathode. Following this line of evidence, the physical appearance of the anode and the cathode indicates that the material transfer took place predominantly during the metallic-phase arc rather than during the ambient-phase arc.

7.2.2 The influence of the operational environment on CR

The presence or absence of oxygen is critical to the occurrence of silver-based electrical contact failure in EMRs as the preceding examination has confirmed. Comparative testing of unsealed and sealed EMRs showed that, despite identical contact construction and material, different deformations may occur. In turn, these different reactions have a direct impact on CR and its relevance as a prognostics performance indicator.

To clarify the subject matter at hand, this section examines research relating to silver-based electrical contacts. By making use of CR measurements from DA2 it will be shown how fluctuations impact the prognostic value of CR. Furthermore, referring to Chapter 3.2, the root causes responsible for different degrees of fluctuations depending on the presence or absence of oxygen are integrated in the argument.

The presence of oxygen affects both the arc in the metallic-phase and the arc in the gaseous-phase. However, they exhibit different deposition mechanisms which are relevant in the context of contact oxidation [13].

Metallic-phase arc The relative short metallic-phase arc removes anode contamination films due to electron sputtering which, in turn, leads to a material transfer from anode to cathode, cf. Chapter 3.2.4.1. In consequence, the cathode surface is tainted during the metallic-phase arc. In the presence of oxygen the minimum arc voltage to sustain a metallic-phase arc is lowered (unsealed EMRs) compared to the minimum arc voltage in a nitrogen atmosphere (sealed EMRs). A reduction of the minimum arc voltage can be observed even for very small concentrations of oxygen. Recall that the metallic-phase arc is operating in a high-pressure zone within the metal-vapour matrix. For this reason, one might expect the type of atmosphere playing a neglectable role in its effect on the metallic-phase arc. However, if the required minimum arc voltage is reduced despite sealing, it has to be concluded that preceding oxidation of the contact surface must already have taken place, e.g. during storage.

Gaseous-phase arc If the load current is high, the duration of the gaseous-phase arc lengthens and degradation mechanisms become predominant during this phase. The direction of material transfer is reversed during the gaseous-phase arc, transporting material from the cathode to the anode, leading to an anode material accumulation. Unlike the metallic-phase arc, the gaseous-phase arc is exposed to higher concentrations of oxygen within the contact gap whenever the contacts are operated in ambient air. Such arcing regimes leave visible, dark oxidation films on the anode surface and the surrounding contact carrier strip forming non-conductive contamination layers, cf. the example of an unsealed EMR in Figure 7.17-(I). In case the

cathode is subjected to ion-sputtering the cathode's surface is cleaned. A local reduction in thickness of the developed oxide contact films promotes the formation of metallic clusters on the contact surface [14]. As the silver-oxides are only stable at lower temperature the majority of silver-oxides stem from silver being oxidised in the extinguishing gaseous-phase arc.

Due to the reduced minimum arc voltage and contact surface contamination effects from silver-oxide layers, EMRs that are equipped with silver-based contacts operating in normal, ambient air are prone to erratic fluctuations of CR. [15] provides a useful point of reference. It is demonstrated that for load currents beyond 0.2 A the CR either starts fluctuating after an initially stable phase or from the start throughout the entire duration of the contact's life. The rate and amplitude of these fluctuations increases with the applied electrical load. This can be linked to the growing fretting rate due to ongoing contact erosion, since arcing becomes the dominant degradation driver over the plastic deformation encountered at lower load currents. If arcing is only of short duration, e.g. the metallic-phase arc is predominant, the amplitude of CR fluctuations is generally low. Whenever shorter arcing changes to extended arcing, an almost instantaneous increase of the amplitude of CR fluctuations can be observed [16]. In that case, high CR values can be typically observed after several, long-duration gaseous-phase arcing events (> 1 ms) [16, 17]. Relying on surface roughness measurements, it has been concluded that such an increase and subsequent decrease of CR is attributed to changes in the contact surface. Non-uniform deposition of transferred material alters the contact morphology and the effective contact area respectively, chaining the contact spots at each making and breaking operation. However, as [14, 17] experimentally demonstrate, these CR fluctuations are dominated by non-conductive oxide film formation on the contact surface surpassing the effects of contact morphology alterations when the contact load increases. An analysis of the contact surface reveals build-up of absorbed oxygen contaminating the contact surface; these oxidised spots are being preferable hit by the arc root [13]. It can be deduced that prominent CR fluctuations are an indication of contact wear mainly due to electrical erosion. Low CR fluctuations further suggest that contact contamination by silver-oxides and, thus, the presence of oxygen plays a subordinate role.

7.2.3 Analysis of DA2 CR and temperature data

In order to illustrate the challenges if solely relying on CR as DI for contact health monitoring, i.e. for diagnostic and prognostic purposes, consider the CR measurements recorded every 10000th actuations during the conducted EMR life-cycle experiments (($AgSnO_2In_2O_3$) plated EMR contact rivets). In Figure 7.19, sealed EMR contacts are switched, isolated through an enclosure to oxygen from the ambient air. Initially, the contacts exhibit high CR fluctuations due to the *burn-in* phase

where small trace amounts of residue oxygen have already been deposited on the surface. This has been confirmed by [18], stating that initial oxide deposition can occur on EMR contacts despite sealing. However, at around 10 % of the EMR's life the CR stabilises and increases continuously till the EOL for the majority of the EMRs without problematic fluctuations ⁵.

On the other hand, consider the temperature profiles of the sealed EMR contacts (to be more specific the EMR enclosure temperature). Figure 7.20 depicts the changes in temperature, whilst the ambient temperature is kept constant at 30 °C. The average operating temperature settles ± 50 °C. A clear upwards trend in the last 20 % of the EMR's life is recognisable. This increases the average operating temperature from ± 50 °C up to ± 60 °C. The heat build-up can be attributed to a set of interrelated mechanisms going hand in hand in with the continuous increase in CR, e.g. contact erosion and debris build-up from fretting restrict the current carrying path. If, through continues erosion of contact material, the contact making and contact breaking becomes increasingly erratic more heat will be emitted during switching. In particular, the increased heat build-up towards the EOL may have a reciprocal effect on the moving contact carrier. A reduction in the stiffness of the moving contact carrier as it heats up reduces the contact force and contact velocity during contact breaking. In turn, this further increases AT and, therefore, accelerates the electrical erosion of the contacts as well as heat emission from arcing events.

The CR measurements relating to silver-based contacts which have been exposed to ambient air, i.e. unsealed EMRs, are displayed in Figure 7.21. In view of the above, the expected significant CR fluctuations of up to 100 $m\Omega$ are observed throughout the entire life of the EMRs, in general exceeding the maximum nominal value of allowed CR. Such *anomalies* are indistinguishable from the final rise in CR prior to the EOL and mask any underlying trend, cf. [19]. These findings further align with results reported [14, 16, 17, 20]. In this context the operational condition can be further analysed based on the contact temperature, cf. Figure 7.22. The tested unsealed EMR contacts settle at an average temperature of around ± 55 °C, approximately ± 5 °C higher than it has been the case for the sealed EMR. The generally higher operating temperature may indicate that the contacts of the unsealed EMRs are more contaminated by silver-oxides than the ones of unsealed EMRs and, therefore, fewer current-carrying paths exists.

⁵Some CR and temperature measurements of DA2-03S and DA2-05U are not available due to a software error relating to the Ohmmeter buffer during the experiments.

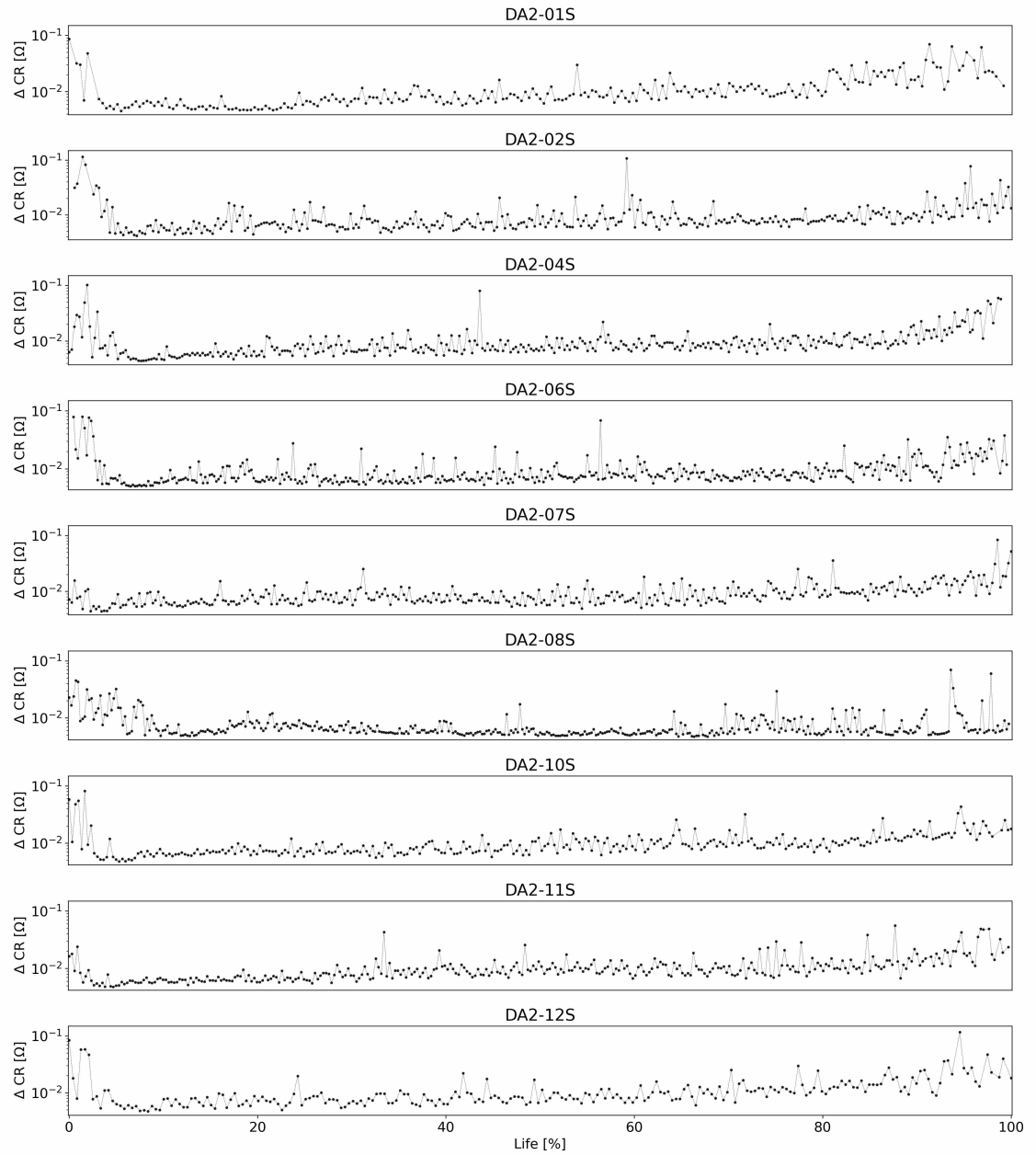


Figure 7.19: DA2 - Changes of sealed EMR CR throughout the life compared to the initial CR.

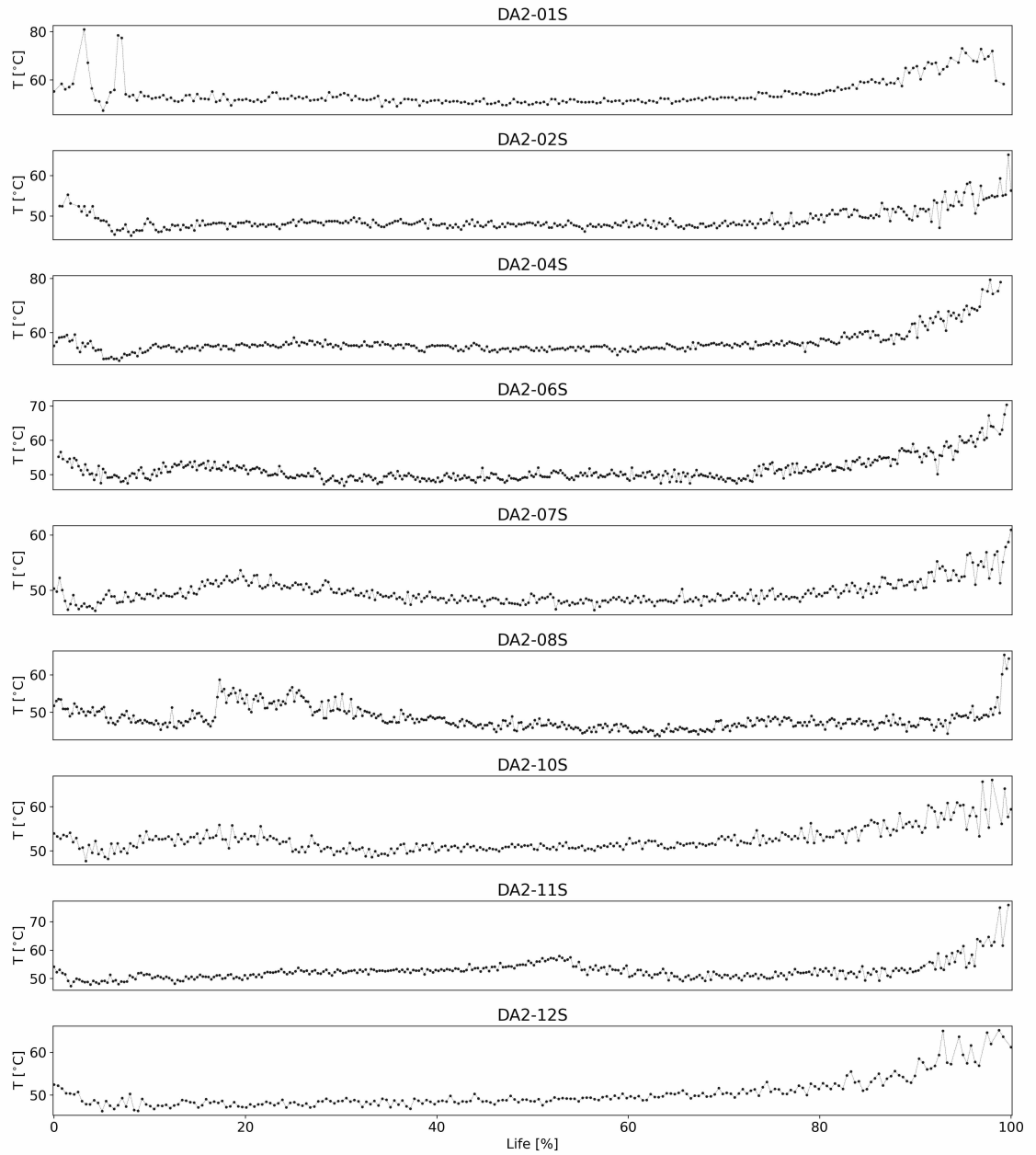


Figure 7.20: DA2 - Sealed EMR contact temperature measurements at constant ambient temperature of 30 °C.

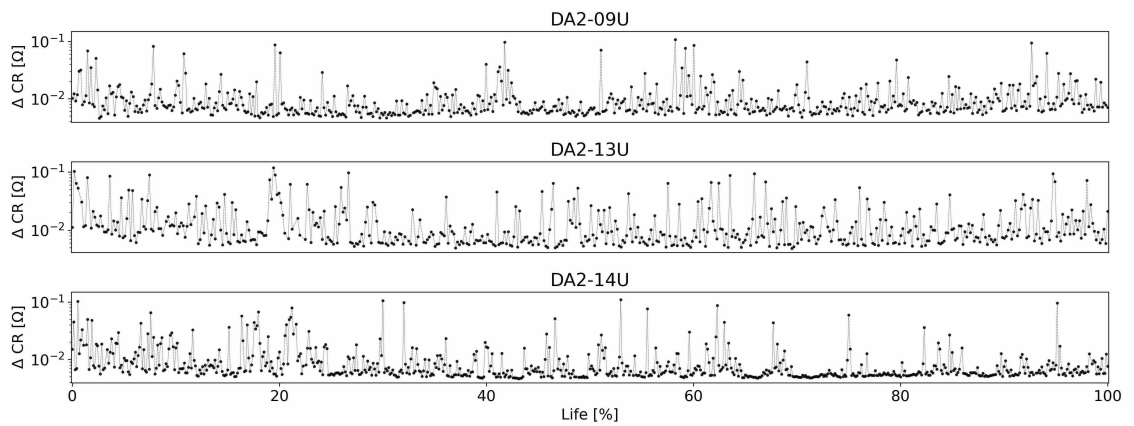


Figure 7.21: DA2 - Changes of unsealed EMR CR measurements throughout the life compared to the initial CR.

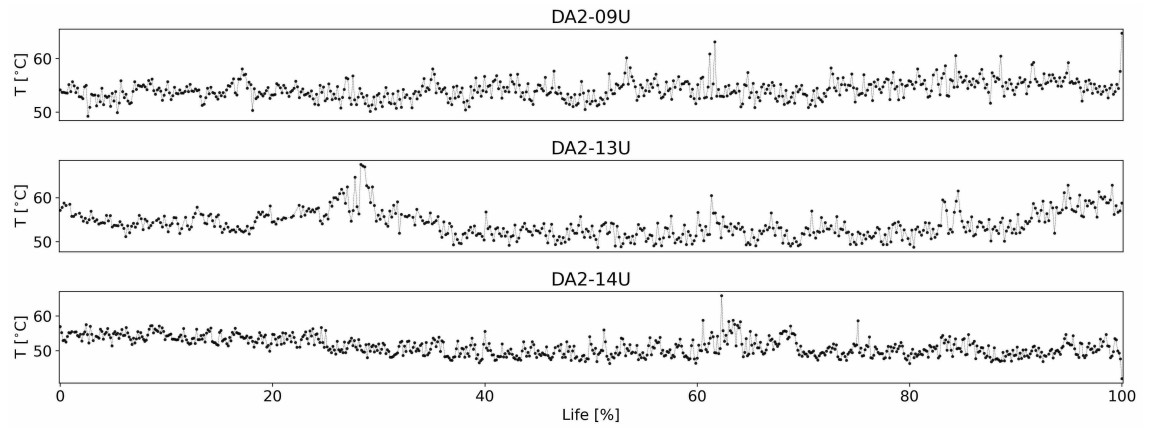


Figure 7.22: DA2 - Unsealed EMR contact temperature measurements at constant ambient temperature of 30 °C.

7.2.4 Implications for the usage of CR

As became clear, the formation of EMR contact failure depends on a multitude of interwoven factors. The manifestation of the failure gradually increases, whereby a significant acceleration of the failure regime appears only in the very last several of hundred actuations. Particularly, in the case of silver-based contacts the presence of oxygen widely controls the degradation regime. This in turn affects the applicability of CR as DI, so that CR measurements for unsealed EMRs are basically without merit, since high amplitude fluctuations will mask any underlying trend. The gross of the CR fluctuation can be attributed to:

1. a repetitive cycle of deposition of contaminating silver-oxide films on the contact surface, increasing the CR and
2. ion-sputtering cleaning the contacts, thus, reducing CR, cf. [14].

Furthermore, CR fluctuations may occur because contact degradation is accelerated by the reduction of conducting paths due to shifting oxide contamination patches as well as the oxide debris deposited on the surface. Mechanical rupture of oxide films is responsible for a sudden decrease of CR, predominantly observed for the unsealed EMRs. [15] remarks that oxide film formation is only dominant when small amounts of carbons are available, i.e. *contact-activation* does not play a significant role. However, if larger amounts of carbon are present the oxide layer will be replaced by a silver carbonate layer. In the case of silver-based EMR contacts that are not exposed to ambient air it has been shown that even the presence of small traces of oxygen present on the contact surface can affect the degradation behaviour initially. Here, oxygen reduces the minimum required arcing voltage and extends the arc duration while producing highlighted spots for such electrical arcing. In addition, the presence of contamination, the oxygen concentration in the ambient environment, the arc-duration, and slower contact velocity promote higher rates of oxidation and are to be linked to CR fluctuations [21, 22]. Finally, fluctuations of

CR can also be caused by silicon-vapours dissolved from, e.g. the EMR enclosure [23].

All things considered, the presence of oxygen alters the degradation behaviour of silver-based contacts. Hence, CR is unsuitable as DI for silver-based contacts exposed to ambient air, i.e. the case of unsealed EMRs, in a diagnostics or prognostics context. Sealed EMRs, on the other hand, are not subject to fluctuations beyond an initial *burn-in* phase. In this case CR can provide some means of assessing the contact health. However, it should be noted that the small increase in the $m\Omega$ range makes accurate measurement in practice difficult (and expensive). Recording the contact temperature may present an alternative. Processing the temperature measurements from sealed EMRs signals an upward trend, especially in the last 20 percent of the components' life. Since the temperature is comparatively easier to obtain than CR, its DI value for prognostics of EMR contacts might be an interesting issue to pursue.

Contrary to what [24] reports, there is no indication that sealed EMRs show longer contact life in comparison to unsealed EMRs. Despite lower CR observed for sealed EMRs in DA2, the observed contact life of sealed EMRs has been considerable shorter compared to unsealed EMRs (cf. Table 6.4 and Table 6.5). Due to the small sample size, one can only speculate about the reasons at this point. However, the static anode and cathode fall region of the sealed contacts are likely to be the root cause for the longer life of the unsealed EMRs. A locally constricted erosion zone leaves a majority of the contact surface virtually unaffected. Whereas the more evenly distributed degradation of the sealed EMR contacts, reduces the contact thickness across the entire surface. This, in turn, leads to earlier impaired contact making and excessive bouncing shortening the life-cycle.

7.3 EMR prognostics pipeline

This section presents and discusses the performance of the proposed EMRUA DL pipeline introduced in Chapter 5.2. The performance is evaluated on DA1, cf. Chapter 6.3.4 ⁶.

7.3.1 Derived feature sets

Three data sets of normalised features are considered, namely F_T , F_S and $F_{T,S}$ in the case of DA1. The features extracted for F_T consist of the mean and variance of time-based reference DIs over a set of actuations which can be extracted solely from the CV and CI waveforms. For the making actuation, the AT during making,

⁶*Parts of this chapter have already been published in my journal or conference article: L. Kirschbaum, et al., (2022).*

the BT, and the PT are considered. As for the breaking actuation, the RT and AT during breaking are studied. These features have been defined in Chapter 7.2 - Table 7.3. Since no CC measurements are provided in DA1, the ST is not considered. A set of statistical features F_S , namely the *min*, *max*, *mean* and *standard deviation* is aggregated differentiating between making and breaking actuations as detailed in Chapter 5.2.6.2. As stated before, in DA1 CV and CI waveforms are recorded respectively at 25 kHz over a duration of 0.04 s which yields 1000 samples per waveform per actuation. Features are derived for each raw waveform over an interval of 0.004 s which reduces the representation to 10 samples per waveform per actuation per feature. Compare Appendix C.1 for F_T features; compare Appendix C.2.1 and C.2.2 for features derived for F_S .

7.3.2 Hyperparameters and sequence selection

The model configuration, cf. Table 5.3, in conjunction with the size of the receptive field r , i.e. the number of elements in the input feature space that relate to the target RUA sequence (cf. Chapter 5.2.4), affects the model performance. While increasing the size of the receptive field r , increasing the kernel-size k or the number of residual blocks R also increases the number of model parameters as can be seen in the bottom of Figure 7.23-(I). The model performance does not necessarily improve with an increase in model complexity. In fact, apparent for data set F_T , an increase of stacked residual blocks R beyond 5 leads to performance deterioration increasing the MAPE respectively. For both, F_S and the combined data set $F_{T,S}$, all tested configurations exhibit similar results; though, some dependency on the architecture can be recognised.

Altering the dilation base d_b to increase the size of the receptive field r without increasing the number of model parameters does not improve performance, cf. Figure 7.23-(II). This might suggest that increasing the receptive field r beyond a certain threshold only improves performance with a parallel increase of trainable parameters. However, care should be taken to avoid overfitting.

As pointed out in the review of related literature, traditional features such as time-based reference DIs have been incorporated in the data set F_T . The performance of F_T incorporating PT, BT, and AT during contact making as well as RT and AT during contact breaking is displayed in Figure 7.24. Across all selection strategies, i.e. GI, LI, and EI the performance varies for the EMRs tested in DA1. However, the selection strategies EI exhibits less performance variance than the others. Comparing LI and EI, the prior one exhibits slightly better RUA estimation performance for the model $LI-k = 7-d_b = 2-R = 6$; model $EI-k = 3-d_b = 2-R = 6$ yields similar results having considerable less trainable parameters due to the smaller kernel-size

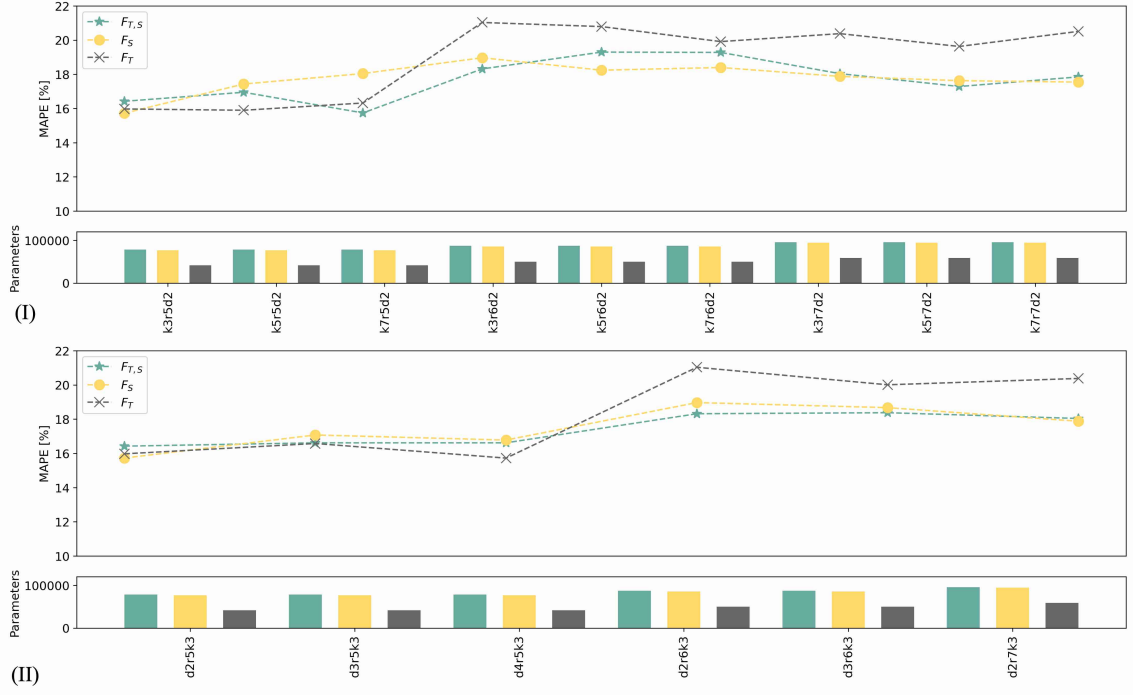


Figure 7.23: (I) Kernel k against Residual-Blocks R with constant dilation-base $d_b = 2$ for sequence-selection strategy GI; (II) Dilation-base d_b against Residual-Blocks R with constant Kernel $k = 3$ for sequence-selection strategy GI; performance averaged over all test samples.

($k = 3$ instead of $k = 7$).

The evaluation of the performance of the statistical data set F_S , displayed in Figure 7.25, yields consistent results for the LI and EI sequence selection strategy, simultaneously reducing the variance in performance compared to F_T . Only F_S -GI exhibits higher levels of variance among all tested configurations compared to F_S -LI and F_S -EI.

Combining the data sets F_T and F_S as a joint feature set $F_{T,S}$ does not generally improve the model performance. On the contrary, adding F_T seems to impair the overall performance in some instances. This is indicated by an increase in performance variance, in particular, for EI, cf. Figure 7.26.

With respect to the results in Figures 7.24, 7.25, and 7.26 examine the RUA estimation performance reported in Table 7.4. The results are averaged over all tested EMR samples in DA1 for different model configurations. The best performing model achieves an average MAPE= 12.4 % using the configuration EI- $k = 7$ - $d_b = 2$ - $R = 6$ trained on $F_{T,S}$ (87169 trainable parameters). With $\alpha_{sc} = 92\%$ the gross of all predictions reside within the $\alpha \pm$ confidence zone. The best performing model using only statistical features F_S is of similar configuration with $k = 5$ instead of $k = 7$

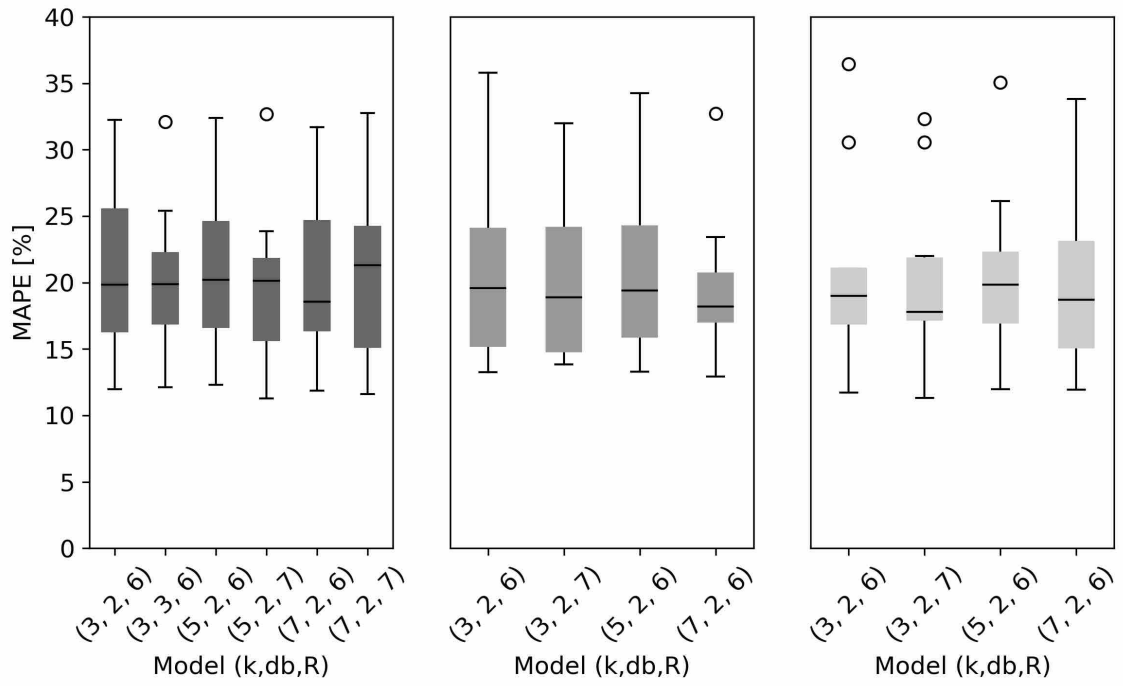


Figure 7.24: Performance results on all test samples of selected model configurations - F_T - left: GI, middle: LI, right: EI.

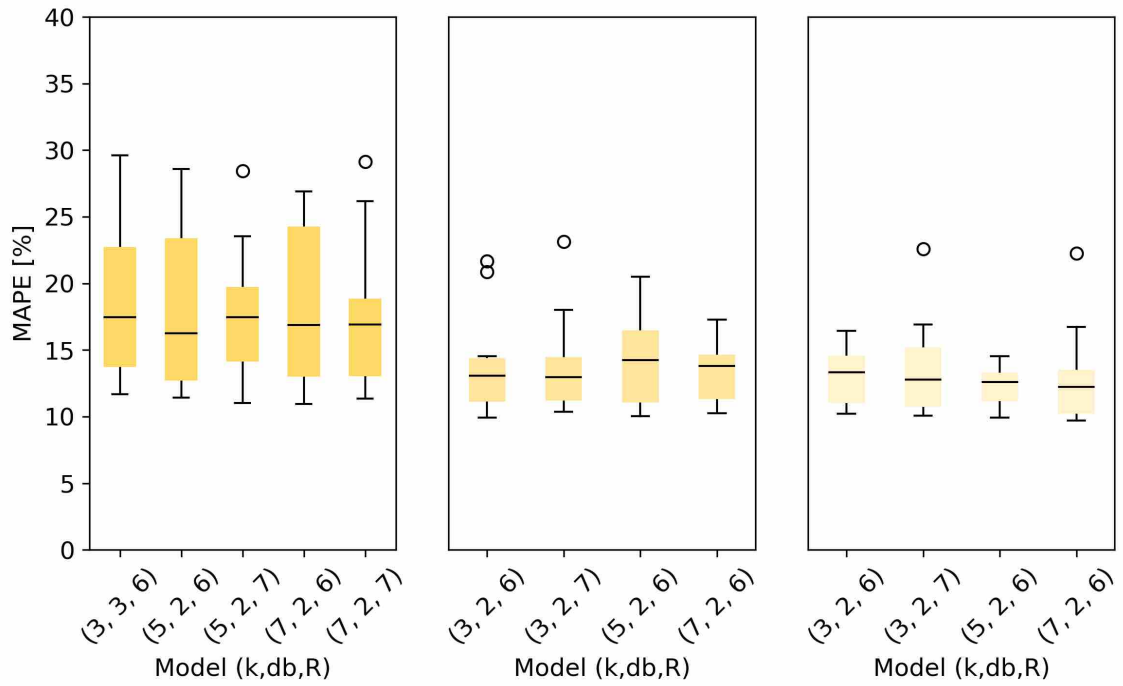


Figure 7.25: Performance results on all test samples of selected model configurations - F_S - left: GI, middle: LI, right: EI.

