

Modellbasierte Simulation und Prädiktion der Herzfrequenz im Ausdauersport zur Einschätzung der Leistungsentwicklung

Inauguraldissertation

zur Erlangung des Akademischen Grades
eines Doktors der Philosophie
(Dr. phil.)

vorgelegt dem Fachbereich 02 – Sozialwissenschaften, Medien und Sport
der Johannes Gutenberg Universität
Mainz

von

Melanie Ludwig

aus

Andernach

Sankt Augustin, 2023

Referent: Prof. Dr. Mark Pfeiffer
Johannes Gutenberg Universität Mainz

Korreferent: Prof. Dr. Alexander Asteroth
Hochschule Bonn-Rhein-Sieg

Tag des Prüfungskolloquiums: 10. Juli 2023

Deutsches Abstract

Eine Überprüfung der Leistungsentwicklung im Radsport geht bis heute mit der Durchführung einer spezifischen Leistungsdiagnostik unter Verwendung vorgegebener Testprotokolle einher. Durch die zwischenzeitlich stark gestiegene Popularität von »wearable devices« ist es gleichzeitig heutzutage sehr einfach, die Herzfrequenz im Alltag und bei sportlichen Aktivitäten aufzuzeichnen. Doch eine geeignete Modellierung der Herzfrequenz, die es ermöglicht, Rückschlüsse über die Leistungsentwicklung ziehen zu können, fehlt bislang.

Die Herzfrequenzaufzeichnungen in Kombination mit einer phänomenologisch interpretierbaren Modellierung zu nutzen, um auf möglichst direkte Weise und ohne spezifische Anforderungen an die Trainingsfahrten Rückschlüsse über die Leistungsentwicklung ziehen zu können, bietet die Chance, sowohl im professionellen Radsport wie auch in der ambitionierten Radsportpraxis den Erkenntnisgewinn über die eigene Leistungsentwicklung maßgeblich zu vereinfachen.

In der vorliegenden Arbeit wird ein neuartiges und phänomenologisch interpretierbares Modell zur Simulation und Prädiktion der Herzfrequenz beim Radsport vorgestellt und im Rahmen einer empirischen Studie validiert. Dieses Modell ermöglicht es, die Herzfrequenz (sowie andere Beanspruchungsparameter aus Atemgasanalysen) mit adäquater Genauigkeit zu simulieren und bei vorgegebener Wattbelastung zu prognostizieren. Weiterhin wird eine Methode zur Reduktion der Anzahl der kalibrierbaren freien Modellparameter vorgestellt und in zwei empirischen Studien validiert. Nach einer individualisierten Parameterreduktion kann das Modell mit lediglich einem einzigen freien Parameter verwendet werden. Dieser verbleibende freie Parameter bietet schließlich die Möglichkeit, im zeitlichen Verlauf mit dem Verlauf der Leistungsentwicklung verglichen zu werden. In zwei unterschiedlichen Studien zeigt sich, dass der freie Modellparameter grundsätzlich in der Lage zu sein scheint, den Verlauf der Leistungsentwicklung über die Zeit abzubilden.

English Abstract

To this day, a review of performance development in cycling goes hand in hand with the performance of specific performance diagnostics using predefined test protocols. The increased popularity of »wearable devices« otherwise makes it effortless nowadays to record heart rate in everyday life and during sporting activities. However, there is still no suitable approach to draw conclusions about long-term performance development from simple heart rate measurements.

The use of heart rate recordings in combination with phenomenologically interpretable modelling to draw conclusions about performance development without specific requirements for training rides offers the chance to substantially simplify the gain of knowledge about one's own performance development—in professional cycling as well as in ambitious cycling practice.

In the present work an alternative and phenomenologically interpretable model for heart rate simulation and heart rate prediction is presented and validated within an empirical study. This model enables to simulate the heart rate (as well as other load parameters from respiratory gas analyses) with sufficient accuracy and to predict them at a given watt load. Furthermore, a method is presented to reduce the number of calibratable free model parameters and validated throughout two empirical studies. After an adjusted parameter reduction, the model can be used with a single free parameter. Finally, this remaining free parameter offers the possibility to be compared to performance development. Two different empirical studies show that the free model parameter seems to be generally able to simulate the trend of performance development over time.

Danksagung

Meinen beiden betreuenden Professoren, Mark Pfeiffer und Alexander Asteroth – für die angenehme Zusammenarbeit, die zahlreichen wissenschaftlichen Diskussionen, die vielen Anregungen zu neuen Forschungsideen, Euer wertvolles Feedback und Eure Zeit.

Alexander Asteroth – zusätzlich besonders für Dein Mentoring, Deine Unterstützung, Dein Vertrauen und Deine Ratschläge, auch in Bezug auf zukünftige Wege, und das schon seit Beginn meines Masterstudiums.

Christian Rasche sowie Katrin Hoffmann, Stefan Endler und Kevin Bach – für Eure unterschiedlichen Blickwinkel und die wertvolle Zusammenarbeit.

Jakob Meng, Aysegül Yasari, Ahmad Drak, Youssef Mahmoud Youssef, Alexander Hagg, Martin Müller – für den Austausch und die Extraportion kleiner Ablenkungen im Alltag.

Dem Graduierten-Institut und der Gleichstellungsstelle der Hochschule Bonn-Rhein-Sieg sowie dem PK NRW – für die ideelle und finanzielle Unterstützung, sowie für die zahlreichen Möglichkeiten zur Weiterqualifizierung in den unterschiedlichsten Themenfeldern.

Dem Land NRW und dem Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen – für die Projektförderung des Projekts »effiziente Transportalternativen« (AZ: 322-8.03.04.02-FH-STRUKTUR 2017/07), im Rahmen dessen große Teile der Forschung für die Dissertation entstanden sind.

Thorsten Nowack und Leila Stöhr – für das stete »Rückenstärken«.

Thoralf Räsch und Michael Meier – für die lehrreiche, prägende und spannende Zeit im *Hausdorff Center for Mathematics* in den frühen Jahren meiner Studienzeit.

Susanne Hilger, Marika Enders, Thekla Kalamán, Tasja V. und Lin Neubauer – für den Rückhalt, das Mitfiebern, Eure Freundschaft und Euer stets offenes Ohr.

Marion Ludwig, Harald G. Grohganzen und Eira – für Eure Geduld, Euer Vertrauen in mich und Eure unermüdliche Unterstützung auf den unterschiedlichsten Ebenen.

Vielen herzlichen Dank!

Inhaltsverzeichnis

Abbildungsverzeichnis	xiii
Tabellenverzeichnis	xvii
Einleitung	1
Zielsetzung	3
Aufbau der Arbeit	7
Verwandte Publikationen	8
Abkürzungsverzeichnis	1
I. Theoretische Hintergründe	9
1. Leistungsphysiologische Grundlagen	15
1.1. Allgemeine physiologische Grundlagen	17
1.1.1. Energiebereitstellung	17
1.1.2. Adaptationen der beteiligten Systeme beim Sauerstofftransport	22
1.1.3. Physiologische Reaktionen und Anpassungen durch Ausdauerbelastung	35
1.2. Allgemeine Grundlagen der Trainingssteuerung im Ausdauersport	38
1.2.1. Belastungs-Beanspruchungs-Konzept	40
1.2.2. Belastungsnormative	42
1.2.3. Messgrößen der Beanspruchung in der Diagnostik	44
1.2.4. Schwellenkonzepte	49
1.2.5. Testverfahren im Ausdauersport (Radsport)	53
1.3. Herzfrequenzverhalten bei körperlicher Belastung	55
1.3.1. Die Herzfrequenz als Indikator: Kritische Aspekte	56
1.3.2. Gemeinsamkeiten akuter und chronischer Reaktionen auf Ausdauerbe- lastung	58
2. Grundlagen zur Modellierung des Herzfrequenzverhaltens im Ausdauersport	61
2.1. Modellierung, Simulation und Vorhersage	63
2.1.1. Zeitreihenanalyse und statistische Parameter	65