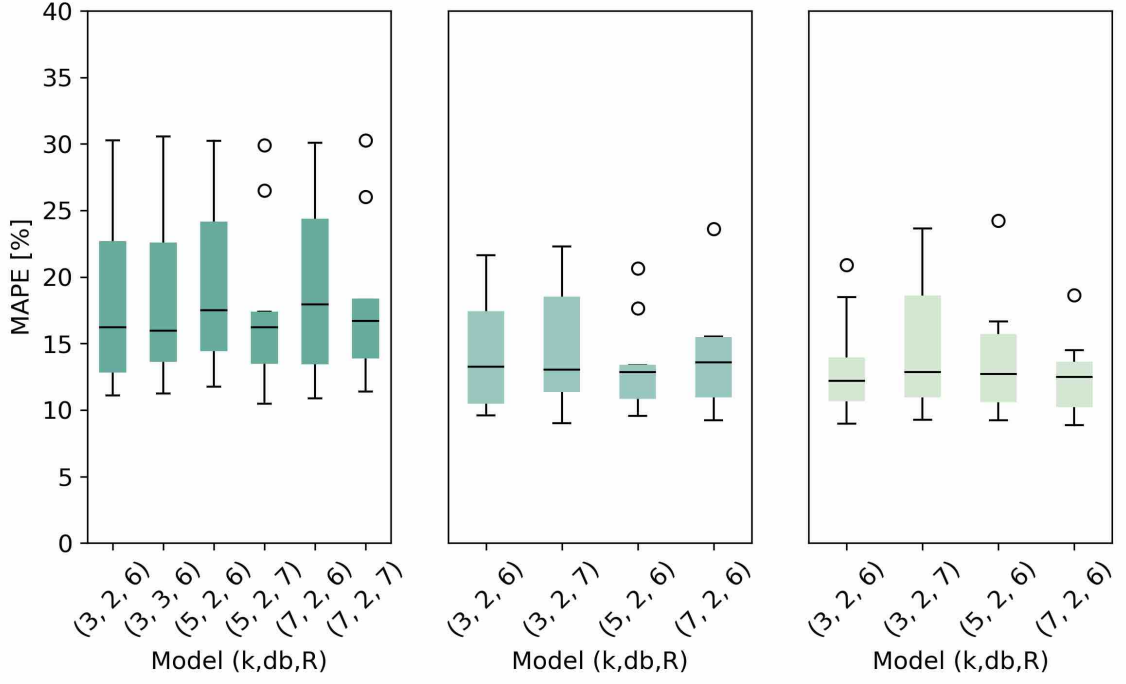


Figure 7.26: Performance results on all test samples of selected model configurations - $F_{T,S}$ - left: GI, middle: LI, right: EI.

$EI-k = 5-d_b = 2-R = 6$. This model requires slightly less parameters due to the reduced size of the input feature-map and kernel size (85889 trainable parameters). As already demonstrated, in general, EI and LI in combination with F_S or $F_{T,S}$ are superior to GI or F_T .

The findings - in terms of the selected feature sets - confirm the challenges of time-based reference DIs in a prognostics scope that have been previously pointed out in the literature review. Even the same batch of EMRs produces highly varying time-based reference DIs, partially due to fluctuations of, e.g. AT, RT, or PT during the fast contact making and breaking. Hence, such features do not necessarily provide the best performance nor robust results. The set of statistical features F_S yields somewhat more stable results for all tested EMRs and TCN configurations. Combining F_T and F_S as a joint feature set $F_{T,S}$ does not lead to a notable improvement. Considering the sequence-selection strategies - the performance of GI is worse than LI or EI. This might be attributed to the comparably smaller history coverage of each sequence point of inputs selected via GI in the TCN. No significant difference in performance between LI and EI can be recognised suggesting that recent changes in the degradation might not contribute significantly more to the average degradation rate and RUA estimation respectively.

Table 7.4: Average results of the best performing models using sub-sampling strategy S - EI or LI; sorted ascending according to MAPE; k - kernel; d_b - dilation base; R - Residual blocks; r - receptive field; F - data set; validation metrics MAE, MAPE, RMSE, and α_{sc} with $\alpha = 0.2$.

S	k	d_b	R	r	F	MAE	MAPE	RMSE	α_{sc}
EI	7	2	6	757	$F_{T,S}$	144.7	12.4	179.0	0.92
	5	2	6	505	F_S	140.3	12.4	174.4	0.92
	3	2	6	253	F_S	147.7	13.0	183.1	0.90
	7	2	6	757	F_S	154.6	13.1	187.7	0.91
	3	2	6	253	$F_{T,S}$	155.1	13.2	189.80	0.91
LI	5	2	6	505	$F_{T,S}$	157.2	13.3	190.6	0.89
	7	2	6	757	F_S	155.0	13.4	190.7	0.88
EI	5	2	6	505	$F_{T,S}$	162.3	13.7	195.3	0.91
	3	2	7	509	F_S	161.5	13.8	196.9	0.91
LI	7	2	6	757	$F_{T,S}$	164.9	13.9	198.3	0.88
	3	2	7	509	F_S	165.5	14.0	198.5	0.87
			6	253	F_S	165.4	14.0	198.7	0.88
					$F_{T,S}$	169.3	14.3	203.3	0.87
			7	509	$F_{T,S}$	176.3	14.5	210.6	0.87
	5	2	6	505	F_S	171.4	14.6	208.3	0.86
EI	3	2	7	509	$F_{T,S}$	177.2	14.7	210.8	0.91
LI	7	2	6	757	F_T	228.7	19.4	272.5	0.67
EI	3	2	7	509	F_T	229.9	20.1	273.6	0.73
	7	2	6	757	F_T	232.6	20.4	276.9	0.69
EI	5	2	6	505	F_T	229.4	20.5	275.8	0.71
LI	3	2	7	509	F_T	238.7	20.6	285.0	0.68
EI	3	2	6	253	F_T	233.8	20.6	278.3	0.72
LI	5	2	6	505	F_T	245.9	21.2	289.4	0.64
	3	2	6	253	F_T	242.3	21.2	287.3	0.62

7.3.3 EMRUA inference

Based on the presented strategy, RUA forecasts can support timely decisions for maintenance scheduling. The performance of different sequence-selection strategies and data sets in terms of RUA prediction is illustrated through an example using the configuration $EI-k = 4-d_b = 2-R = 6$. Forecasting at different stages of the EMR's life for F_T , F_S and $F_{T,S}$ is presented in Figure 7.27, 7.28, and 7.29. Again, purely relying on traditional time-based reference DIs might produce misleading results. The performance deteriorates closer to the EOL, cf. Figure 7.27. Further results for other EMRs and sequence selection strategies are presented in Appendix C.3.

A closer look at the number of forward passes N during EMRUA inference reveals their impact on the estimation of uncertainty. Consider the model $EI-k = 7-d_b = 2-R = 6$ and data set $F_{T,S}$ evaluated on EMR DA1-16S. Figure 7.30 visualises the inference phase of the EMRUA pipeline which can provide timely RUA estimation in order to be embedded in a maintenance solution, e.g. PdM. The RUA is estimated at three different a_t using MCD which provides the distribution of the RUA estimates. With the EMR approaching its EOL the RUA forecast improves. As mentioned before, the number of samples N equivalent to the batch size does not

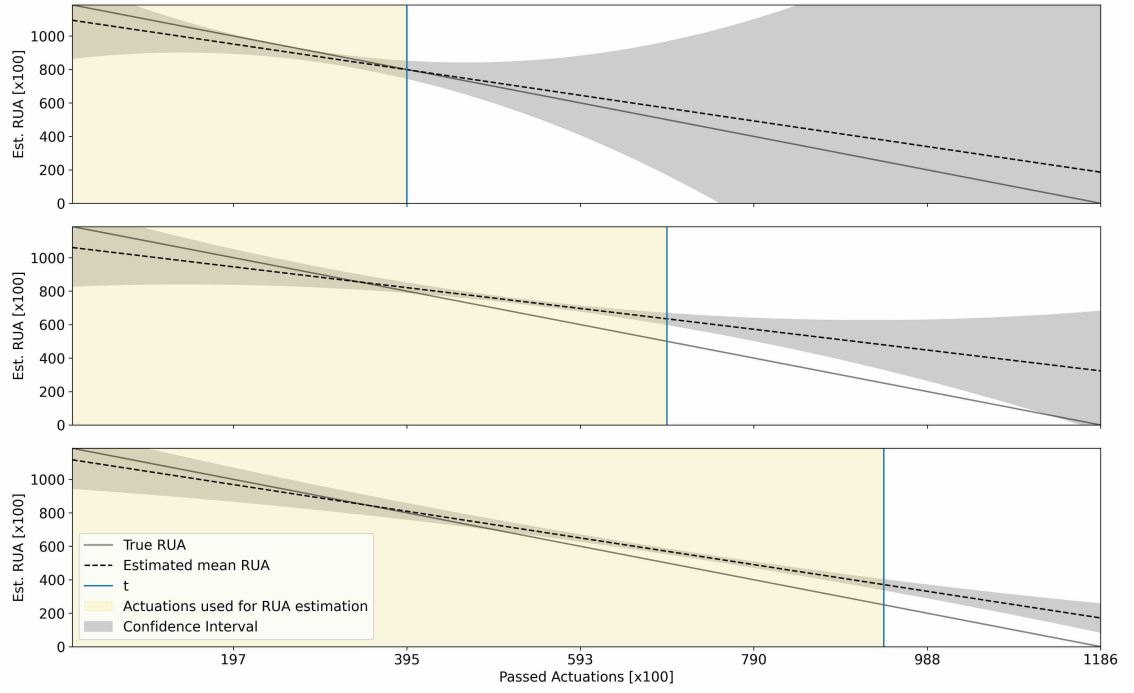


Figure 7.27: DA2-01U - F_T - EI - $k = 4 - d_b = 2 - R = 6$; prediction of RUA at actuation $a_t =$ (top) 40000, (centre) 70000 and (bottom) 90000.

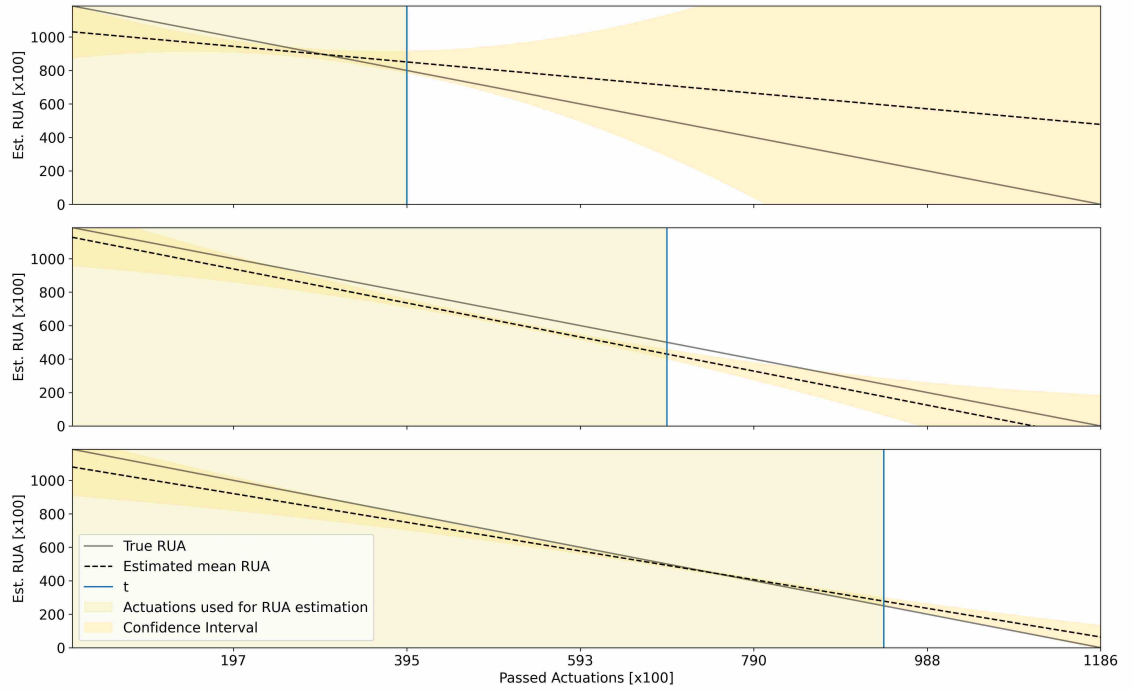


Figure 7.28: DA2-01U - F_S - EI - $k = 4 - d_b = 2 - R = 6$; prediction of RUA at actuation $a_t =$ (top) 40000, (centre) 70000 and (bottom) 90000.

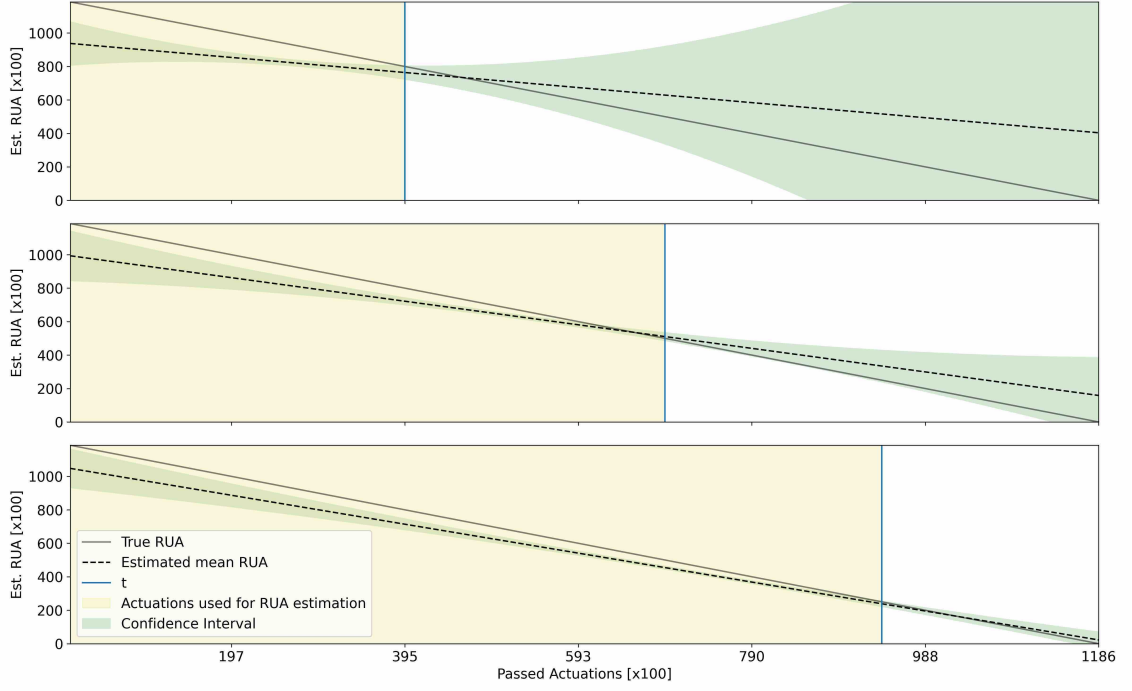


Figure 7.29: DA2-01U - $F_{T,S}$ - EI - $k = 4 - d_b = 2 - R = 6$; prediction of RUA at actuation $a_t =$ (top) 40000, (centre) 70000 and (bottom) 90000.

need to be large to retrieve a sufficient uncertainty representation. For further values of $N = [50, 500, 50000]$ refer to Appendix C.4 for a comparison. At $a_t = 60000$, i.e. 50000 actuations prior to failure, the estimated RUA exhibits wide confidence intervals as can be seen in Figure 7.30-(top) using $N = 5000$. The forecast improves distinctly 25000 actuations later at $a_t = 85000$, cf. Figure 7.30-(centre). Lastly, consider $a_t = 106000$ in Figure 7.30-(bottom), a_t lies within the estimated confidence interval of the RUA. Now, the predicted mean RUA is only 4900 actuations ahead. The example given underlines the interdependence of the forecast and the associated uncertainty. Therefore, N is an important hyperparameter to consider for any subsequent maintenance action; however, for this application case the selection of $N = 5000$ appears to be a reasonable choice.

7.3.4 Performance comparison with LSTM

In order to evidence the performance gains through TCN, the proposed DL architecture is compared to a reference LSTM model which has previously been employed in RUL forecasting proposed by [25]. The results are displayed in Table 7.5. TCN improves the RUA performance as the feature space increases, i.e. using F_S or $F_{T,S}$, despite using less parameters than the LSTM.

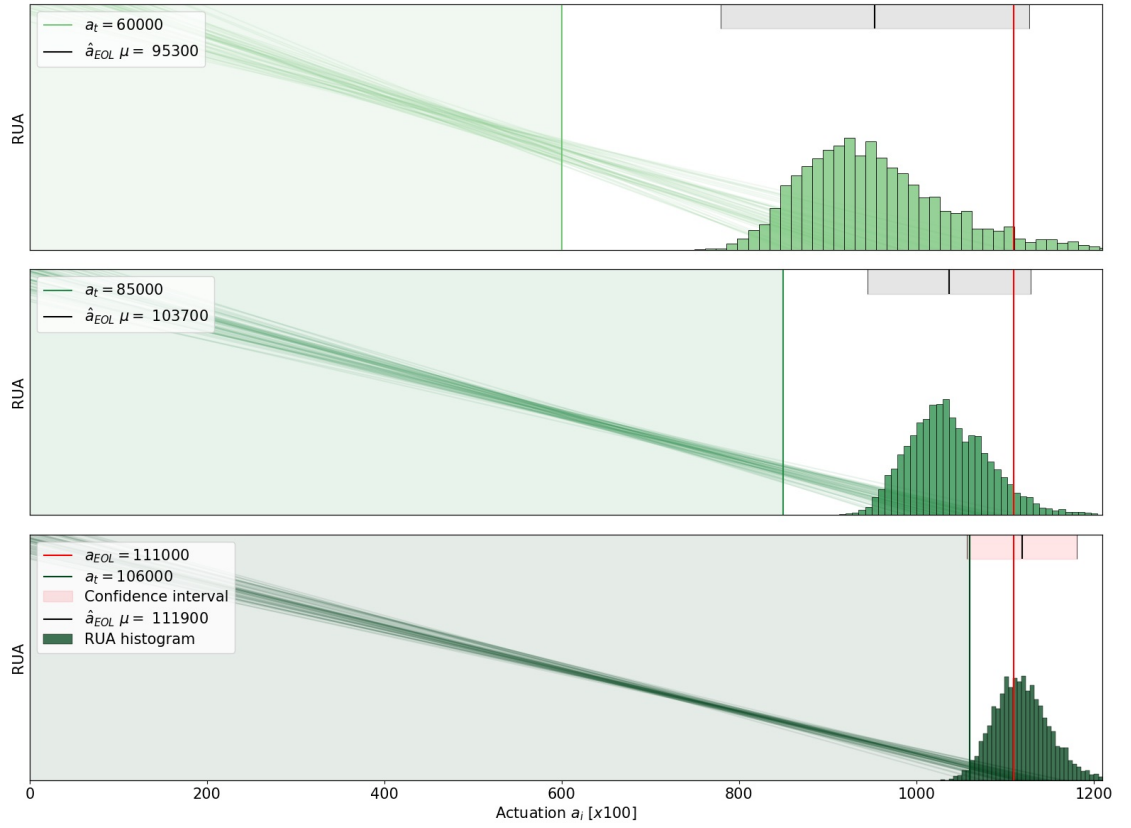


Figure 7.30: Exemplary predicting RUA for DA1-16U with $N = 5000$. The EI sequence selection strategy and the data set $F_{T,S}$ are used.

Table 7.5: Comparison of the best performing TCN and a reference LSTM architecture for the LI sub-sampling strategy.

Model	Data set	Parameters	MAE	MAPE	RMSE
TCN	F_T	50305	228.7	19.4	272.5
LSTM	F_T	54601	198.5	16.9	230.9
TCN	F_S	85889	155.0	13.4	190.7
LSTM	F_S	165801	223.4	18.8	276.0
TCN	$F_{T,S}$	87169	157.2	13.3	190.6
LSTM	$F_{T,S}$	169801	272.1	21.3	327.1

7.3.5 Discussion and research prospects

The literature reviewed in Chapter 3.2 and Chapter 4.2 as well as the failure analysis of DA2 in Chapter 7.2 imply that failure precursors are subject to specific failure mechanisms and, in turn, responsible for different failure modes. The analysis in this chapter focused on a particular failure mode, namely *stuck-closed* failure in DA1. The results demonstrate that high variance is present among the individual EMR samples in the data set. Given that sources of uncertainty could be further reduced through collection of additional data, the proposed configuration of TCN shows promising results.

A possible direction for future research relates to methods that will improve EMRUA through alternative ways of feature extraction, extensively discussed in Chapter 5.2.2. To name some possible approaches:

- Fast-Fourier-Transform;
- automated feature extraction through CAE;
- specific techniques to address the challenge of the highly localised information-holding section of the contact switching waveform that varies in the time domain, e.g. time series Motifs, cf. [26].

Training individual models for different failure modes might be necessary. In addition, further investigation of the method's robustness to changes regarding the volume and granularity of available training data and the sampling rate is required. Optimisation of the hyperparameter selection through, e.g. grid-search or randomised-search as it has been deployed for the BHA-PCBA in this dissertation should be considered.

Transfer-learning is another area of interest, bearing the potential to reduce training time not only among models pertinent to different failure modes but also in respect to a reduction of the required training data when changing the EMR type, e.g. a different contact material or design. Moreover, the value of alternative measurements for EMR health estimation could be investigated, e.g. the contact temperature using an infrared temperature sensor as already shortly alluded to in Chapter 7.2. Sensing such waveforms and extracting meaningful features might improve the model's performance, though the practicality and retrofittability of such method is limited.

7.3.6 Computational hardware and software

The LSTM has been implemented using `tensorflow` (version 2.6.0) [27] and `keras` (version 2.6.0) [28]. The TCN has been implemented using `tensorflow` (version 2.6.0) [27] and `keras-tcn` (version 3.4.0) [29]. A NVIDIA GeForce GTX 1060 (3GB) graphic card and CUDA [30] has been used for training. Training was conducted over a maximum of 1000 epochs, though a `EarlyStopping` callback has been used

to stop training if the validation error did not improve over 10 actuations. The learning rate has been adjusted during training using the `LearningRateSchedule` callback. Due to the size of the input data, `generators` were employed to provide the data to the model during training. The maximum encountered training time for either TCN or LSTM did not exceed 24 hours.

Summary

As this chapter has made clear, both, ML and DL have the potential to provide viable solutions for data-driven maintenance support. It has become evident that methods must be tailored to their respective background and, most importantly, take into account the data requirements and the applicability of the prognostic feedback.

The BHA-PCBA framework provides estimates of the susceptibility to failure of electronics installed in downhole tools. However, limited by the resolution of the data, it is not possible to isolate the exact location of a fault. Nevertheless, such an approach supports the maintenance process and enables data-driven decisions making.

EMRUA as a prognostics method can drive PdM, complementing existing EMR maintenance decision making paradigms. The method offers additional confidence in EMR performance evaluation, important within applications where EMRs play a mission or safety critical role. EMRUA has the capability to provide real-time insights of the individual EMR health with the associated uncertainty estimate, rather than blindly relying on conventional maintenance measures. To this end, EMRUA only relies on CV and CC measurements, recordings that are already commonly collected in many safety critical systems.

References

- [1] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [2] Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 785–794, New York, NY, USA, 2016. ACM.
- [3] L Toloşi and T Lengauer. Classification with correlated features: unreliability of feature ranking and solutions. *Bioinformatics*, 27(14):1986–1994, 2011. doi:10.1093/bioinformatics/btr300.
- [4] J Sheppard, M Kaufman, and T Wilmering. IEEE standards for prognostics and health management. In *IEEE Aerospace and Electronic Systems Magazine*, volume 24, pages 97–103. IEEE, 2009. doi:10.1109/MAES.2009.5282287.
- [5] L Morin, N Jemaa, D Jeannot, J Pinard, and L Nedelec. Contacts materials performances under break arc in automotive applications. *IEEE Transactions on Components and Packaging Technologies*, 23(2):367–375, 2000. doi:10.1109/6144.846776.

- [6] A Ksiazkiewicz and J Janiszewski. Welding tendency for selected contact materials under different switching conditions. *Eksploatacja i Niezawodność – Maintenance and Reliability*, 21:237–245, 2019.
- [7] T Mützel and R Niederreuther. The influence of switching arcs on contact resistance of Ag/SnO₂ materials. In *61st IEEE Holm Conference on Electrical Contacts*, pages 171–175. IEEE, 2015. doi:10.1109/HOLM.2015.7355092.
- [8] D McDonnell, J Gardener, and J Gondusky. Comparison of the switching behaviour of silver metal oxide contact materials. In *IEEE Holm Conference on Electrical Contacts*, pages 37–43. IEEE, 1993. doi:10.1109/HOLM.1993.489658.
- [9] L Yi and N Lu. Evaluation and comparison of the arc erosion characteristic of Ag/SnO₂/In₂O₃ contact materials used for an automotive relay. In *64th IEEE Holm Conference on Electrical Contacts*, pages 387–393. IEEE, 2018. doi:10.1109/HOLM.2018.8611673.
- [10] C Leung. Arcing contact materials, silver refractory metals. *Encyclopedia of Tribology*, Springer US, pages 104–107, 2013. doi:10.1007/978-0-387-92897-5_404.
- [11] J Swingler and J McBride. The erosion and arc characteristics of AgCdO and AgSnO₂ contact materials under DC break conditions. *IEEE Transactions on Components, Packaging, and Manufacturing Technology: Part A*, 19(3):404–415, 1996. doi:10.1109/95.536842.
- [12] J Swingler and A Sumption. Arc erosion of AgSnO₂ electrical contacts at different stages of a break operation. *Rare Metals*, 29(3):248–254, 2010. doi:10.1007/s12598-010-0043-1.
- [13] D Grogg and C Schrank. Impact of the gas environment on the electric arc. In *62nd IEEE Holm Conference on Electrical Contacts*, pages 125–128. IEEE, 2016. doi:10.1109/HOLM.2016.7780019.
- [14] Z Chen and K Sawa. Effect of oxide films and arc duration characteristics on Ag contact resistance behavior. *IEEE Transactions on Components, Packaging, and Manufacturing Technology: Part A*, 18(2):409–416, 1995. doi:10.1109/95.390325.
- [15] P Slade. *Electrical contacts: principles and applications*. CRC press, 2017.
- [16] H Sone, H Sugimoto, and T Takagi. A measurement on contact resistance and surface profile of arcing ag contacts. In *IEEE Holm Conference on Electrical Contracts*, pages 105–109. IEEE, 1994. doi:10.1109/HOLM.1994.636826.
- [17] Z Chen, H Mizukoshi, and K Sawa. Contact resistance characteristics of Ag material in breaking low-load DC arcs. *IEEE Transactions on Components, Packaging, and Manufacturing Technology: Part A*, 17(1):113–120, 1994. doi:10.1109/95.296376.
- [18] Z Wang, W Li, K Chen, Z Li, and S Shang. Storage failure mechanism analysis and reliability improvement measures for electromagnetic relay. In *Conference Series: Materials Science and Engineering*, volume 1043, page 042058. IOP Publishing, 2021. doi:10.1088/1757-899X/1043/4/042058.
- [19] A Wileman and S Perinpanayagam. Integrated vehicle health management: An approach to dealing with lifetime prediction considerations on relays. *Microelectronics Reliability*, 55(9-10):2165–2171, 2015. doi:10.1016/j.microrel.2015.06.013.
- [20] K Li, F Yao, J Lu, and Z Li. Test and analysis of reliability for electromagnetic relay. In *46th IEEE Holm Conference on Electrical Contacts*, pages 79–82. IEEE, 2000. doi:10.1109/HOLM.2000.889915.
- [21] A Wileman and S Perinpanayagam. A prognostic framework for electromagnetic relay contacts. In *PHM Society European Conference*, volume 2, pages 1–7. PHM Society, 2014. doi:10.36001/phme.2014.v2i1.1531.
- [22] X Cui, X Zhou, G Zhai, Z Fang, and F Liu. Electrical lifespan prediction of HVDC relay based on the accumulated arc erosion mass. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 8(3):356–363, 2017. doi:10.1109/TCPMT.2017.2768222.
- [23] M Hasegawa, N Kobayashi, and Y Kohno. Contact resistance characteristics of relays operated in silicone-vapor-containing and non-silicone atmospheres with different electrical load conditions. In *57th Holm Conference on Electrical Contacts (Holm)*, pages 1–7. IEEE, 2011. doi:10.1109/HOLM.2011.6034787.
- [24] G Witter and L Polevoy. Contact erosion and material transfer for contacts in automotive relays. In *42nd IEEE Holm Conference on Electrical Contacts*, pages 223–228. IEEE, 1996. doi:10.1109/HOLM.1996.557201.
- [25] S Zheng, K Ristovski, A Farahat, and C Gupta. Long short-term memory network for remaining useful life estimation. In *2017 IEEE international conference on prognostics and health management (ICPHM)*, pages 88–95. IEEE, 2017. doi:10.1109/ICPHM.2017.7998311.

- [26] A Mueen, E Keogh, Q Zhu, S Cash, and B Westover. Exact discovery of time series motifs. In *SIAM international conference on data mining*, pages 473–484. SIAM, 2009. doi:10.1137/1.9781611972795.41.
- [27] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
- [28] François Chollet et al. Keras. <https://keras.io>, 2015.
- [29] Philippe Remy. Temporal convolutional networks for keras, 2020.
- [30] NVIDIA, Péter Vingelmann, and Frank H.P. Fitzek. Cuda, release: 10.2.89, 2020.

Chapter 8

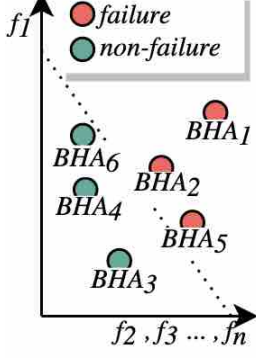
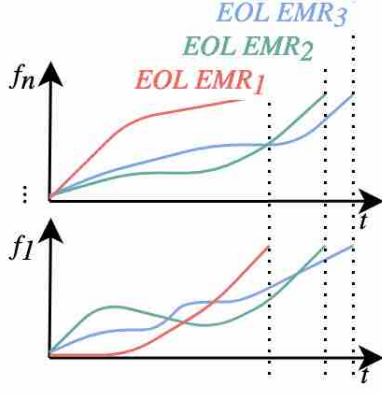
Conclusion

Recall the proposed research question. From this point of view, I explored and created methods that, despite reliance on AI, underpin the criticality of custom-tailoring maintenance strategies for industrial applications. Thus, a thorough understanding of the relevant degradation is an indispensable prerequisite, be it in order to develop a suitable life-cycle experiment or to appraise the forecast capabilities of a model. Given the limitations of traditional maintenance strategies, my thesis evidences how ML and DL pipelines fused into a prognostics concept advance the state of the art for the given applications. It is illustrated that prognostics for individual components can yield targeted, highly expressive forecasts, though this approach may not be feasible in instances where data resolution prevents such a procedure. Therefore, two distinct data-driven frameworks are proposed. Their adoption supports current maintenance procedures, facilitates decision making and, consequently, de-risks active operations.

Reliability, maintenance, diagnostics, and prognostics have been discussed highlighting that industries are actively shifting from classical reliability paradigms towards data-driven PHM in order to continue to meet stringent operational and safety requirements of increasingly complex industrial systems. It has been argued that pursuing the digital transformation is crucial, in particular for the oil and gas industry given the prevailing economic situation. Especially in the upstream sector, the intricate nature of drilling systems and the dynamic operating environment combined with a competitive market and prolonged tool design cycles poses a challenge to present AMO processes. A more detailed examination of current maintenance approaches for mission-critical electronic components and assemblies used in drilling tools has made this problem particularly clear. The inadequacy of the presently applied methods, illustrated by the high failure rates of electronics, contrasts with the optimisations that can be achieved through the implementation of data-driven techniques. Moreover, it has been shown that the potential of diagnostics and prognostics to accelerate future cost savings is readily acknowledged by both the relevant

research community and industry.

Table 8.1: Methodological comparison of operational needs of the developed BHA-PCBA and EMRUA frameworks

	BHA-PCBA	EMRUA
Aim	Support the AMO process through PdM integration by classifying whether BHA-PCBA circuitry has failed in a previous operation and, hence, maintenance on the PCBAs should be conducted.	EMR-RUA estimation throughout the component's life considering the forecast uncertainty.
Level	Fleet-level for BHA tools each holding multiple electrical assemblies within the sub-modules.	Component-level for an individual electromechanical component i.e. EMR.
Model	Supervised classification: RFC, XG-Boost Classifier	Supervised regression: TCN
Data	1D feature vector with n features holding summarised operation statistics and BHA diagnostic information for each BHA tool; the label corresponds with an operational class (<i>failure</i> and <i>no-failure</i>).	2D feature vector comprising processed MVTD from CV and CI waveforms sampled for every making and breaking actuation; label corresponds to the RUA.
Train		
Predict	Predict class probability of BHA-PCBA <i>failure</i> and <i>no-failure</i> class for a single BHA tool under consideration of possible failure costs and fleet reliability levels.	Predict RUA for one EMR at t using the streamed MVTD for the interval $[0, t]$ (multiple forward passes in one batch according to MCD). Determine the best linear-fit for each predicted RUA sequence and extrapolate to the estimated EOL to derive confidence intervals for the RUA estimate.
Chapters	3.1, 4.1, 5.1, 7.1	3.2, 4.2, 5.2, 6, 7.2, 7.3

Following the identification of these research gaps, two tailored data-driven methodologies are put forward: (1) estimation of BHA-PCBA failures; (2) forecasting EMR-RUA. Table 8.1 presents a methodological comparison of both approaches which highlights the distinct differences in the fleet-level and component-level use case scenarios in regard to the predictive targets and, therefore, the maintenance support capabilities. An ALT experiment design has been proposed and implemented in order to address the limited availability of EMR life-cycle data. It facilitates the generation of high resolution *run-to-failure* data. The proposed EMRUA method

has been trained and tested using this data. Likewise, encountered failure mechanisms were analysed and the emergence of individual failure modes evaluated. The remainder of this chapter details the individual contributions and proposes strategies to further advance them.

8.1 BHA-PCBA maintenance support framework

Following a thorough review of health management strategies for downhole drilling tools, the electrical assemblies within are identified as one of the major critical subsystem in the BHA due to the reported high failure rates. Determining failure precursors and the need for maintenance of individual electrical components is challenging, time intensive, costly, and often infeasible within circuits holding many components such as the BHA-PCBAs. In addition, it has been shown that current maintenance practices are inefficient owing to extreme environmental and dynamic operating conditions. The applicability of real-time monitoring and, in consequence, online prognostics is constrained, since data-transfer is limited during the drilling process. Furthermore, retrofitting sensing capabilities to legacy drilling tools in order to allow for targeted detection, isolation, and identification of electronic failures within PCBAs is not attainable because of the restriction imposed by space and hardware constraints. Moreover, exaggerated by the imposed governmental data-policies, the locally constrained storage of relevant data has a detrimental effect on data availability and, therefore, implementability. Despite all those reasons impeding the development of data-driven maintenance support for downhole tool electronics, this thesis proposes a novel approach.

Indifferent to the failure location or mode, a data-driven ML framework is introduced for BHA-PCBA fleet maintenance optimisation and operational support that is implemented and ready to be deployed in-field or during AMO. It is based on commonly collected, but so far not utilised BHA memory data. To increase robustness towards outliers and prevent overfitting, in a first instance the data is augmented by introducing adversarial samples in order to provide additional (simulated) training data. An analysis of the performance of two classification algorithms, RFC and XGBoost-Classifer, is conducted. Perturbation values of $\epsilon = 0.02$ or $\epsilon = 0.05$ served as the best augmentation subsets in order to distinguish between the *need for maintenance - (failure)* and the *availability for a re-run - no-failure* with a high accuracy of 90 %. Furthermore, it has been shown that increasing the class decision threshold reduces operational costs in high-risk applications which illustrate that the selection of an algorithm and its parameters should depend upon the distinct business needs. A tailored solution can be selected based on the need to increase revenue (lower NPT, lower number of failures, higher reliability) or to improve margins (lower AMO costs).

8.2 EMR life-cycle data and prognostics pipeline

EMRs are ubiquitous in electrical systems, thus, significant for electronic-rich industrial equipment such as that used in the upstream industry for deep drilling tools. Here, a data-driven maintenance paradigm has the potential to improve the reliability of systems containing EMRs. To facilitate the development of a prognostics method common EMR failure modes have been analysed, e.g. *high CR*, *stuck-closed* contacts or *stuck-open* contacts. To facilitate this analysis, failure mechanisms that act predominantly on the electrical contacts and lead to EMR fatigue have been studied, e.g. electrical arcing, contact welding, contact contamination, and fretting. Current practices for monitoring EMR contact health such as the CR are often constrained by the complexity of obtaining these measurements and, more importantly, by the environment EMRs are operated under. This has been evidenced for silver-based EMR contacts. In case the contacts are exposed to oxygen, recording CR for health monitoring purposes is superfluous due to significant fluctuations caused by the repetitive build-up and destruction of contaminating oxide layers on the contact surfaces. If the contacts are not exposed to oxygen, the CR, as has been shown, may hold some value for diagnostics or prognostics. It has been further demonstrated that assessing EMR reliability solely on the basis of time-based reference DIs is highly application dependent and limited in its generalisation.

Based on this state-of-the-art analysis, a method is put forward that mitigates the need for strenuous CR measurements and reduces the dependency on time-based reference DIs or CR while enabling continuous forecasting of the expected EMR's EOL. Although the methodology presents a novel approach which is evaluated for EMRs, the general concept is not limited to EMRs. The method is aligned to the challenge and opportunities of high volumes of MVTD. The proposed DL pipeline, termed EMRUA, utilises the aggregated EMR life-cycle data in order to provide an accurate RUA estimation, only relying on CV and CI switching waveforms. The use of this type of measurements is desirable because such condition monitoring data is already being recorded in many industrial settings where EMRs play a safety critical role.

TCN has been adopted as autoregressive DL strategy incorporating MCD based uncertainty quantification as a computationally efficient implementation deployed during inference. To support online maintenance decision making, the trade-off between model complexity and model performance has been examined. Therefore, the effects of specific TCN model hyperparameters (*dilation*, *dilation base*, *receptive field*, *number of residual blocks*, *kernel-size*) on the predictive performance under consideration of the amount of trainable parameters were investigated. Additionally, three different feature sets (time-based features, statistical features, and a combination of both) and sub-sampling strategies (GI, LI, EI) have been explored. The

results indicate that TCN achieves the best performance and the lowest forecasting error ($\text{MAPE} = \pm 12\%$) using solely statistical features or in combination with time-based reference DIs and either LI or EI sub-sampling. Furthermore, it was shown that in some instances classical, time-based reference DIs can adversely affect the forecasting performance.

In summary, based on the collected EMR life-cycle data, it has been demonstrated how TCN can be fused into a prognostic method. The proposed approach emphasises the potential adoption of DL for PdM given high volumes of MVTD. This will aid future research utilising DL - in particular for EMR - to develop novel, data-driven maintenance solutions.

8.3 General suggestions for future work

Given the conditional constraints, this work can be extended in several directions for both case studies. For one, aspects of generalisation, transfer, verification, and robustness evaluation, as well as the economics of each respective approach are to be further examined. Foremost, extending the number of considered samples subjected to different failure modes and a wider spectrum of possible operating parameters would be beneficial to improve performance. Likewise, considering the means of explainability on model reasoning and leveraging the concept of, e.g. humble AI could support building trust in the potential of the proposed methods. Finally, the requirements for the implementation of such forecasts into an automated maintenance action in the form of PrM remain an open question.

Appendices

Appendix A

Experimental design

A.1 ALT-EMR test PCB

The designed test board consists of (a) the attachable PCB (the test EMR is soldered onto this PCB); (b) the main board with terminals to connect sensors, loads and power supplies; (c) the DPDT-EMR to switch in the CR measurement circuit; (d) the current sensor for CI measurements.

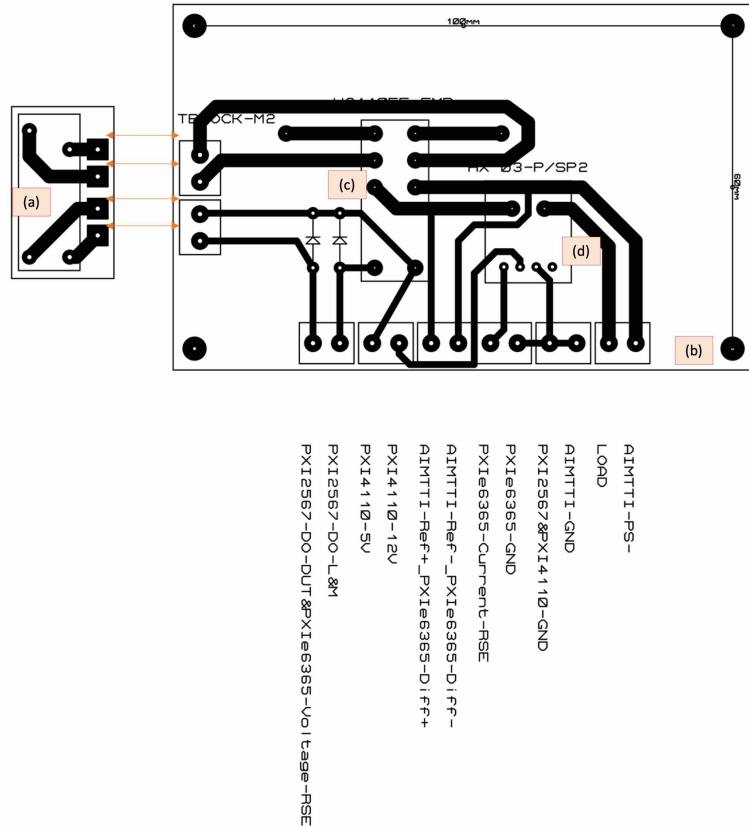


Figure A.1: PCB designed to test facilitate the EMR life-cycle test.

A.2 LabVIEW user interface

As displayed, the settings of the DAQ are controlled (a) on the H-PC. The user can specify whether CR measurements are taken as well as the switching frequency for the EMR at a 50/50 % duty cycle; (b) the data streamed back from the DAQ indicating the current operating mode; (c) indication of the number of failed, accumulated switching actuations and the operational state of the EMR; (d) immediate, safe shutdown of the experiment.

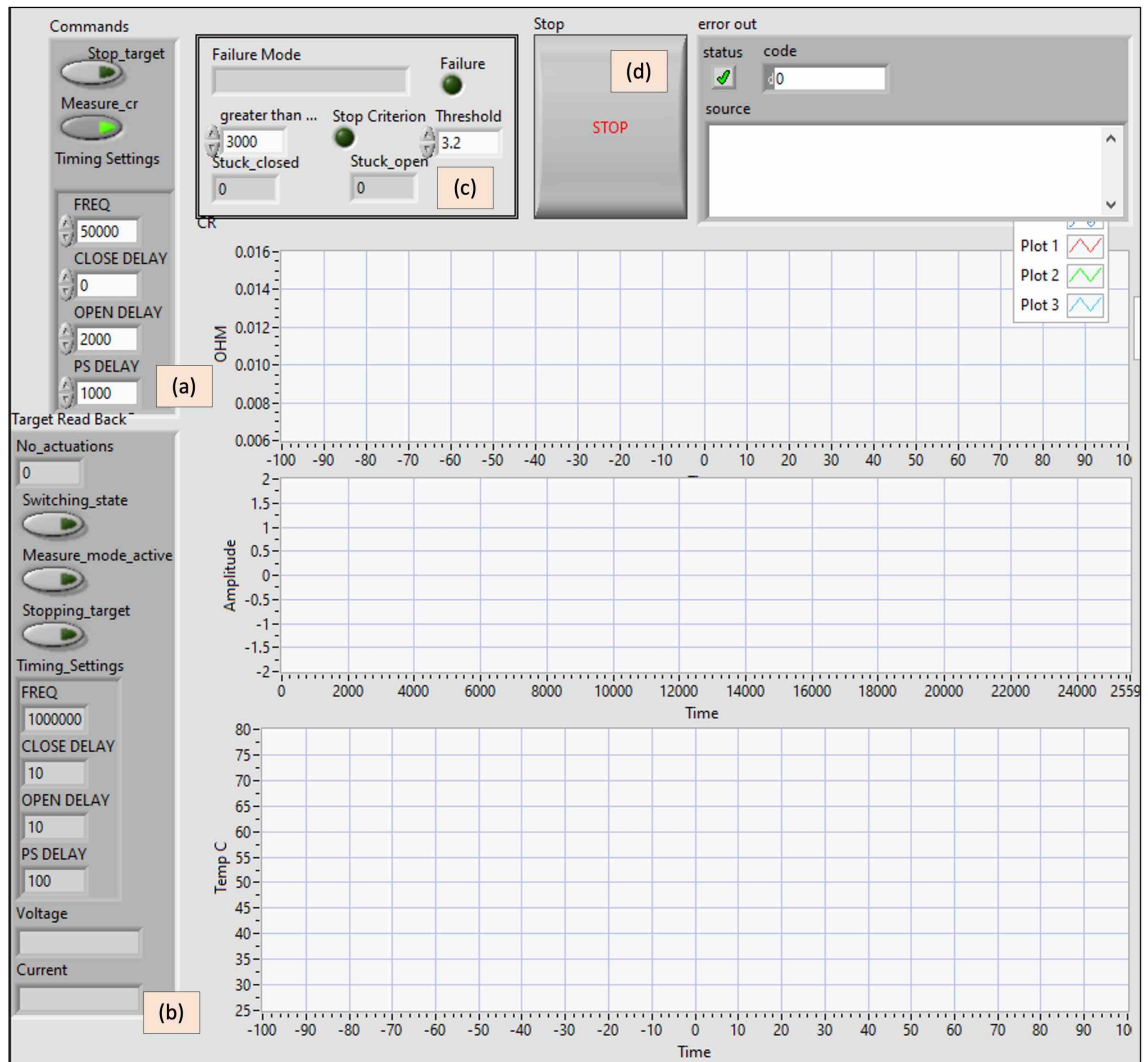


Figure A.2: The user interface for the EMR experiments.

A.3 H-PC LabVIEW main program

In Figure A.3, (a) the main program is initialised and the connectivity to the DAQ is established via the network connection and pre-tests of the experimental setup commences (e.g. initialise the power supplies); (b) the DAQ awaits new commands which can be entered through the user interface and are aggregated and streamed to the DAQ; (c) initial CR readings of the test EMR are taken; (d) once complete a loop which reads back each actuation from the DAQ writes the data to file, and controls the experiment state; (e) in each loop iteration the state of the buffer is first checked for available data; (f) if data is available its read and logged to different files, e.g. *.hdf5*; (g) the failure criteria for *stuck-closed*, *stuck-open*, excessive *opening time* and *closing time*, and *high CR* are checked; (h) if, e.g. contact sticking occurs it is recorded and checked against a failure threshold, i.e. the maximum number of allowed failed actuations; (i) every 10000th actuations the CR is measured which requires to pause the switching of the EMR; (j) CR measurements are taken through the Ohmmeter and send back directly to the H-PC; (k) if the test EMR has not failed, the next switching cycle at (d) commences, otherwise an automatic *shutdown* command is send to the DAQ.

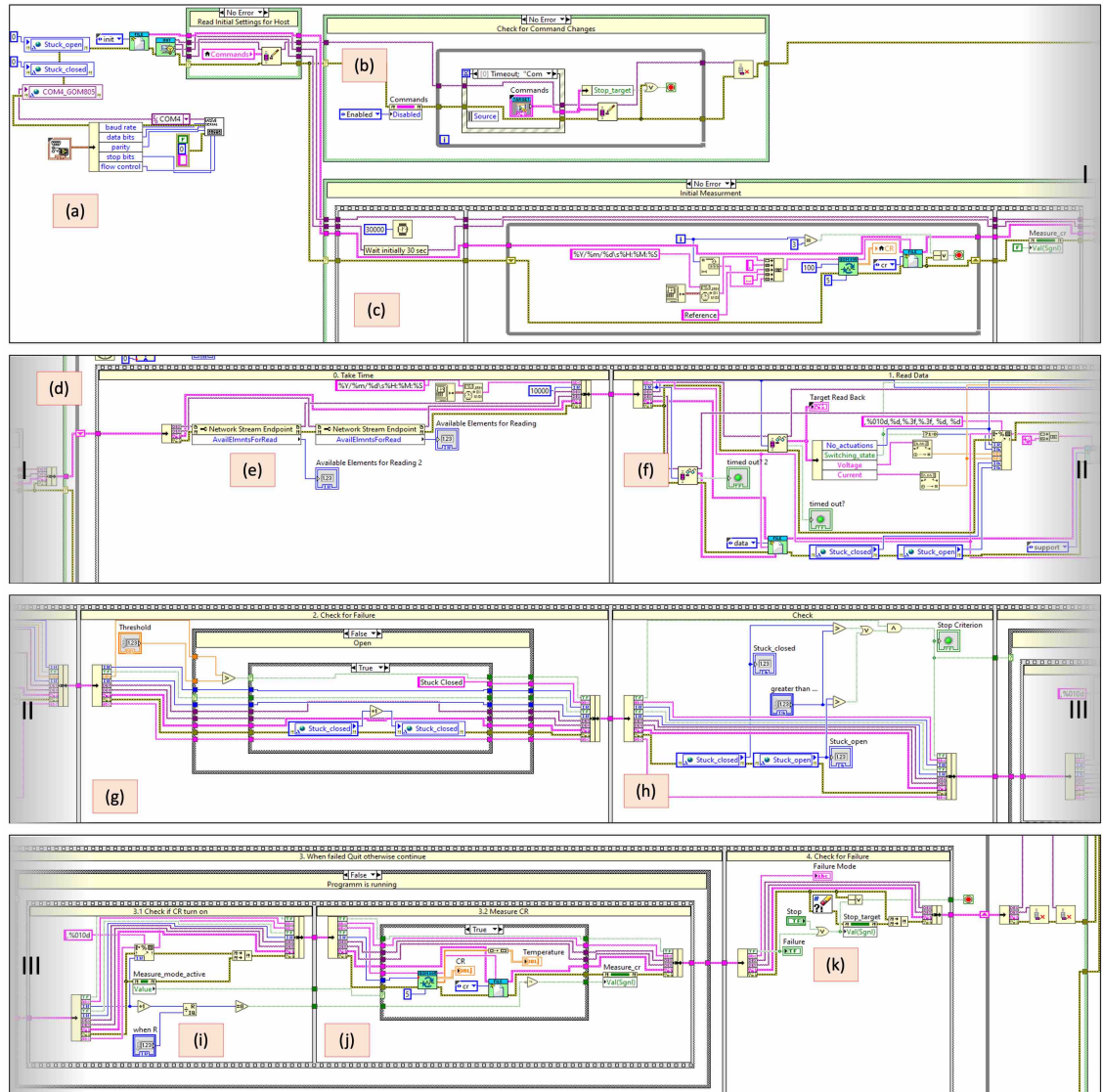


Figure A.3: LabVIEW block diagram of the main host program running on the H-PC; split up in four sections.

A.4 DAQ LabVIEW EMR switching loop

As displayed in Figure A.4, the following consecutive steps are performed for each making and breaking actuation. (a) Starting a new execution of the deterministic loop running at MHz clock speed (PXIe8880 internal clock); (b) reading the switching state of the previous operation (*opened* or *closed*) and updating the new switching state; (c) initialising the data acquisition of CV, CC, and CI; (d) sending the trigger to switch the EMR under test; (e) in case of DA2 sampling the data at 150 kHz for a 0.02 ms interval; (f) stopping the data acquisition; (g) reading back reference voltage and current measurements from the power supply's internal sensors verifying a successful switching operation; (h) transferring data from the real-time buffer to (i) enter the queue buffer for being streamed back to the H-PC; (j) verifying to continue to the next switching operation or to interrupt the switching either due to a *shutdown* command received from the H-PC or in order to (k) take a CR and temperature reading. In latter case, the CR measurement circuit is switched in via the DPDT-EMR.

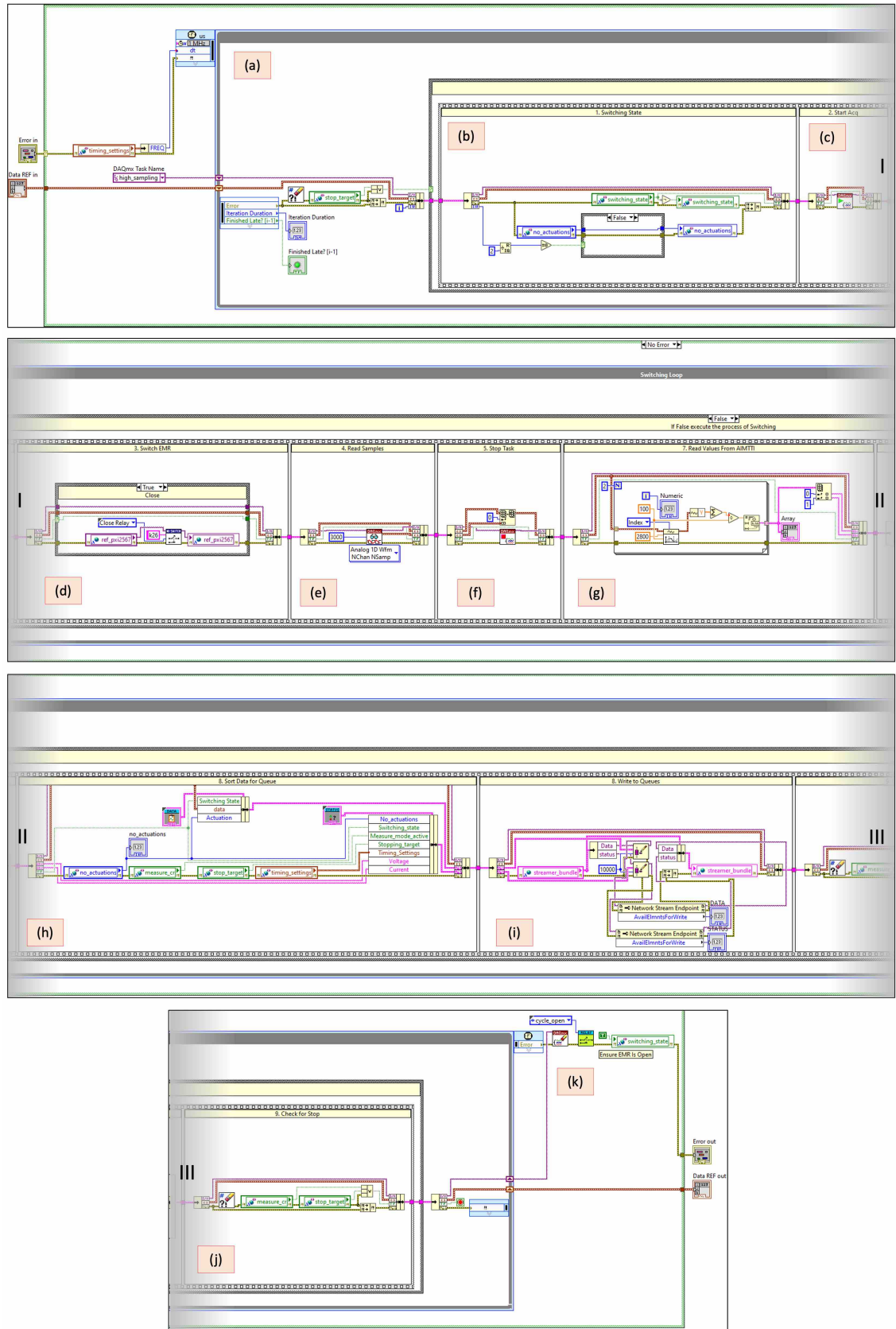


Figure A.4: LabVIEW block diagram of EMR deterministic switching loop running on the DAQ, opening and closing the test EMR.

A.5 DA1 actuations till failure

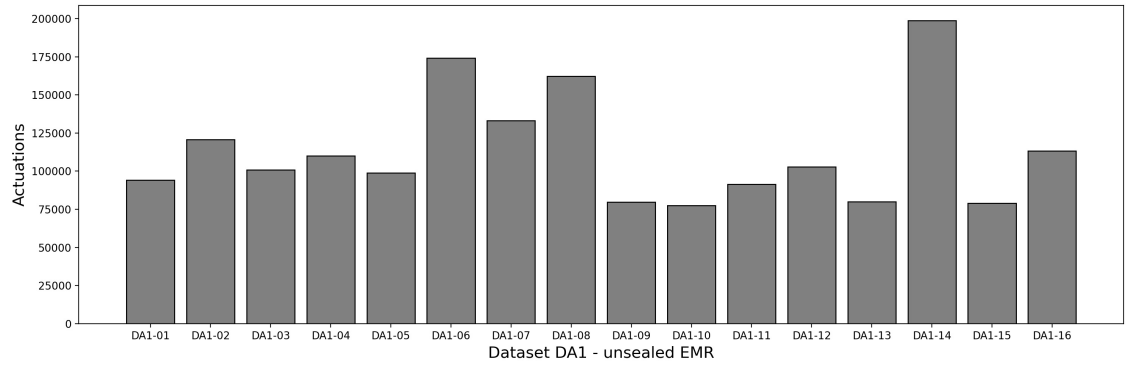


Figure A.5: Actuations till failure of the 16 tested unsealed EMRs contained in data set DA1 (tested at 60 °C) and a switching frequency of 0.25 Hz).

A.6 DA2 actuations till failure

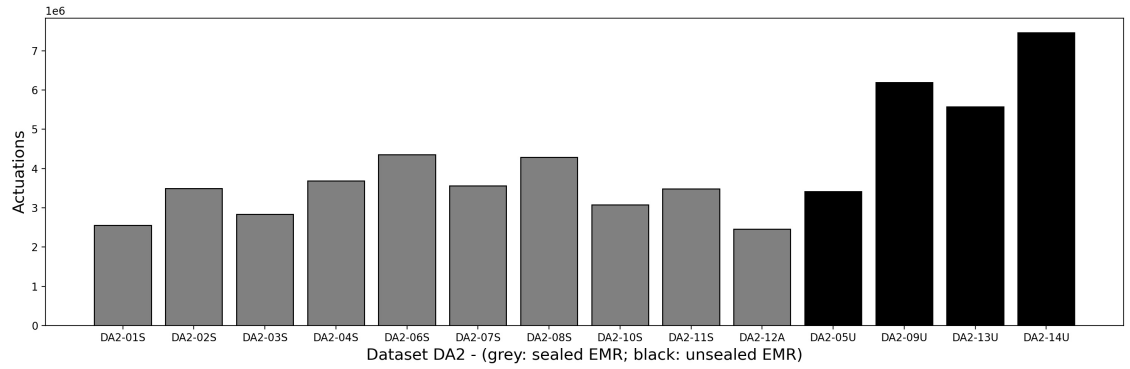


Figure A.6: Actuations till failure of the 14 tested EMRs contained in data set DA2 (10 sealed EMRs and 4 unsealed EMRs, tested at 30 °C) and a switching frequency of 25 Hz).

Appendix B

Supplementary EMR failure analysis results

B.1 DA2 sealed EMR making waveform

The contact making waveform of (left) CV, (centre) CI and (right) CI sub-sampled from 150 kHz to 5 kHz. The graphs show the changing waveform shape over the EMR life where as a lighter colour symbolises a new EMR (0 % life) and as it becomes gradually darker the EMR is approaching EOL (100 % life).

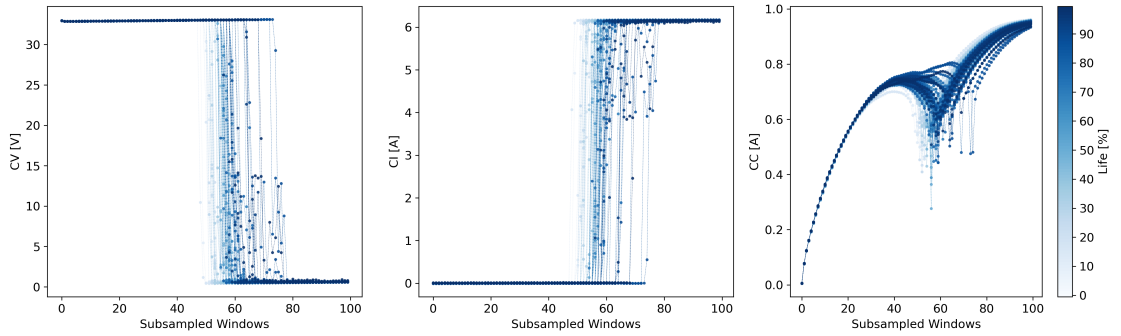


Figure B.1: DA2-03S - making.

B.2 DA2 sealed EMR breaking waveform

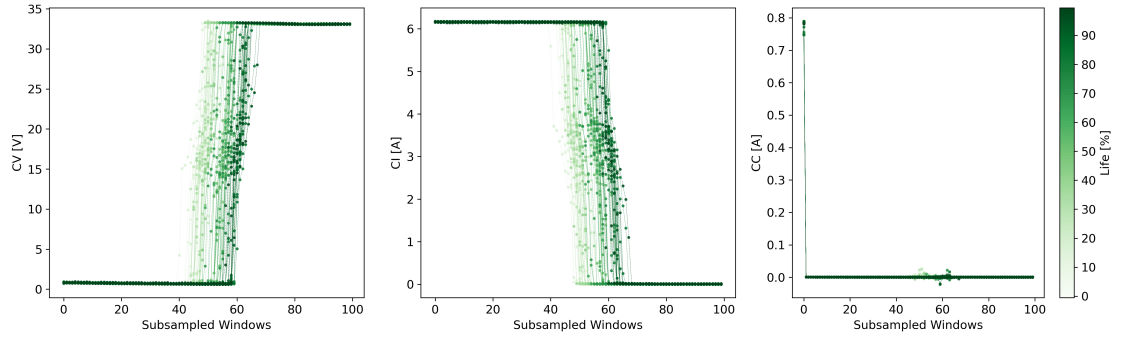


Figure B.2: DA2-03S - breaking.

B.3 DA2 unsealed EMR making waveform

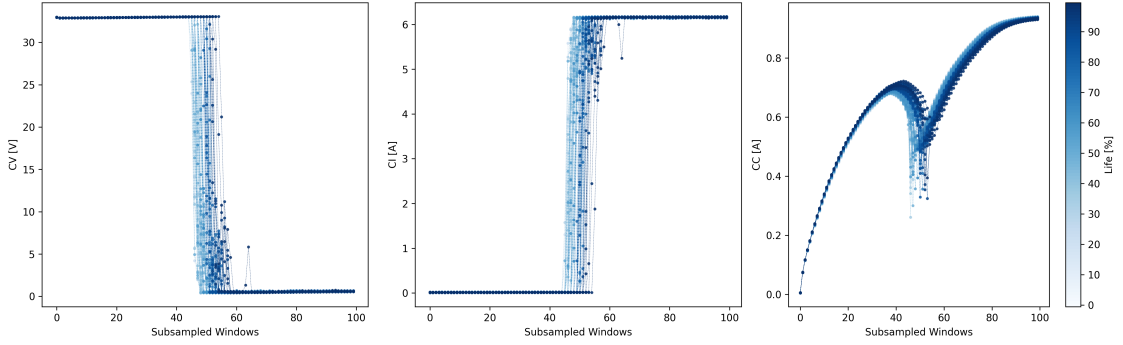


Figure B.3: DA2-09U - making.