2.1.2. Visualisierung	73
2.1.3. Kritische Betrachtung statistischer Parameter	80
2.2. Modellierung der Herzfrequenz als akute Anpassung	91
2.2.1. Anforderungen an Modelle zur Herzfrequenzmodellierung	94
2.2.2. Herzfrequenzmodellierung: Stand der Forschung	95
2.3. Zusammenhang zwischen Herzfrequenz und Leistung in der Modellierung: Eine Einordnung	106
3. Zusammenfassung des Forschungsstands und Fragestellungen	109
II. Empirische Studien	115
4. Datenmaterial	119
4.1. Datensatz 1: Daten aus einer Studie von Gonzalez et al.	121
4.2. Datensatz 2: Daten aus einer Studie von Schrot mit Profi-Athleten	122
4.3. Datensatz 3: Daten von Hobby-Sportlern	124
5. Das Faltungsmodell: Ein neuer Ansatz zur Simulation und Vorhersage der Herzfrequenz im Training	127
5.1. Faltungsmodell	129
5.2. Analyse des Faltungsmodells: Modellvergleich	131
5.2.1. Experimenteller Aufbau und Methoden	133
5.2.2. Ergebnisse	134
5.2.3. Interpretation und Diskussion	137
5.2.4. Fazit zur Studie	140
5.3. Mathematische Betrachtung und Vergleich zu ähnlichen Modelltypen	141
5.3.1. Das Faltungsmodell als Differentialgleichungsmodell	141
5.3.2. Das Faltungsmodell als Wiener-Modell	146
5.4. Zusammenfassung und Fazit	146
6. Verbesserung des Faltungsmodells: Parameterreduktion	149
6.1. Parameterreduziertes Faltungsmodell	152
6.2. Konstruktion eines einparametrischen Faltungsmodells: Studiendaten von Profi-Sportlern	159
6.2.1. Durchführung der Parameterreduktion und Ergebnisse	160
6.2.2. Interpretation und Diskussion	175
6.2.3. Fazit zur Studie	180
6.3. Konstruktion eines einparametrischen Faltungsmodells: Studiendaten von Hobby-Sportlern	180
6.3.1. Durchführung der Parameterreduktion und Ergebnisse	180
6.3.2. Interpretation und Diskussion	190

6.3.3. Fazit zur Studie	192
6.4. Zusammenfassung und Fazit	193
7. Nutzen des parameterreduzierten Faltungsmodells bei der Leistungsbeurteilung	195
7.1. Analyse zur Abschätzung der Leistungsveränderung auf Basis von Herzfrequenz-Simulationen: Profi-Sportler	197
7.1.1. Experimenteller Aufbau und Methoden	198
7.1.2. Ergebnisse	201
7.1.3. Interpretation und Diskussion	206
7.1.4. Fazit zur Studie	211
7.2. Analyse zur Abschätzung der Leistungsveränderung auf Basis von Herzfrequenz-Simulationen: Hobby-Sportler	212
7.2.1. Experimenteller Aufbau und Methoden	212
7.2.2. Ergebnisse	214
7.2.3. Interpretation und Diskussion	221
7.2.4. Fazit zur Studie	227
7.3. Zusammenfassung und Fazit	228
8. Zusammenfassung und Ausblick	231
Literaturverzeichnis	235
Anhang	249

Abbildungsverzeichnis

1.	Schematische Darstellung der Gliederung der Dissertation	6
2.	Übersicht über die einzelnen Themenbereiche, Grundlagenteil 1	11
1.1.	Energiebereitstellung im Muskel	18
1.2.	Energiebereitstellung unterschiedlicher physiologischer Systeme	21
1.3.	Schematische Darstellung des Herz-Lungen-Kreislaufs	23
1.4.	Aufbau des Herzens	24
1.5.	Phasen der Herztätigkeit	25
1.6.	Lunge und Atmungssystem	28
1.7.	Typen von Muskelgewebe	30
1.8.	Aufbau eines Skelettmuskels	31
1.9.	Entspannung und Kontraktion im Sarkomer	32
1.10.	Ausdauerbereiche und Energiebereitstellung	40
1.11.	Belastungs-Beanspruchungs-Konzept (Adaptiert nach Hottenrott und Neumann, 2016, S. 268-270)	41
1.12.	Belastungsbereiche für Geschwindigkeiten beim Langstreckenlauf in Abhängigkeit des Deflektionspunktes	52
2.1.	Schematische Darstellung zur Modellbildung	62
2.2.	Schematische Darstellung eines Modellierungsprozesses	64
2.3.	Übersicht über die möglichen Ergebnisse eines statistischen Tests im Vergleich zur Realität	68
2.4.	Übliches Schema zur Auswahl eines statistischen Tests bei verbundenen Stichproben	69
2.5.	Schematische Darstellung zur Methode der kleinsten Quadrate	72

2.6. Schematische Darstellung eines Boxplots	74
2.7. Unterschiedlich verteilte Daten dargestellt als Streudiagramm, Histogramm und Boxplot	74
2.8. Vorgehensweise zur Erstellung eines Violinenplots	75
2.9. Violinenplots zu unterschiedlich verteilten Datensätzen	76
2.10. Visualisierung: Vergleich zwischen modellierten und empirischen Daten	77
2.11. Beispiele für QQ-Plots unterschiedlicher Verteilungen	78
2.12. Beispiel für Heatmaps	80
2.13. Graphische Visualisierung der vier Mengen aus dem Anscombe-Quartett mit ihrer jeweiligen linearen Regression	82
2.14. Plots zu unterschiedlichen Zufallszahlen	87
2.15. Trendlinie und Trendbeispiel zur Korrelationsanalyse	87
2.16. Trendbeispiele zur Korrelationsanalyse	88
2.17. Box-Diagramm für Hammerstein-Modelle	93
2.18. Box-Diagramm für Wiener-Modelle	93
3. Übersicht über die einzelnen Fragestellungen, Empirischer Teil	117
4.1. Exemplarische Darstellung der vier Testprotokolle	122
5.1. Vergleich der Prognosegüte zweier Modelle mit dem Faltungsmodell	132
5.2. Simulation der Testprotokolle mit dem Faltungsmodell, Studie 1	134
5.3. Vorhersage von Testprotokoll 04, Studie 1	136
5.4. Vorhersage der Herzfrequenz (HF) von Proband #05, Studie 1	138
5.5. Beispiel eines Simulationsergebnisses aus der Vorhersage, Studie 1	138
6.1. Heatmapbeispiel zur Parameteranalyse: Median-Beispiel	157
6.2. Heatmap zur Parameteranalyse: Median, Studie 2	161
6.3. Heatmap für die Parameterkombination a_2, a_4 und Curvefitting, Studie 2	163
6.4. Violinenplot: Vergleich Faltungsmodell 4 Parameter und 3 Parameter, Studie 2	163
6.5. Heatmap für die Parameterkombination a_1, a_2, a_3 , Studie 2	165
6.6. Heatmap einer Detail-Rastersuche für a_2, a_3 , Studie 2	166