B.4 DA2 unsealed EMR breaking waveform

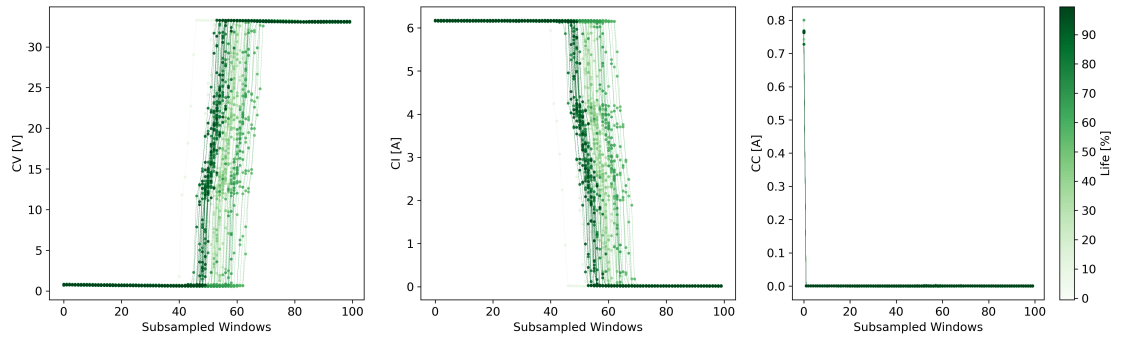


Figure B.4: DA2-09U - breaking.

B.5 EDX analysis

This EDX analysis compares an unsealed and a sealed EMR operated for 70000 actuations. Clearly recognisable is the silver-oxide splattering which is deposited on the anode and cathode contact carrier strips, i.e. material dispersed through electrical erosion.

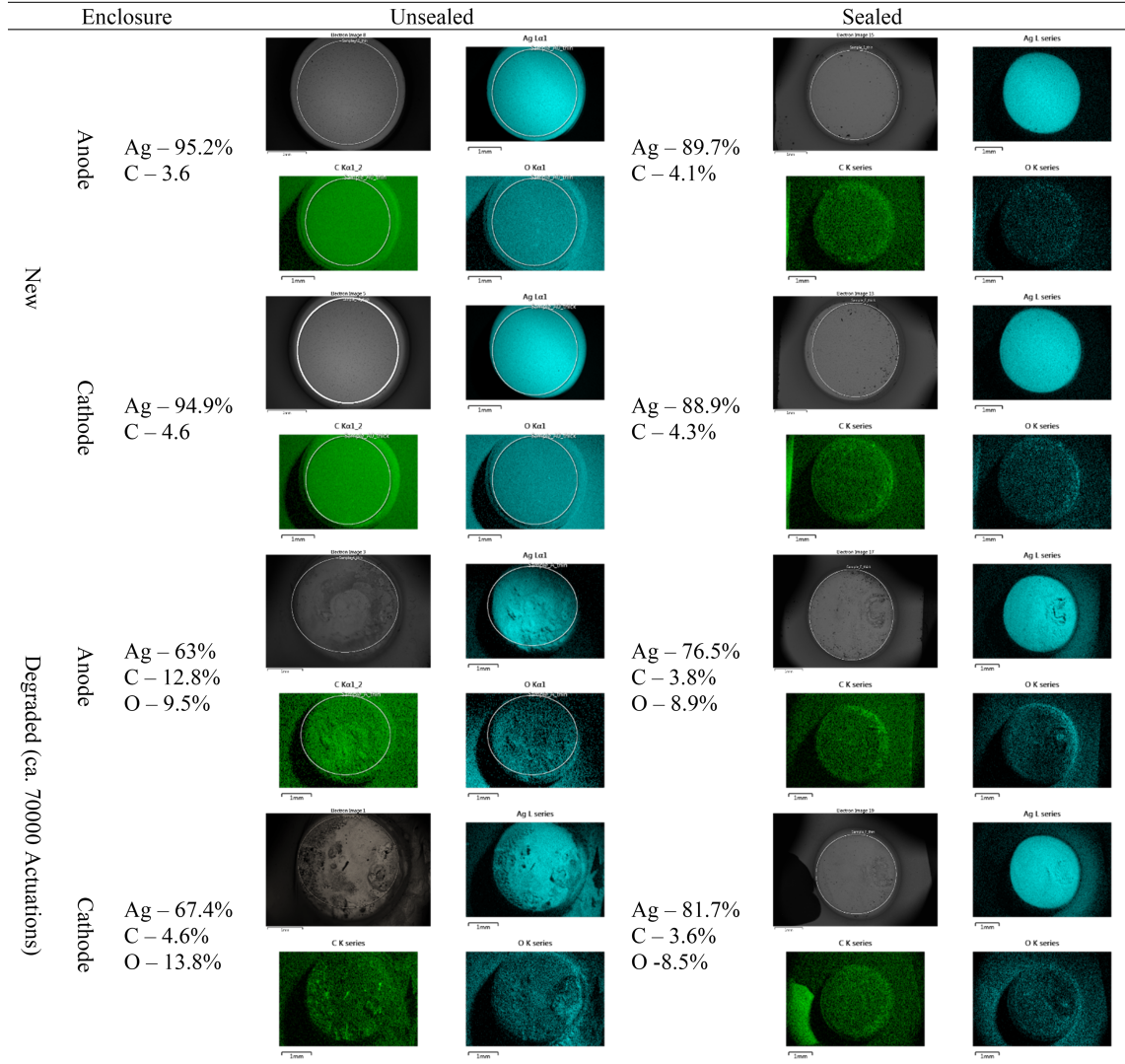


Figure B.5: EDX analysis.

Appendix C

Supplementary EMRUA results

C.1 DA1 time-based feature set

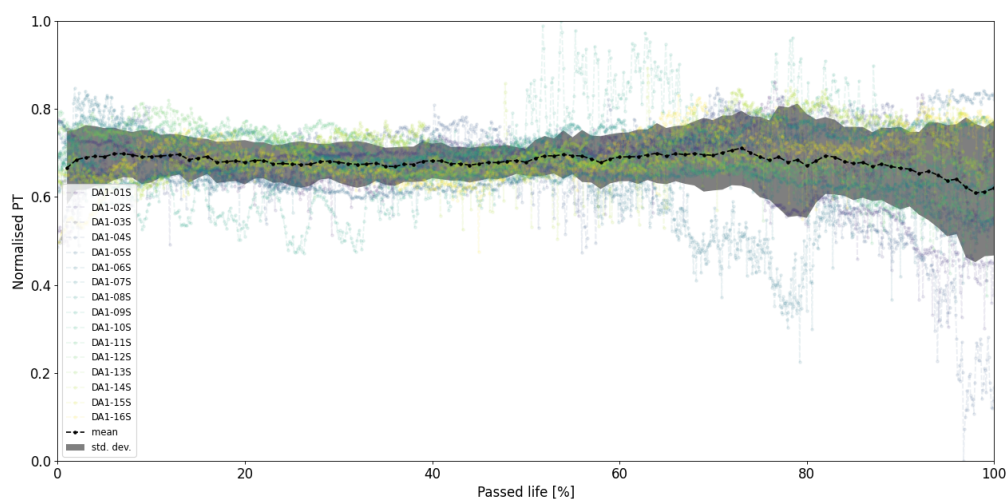


Figure C.1: DA1 - Pick-Up Time (PT).

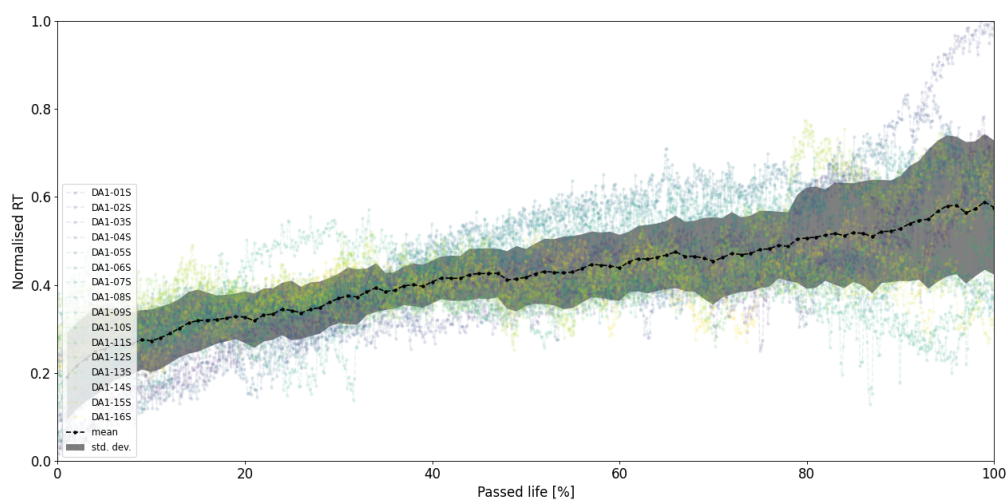


Figure C.2: DA1 - Release Time (RT).

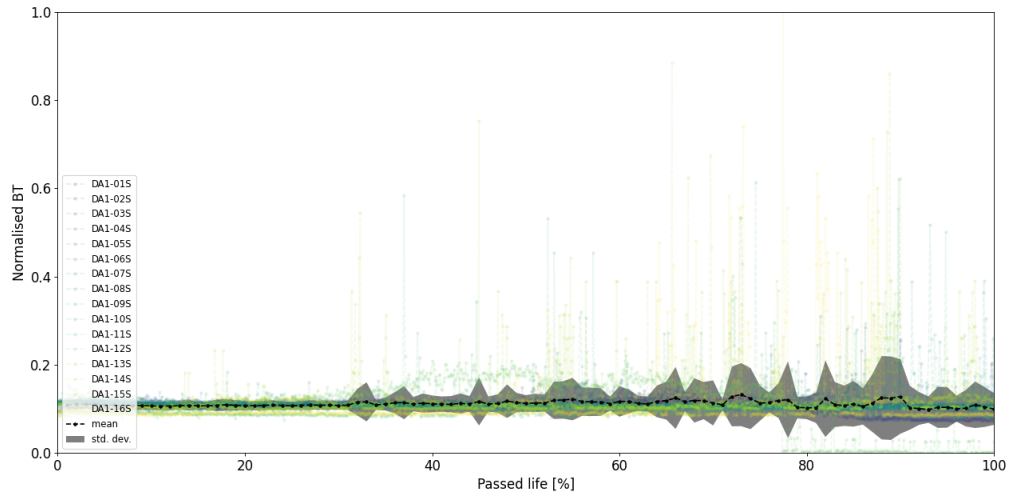


Figure C.3: DA1 - Bounce Time (BT).

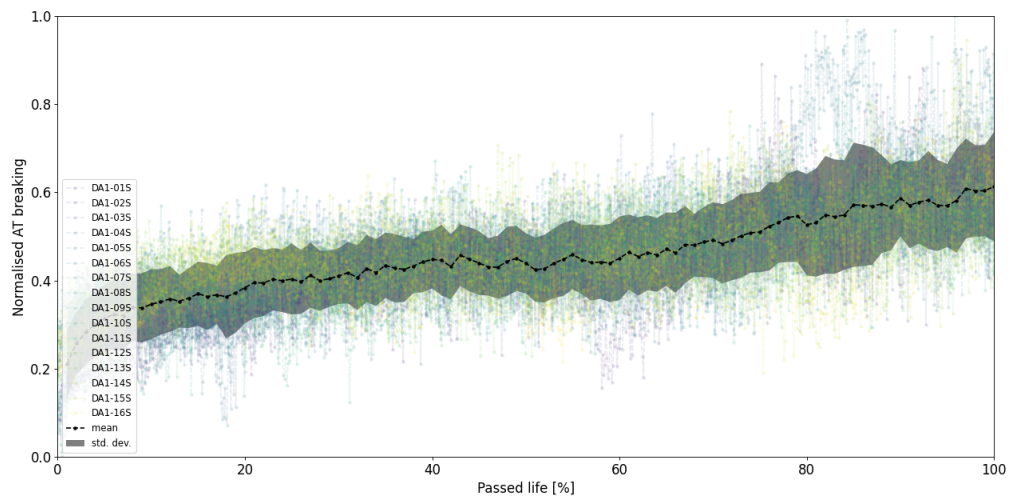


Figure C.4: DA1 - Arcing Time (AT) during breaking.

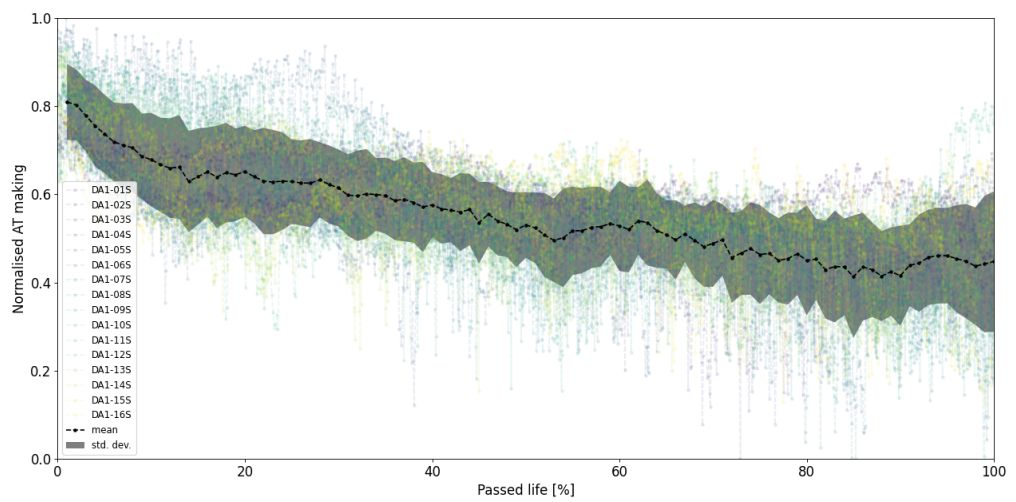


Figure C.5: DA1 - Arcing Time (AT) during making.

C.2 DA1 statistical feature set

C.2.1 Making actuation

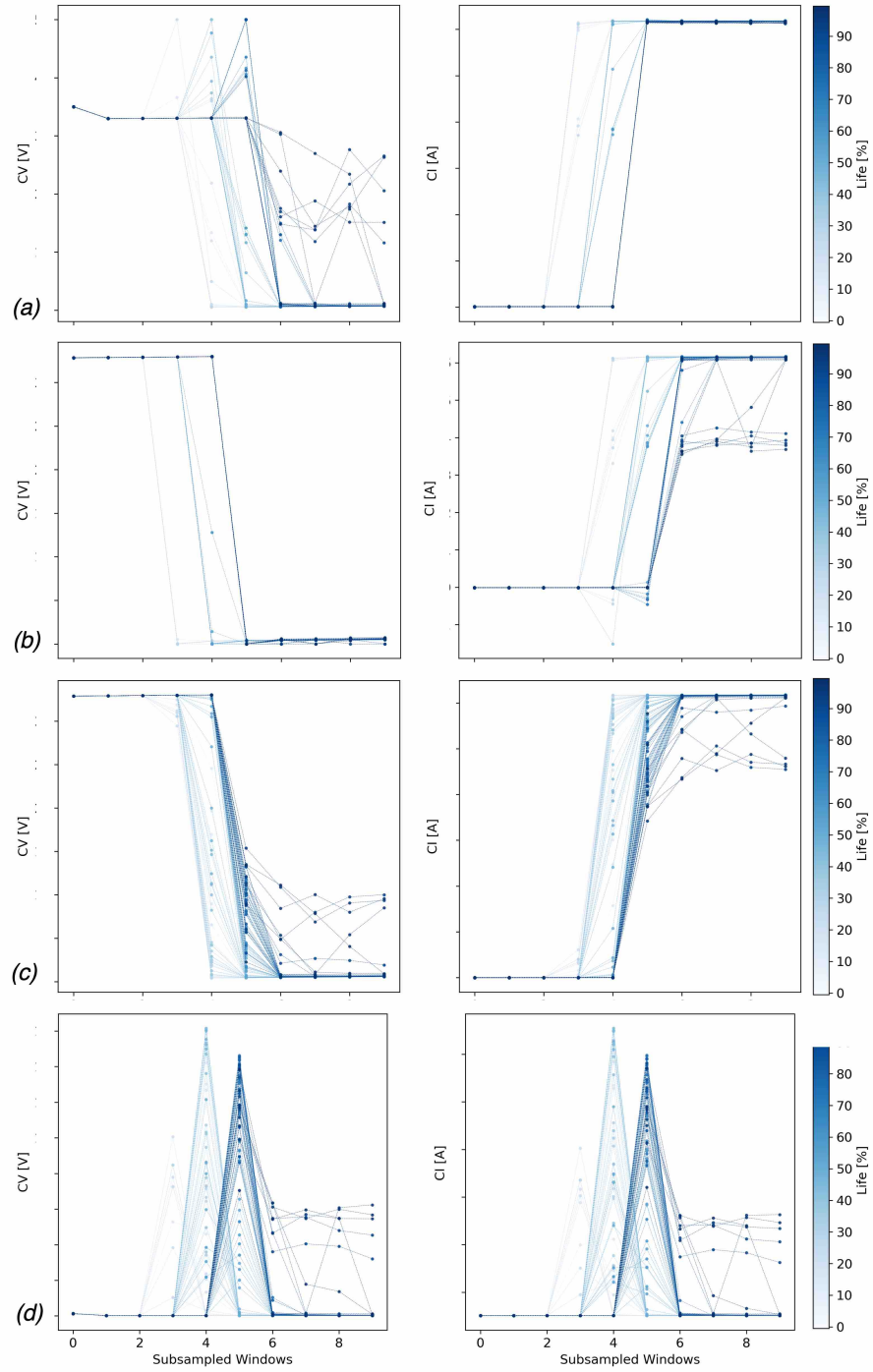


Figure C.6: DA1-01S F_S extracted from the making actuation, (max, min, mean, std. dev.).

C.2.2 Breaking actuation

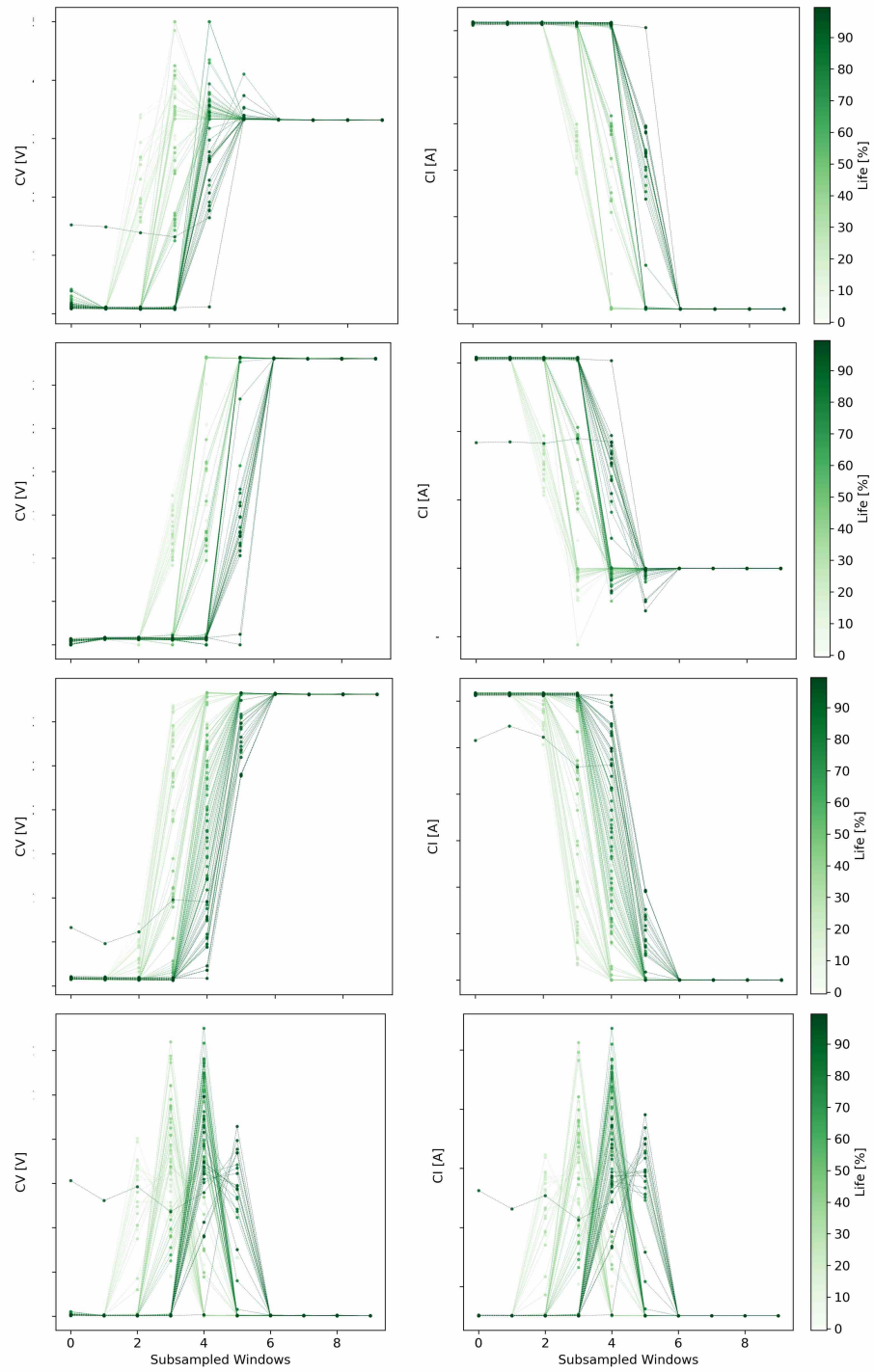


Figure C.7: DA1-01S F_S extracted from the breaking actuation, (max, min, mean, std. dev.).

C.3 DA1 inference performance evaluation

C.3.1 DA1-01U

C.3.1.1 LI

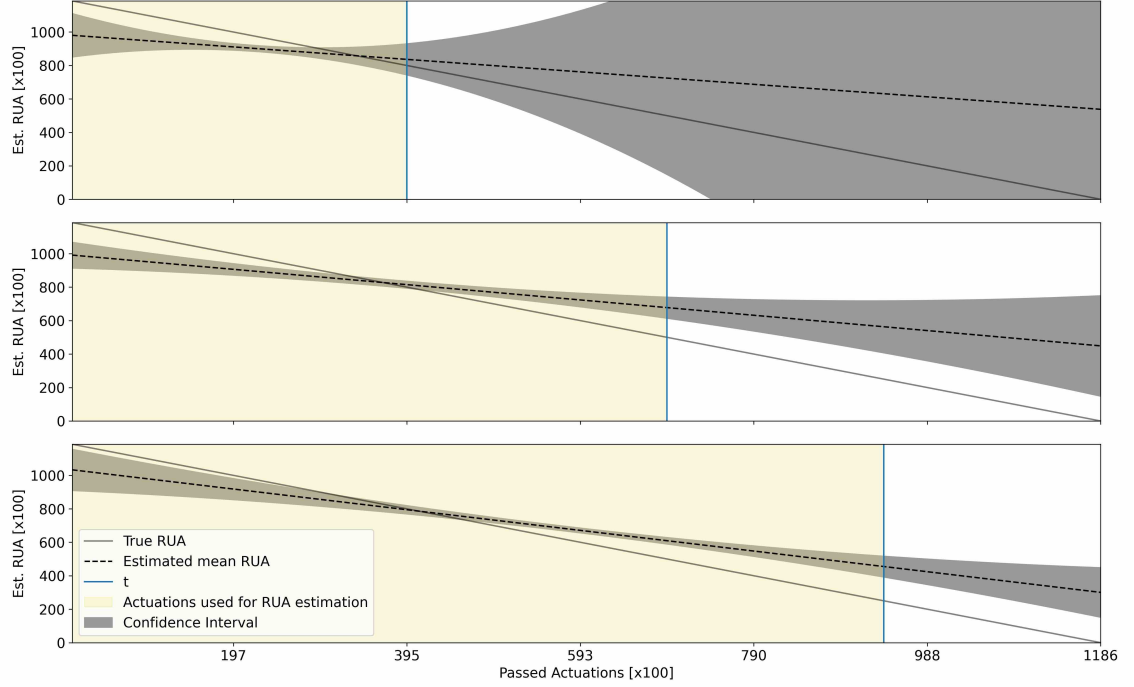


Figure C.8: DA1-01U - F_T - LI - $k = 4 \cdot d_b = 2 \cdot R = 6$.

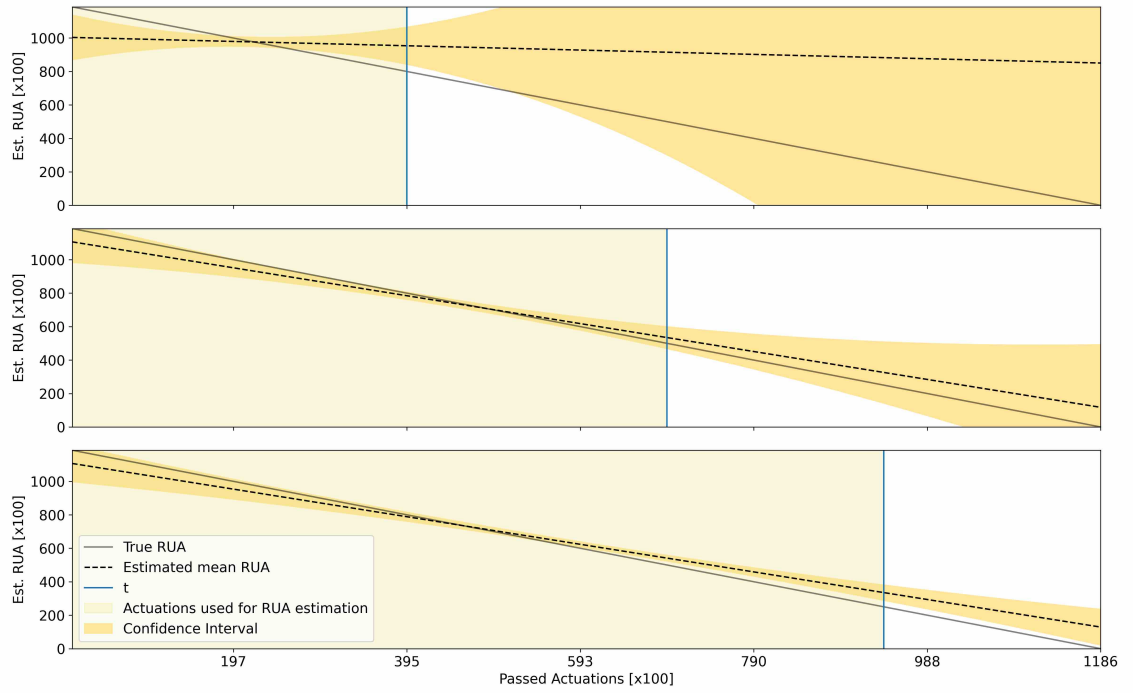


Figure C.9: DA1-01U - F_S - LI - $k = 4 \cdot d_b = 2 \cdot R = 6$.

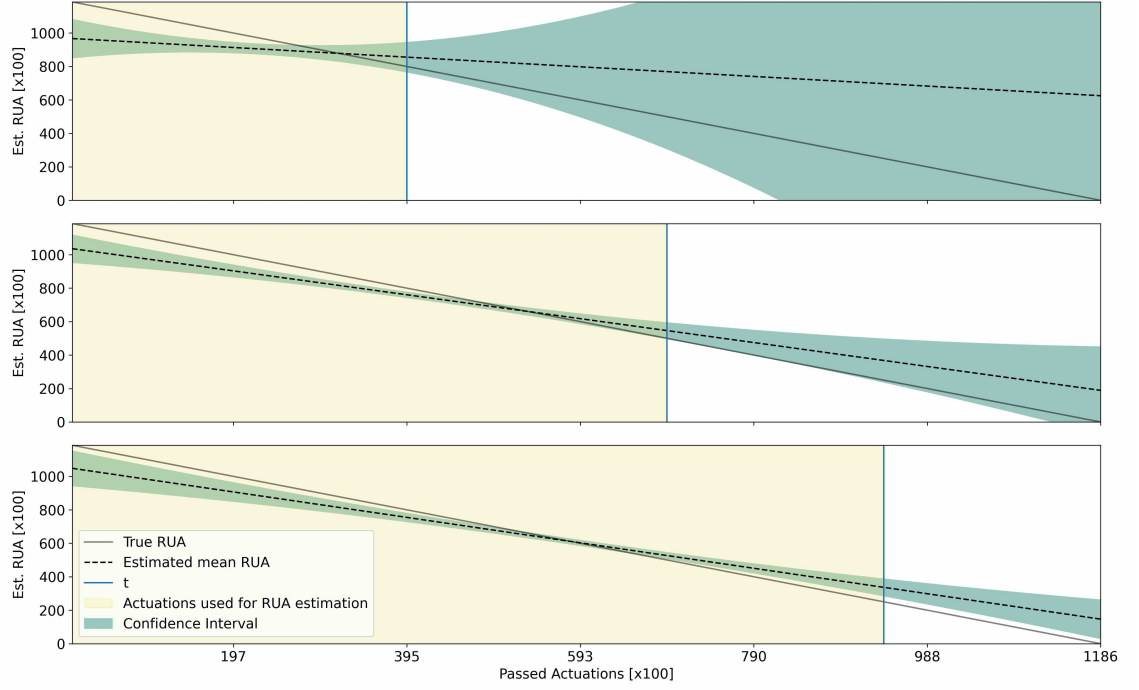


Figure C.10: DA1-01U - $F_{T,S}$ - LI - $k = 4 - d_b = 2 - R = 6$.

C.3.1.2 GI

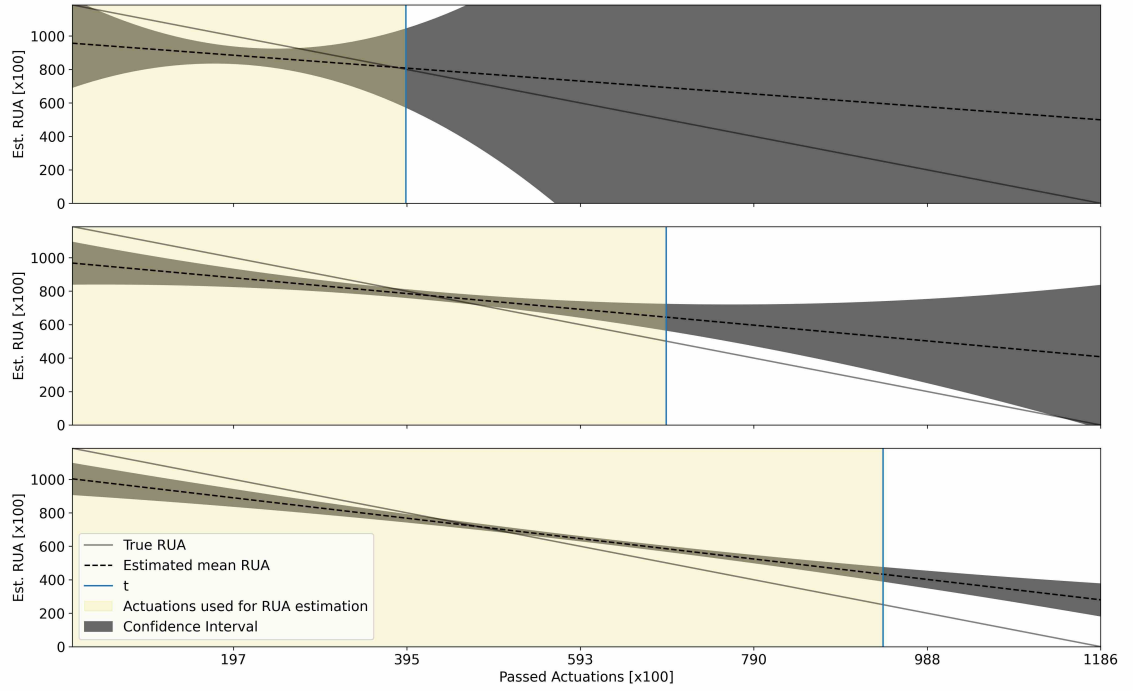


Figure C.11: DA1-01U - F_T - GI - $k = 4 - d_b = 2 - R = 6$.