6.7. Violinenplot: Vergleich mehrerer Varianten des Faltungsmodells, Studie 2 . . .	167
6.8. Violinenplot: Vergleich mehrerer Varianten des Faltungsmodells, Studie 2 . . .	168
6.9. Violinenplot: Vergleich mehrerer Varianten des Faltungsmodells, Studie 2 . . .	169
6.10. Evaluation der Parameterreduktion auf Outdoor-Trainingsdaten, Violinenplot, Studie 2	170
6.11. Evaluation der Parameterreduktion auf Outdoor-Trainingsdaten, QQ-Plot, Studie 2	171
6.12. Exemplarische Vorhersagesimulationen auf den Outdoor-Trainingsdaten, Studie 2	172
6.13. Evaluation der virtuellen Leistungsdiagnostik, Violinenplot, Studie 2	173
6.14. Exemplarische Vorhersagesimulationen der virtuellen Leistungsdiagnostik, Studie 2	174
6.15. Illustration des bifurkalen Zusammenhangs von a_2 und a_3 , Studie 2	178
6.16. Heatmap zur Parameteranalyse: Median, Studie 3	183
6.17. Heatmaps für die Parameterkombination a_2, a_3 , Studie 3	184
6.18. Heatmap für die Parameterkombination a_2, a_4 und Curvefitting, Studie 3	184
6.19. Violinenplot: Vergleich Faltungsmodell 4 Parameter und 3 Parameter, Studie 3 .	185
6.20. Violinenplot: Vergleich mehrerer Varianten des Faltungsmodells, Studie 3 . . .	186
6.21. Violinenplot: Vergleich mehrerer Varianten des Faltungsmodells, Studie 3 . . .	187
6.22. Violinenplot: Vergleich mehrerer Varianten des Faltungsmodells, Studie 3 . . .	188
6.23. Violinenplot: Vergleich mehrerer Varianten des Faltungsmodells, Studie 3 . . .	188
6.24. Illustration des bifurkalen Zusammenhangs von a_2 und a_3 , Studie 3	192
7.1. Vorgehensweise zur Korrelationsanalyse zwischen der Leistung und dem Para- metern der HF-Modellierung bei Profi-Athleten	198
7.2. Vergleich 1a: Parameter und Leistung, Studie 4	202
7.3. Vergleich 1b: Parameter und Leistung in der Trendanalyse, Studie 4	203
7.4. Vergleich 2a: Parameter und Leistung, Studie 4	204
7.5. Vergleich 2b: Parameter und Leistung in der Trendanalyse, Studie 4	205
7.6. Vorgehensweise zur Korrelationsanalyse zwischen der Leistung und dem Para- metern der HF-Modellierung bei Hobby-Athleten	213

7.7. Vergleich: Parameter des Faltungsmodells und Leistung anhand unterschiedlicher Schwellwerte bezüglich der geschätzten IAS, Modell 1, Studie 5	215
7.8. Vergleich: Parameter des Faltungsmodells und Leistung anhand unterschiedlicher Schwellwerte bezüglich der geschätzten IAS, Modell 1, Studie 5	216
7.9. Vergleich: Parameter des Faltungsmodells und Leistung anhand unterschiedlicher Schwellwerte bezüglich der geschätzten IAS, Modell 1, Studie 5	219
7.10. Vergleich: Parameter des Faltungsmodells und Leistung anhand unterschiedlicher Schwellwerte bezüglich der geschätzten IAS, Modell 1, Studie 5	220
7.11. Vergleich: Wattleistung an der geschätzten IAS und Blutlaktatkonzentration an der geschätzten IAS, Studie 5	224
7.12. Vergleich: Parameterverlauf bei unterschiedlichen konstanten Parameterwerte für a_1 bzw. a_2 , Studie 5	227
A.1. Heatmap zur Parameteranalyse: Median, Studie 2	251
A.2. QQ-Plots, Studie 2	252
A.3. Einzelne Heatmaps von Proband 1 (Studie 3)	253
A.4. Einzelne Heatmaps von Proband 2 (Studie 3)	254
A.5. Einzelne Heatmaps von Proband 3 (Studie 3)	254
A.6. QQ-Plots, Studie 3	255
A.7. Ergebnis des Ergonizers: Proband 1	256
A.8. Ergebnis des Ergonizers: Proband 2	257
A.9. Ergebnis des Ergonizers: Proband 3	258
A.10. Vergleich: Wattleistung an der geschätzten IAS und Herzfrequenz an der geschätzten IAS	259
A.11. Vergleich: Blutlaktatkonzentration an der geschätzten IAS und Herzfrequenz an der geschätzten IAS	259
A.12. Vergleich: verschiedene Schwellwerte an der IAS im Trendverlauf	260
A.13. Vergleich: Parameterverlauf bei unterschiedlichen konstanten Parameterwerte für a_1 bzw. a_2	261

Tabellenverzeichnis

1.1. Eigenschaften der verschiedenen Wege der Energiebereitstellung	18
1.2. Eigenschaften der einzelnen Muskelfasertypen	34
1.3. Anpassung einzelner Organe an ein Ausdauertraining	35
1.4. Übersicht über unterschiedliche Ausdauerbereiche und ihre Stoffwechsellage .	39
1.5. Übersicht über die fünf Belastungsnormative; angelehnt an Hohmann et al. (2020) (S. 54)	42
2.1. Rohdaten zum Anscombe-Quartett	81
2.2. Statistische Merkmale zum Anscombe-Quartett	81
2.3. Übersicht über verschiedene Spezifikationen von HF-Modellen	104
4.1. Übersicht der für die einzelnen empirischen Arbeiten verwendeten Datensätze	120
4.2. Charakteristiken der Teilnehmer der Studie, Datensatz 1	121
4.3. Charakteristiken der Teilnehmer der Studie, Datensatz 2	123
4.4. Charakteristiken der Teilnehmer der Studie, Datensatz 3	124
5.1. Durchschnittliche Simulationsergebnisse, Studie 1	135
5.2. Prädiktion verschiedener Beanspruchungsmaße, Studie 1	136
5.3. Überblick über die Spezifikationen der Unterfunktionen in den betrachteten DGL-Modellen	144
5.4. Eigenschaften verschiedener DGL-Modelle	145
6.1. Beispielparameter über mehrere Datensätze	151
6.2. Beispiel: Bereiche und Normalisierung der Parameter	151

6.3. Durchschnittliche Vorhersageergebnisse für drei physiologische Beanspruchungsmaße	155
6.4. Bereiche und Normalisierung der Parameter: Beispiel	156
6.5. Bereiche und Normalisierung der Parameter: Studie 2	160
6.6. Bereiche und Normalisierung der Parameter: Studie 3	181
7.1. Statistische Maße der vier Probanden, Studie 4	206
7.2. Statistische Maße der vier Probanden, Studie 4	206
7.3. Statistische Maße der drei Probanden, Studie 5, Modell 1	217
7.4. Statistische Maße der drei Probanden, Studie 5, Modell 1, Trend	217
7.5. Statistische Maße der drei Probanden, Studie 5, Modell 2	221
7.6. Statistische Maße der drei Probanden, Studie 5, Modell 2, Trend	221
7.7. Statistische Maße der drei Probanden, Schwellwerte untereinander, Studie 5	224
7.8. Statistische Maße der drei Probanden, Schwellwerte untereinander im Trendverlauf, Studie 5	224

Abkürzungsverzeichnis

HF Herzfrequenz.....	xiv
HRV Herzfrequenzvariabilität [engl. <i>Heart Rate Variability</i>]	3
HRR Erholungsherzfrequenz [engl. <i>Heart Rate Recovery</i>]	58
$\dot{V}O_2$ Sauerstoffaufnahme.....	48
$\dot{V}O_{2max}$ maximale Sauerstoffaufnahme.....	29
$\dot{V}CO_2$ Kohlendioxidabgabe.....	45
MAPE <i>mean absolute percentage error</i>	65
WAPE <i>weighted average percentage error</i>	65
RMSE <i>root-mean-square error</i>	65
MSE <i>mean-square error</i>	65
ICC Intra-Klassen-Korrelation.....	65
R² Determinationskoeffizient.....	65
IQA Interquartilsabstand.....	73

PCA Hauptkomponentenanalyse [engl. *Principal Component Analysis*] 154

ATP Adenosintriphosphat 17

ADP Adenosindiphosphat 17

AMP Adenosinmonophosphat 17

CrP Kreatinphosphat 17

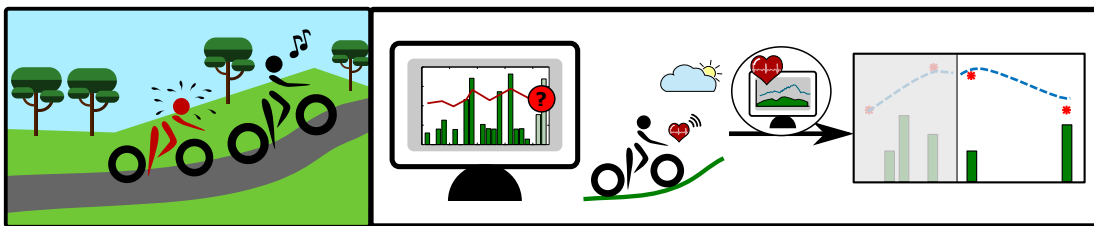
CK Kreatinkinase 46

Acetyl-CoA Acetyl-Coenzym-A 19

DGL Differentialgleichung 91

DGL-Modell Differentialgleichungsmodell 91

Einleitung



Abstract.

In diesem einführenden Kapitel wird zunächst das Problem dargestellt, das innerhalb dieser Arbeit behandelt wird. Dabei werden vor allem das zentrale Ziel sowie daraus abgeleitete Fragestellungen vorgestellt. Die Inhalte der einzelnen Kapitel und der Aufbau dieser Arbeit werden kurz erläutert.

Eine Überprüfung der Leistungsentwicklung im Radsport geht bis heute mit der Durchführung spezifischer leistungsdiagnostischer Ergometertests einher. Besonders weit verbreitet sind dabei Testprotokolle bis zur maximalen Ausbelastung oder spezifische Protokolle zur Blutlaktatdiagnostik. Ein bedeutender Nachteil von Ausbelastungstests ist die intensive Belastung, weswegen die Testroutine vor Wettkämpfen häufig reduziert wird. Die Reduktion der Testroutine soll die andernfalls kurzfristig auftretenden, negativen Auswirkungen auf die Wettkampfleistung vermeiden.

Grundsätzlich gehen die meisten Leistungsergometrien mit einem gewissen messtechnischen und organisatorischen Aufwand einher. Stufentests, Rampentest, All-Out-Tests und die meisten anderen Protokolle erfordern spezifische Messgeräte und ein geeignetes Ergometer, mit dem die spezifische Belastung für das jeweilige Protokoll eingestellt werden kann. Sollen zusätzlich die Blutlaktat- oder Atemgaskonzentration erhoben werden, sind entsprechende weitere Messgeräte unabdingbar. Eine Ausnahme bilden hier »Time Trail Tests« (z.B. der »5 Minuten All-Out-Test«), die auf einem handelsüblichen magnetgebremsten Rollentrainer in Kombination mit einem mobilen Leistungsmesssystem durchgeführt werden können. Eine einheitliche, standardisierte Testroutine (Aufwärmphase, Leistungstest, Regenerationsphase) ist dabei unabhängig vom gewählten Testprotokoll für valide und über die Zeit vergleichbare Messergebnisse notwendig.

Auf Grund der intensiven Belastung durch Ausbelastungstests werden seit mehreren Jahren auch submaximale Protokolle auf ihre Anwendbarkeit bei der Leistungsüberprüfung analysiert. Während viele submaximale Testprotokolle eine geringere Genauigkeit in der Bewertung der Leistungsfähigkeit aufweisen als Tests mit maximaler Ausbelastung, konnten in den vergangenen Jahrzehnten auch valide submaximale Testprotokolle entwickelt werden (Lamberts, 2009; Lamberts et al., 2011). Wenngleich die Belastung bei submaximalen Testprotokollen deutlich geringer ausfällt als bei Ausbelastungstests, ist auch hierzu die Durchführung von Ergometertests mit spezifisch vorgegebenen Belastungen notwendig.

Für Hobby-Athlet*innen in der ambitionierten Radsportpraxis stellen leistungsdiagnostische Tests zusätzlich zur physischen Belastung und dem organisatorischen sowie zeitlichen Aufwand auch einen hohen Kostenfaktor dar. Hier bieten sich häufig Heuristiken und Modelle wie die Analysen aus »Golden Cheetah« – ein Analyseprogramm für Radsportler – an, um die eigene Leistungsentwicklung nachverfolgen zu können. Doch auch für die dort verwendeten Heuristiken ist es notwendig, spezifische Radfahr-Protokolle in das eigene Training zu integrieren. Ein Beispiel hierfür ist die dort verwendete Heuristik zur »Critical Power« (CP), zu deren Berechnung es notwendig ist, sich regelmäßig über unterschiedliche Zeitintervalle auszubelasten. Um eine möglichst akkurate CP-Kurve modellieren zu können, müssen also in regelmäßigen Abständen Ausbelastungseinheiten über beispielsweise 1 min, 4 min, 5 min, 10 min, 20 min etc. durchgeführt werden.

Gleichzeitig haben in den vergangenen Jahrzehnten »wearable devices« mehr und mehr an Popularität gewonnen, sodass es heutzutage sehr einfach ist, die Herzfrequenz im Alltag und bei sportlichen Aktivitäten aufzuzeichnen. Die Herzfrequenzaufzeichnungen zu nutzen, um auf direkte Weise und möglichst ohne spezifische Anforderungen an die Trainingsfahrten Rückschlüsse über die Leistungsentwicklung ziehen zu können, bietet die Chance, sowohl im professionellen Radsport wie auch in der ambitionierten Radsportpraxis den Erkenntnisgewinn über die eigene Leistungsentwicklung zu vereinfachen. Bereits Lamberts et al. (2010) haben die Verwendung der »Heart Rate Recovery« (HRR) – die Erholungsrate der Herzfrequenz, die aus der Differenz zwischen der Herzfrequenz bei Beginn und Ende eines Trainings ermittelt wird – hinsichtlich ihrer Vorhersagekraft der Entwicklung von Erschöpfung und Leistungsfähigkeit hin untersucht. In ihren Analysen haben sie herausgefunden, dass ein Absinken der HRR anscheinend mit einer geringeren Verbesserung der Leistung einhergeht. Sie schlussfolgerten, dass die HRR das Potenzial hat, eine trainingsinduzierte Ermüdung und damit die Unfähigkeit vorherzusagen, eine spezifische Trainingsleistung erbringen zu können. Auch Buchheit et al. (2008) haben sich mit spezifischen Parametern der Herzfrequenz hinsichtlich der Leistungsentwicklung befasst. In ihren Analysen über die HRR und die Herzfrequenzvariabilität [engl. *Heart Rate Variability*] (HRV) fanden sie heraus, dass die HRR lediglich mit der wöchentlichen Trainingsbelastung, nicht aber mit einer längerfristigen Leistungsentwicklung zusammen hänge. Einen spezifischen Index der HRV hingegen konnten sie mit der maximalen Sauerstoffaufnahme als Leistungsindikator korrelieren. Sie stellten jedoch auch heraus, dass für eindeutige Schlussfolgerungen in dieser Hinsicht weitere Untersuchungen notwendig seien.

Während die HRR folglich nicht geeignet erscheint, um Rückschlüsse auf die längerfristige Leistungsentwicklung zu ziehen, benötigt die Verwendung der HRV geeignete Messmethoden, eine vergleichsweise aufwändige Datenaufbereitung und entsprechende Kenntnisse bei der Datenverarbeitung und Analyse. Die meisten »wearable devices« zeichnen in der Regel zudem lediglich die Herzfrequenz, nicht jedoch die HRV während des Trainings auf.