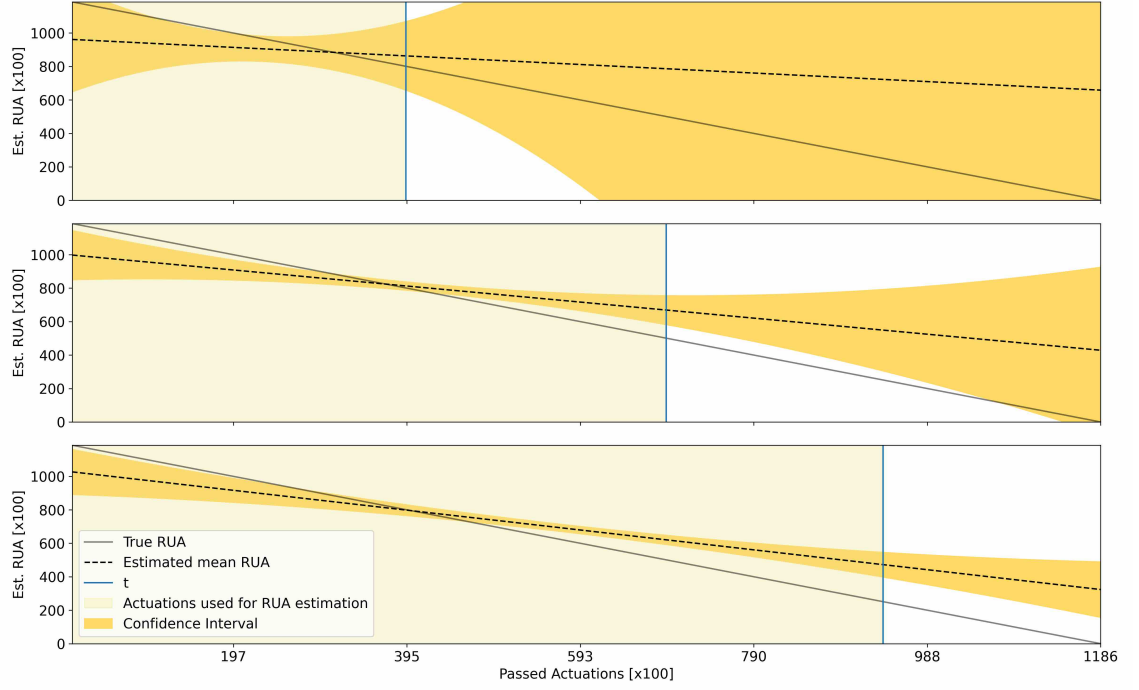


Figure C.12: DA1-01U - F_S - GI - $k = 4-d_b = 2-R = 6$.

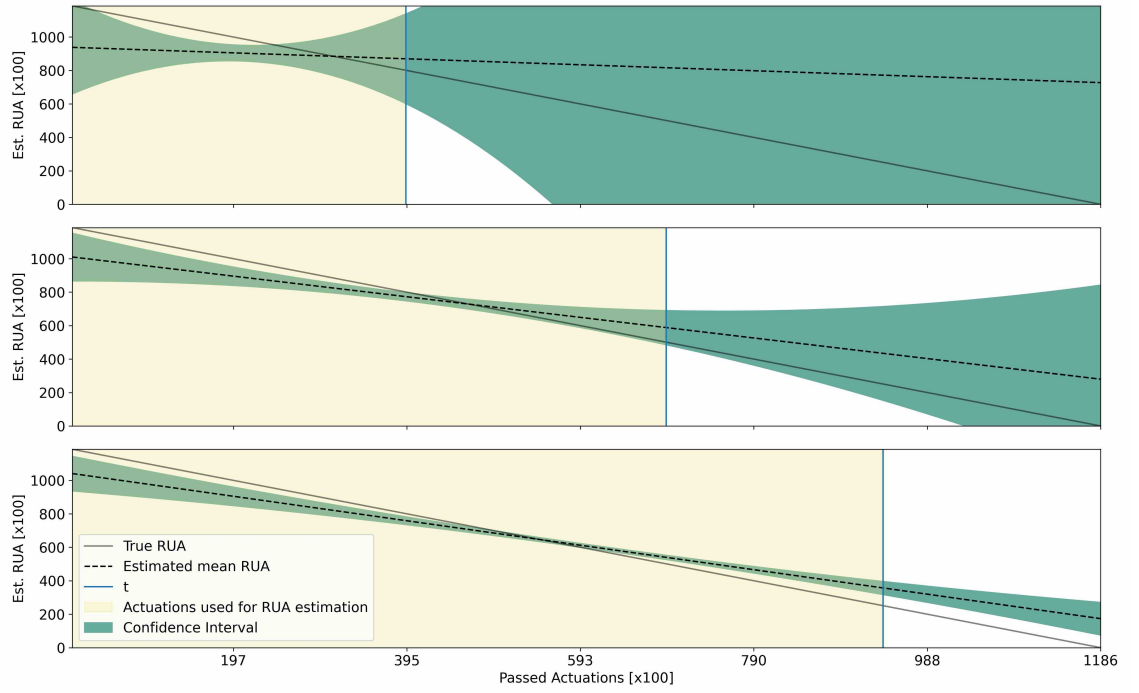


Figure C.13: DA1-01U - $F_{T,S}$ - GI - $k = 4-d_b = 2-R = 6$.

C.3.2 DA1-02U

C.3.2.1 EI

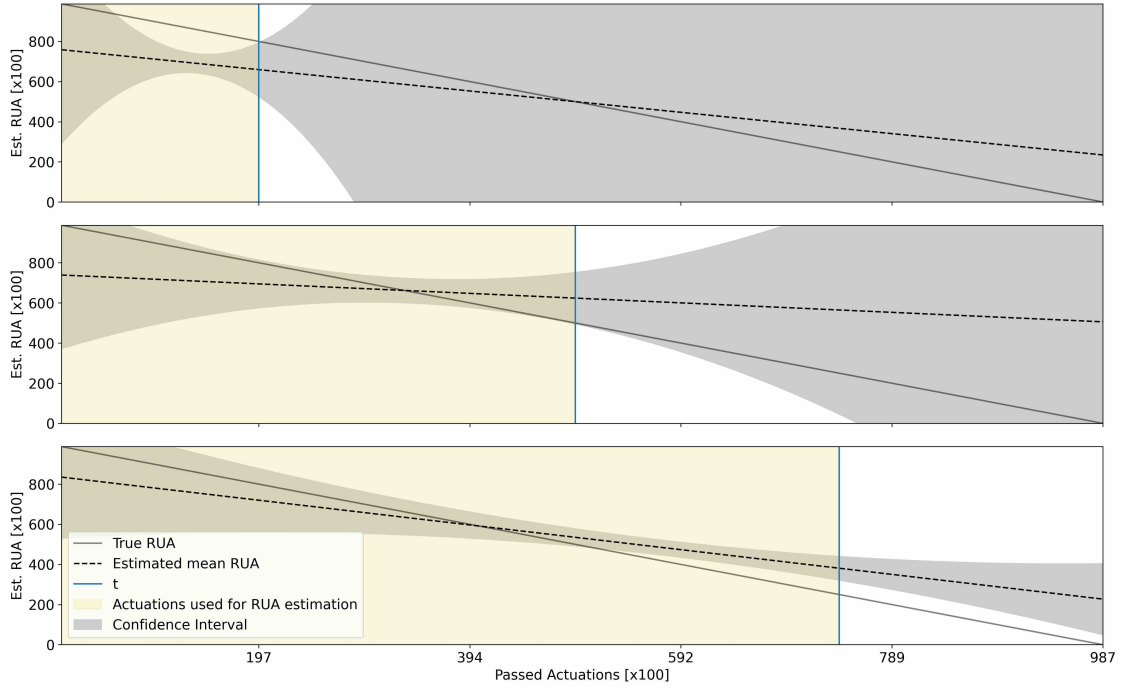


Figure C.14: DA1-02U - F_T - EI - $k = 4 - d_b = 2 - R = 6$.

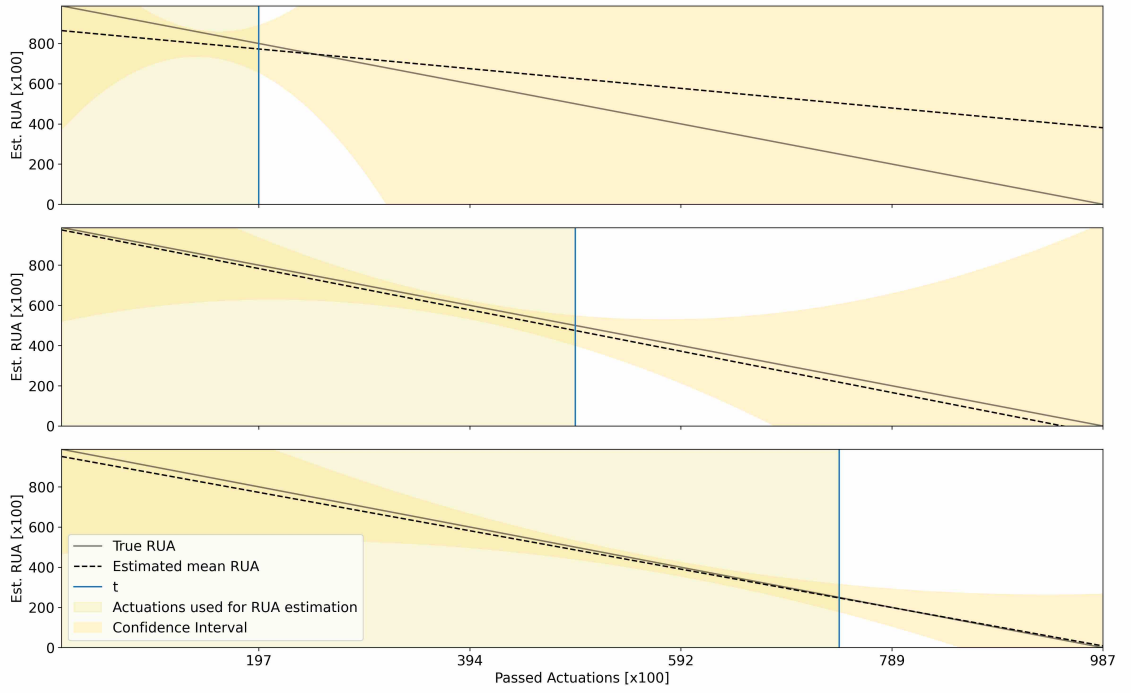


Figure C.15: DA1-02U - F_S - EI - $k = 4 - d_b = 2 - R = 6$.

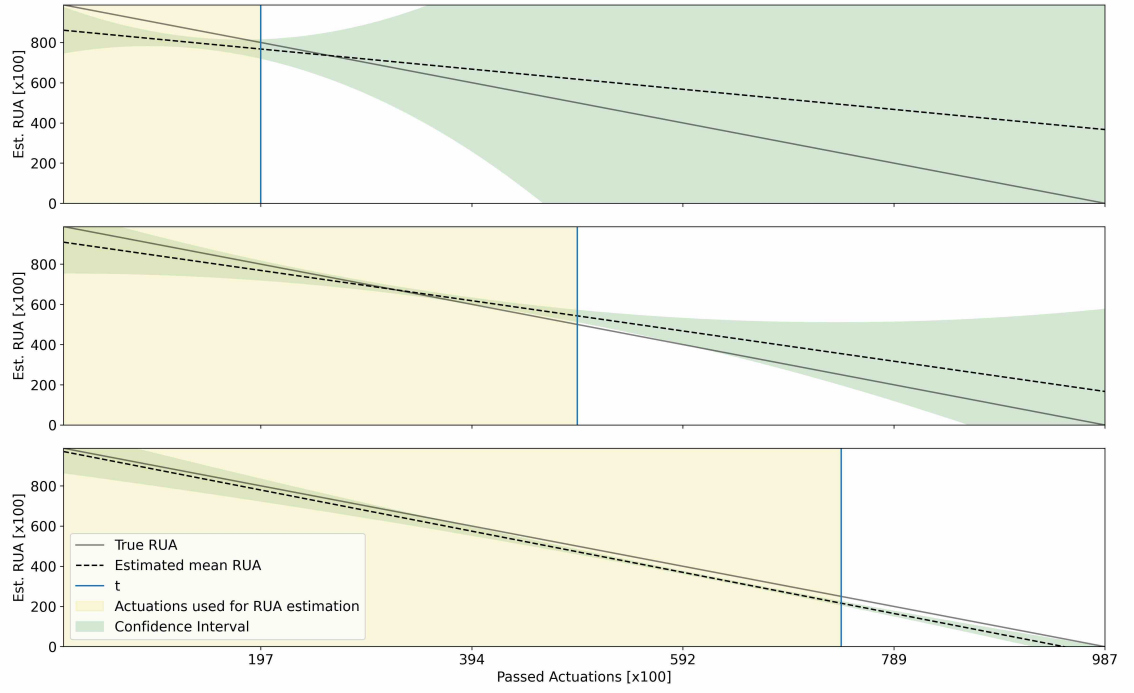


Figure C.16: DA1-02U - $F_{T,S}$ - EI - $k = 4-d_b = 2-R = 6$.

C.3.2.2 LI

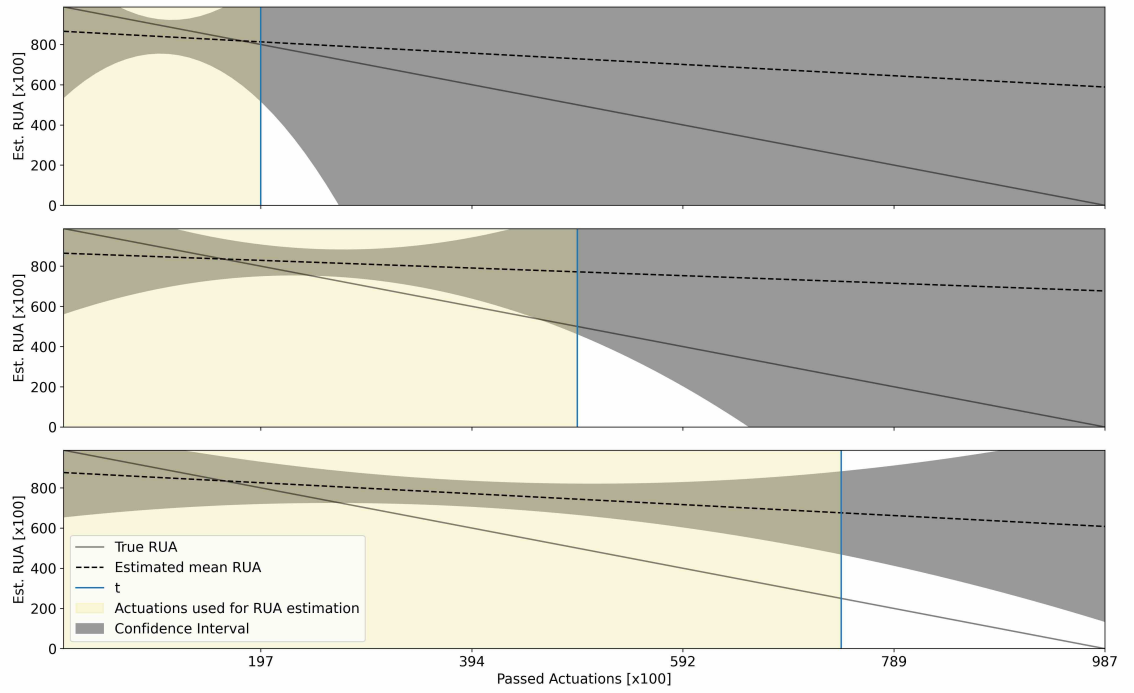


Figure C.17: DA1-02U - F_T - LI - $k = 4-d_b = 2-R = 6$.

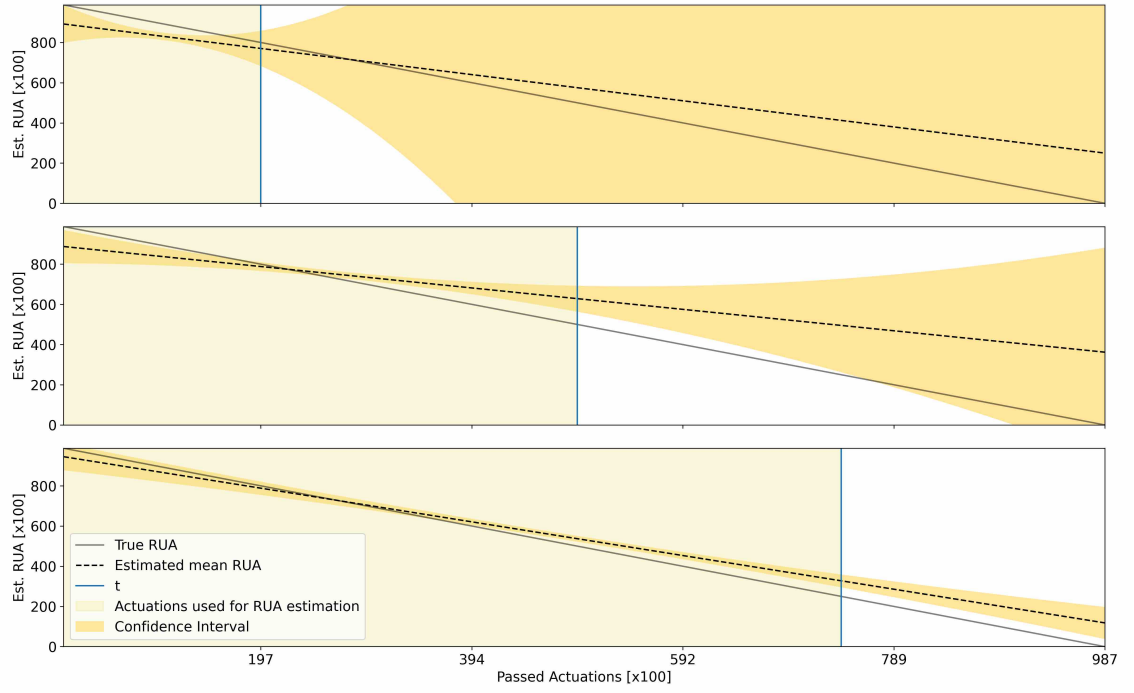


Figure C.18: DA1-02U - F_S - LI - $k = 4 - d_b = 2 - R = 6$.

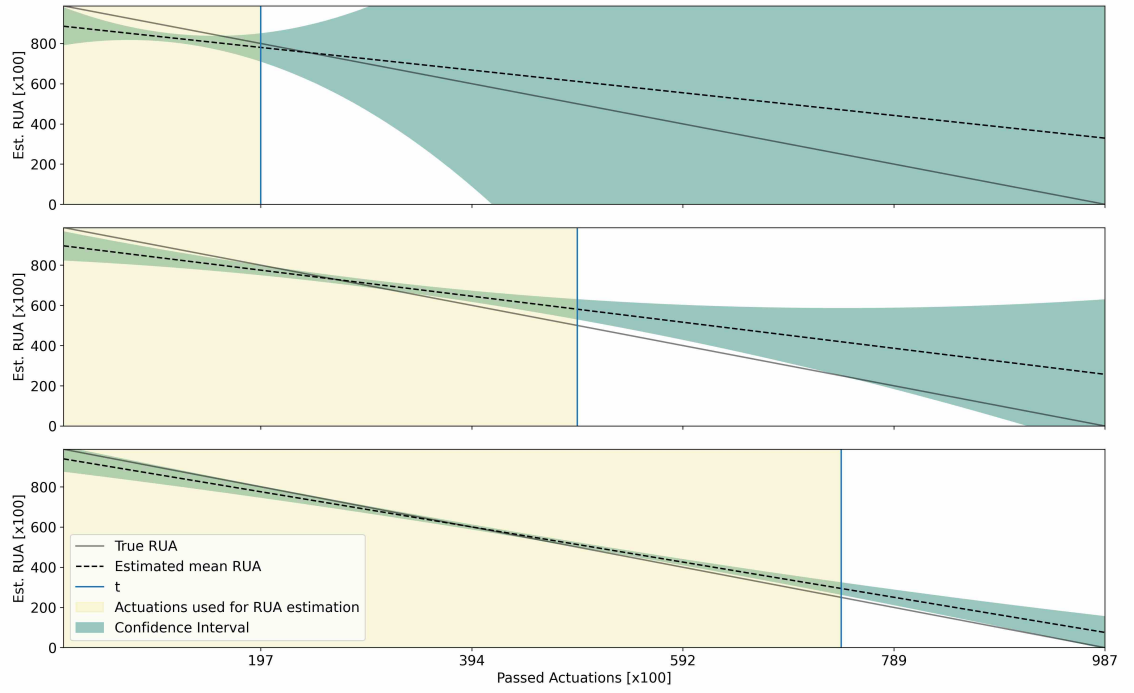


Figure C.19: DA1-02U - $F_{T,S}$ - LI - $k = 4 - d_b = 2 - R = 6$.

C.3.2.3 GI

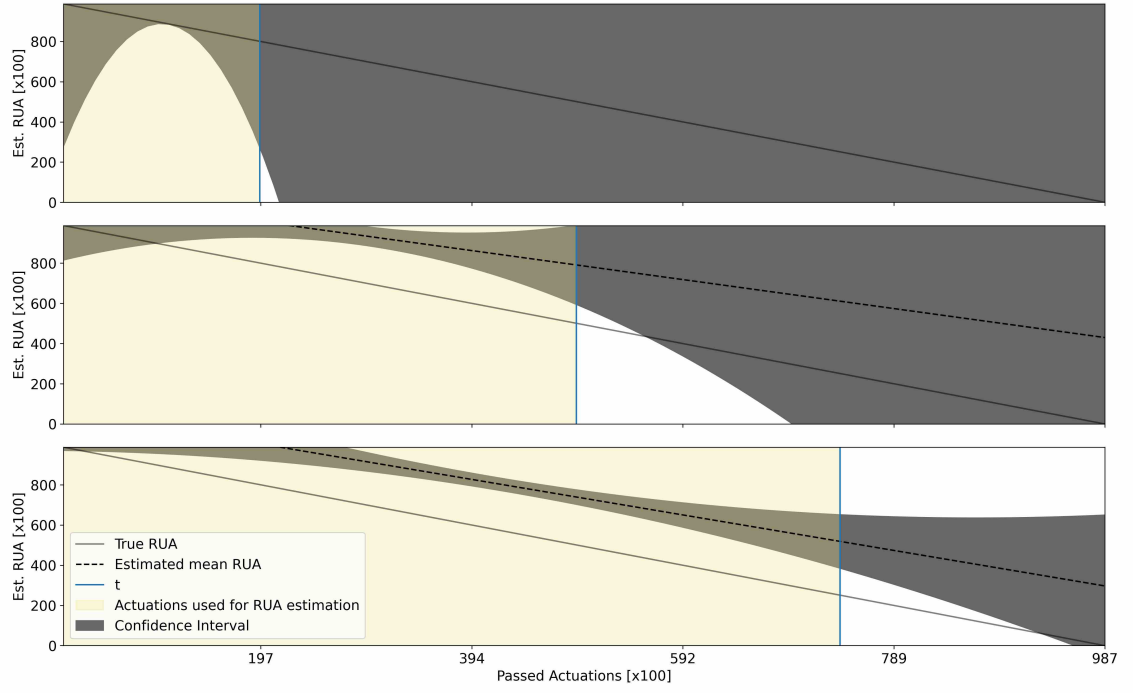


Figure C.20: DA1-02U - F_T - GI - $k = 4 - d_b = 2 - R = 6$.

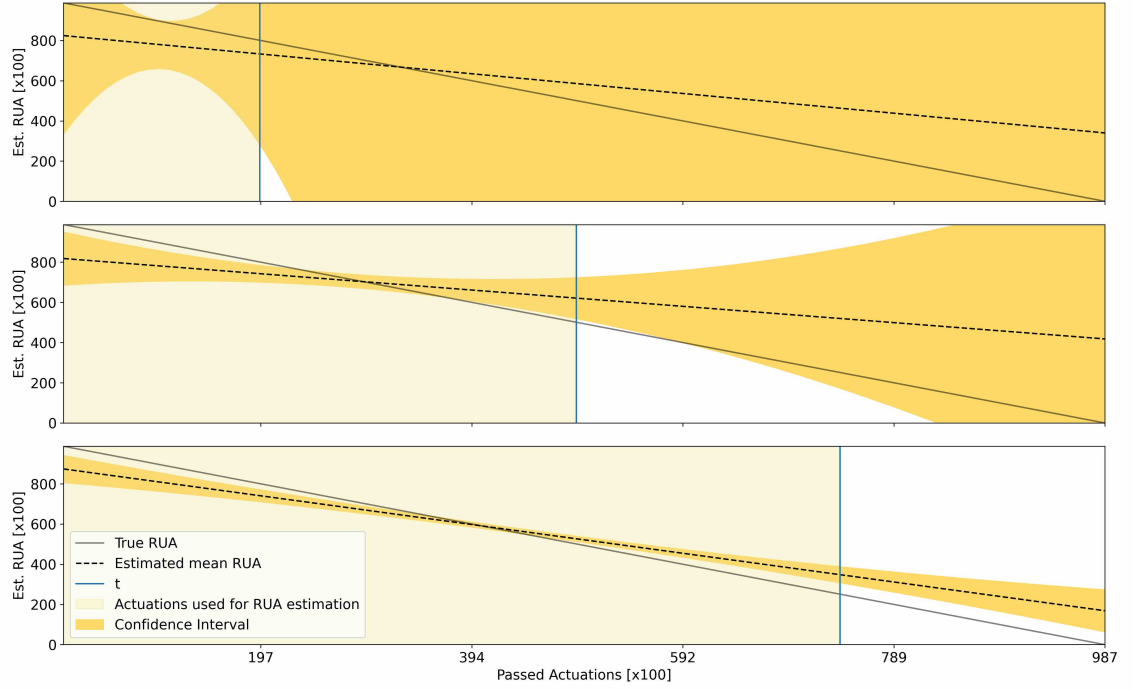


Figure C.21: DA1-02U - F_S - GI - $k = 4 - d_b = 2 - R = 6$.

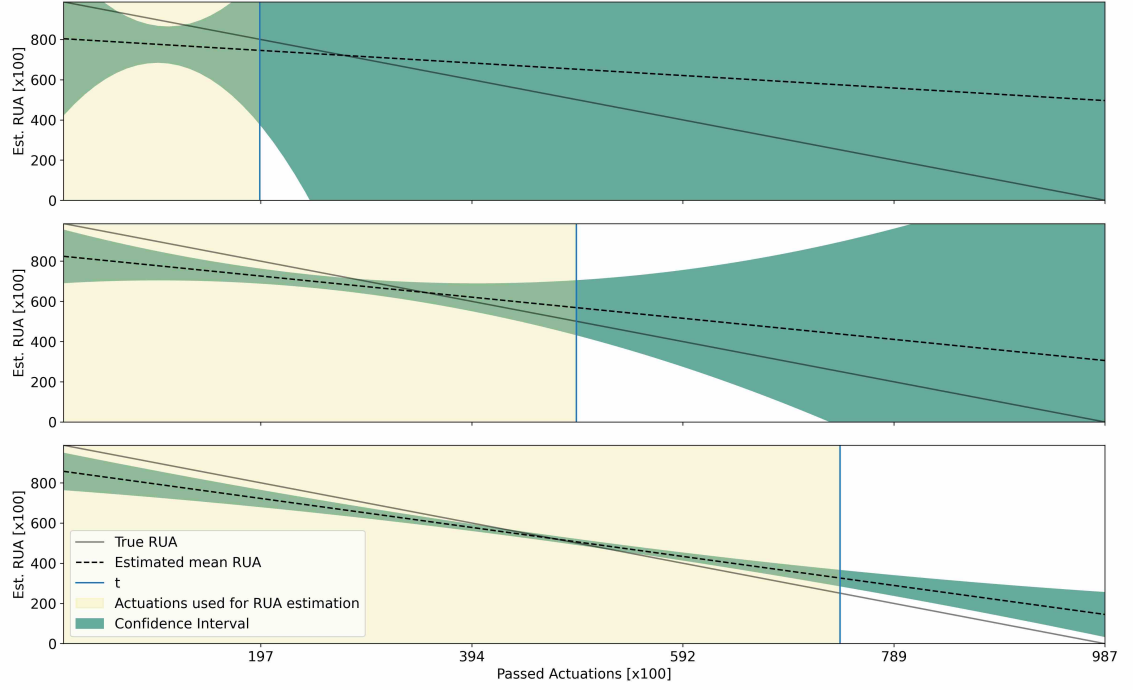


Figure C.22: DA1-02U - $F_{T,S}$ - GI - $k = 4 - d_b = 2 - R = 6$.

C.3.3 DA1-03U

C.3.3.1 EI

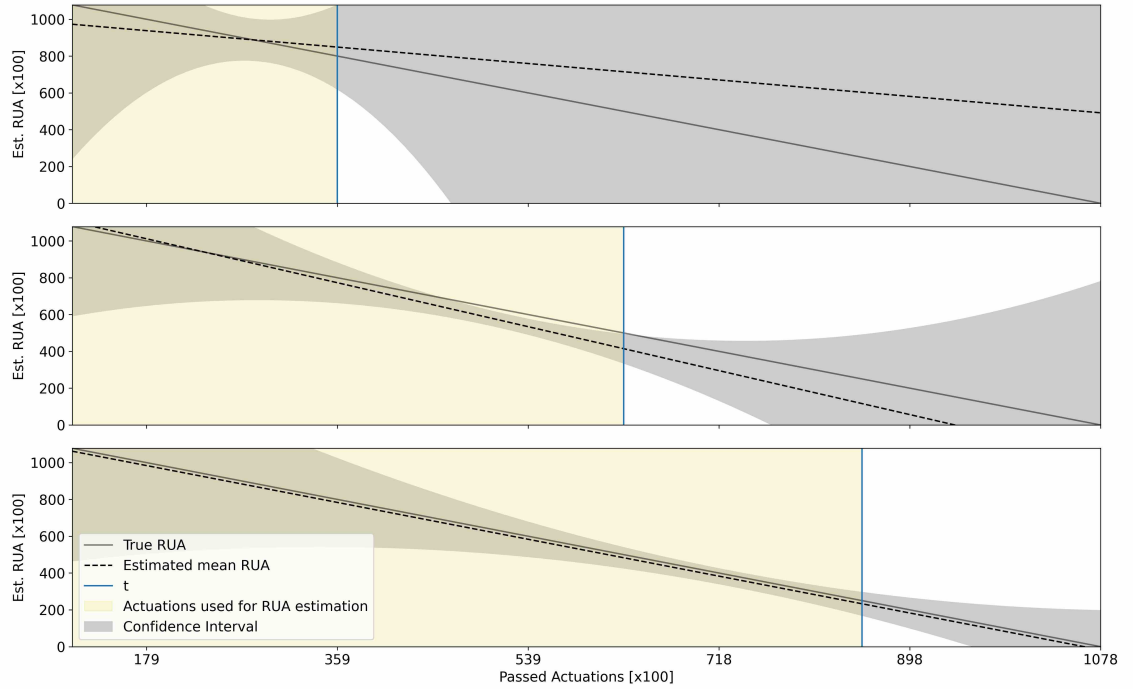


Figure C.23: DA1-03U - F_T - EI - $k = 4 - d_b = 2 - R = 6$.

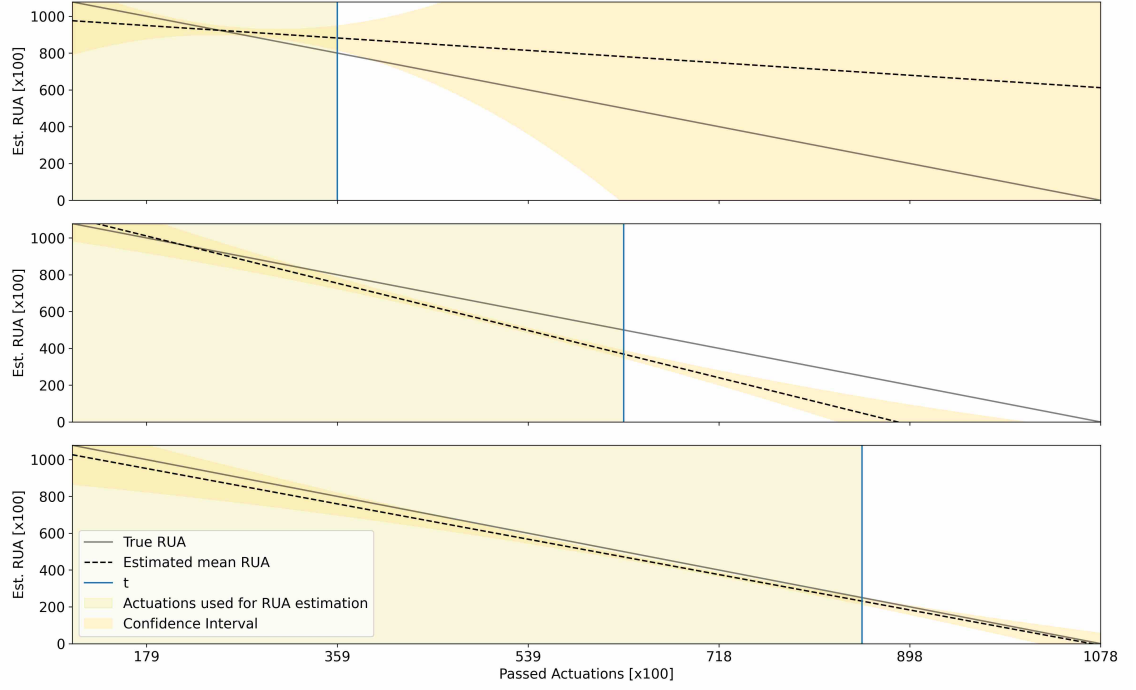


Figure C.24: DA1-03U - F_S - EI - $k = 4-d_b = 2-R = 6$.