Zum aktuellen Zeitpunkt fehlt eine geeignete und physiologisch oder phänomenologisch interpretierbare Modellierung, die eine Verwendung der reinen Herzfrequenz zur Simulation der Leistungsentwicklung basierend auf regulären (Radsport-)Trainingseinheiten ermöglicht.

Zielsetzung

Ziel dieser Arbeit ist es, zu untersuchen, inwieweit es mit Hilfe eines möglichst einfachen mathematisch-phänomenologischen Modells möglich ist, basierend auf der Herzfrequenzmessung bei regulären Trainingseinheiten im Radsport Rückschlüsse auf die Leistungsentwicklung der Sportlerin oder des Sportlers zu ziehen. Dazu wird zunächst ein Modell (das

»Faltungsmodell«) zur Simulation und Prognose des Herzfrequenzverhaltens im Ausdauersport entwickelt und evaluiert. Nach einer Verbesserung dieses Modells, bei der die Anzahl der freien Parameter auf einen freien Parameter reduziert werden kann, wird schließlich untersucht, ob dieser verbleibende Parameter Rückschlüsse auf die Leistungsentwicklung zulässt.

Die Arbeit behandelt daher die folgende zentrale

Fragestellung :

Trägt der aus der Herzfrequenzmodellierung stammenden Parameter des parameterreduzierten Faltungsmodells basierend auf alltäglichem Training zur Gewinnung von Rückschlüssen auf die Leistungsentwicklung einer Athletin oder eines Athleten im Ausdauersport bei?

Damit Daten aus alltäglichem Training genutzt werden können, ist es zunächst notwendig, geeignete empirische Messwerte zur Modellierung zu identifizieren. Obwohl die Herzfrequenz als einzelner Beanspruchungsparameter des Organismus kritisch betrachtet werden muss, bietet sie den großen Vorteil, dass sie beispielsweise über einen Brustgurt unkompliziert auch bei täglichem Training erfasst werden kann. Hottenrott und Neumann (2016) äußern dazu:

»Die Herzfrequenz (HF) bei Belastungen ist im Sport seit Langem ein zuverlässiger Indikator für die Beurteilung der Trainingswirkung, der Ermüdung oder des Gesundheitszustandes. (...) Seit Einführung exakt arbeitender HF-Messgeräte (»Pulsuhren«) ist bekannt, dass die HF zu den empfindlichsten Messgrößen gehört, welche die Wirkung des Trainings oder die Veränderung der Belastung (Geschwindigkeit, Streckenprofil) erfassbar macht.«

Auch sei inzwischen »gesichert, dass die trainingsbedingten Veränderungen der HF-Regulation bereits nach wenigen Tagen nachweisbar sind« (Hottenrott und Neumann, 2016, S. 62). Grundlagen der Herzfrequenzreaktion sowie kritische Aspekte werden in Abschnitt 1.3 betrachtet. Auf Grund ihrer zuverlässigen und einfachen Erfassbarkeit stellt die Herzfrequenz im Rahmen dieser Arbeit den zentralen Bestandteil der Untersuchungen, Modellierungen und Analysen dar.

Zur Analyse der Fragestellung, ob das Faltungsmodell auch ohne vorhergehende ergometriebasierte Leistungsdiagnostik und stattdessen basierend auf alltäglichem Training Rückschlüsse auf die Leistungsentwicklung beim Radfahren zulässt, werden drei Themenkomplexe betrachtet:

1. Es wird ein Modell zur Herzfrequenzvorhersage entwickelt, das möglichst wenige und phänomenologisch interpretierbare Parameter beinhaltet: das Faltungsmodell.

Es wird sich zeigen, dass mit diesem Modell vollständige Trainingseinheiten modelliert und prognostiziert werden können.

2. Das Faltungsmodell wird weiter vereinfacht und die Anzahl der Parameter wird reduziert. Auf diese Weise können Wechselwirkungen der Parameter untereinander reduziert und so die Stabilität des Modells erhöht werden. Es wird sich zeigen, dass es dabei möglich ist, die Anzahl der Parameter auf einen einzigen verbleibenden Parameter zu reduzieren.
3. Der verbleibende Parameter des Faltungsmodells wird hinsichtlich seiner Veränderung über die Zeit untersucht und es wird analysiert, ob diese Veränderungen Rückschlüsse auf die Leistungsentwicklung erlauben.

Für die Modellierung ist daher besonders relevant, dass die Herzfrequenz beim Ausdauertraining mit guter Genauigkeit über vollständige Trainingseinheiten modelliert und prognostiziert werden kann. Abbildung 1 zeigt schematisch den Aufbau dieser Arbeit. Um die Beantwortung der zentralen Fragestellung zu ermöglichen, werden für jeden der drei Themenkomplexe empirische Studien durchgeführt: Eine Studienanalyse über die Modellierung und Prognosegüte der Herzfrequenz mit dem phänomenologischen Modell, zwei Studienanalysen über die Möglichkeiten einer Parameterreduzierung innerhalb des Modells ohne signifikanten Verlust der Prognosegüte, und zwei Studienanalysen über das Verhalten des Modellparameters im zeitlichen Verlauf hinsichtlich der langfristigen Leistungsentwicklung.

Damit all diese Analysen durchgeführt werden können, sind im Vorfeld unterschiedliche Betrachtungen und Grundlagen notwendig: Auf physiologischer Ebene ist es erforderlich, die beim Ausdauersport besonders beteiligten Systeme des Körpers zu betrachten, um anhand derer mögliche Parameter der akuten Beanspruchungsreaktion betrachten und einschätzen zu können. Dabei steht das Ziel im Vordergrund, diese für die zentrale Fragestellung mit der Leistungsentwicklung zu kombinieren. Entsprechend spielen auch Grundlagen der Ausdauerleistung, der Testdiagnostik und des Zusammenhangs zwischen akuten und chronischen Anpassungsreaktionen eine fundamentale Rolle. Im Kontext der Modellierung und im Hinblick auf die Studienanalysen sind sowohl einige modelltheoretische Grundlagen, als auch insbesondere statistische Grundlagen notwendig. Mit diesen können schließlich die Ergebnisse bewertet und entsprechende Rückschlüsse für die Kombination aus akuter und chronischer Beanspruchung gezogen werden. Eine zusammenfassende Diskussion (Abschnitt 7.3) anschließend an die letzte empirische Studie und eine abschließende Zusammenfassung mit Fazit (Kapitel 8) runden die Thematik ab und geben einen Ausblick auf mögliche nächste Schritte.