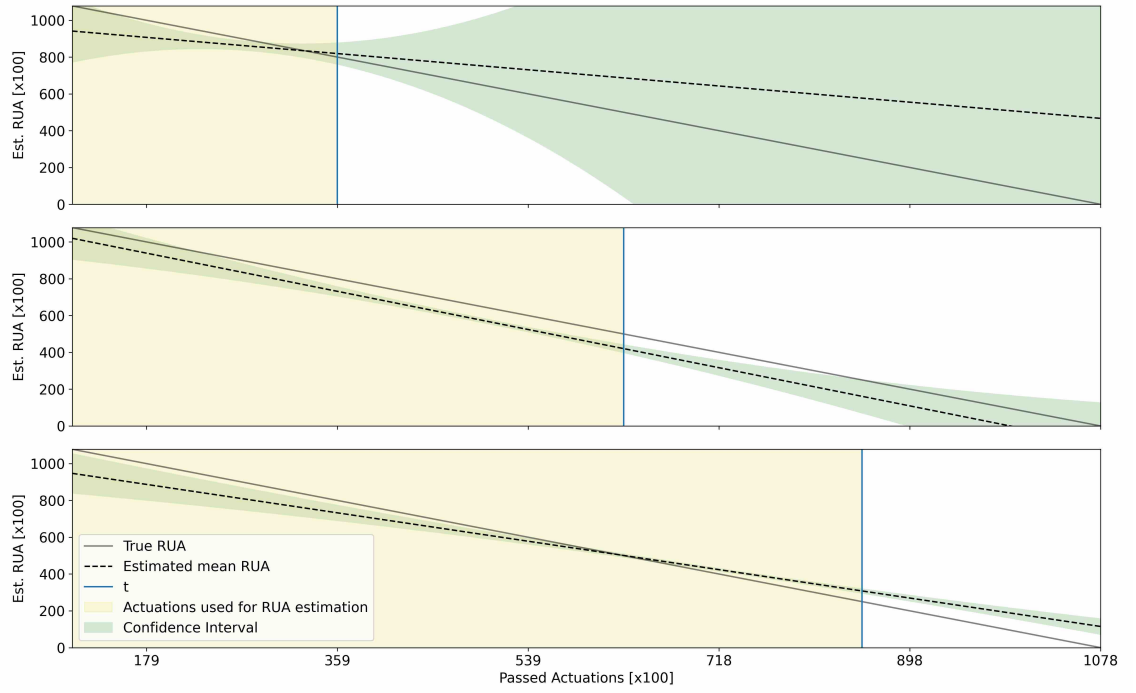


Figure C.25: DA1-03U - $F_{T,S}$ - EI - $k = 4-d_b = 2-R = 6$.

C.3.3.2 LI

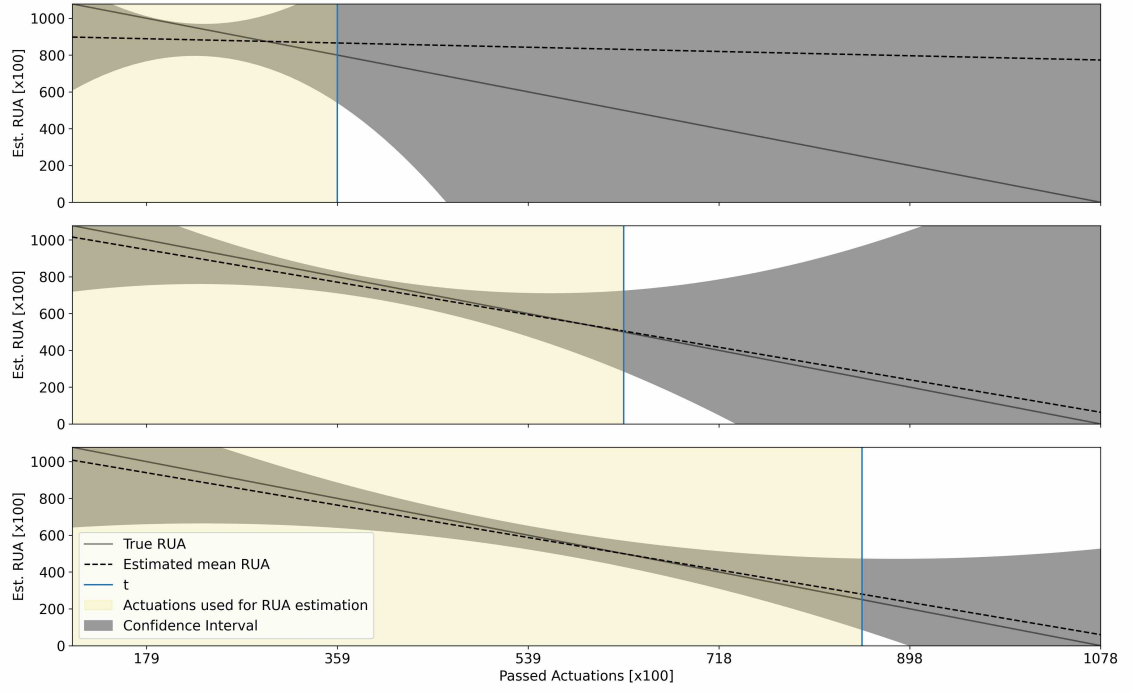


Figure C.26: DA1-03U - F_T - LI - $k = 4-d_b = 2-R = 6$.

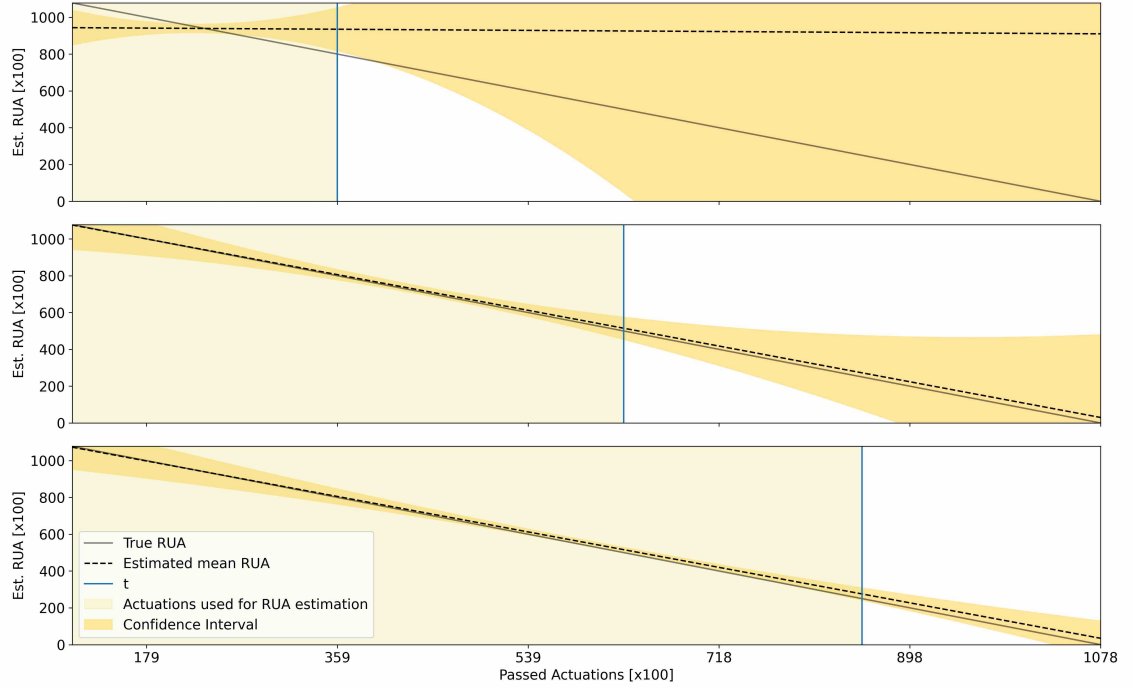


Figure C.27: DA1-03U - F_S - LI - $k = 4-d_b = 2-R = 6$.

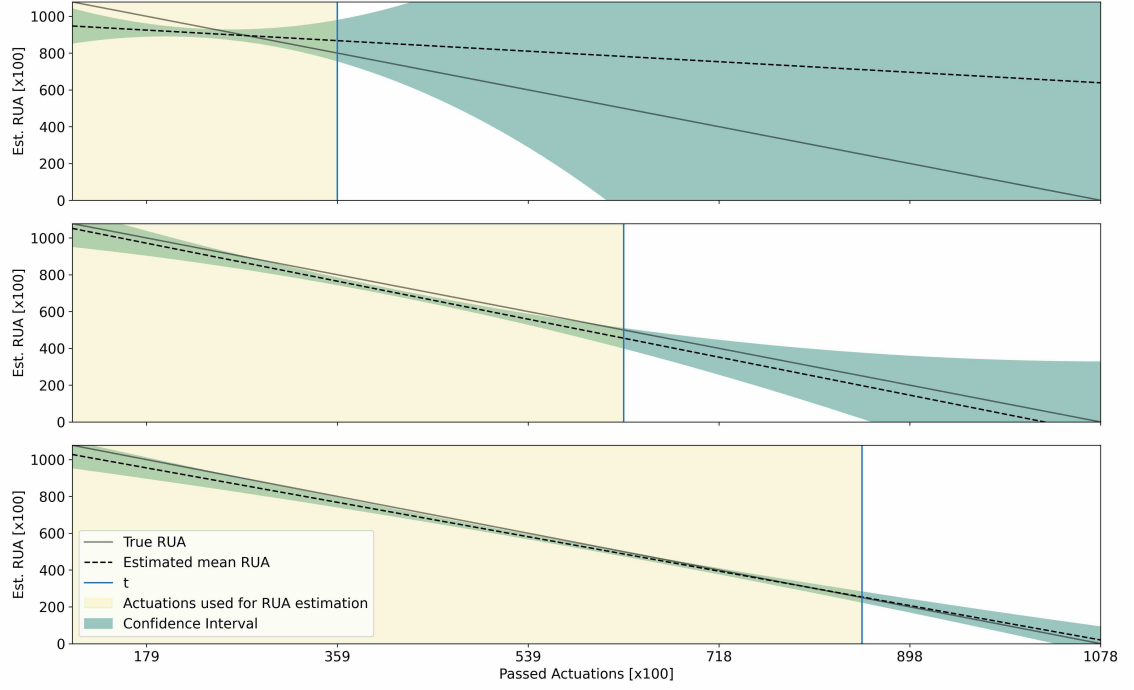


Figure C.28: DA1-03U - $F_{T,S}$ - LI - $k = 4 - d_b = 2 - R = 6$.

C.3.3.3 GI

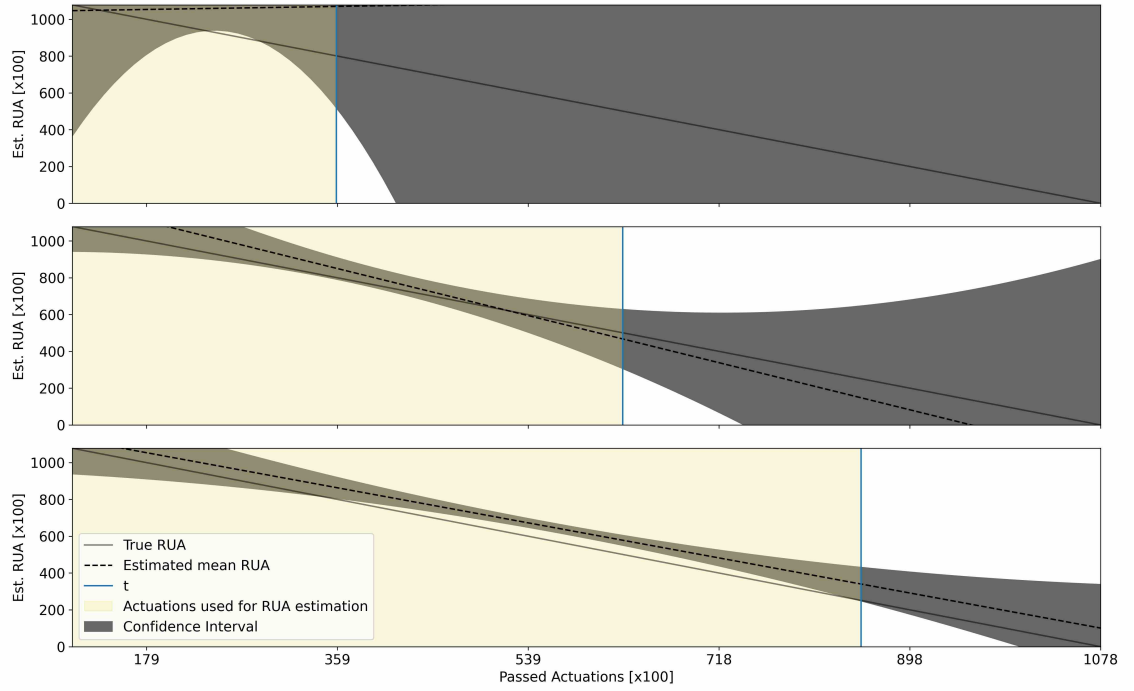


Figure C.29: DA1-03U - F_T - GI - $k = 4 - d_b = 2 - R = 6$.

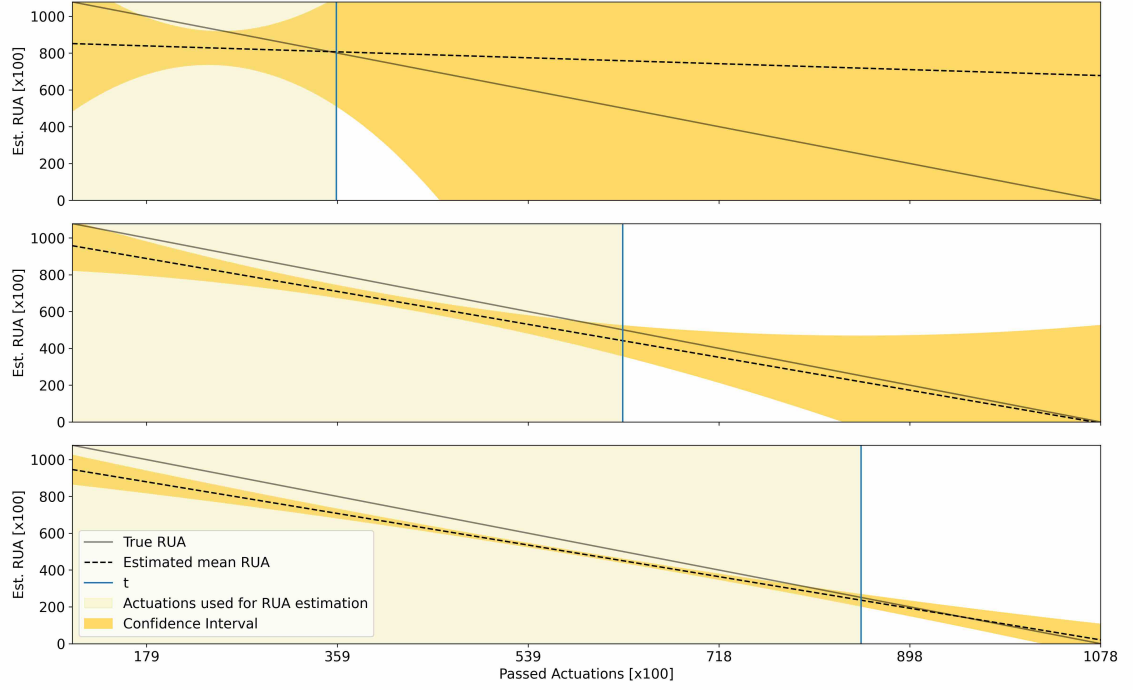


Figure C.30: DA1-03U - F_S - GI - $k = 4 - d_b = 2 - R = 6$.

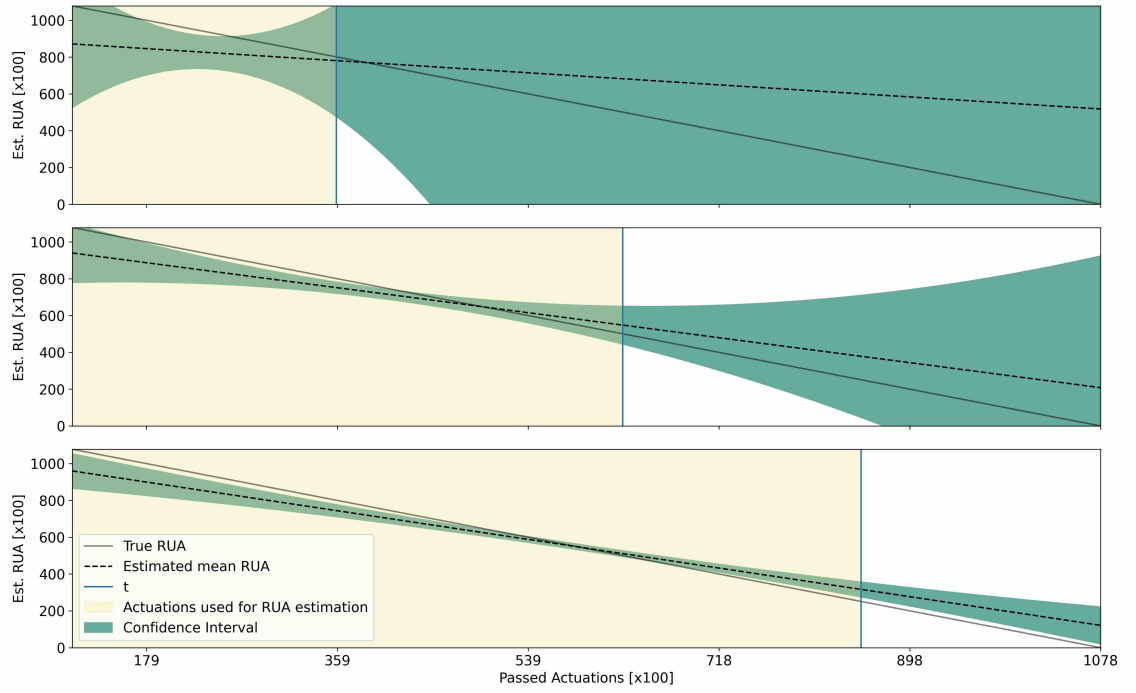


Figure C.31: DA1-03U - $F_{T,S}$ - GI - $k = 4 - d_b = 2 - R = 6$.

C.3.4 DA1-04U

C.3.4.1 EI

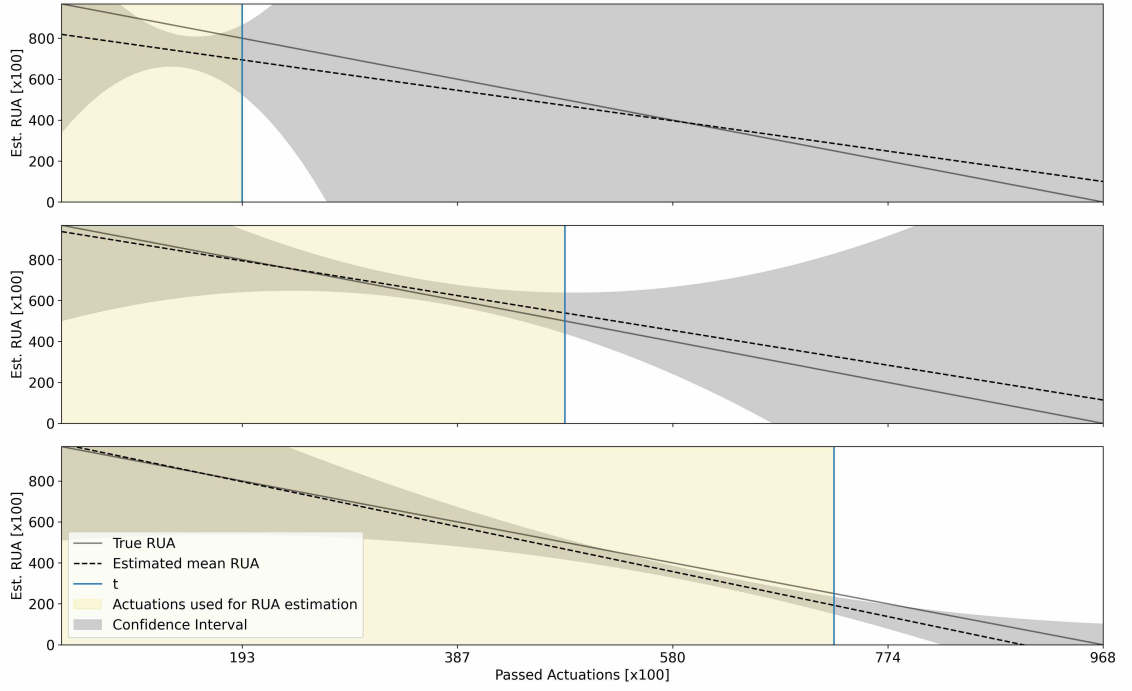


Figure C.32: DA1-04U - F_T - EI - $k = 4 - d_b = 2 - R = 6$.

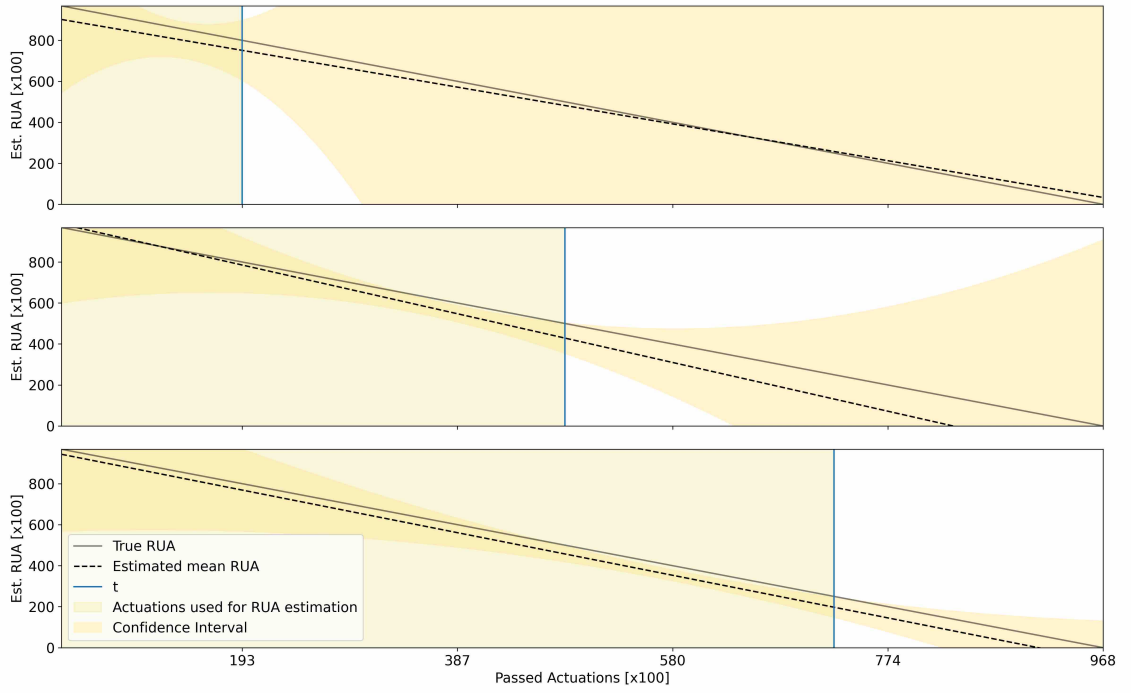


Figure C.33: DA1-04U - F_S - EI - $k = 4 - d_b = 2 - R = 6$.

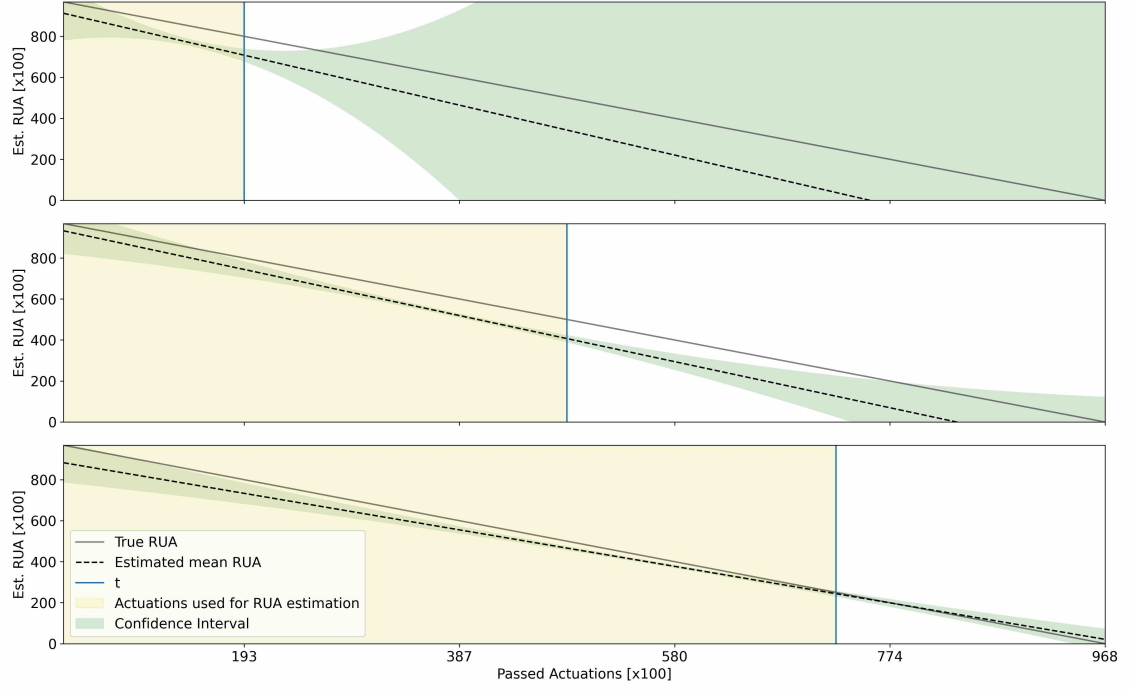


Figure C.34: DA1-04U - $F_{T,S}$ - EI - $k = 4-d_b = 2-R = 6$.

C.3.4.2 LI

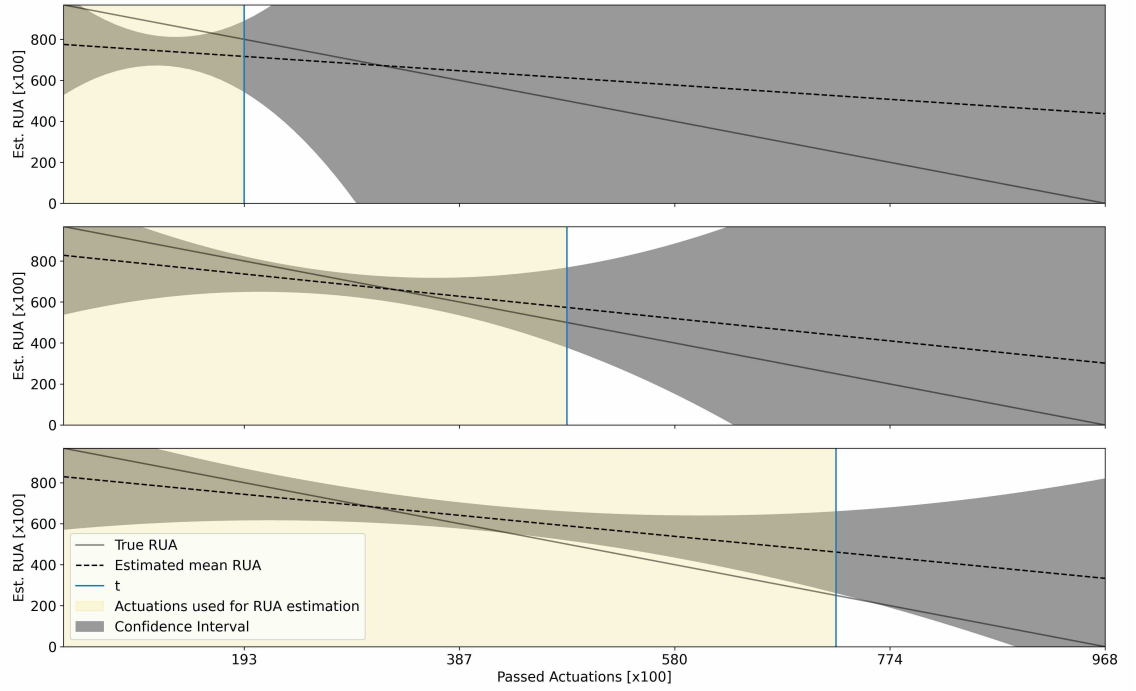


Figure C.35: DA1-04U - F_T - LI - $k = 4-d_b = 2-R = 6$.

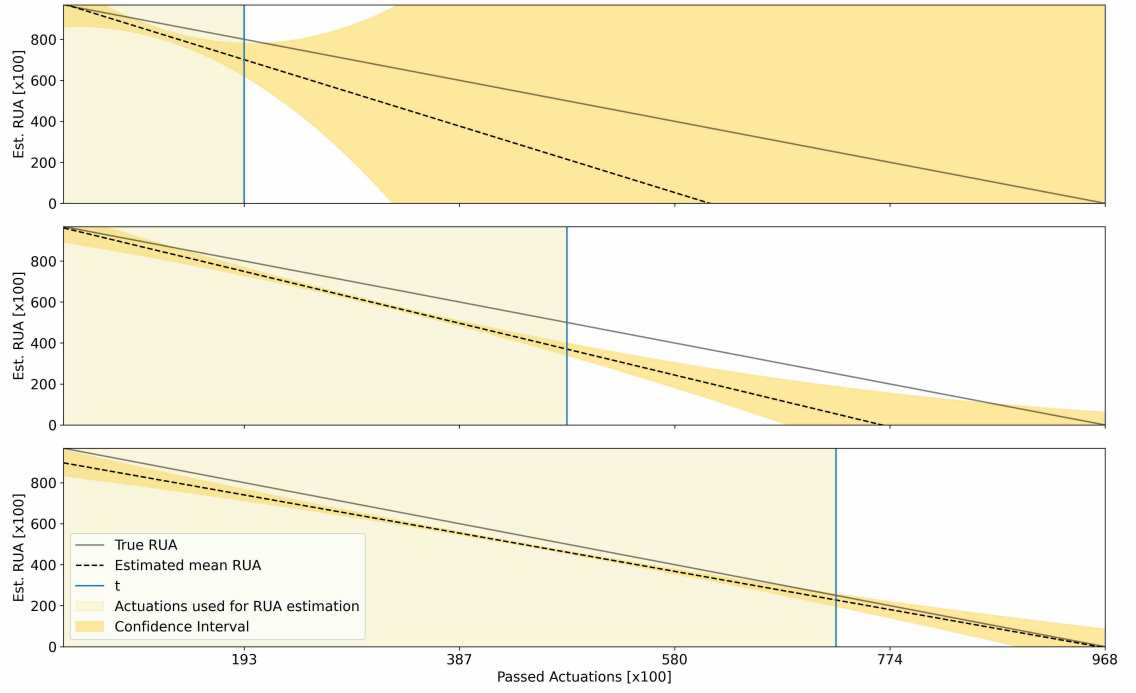


Figure C.36: DA1-04U - F_S - LI - $k = 4-d_b = 2-R = 6$.