Umgekehrt bedeutet das, dass basierend auf den physiologischen (Kapitel 1) und modellierungstheoretischen Grundlagen (Kapitel 2) sowie dem zugehörigen aktuellen Stand der Forschung für die Beantwortung der zentralen Frage zwei vorbereitende Fragestellungen de-

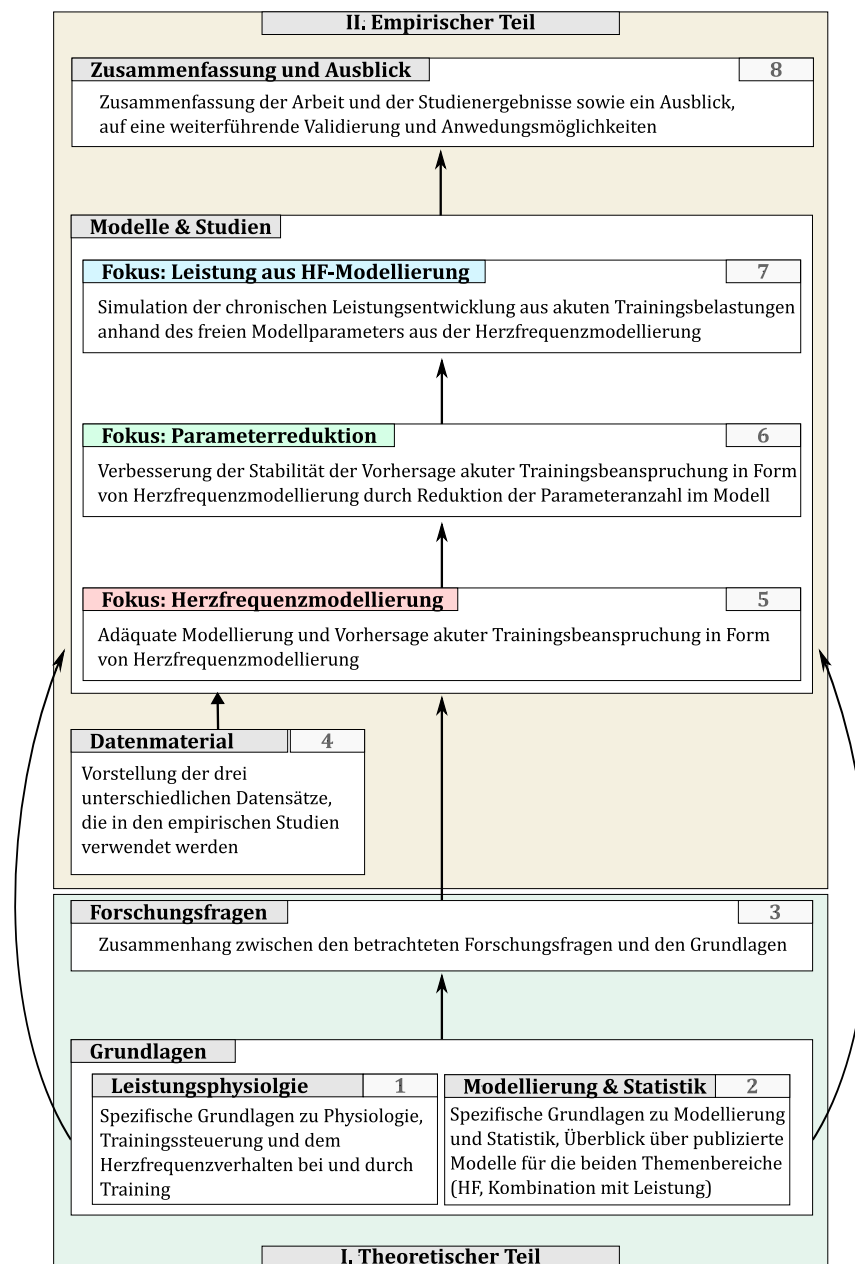


Abbildung 1.: Schematische Darstellung der Gliederung der Dissertation. Die Teil-Kästen enthalten neben dem Thema und einem groben Überblick der Kapitel-Inhalte die Kapitelnummer (Zahl in der jeweils rechten oberen Ecke). Um die zugrundeliegende Fragestellung beantworten zu können, werden sowohl unterschiedliche Teilaspekte betrachtet, als auch die notwendigen Grundlagen zusammengefasst

finiert werden können: Diese Fragestellungen befassen sich mit der Herzfrequenzmodellierung allgemein und dem parameterreduzierten Modell zur Herzfrequenzmodellierung. Der nach einer Parameterreduktion verbleibende Parameter wird schließlich zur Ermittlung der Leistungsfähigkeit anhand der Herzfrequenzdaten und damit zur Beantwortung der zentra-

len Fragestellung analysiert; die Relevanz der drei Fragestellungen wird basierend auf den Grundlagen und dem Stand der Forschung in Kapitel 3 dargestellt.

Zu jeder der insgesamt drei Fragestellungen werden in Kapitel 5 (»Das Faltungsmodell: Ein neuer Ansatz zur Simulation und Vorhersage der Herzfrequenz im Training«), Kapitel 6 (»Verbesserung des Faltungsmodells: Parameterreduktion«) und Kapitel 7 (»Nutzen des parameterreduzierten Faltungsmodells bei der Leistungsbeurteilung«) Studienanalysen durchgeführt. Innerhalb von Kapitel 7 wird die zentrale Fragestellung analysiert und diskutiert.

Insgesamt können die Hauptbeiträge dieser Arbeit damit wie folgt zusammengefasst werden:

1. Entwicklung eines neuen Modells zur Herzfrequenzvorhersage (»Faltungsmodell«) mit phänomenologischen Interpretationsmöglichkeiten;
2. Evaluation des Faltungsmodells und Vergleich seiner Simulations- und Vorhersagegüte mit bekannten Modellen aus der Literatur;
3. Verbesserung des neu entwickelten Modells und ausführliche Evaluation einzelner Verbesserungsschritte;
4. Anwendung des Faltungsmodells auf den Bereich der Trainingswirkungsanalyse durch die Analyse eines spezifischen freien Modellparameters hinsichtlich seiner Nutzbarkeit zur Einschätzung der Leistungsentwicklung.

Aufbau der Arbeit

Die vorliegende Arbeit unterteilt sich in einen grundlegenden theoretischen und einen empirischen Teil.

Der theoretische Teil umfasst drei Themenschwerpunkte (Kapitel 1 bis 3): In Kapitel 1 werden zunächst die physiologischen und trainingswissenschaftlichen Grundlagen dargestellt, die im Rahmen der späteren Analysen relevant sind. Dabei werden mögliche physiologische Parameter zur Leistungsüberprüfung diskutiert und kritische Aspekte der später verwendeten Herzfrequenz beleuchtet. Kapitel 2 fasst die modelltheoretischen sowie mathematisch-statistischen Grundlagen zusammen. Hierzu werden typische statistische Kennzahlen hinsichtlich ihrer Anwendbarkeit auf die empirischen Aspekte dieser Arbeit kritisch hinterfragt und mit Hilfe aktueller wissenschaftlicher Erkenntnisse aus der Literatur bewertet. Innerhalb von Kapitel 3 werden schließlich die drei Forschungsfragen vorgestellt, die im nachfolgenden empirischen Teil beantwortet werden sollen.

Der empirische Teil ist in fünf Themenschwerpunkte (Kapitel 4 bis 8) untergliedert: Kapitel 4 gibt einen Überblick über die anschließend in den empirischen Studien verwendeten Daten. In Kapitel 5 wird das Faltungsmodell vorgestellt. Im Rahmen einer ersten empirischen Stu-

die findet eine grundlegende Validierung und Evaluation des Faltungsmodells statt. Ergänzend wird das Faltungsmodell strukturell genauer untersucht und hinsichtlich unterschiedlicher Eigenschaften mit ausgewählten Modellen zur Herzfrequenzsimulation aus der Literatur verglichen. Eine Möglichkeit zur Verbesserung des Faltungsmodells durch die Reduktion der Parameteranzahl wird in Kapitel 6 vorgestellt. Basierend auf zwei empirischen Studien mit unterschiedlichen Datensätzen wird jeweils eine für den jeweiligen Datensatz individuelle Parameterreduktion durchgeführt und sowohl qualitativ wie auch quantitativ mit dem originalen Faltungsmodell verglichen. Dabei lässt sich die Anzahl der freien Parameter in beiden empirischen Studien auf einen einzigen freien Parameter reduzieren. Basierend auf der durchgeführten Parameterreduktion wird in Kapitel 7 schließlich die zentrale Fragestellung untersucht. Dabei wird mit den gleichen Probandendaten wie im vorhergehenden Kapitel untersucht, inwieweit sich ein Zusammenhang zwischen dem identifizierten freien Parameter des Faltungsmodells und der Leistungsentwicklung der Probanden herstellen lässt. In Kapitel 8 wird die Arbeit abschließend zusammenfassend dargestellt.