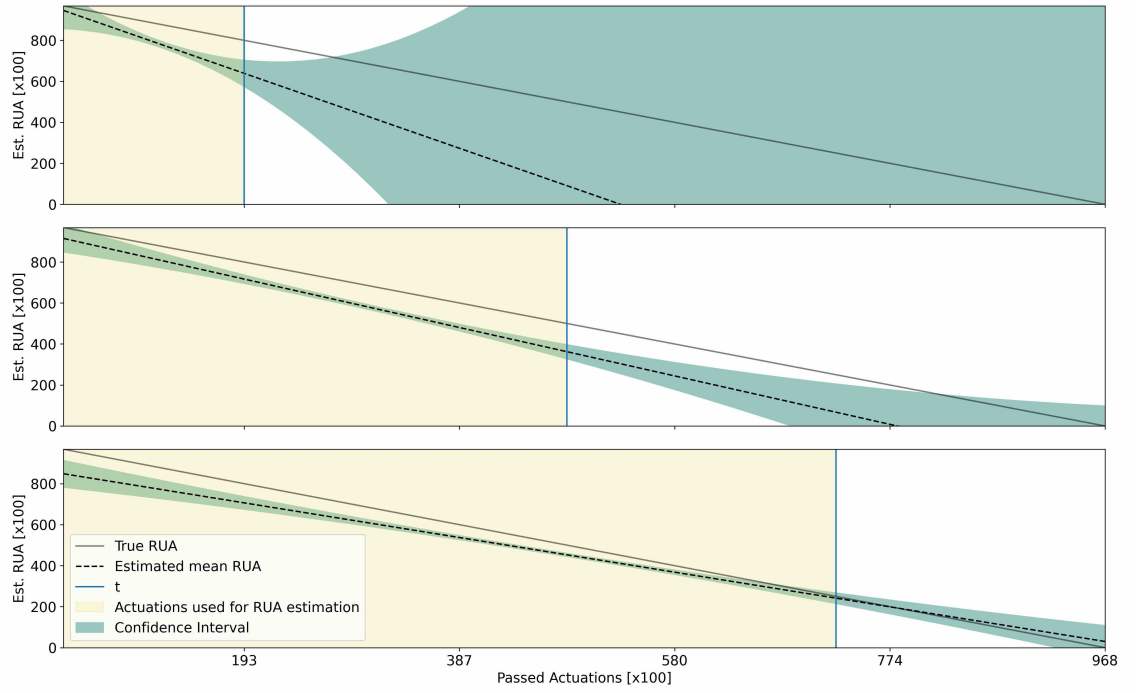


Figure C.37: DA1-04U - $F_{T,S}$ - LI - $k = 4-d_b = 2-R = 6$.

C.3.4.3 GI

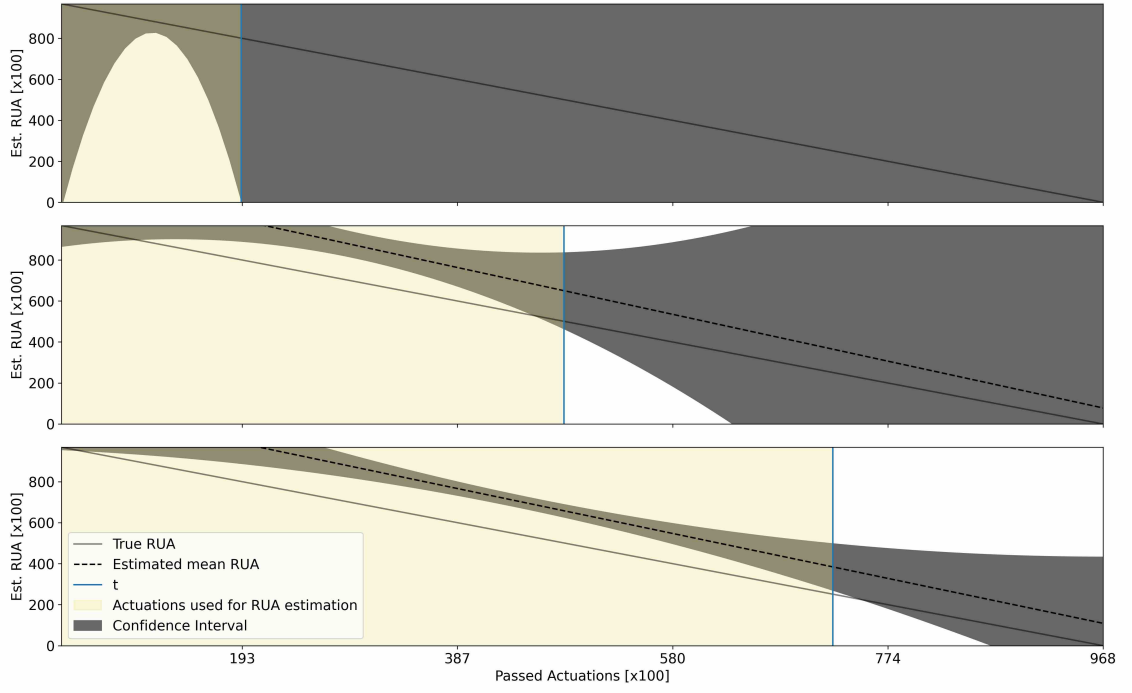


Figure C.38: DA1-04U - F_T - GI - $k = 4 - d_b = 2 - R = 6$.

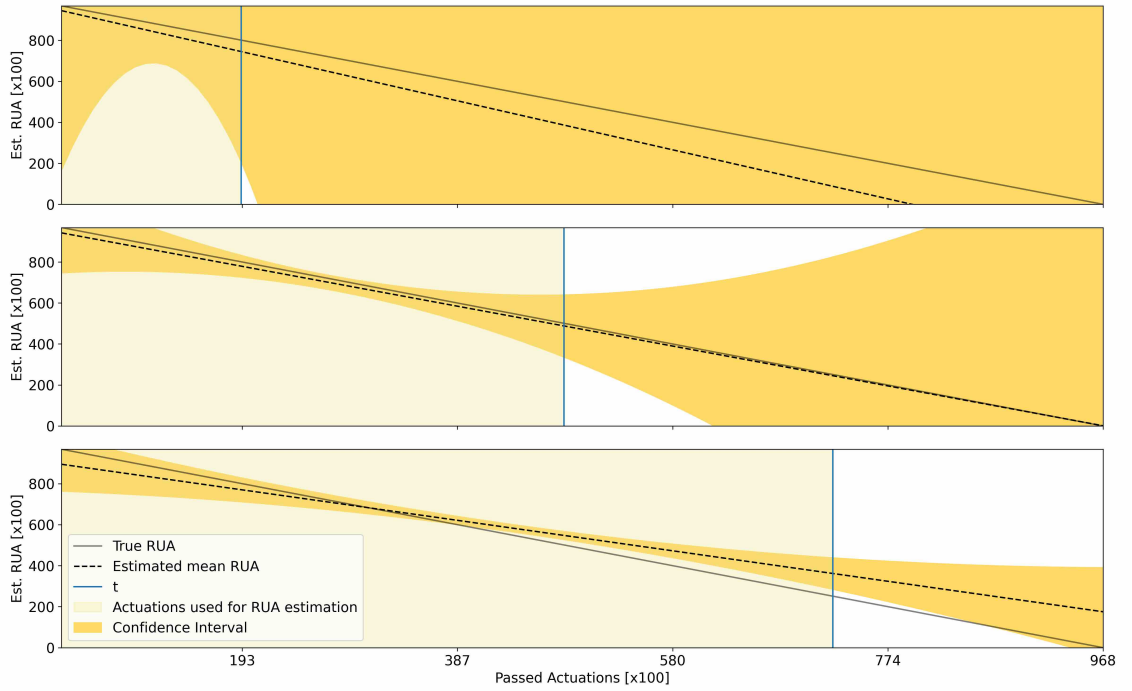


Figure C.39: DA1-04U - F_S - GI - $k = 4 - d_b = 2 - R = 6$.

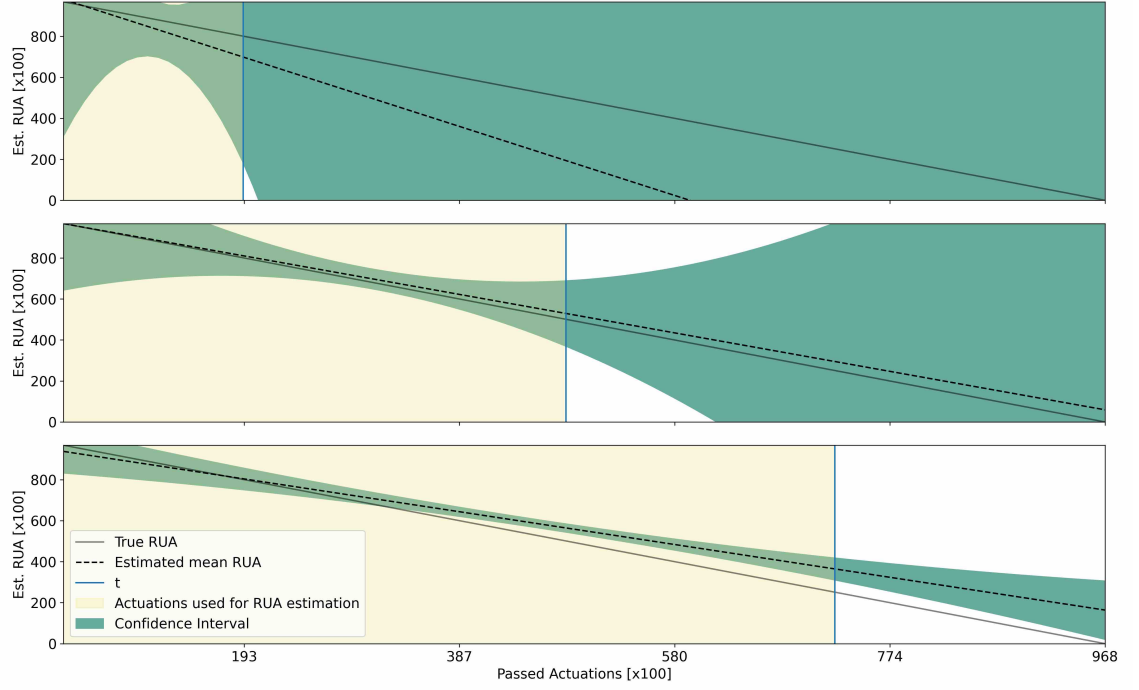


Figure C.40: DA1-04U - $F_{T,S}$ - GI - $k = 4 - d_b = 2 - R = 6$.

C.3.5 DA1-05U

C.3.5.1 EI

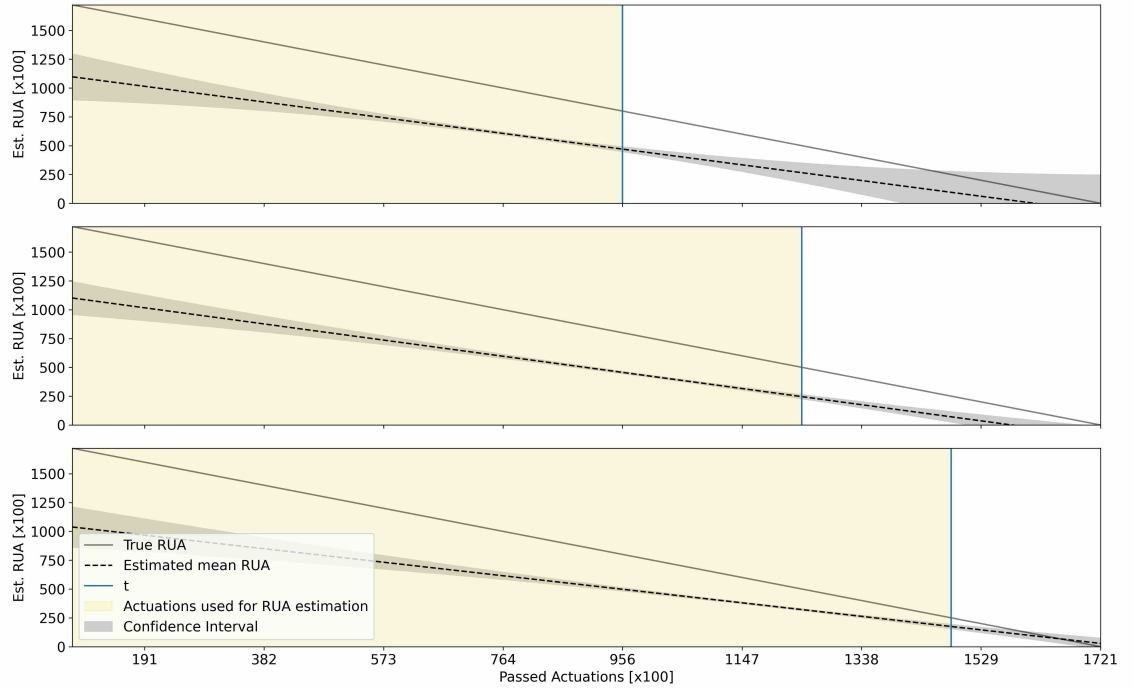


Figure C.41: DA1-05U - F_T - EI - $k = 4 - d_b = 2 - R = 6$.

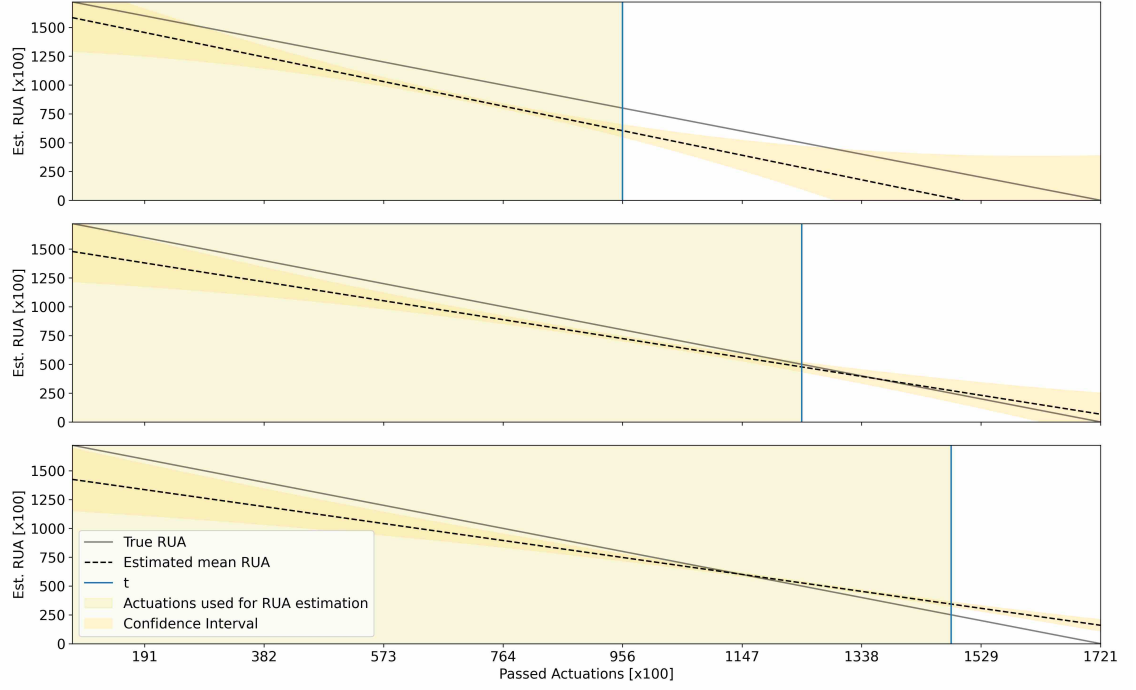


Figure C.42: DA1-05U - F_S - EI - $k = 4-d_b = 2-R = 6$.

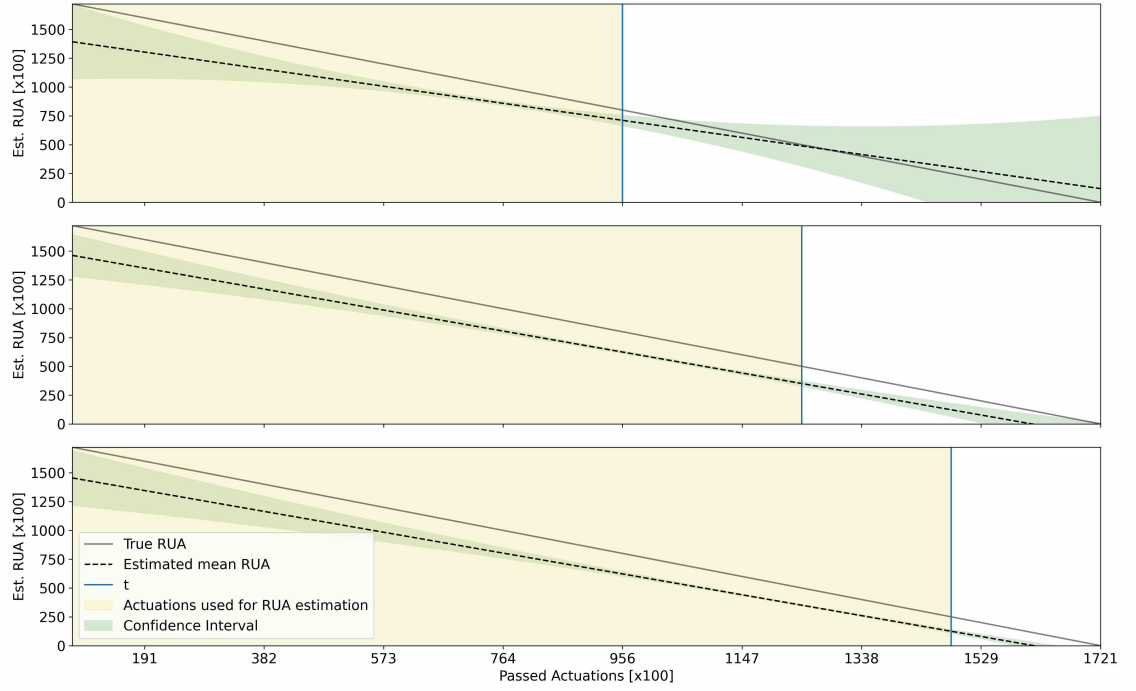


Figure C.43: DA1-05U - $F_{T,S}$ - EI - $k = 4-d_b = 2-R = 6$.

C.3.5.2 LI

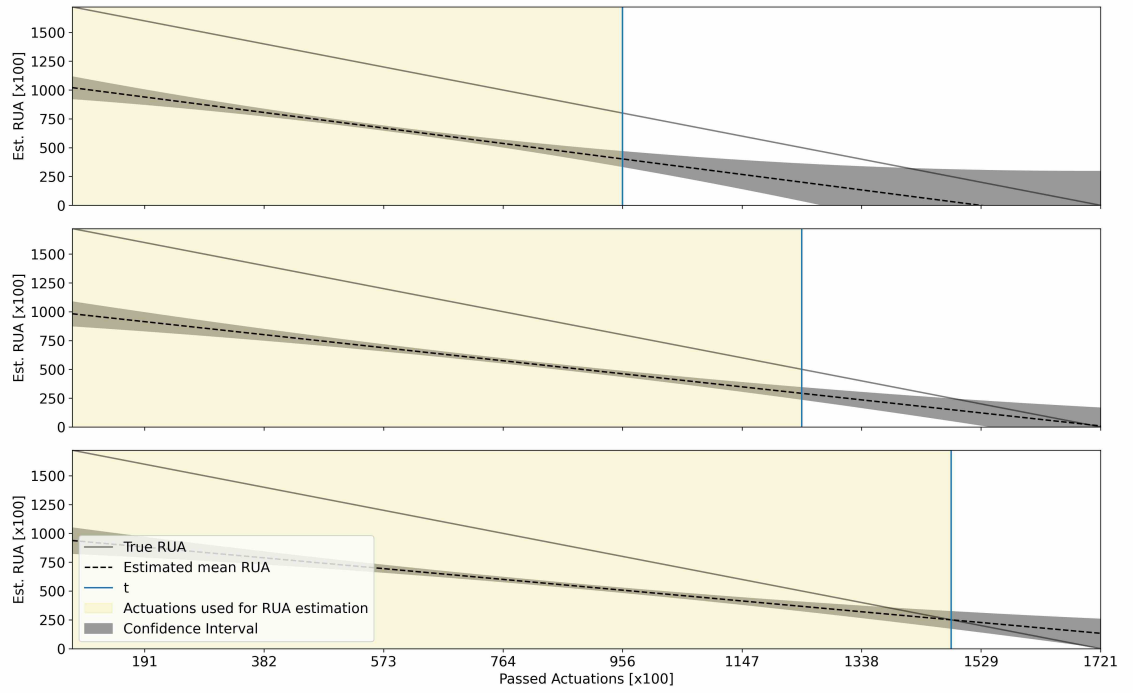


Figure C.44: DA1-05U - F_T - LI - $k = 4-d_b = 2-R = 6$.

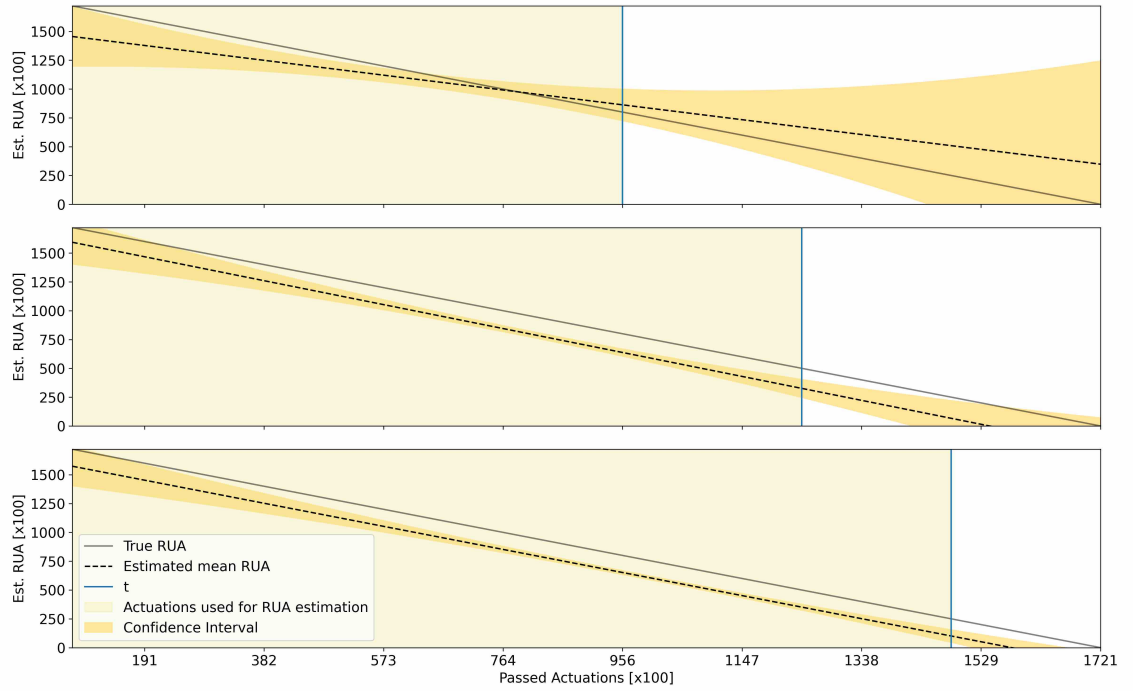


Figure C.45: DA1-05U - F_S - LI - $k = 4-d_b = 2-R = 6$.

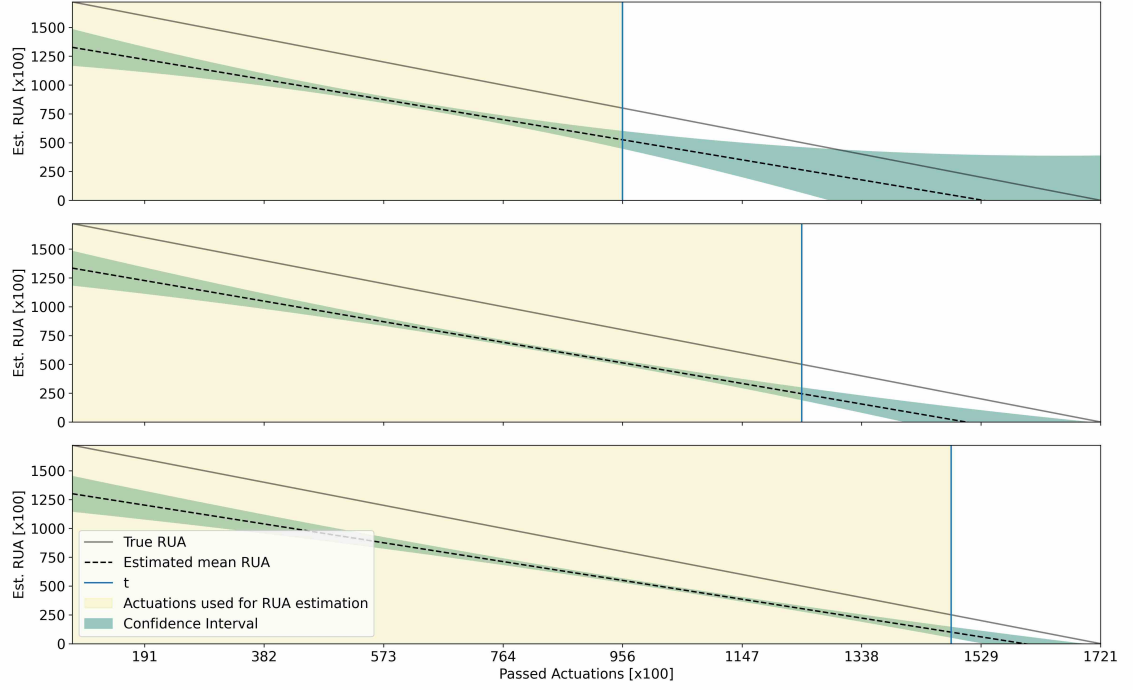


Figure C.46: DA1-05U - $F_{T,S}$ - LI - $k = 4-d_b = 2-R = 6$.

C.3.5.3 GI

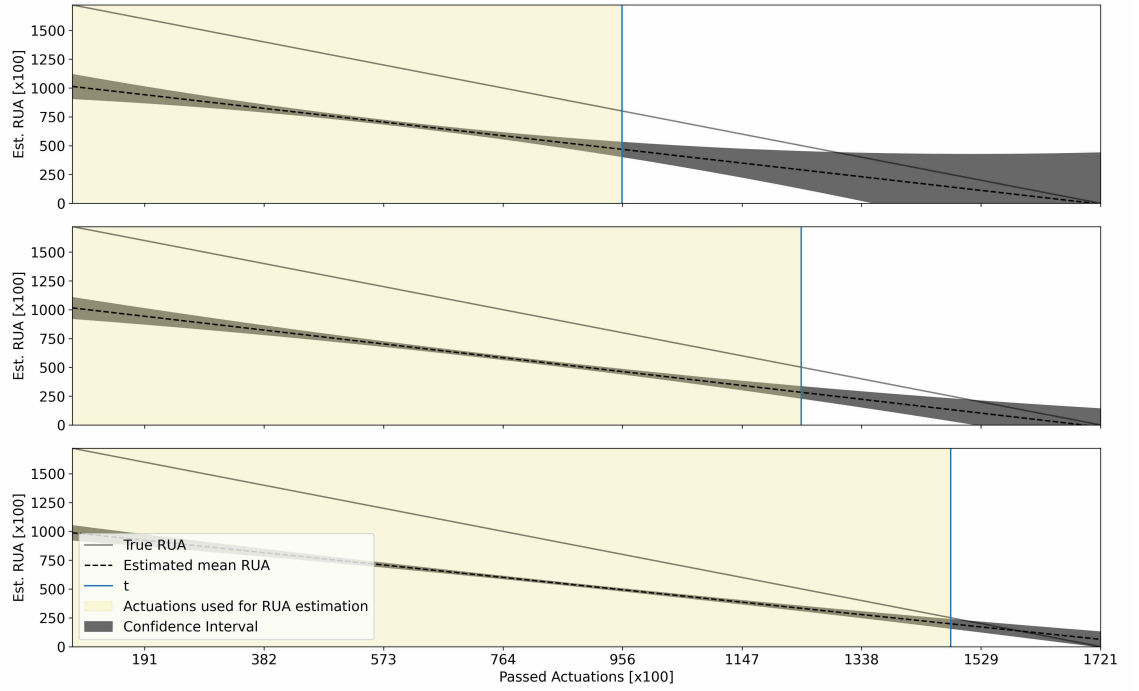


Figure C.47: DA1-05U - F_T - GI - $k = 4-d_b = 2-R = 6$.

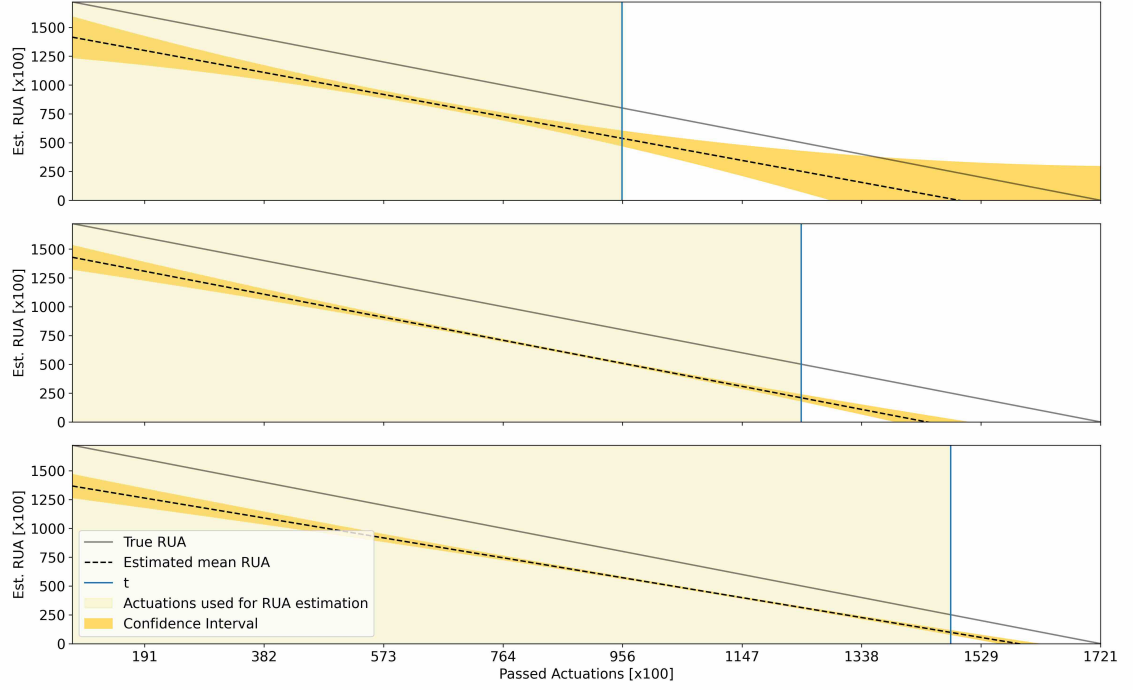


Figure C.48: DA1-05U - F_S - GI - $k = 4 - d_b = 2 - R = 6$.

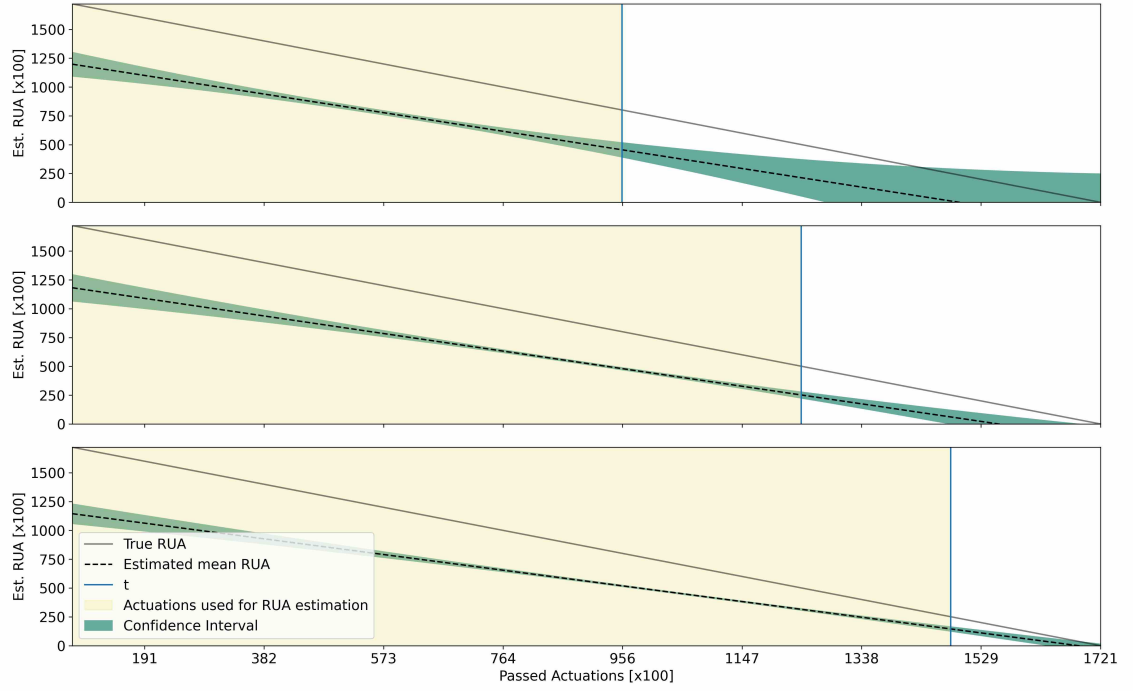


Figure C.49: DA1-05U - $F_{T,S}$ - GI - $k = 4 - d_b = 2 - R = 6$.

C.3.6 DA1-06U

C.3.6.1 EI

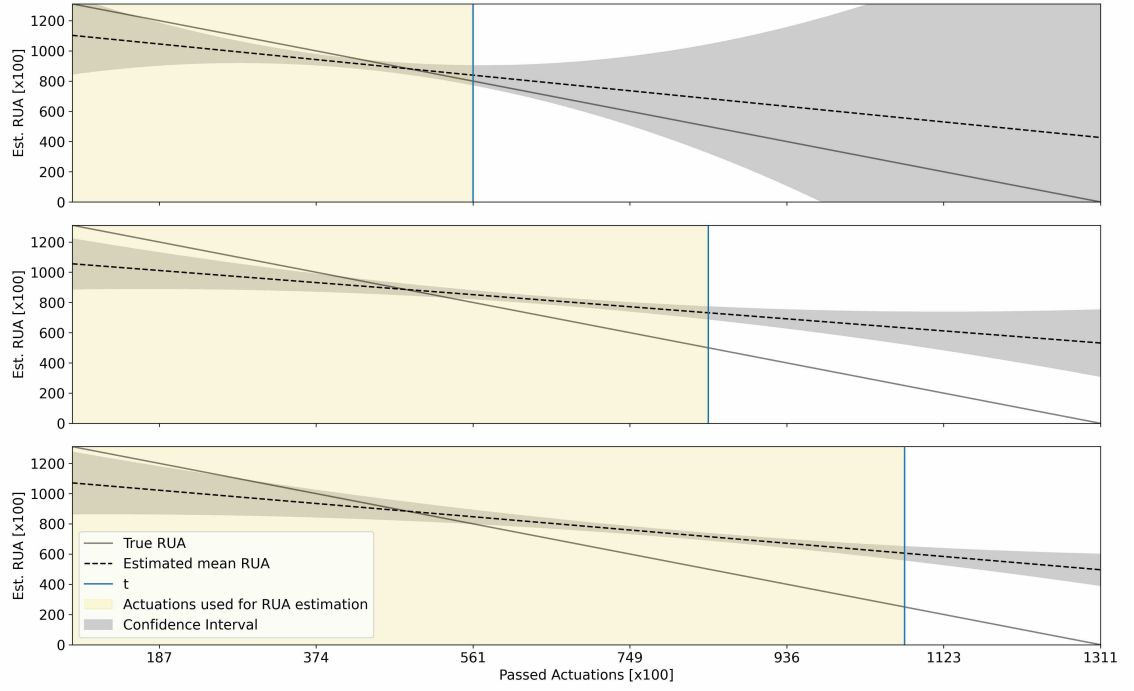


Figure C.50: DA1-06U - F_T - EI - $k = 4 - d_b = 2 - R = 6$.

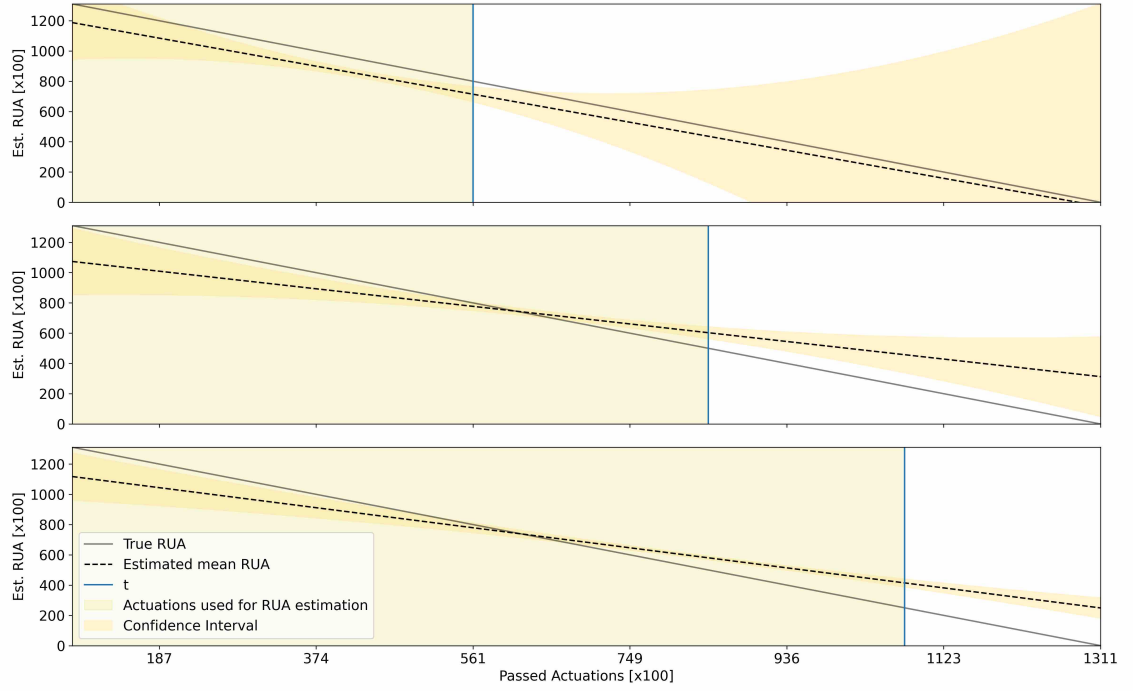


Figure C.51: DA1-06U - F_S - EI - $k = 4 - d_b = 2 - R = 6$.

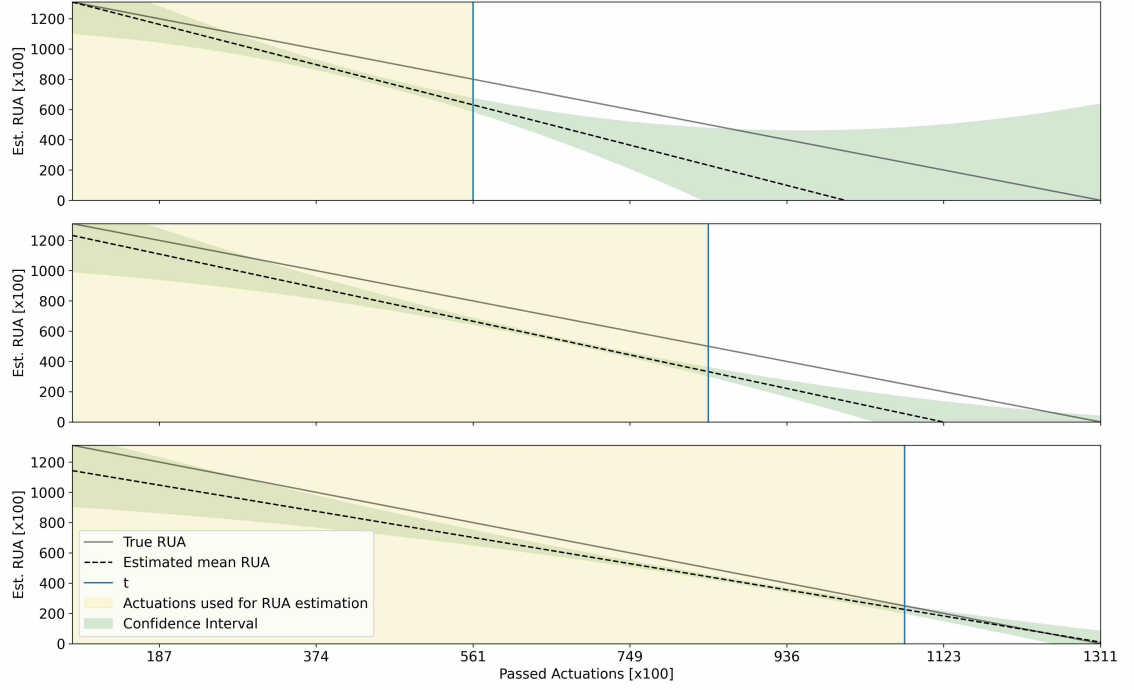


Figure C.52: DA1-06U - $F_{T,S}$ - EI - $k = 4-d_b = 2-R = 6$.

C.3.6.2 LI

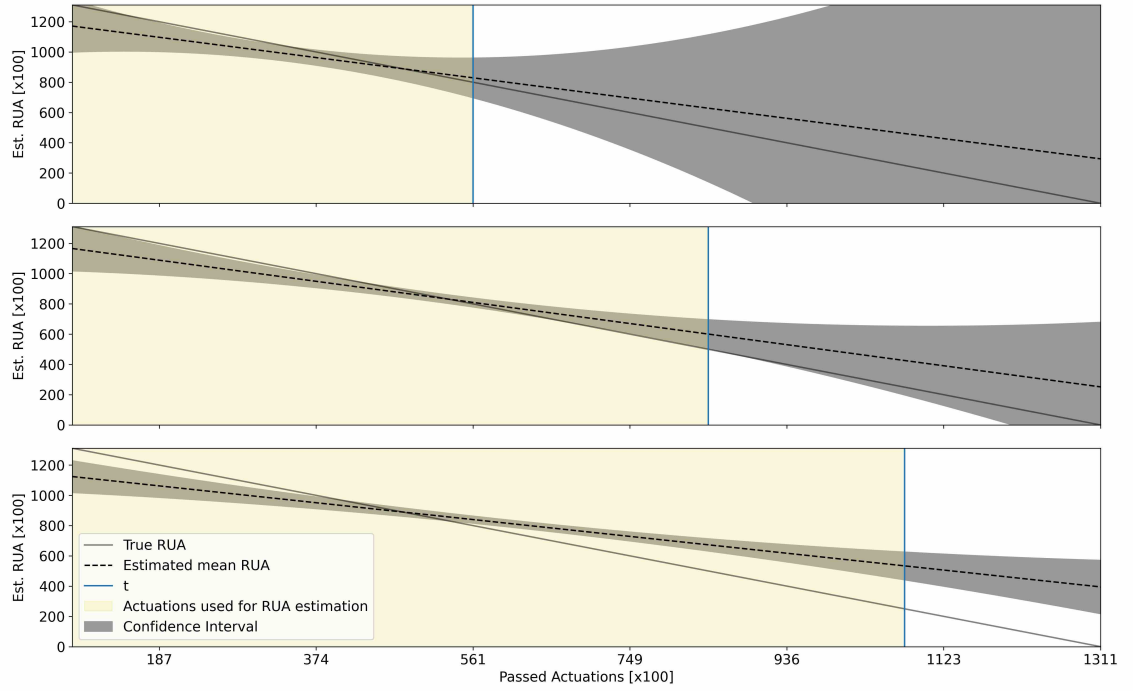


Figure C.53: DA1-06U - F_T - LI - $k = 4-d_b = 2-R = 6$.

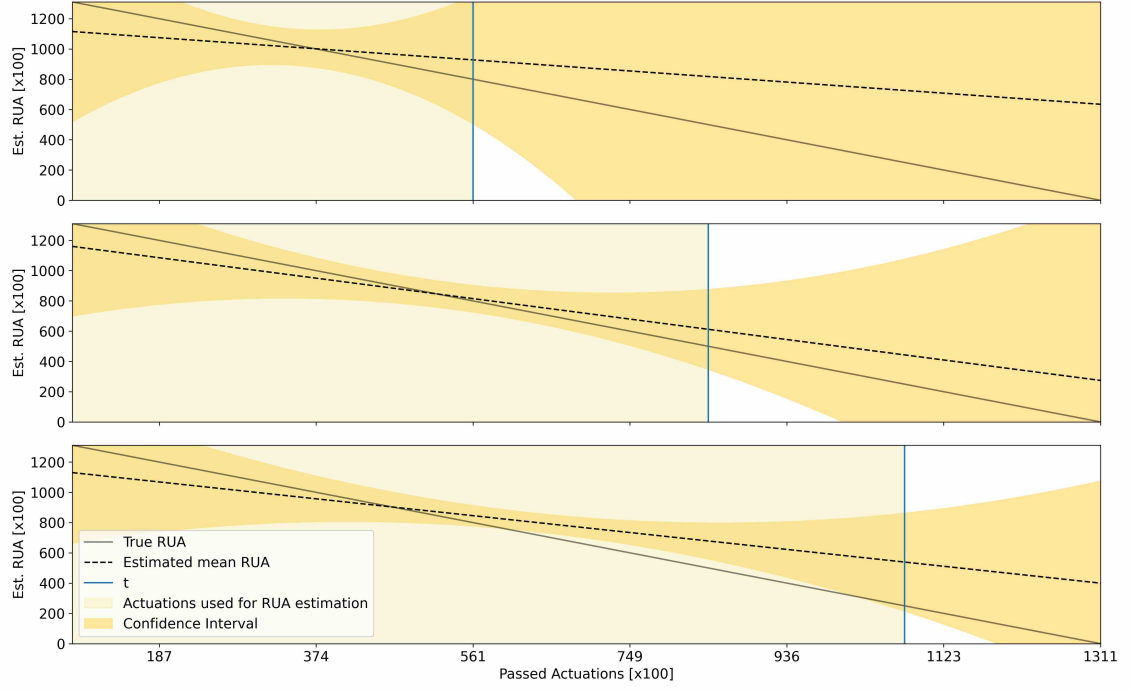


Figure C.54: DA1-06U - F_S - LI - $k = 4-d_b = 2-R = 6$.

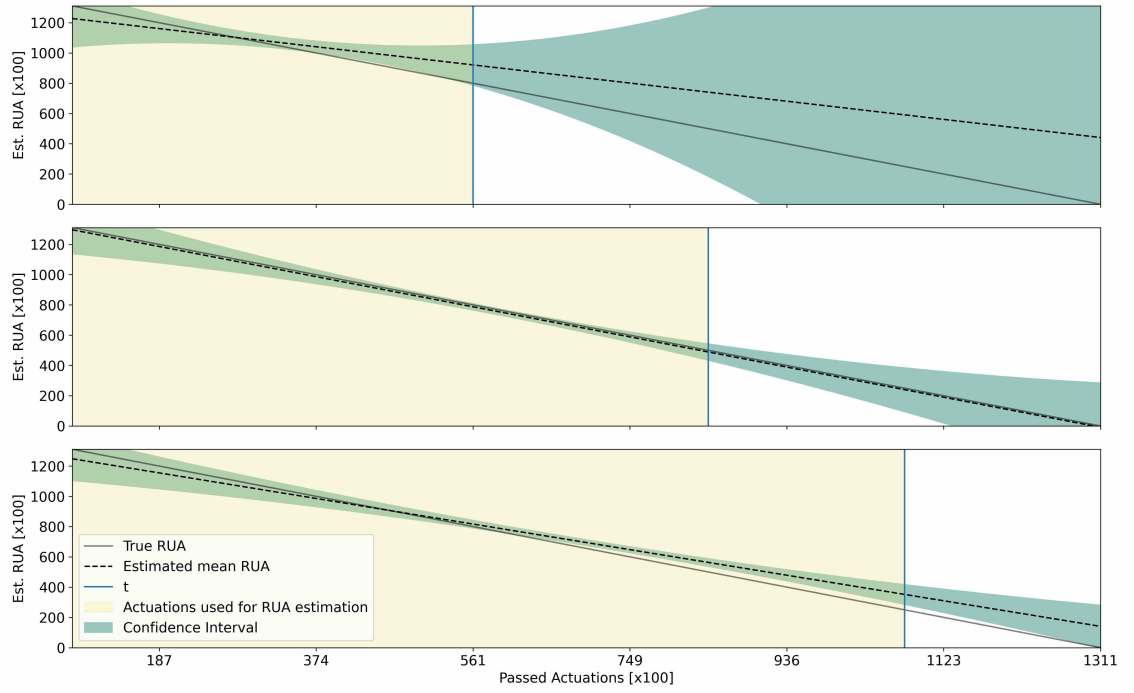


Figure C.55: DA1-06U - $F_{T,S}$ - LI - $k = 4-d_b = 2-R = 6$.

C.3.6.3 GI

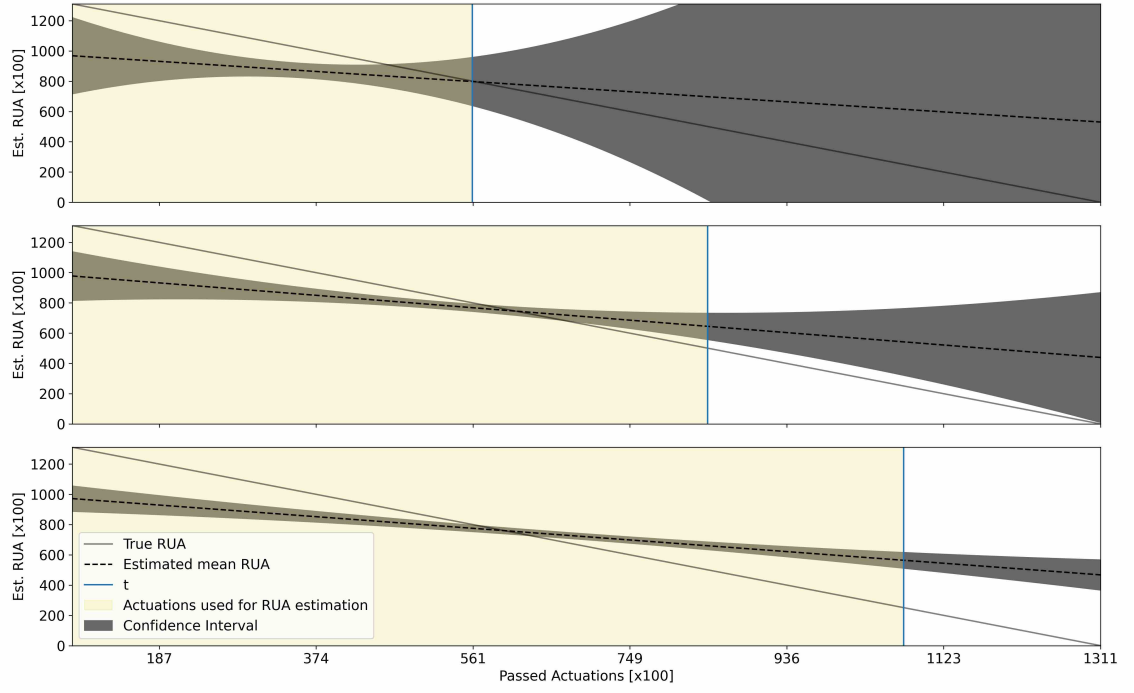


Figure C.56: DA1-06U - F_T - GI - $k = 4-d_b = 2-R = 6$.

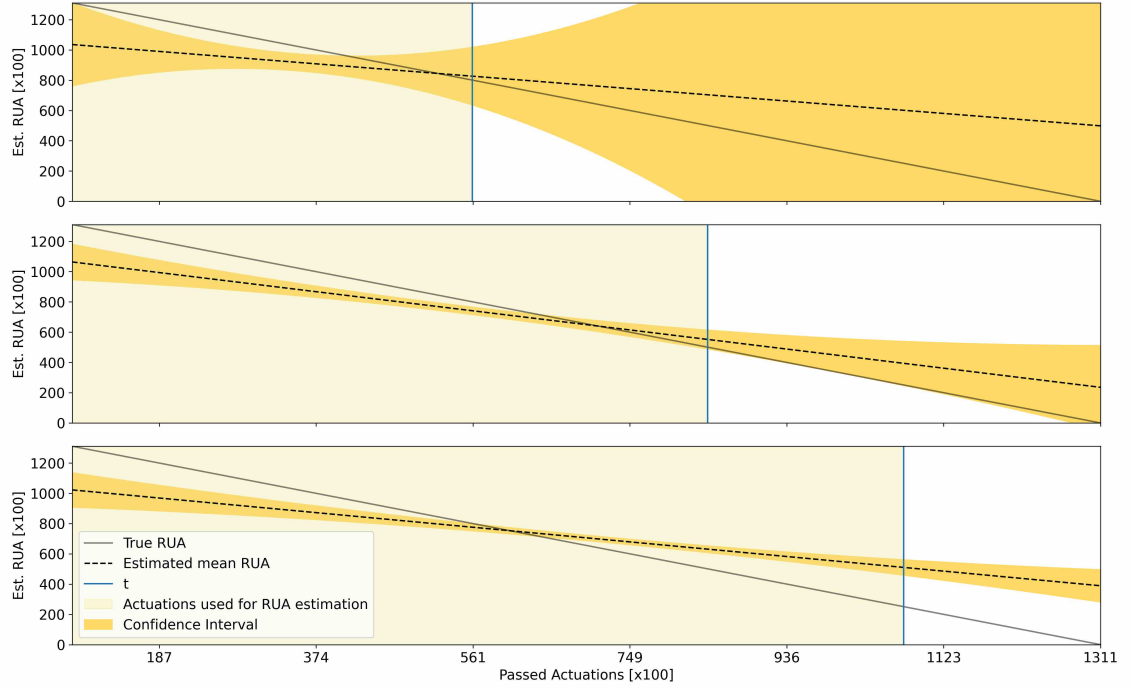
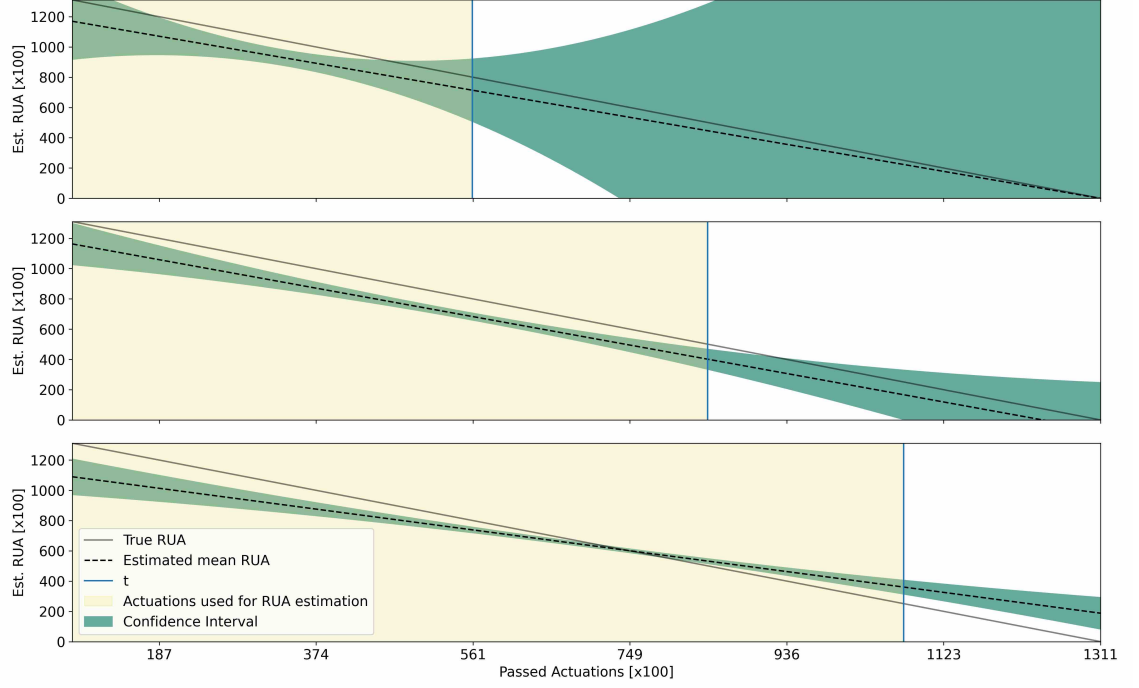


Figure C.57: DA1-06U - F_S - GI - $k = 4-d_b = 2-R = 6$.

Figure C.58: DA1-06U - $F_{T,S}$ - GI - $k = 4 - d_b = 2 - R = 6$.

C.4 Batch size effects on uncertainty estimation

Exemplary predicting RUA for DA1-16U. The EI sequence selection strategy and the data set $F_{T,S}$ are used.

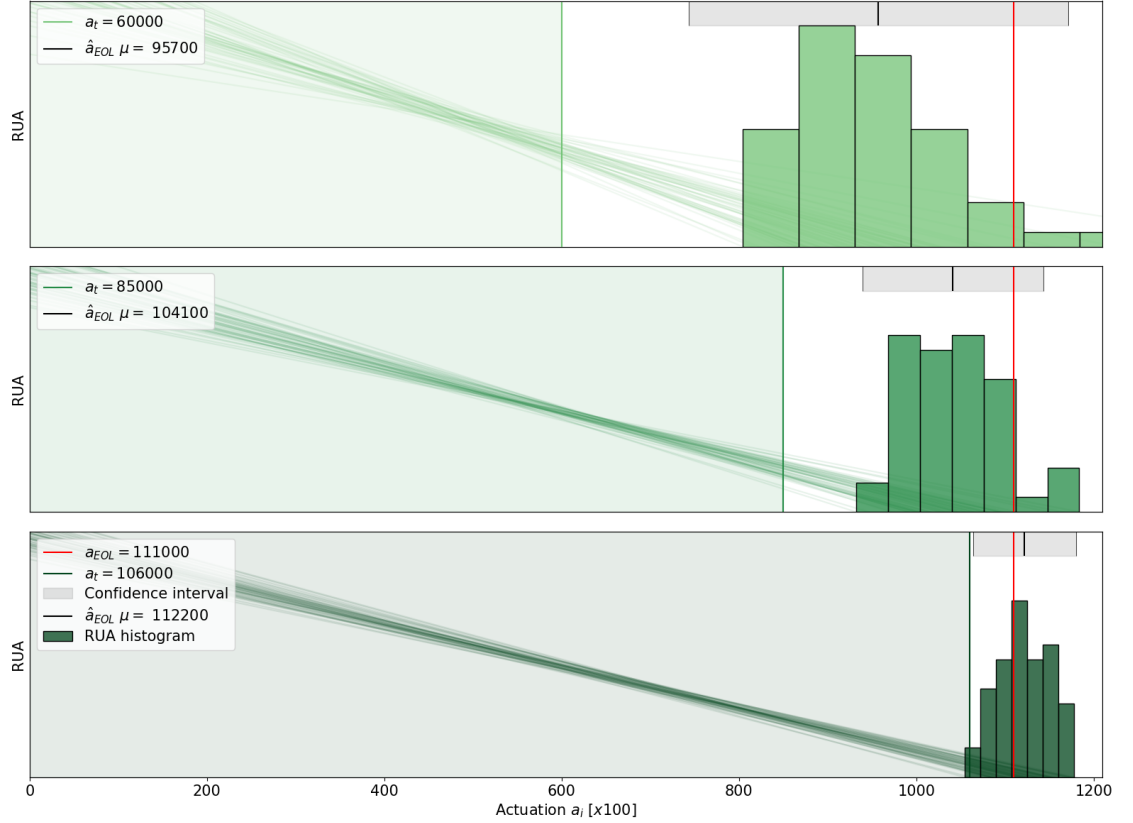


Figure C.59: DA1-16U - $N = 50$.

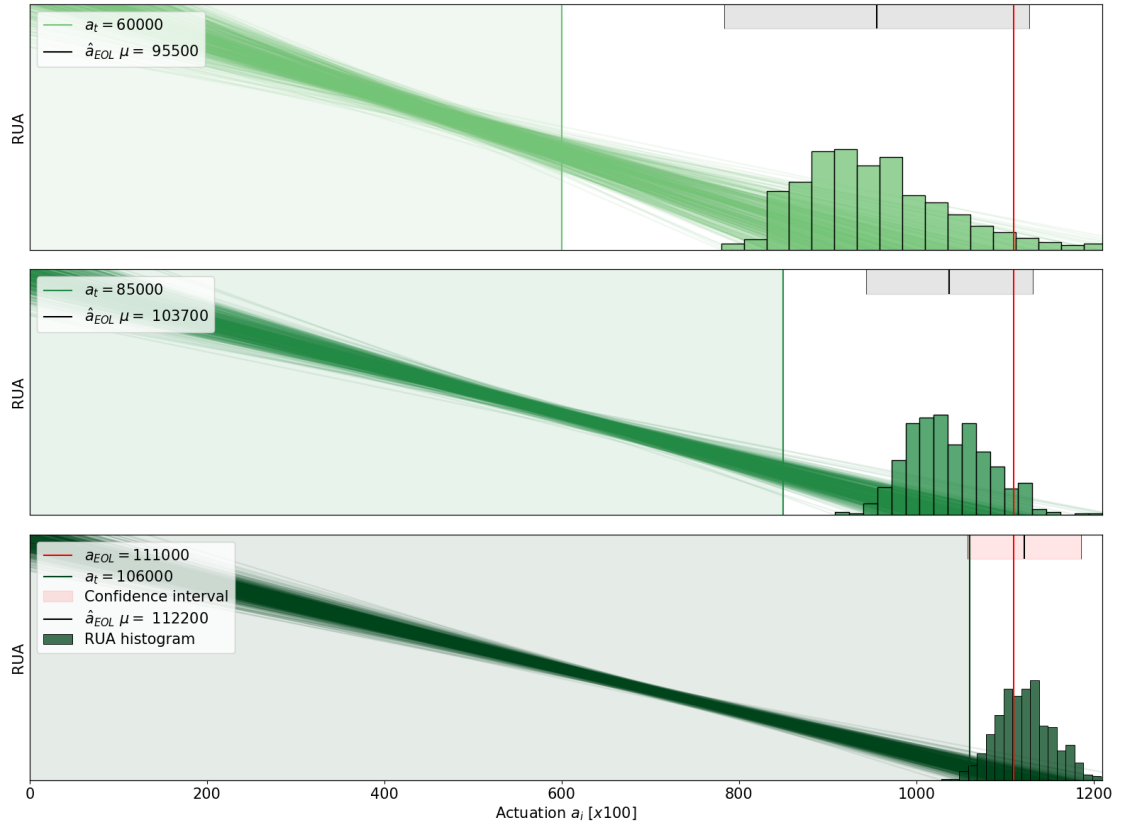


Figure C.60: DA1-16U - $N = 500$.

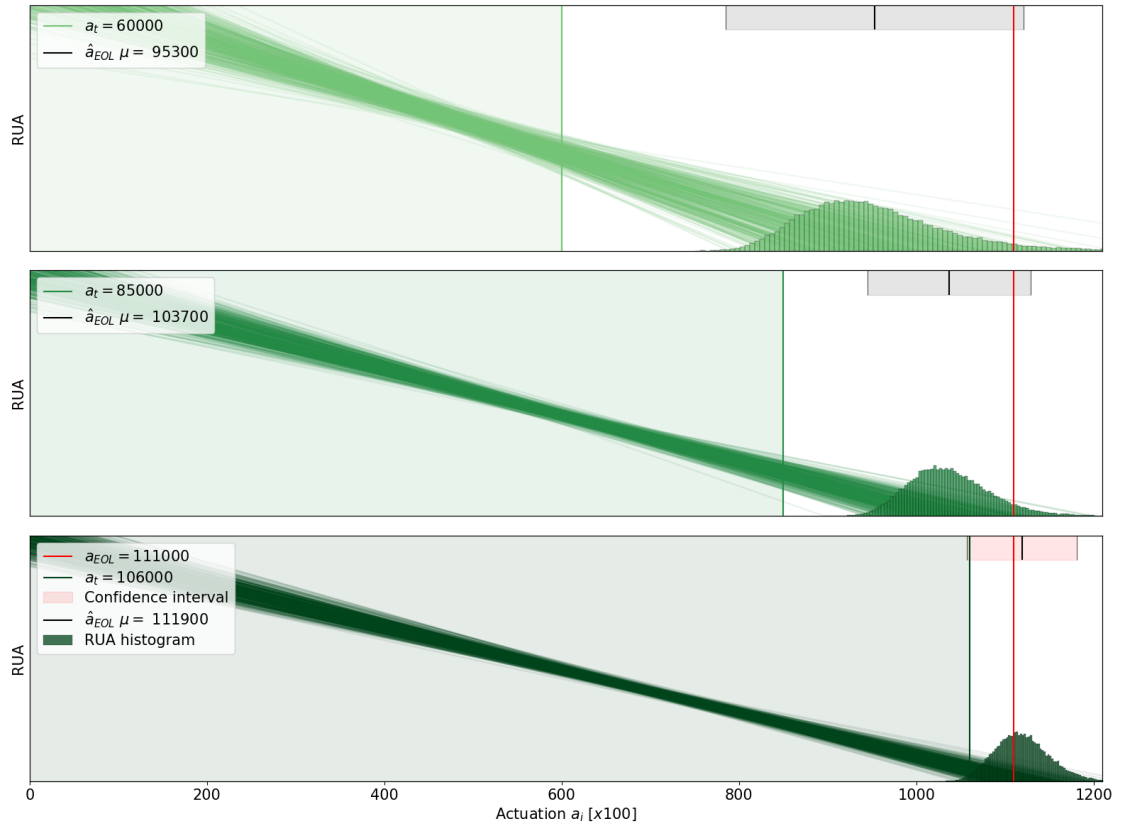


Figure C.61: DA1-16U - $N = 50000$.