Verwandte Publikationen

Einige Teile dieser Arbeit wurden bereits bei verschiedenen Konferenzen oder in Fachzeitschriften veröffentlicht. Im Folgenden sind die Publikationen der Autorin mit Bezug zu den Inhalten dieser Arbeit in chronologischer Reihenfolge aufgeführt:

- Ludwig, M., Grohgan, H. G., und Asteroth, A. (2016). A Convolution Model for Heart Rate Prediction in Physical Exercise. In *Proceedings of the 4th International Congress on Sport Sciences Research and Technology Support (icSPORTS 2016)*, Seiten 157–164.
- Ludwig, M., Hoffmann, K., Endler, S., Asteroth, A., und Wiemeyer, J. (2018). Measurement, Prediction, and Control of Individual Heart Rate Responses to Exercise - Basics and Options for Wearable Devices. *Front. Physiol.*, 9(778).
- Ludwig, M., Grohgan, H. G., und Asteroth, A. (2019). A Convolution Model for Prediction of Physiological Responses to Physical Exercises. In *Sport Science Research and Technology Support Information (icSPORTS 2016/2017)*, Band 975 in *Communications in Computer and Information Science*, Seiten 18–35. Springer Nature Switzerland AG.

Teil I.

Theoretische Hintergründe

Bevor die zentrale Fragestellung dieser Arbeit untersucht werden kann, werden nachfolgend die dazu relevanten Grundlagen aus der Physiologie, der Trainingswissenschaft sowie aus der Modelltheorie und Statistik dargestellt.

Innerhalb von Kapitel 1: »Leistungsphysiologische Grundlagen« werden zunächst die zum Verständnis der im empirischen Teil folgenden Modellierungen und Fragestellungen notwendigen physiologischen und trainingswissenschaftlichen Grundlagen zusammengefasst. Innerhalb von Teil II werden das Herzfrequenzverhalten, unterschiedliche Messgrößen zur Leistungsüberprüfung aus der Leistungsdiagnostik sowie der Zusammenhang zwischen dem Herzfrequenzverhalten und chronischen physiologischen Anpassungen durch Ausdauersport im Vordergrund stehen – entsprechend sind hinsichtlich der Grundlagen primär Hintergründe zur Herzfrequenz im Ausdauer- bzw. Radsport, unterschiedliche Schwellenkonzepte und Testverfahren bei der Leistungsdiagnostik, sowie die Zusammenhänge zwischen der Herzfrequenz und chronischen Anpassungen des Körpers von Interesse. Für ein besseres Verständnis werden ergänzend einige Grundlagen dieser drei Themenbereiche aufbereitet dargestellt – die Zusammenhänge sind in Abbildung 2 illustriert.

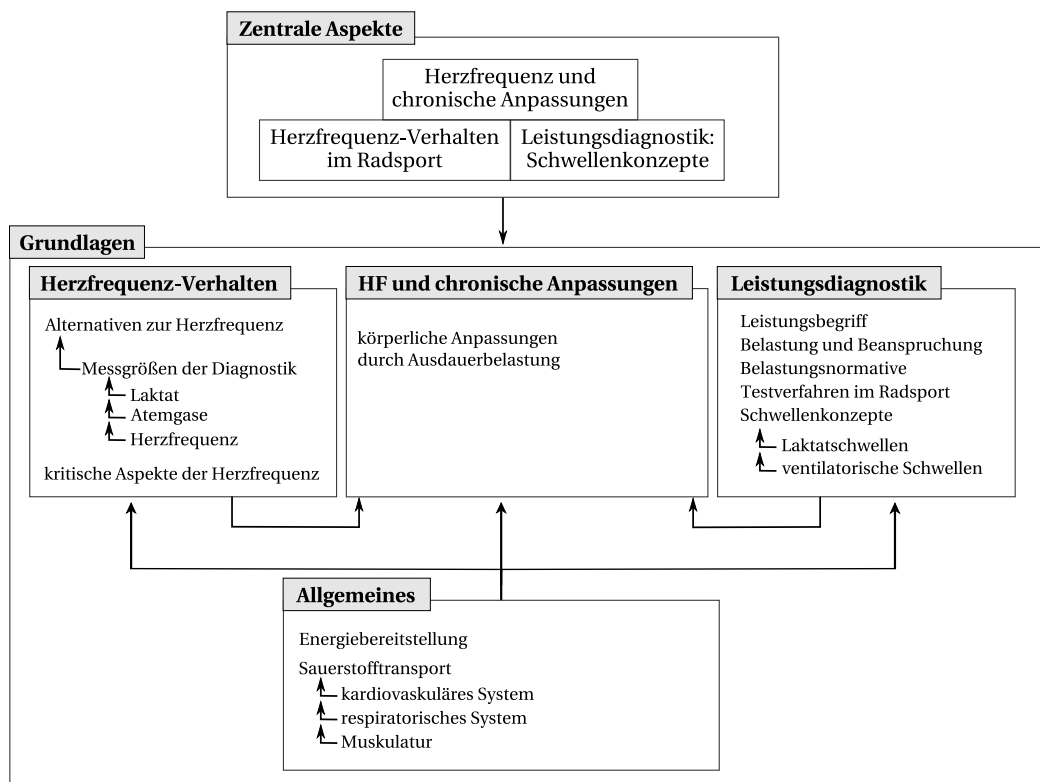


Abbildung 2.: Übersicht über die einzelnen Themenbereiche aus den physiologischen und trainingswissenschaftlichen Grundlagen und ihrer Zusammenhänge

Im Themenkomplex zum Herzfrequenzverhalten werden alternative Möglichkeiten diskutiert, die zur späteren Modellierung verwendet werden könnten. Dabei wird zusammengefasst dargestellt, welche Besonderheiten bei den einzelnen Messgrößen der Diagnostik vor-

liegen und warum letztlich die Herzfrequenz in den durchgeführten Analysen verwendet wird. Da keine der Messgrößen ohne äußere Einflussfaktoren betrachtet werden kann, werden in diesem Kontext auch kritische Aspekte speziell zur Verwendung der Herzfrequenz beleuchtet. Für diese Betrachtungen sind zudem einige allgemeine Grundlagen über das kardiovaskuläre und das respiratorische System nützlich, die daher im Vorfeld innerhalb von Unterabschnitt 1.1.2: »Adaptationen der beteiligten Systeme beim Sauerstofftransport« zusammengefasst werden.

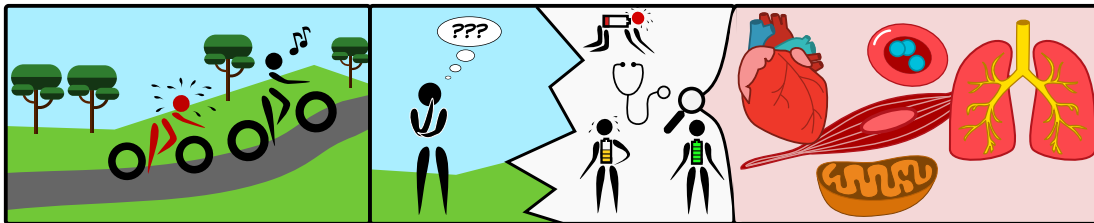
Für die relevanten trainingswissenschaftlichen Grundlagen hinsichtlich der Leistungsdiagnostik sind neben dem Leistungsbegriff und dem Belastungs-Beanspruchungs-Konzept vor allem die Quantifizierung der Belastung und Beanspruchung während eines Trainings zum Verständnis der empirischen Studien von Interesse. Außerdem beinhalten die im späteren Verlauf dieser Arbeit verwendeten empirische Daten unterschiedliche Messungen aus leistungsdiagnostischen Tests sowie verschiedene Parameter zur Leistungsbeurteilung. In diesem Kontext werden daher ergänzend unterschiedliche Testverfahren sowie verschiedene Schwellenkonzepte (basierend auf der Blutlaktatkonzentration und Atemgasanalysen) innerhalb von Abschnitt 1.2: »Allgemeine Grundlagen der Trainingssteuerung im Ausdauersport« zusammengefasst. Für diese Betrachtungen sind zudem einige allgemeine Grundlagen über das respiratorische System sowie hinsichtlich der Blutlaktatbildung und über die allgemeine Energiebereitstellung im Körper nützlich, die daher im Vorfeld nacheinander zu Beginn von Abschnitt 1.1: »Allgemeine physiologische Grundlagen« zusammengefasst werden.

Die Zusammenhänge zwischen dem Herzfrequenzverhalten und chronischen physiologischen Anpassungen basieren letztlich auf allgemeinen körperlichen Anpassungsreaktionen durch (Ausdauer-)Belastung und den thematisierten Grundlagen über die Herzfrequenz und die Leistungsentwicklung. Kapitel 1 schließt daher die physiologischen und trainingswissenschaftlichen Grundlagen mit einem Überblick über diese Zusammenhänge ab.

In Kapitel 2: »Grundlagen zur Modellierung des Herzfrequenzverhaltens im Ausdauersport« werden zunächst die im Rahmen dieser Arbeit relevanten Begriffe aus der Modelltheorie definiert. In diesem Kontext wird zudem ein Überblick über geeignete und verwendete statistische Kennzahlen, Testverfahren und unterschiedliche Möglichkeiten geeigneter Visualisierungen gegeben. Diese werden ergänzend hinsichtlich der Nutzbarkeit für die durchgeführten Analysen kritisch diskutiert. Dabei werden relevante Aspekte der Zeitreihenanalyse besonders berücksichtigt. Anschließend wird für die Modellierung der Herzfrequenz umfassend dargestellt, welche Anforderungen an die Modellierung bestehen und welche Modelle zur Herzfrequenzanalyse im Ausdauersport in der Literatur typischerweise verwendet werden. Ein besonderes Augenmerk liegt hinsichtlich der zentralen Fragestellung abschließend auf dem Zusammenhang zwischen der Herzfrequenz und der Leistungsentwicklung innerhalb der Modellierung.

In Kapitel 3: »Zusammenfassung des Forschungsstands und Fragestellungen« werden schließlich die Lücken innerhalb der bisherigen Forschung aufgezeigt, auf deren Basis sich drei aufeinander aufbauende Fragestellungen ergeben. Mit diesen werden schließlich im darauffolgenden empirischen Teil dieser Arbeit das eingangs formulierte Ziel untersucht, Rückschlüsse auf die Leistungsentwicklung mit Hilfe eines möglichst einfachen mathematisch-phänomenologischen Modells und basierend auf regulären Trainingseinheiten und der Herzfrequenzmessung zu ziehen.

Leistungsphysiologische Grundlagen



Abstract.

Die Betrachtung der Leistungsentwicklung beim Radfahren setzt auf leistungsphysiologischer und sportwissenschaftlicher Ebene die Kenntnis unterschiedlicher Faktoren voraus. Schon allein die Betrachtung der Leistungsentwicklung setzt ein einheitliches Verständnis des Begriffs »Leistung« sowie Kenntnisse über die zahlreichen physiologischen Systeme voraus, die die Leistung beeinflussen und eine Entwicklung überhaupt erst ermöglichen. Gleichzeitig lohnt sich ein Blick in die sportwissenschaftlichen Methoden, mit denen die Leistung insbesondere im Ausdauersport üblicherweise festgestellt wird (»status quo«). Nur die parallele Verwendung des »status quo« ermöglicht eine fundierte Bewertung, inwieweit und unter welchen Bedingungen eine Prognose zur Leistungsentwicklung basierend auf alltäglichem Training realisiert werden kann.

In diesem Kapitel werden die Aspekte der physiologischen und sportwissenschaftlichen Grundlagen betrachtet, die für die betrachteten Leistungsdiagnostiken sowie verwendeten Parameter notwendig sind. Dabei wird hergeleitet, warum die Herzfrequenz zur Analyse und Beantwortung der zentralen Fragestellung eine zentrale Rolle einnimmt. Auch die sich daraus ergebenden Nachteile werden kritisch hinterfragt und bewertet.

Ein für den Organismus zentraler Aspekt körperlicher Belastung und insbesondere Ausdauerbelastung ist der erhöhte Sauerstoff- und Energiebedarf. Im physiologischen Kontext wird Ausdauer häufig auch als Ermüdungswiderstand bezeichnet und somit als Fähigkeit definiert, eine physische Leistung über einen gewissen Zeitraum erbringen zu können (Hohmann et al., 2020; Hottenrott und Neumann, 2016; Schnabel et al., 1994).

Definition 1.1: Ausdauer.

»Die Ausdauer ist eine konditionelle Fähigkeit, die eine belastungsadäquate Energieversorgung des Organismus sichert, ermüdungsbedingte Leistungs- oder Geschwindigkeitsabnahmen bei sportlicher Belastung verzögert und die Erholungsfähigkeit beeinflusst.« (Hottenrott und Neumann, 2016, S. 112)

Dabei lässt sich Ausdauer sowohl in unterschiedliche Kategorien unterteilen, als auch bezüglich der Dauer der Belastung betrachten. Eine Unterteilung in Kategorien (allgemeine / lokale Muskelausdauer, aerobe / anerobe Ausdauer, statische / dynamische Ausdauer) ist zur Betrachtung verschiedener Kategorien nur bedingt nutzbar, da sich nach Dickhuth et al. (2007) die einzelnen Kategorien nicht strikt voneinander trennen lassen und die Übergänge teilweise fließend sind. Von größerem praktischen Nutzen ist daher eine Aufteilung nach Art der Energiebereitstellung, das heißt in Abhängigkeit der Dauer einer Belastung und damit in Relation zur Art der Energiebereitstellung. Eine Analyse der Energiebereitstellung ermöglicht die Betrachtung leistungslimitierender Faktoren und umgekehrt die Betrachtung leistungsbeeinflussender Systeme. Dabei determiniert die Fähigkeit der Energiebereitstellung die Dauer und Intensität der Durchführung einer Belastung. Die Individualität der Leistungsfähigkeit basiert auf individuellen Limitationen der beteiligten Systeme. Sie basiert damit insbesondere auf der Zusammensetzung der Muskelfasertypen in der beanspruchten Muskulatur sowie den am Sauerstofftransport beteiligten Systemen. Hierzu zählen die Lungendiffusionskapazität, das maximale Herzzeitvolumen, die Sauerstofftransportkapazität des Blutes, sowie die Kapillarisation der belasteten Muskulatur. Neben der Muskulatur und den am Sauerstofftransport beteiligten Systemen hängt die Energiebereitstellung zusätzlich von den stattfindenden Stoffwechselprozessen innerhalb der Mitochondrien ab. (Dickhuth et al., 2007)

Auch Markworth (2010) nennt insbesondere für Ausdauersportarten mit Beanspruchung größerer Muskelgruppen die Sauerstofftransportfähigkeit des Herz-Kreislauf-Systems als wesentlichen begrenzenden Faktor der Leistung. Im Kontext dieser Arbeit wird daher besonderes Augenmerk auf das kardiovaskuläre und das kardiopulmonale System gelegt. Im Folgenden wird in Abschnitt 1.1 ein Überblick über die die Ausdauer limitierenden Systeme und in Abschnitt 1.2 die im Ausdauersport relevanten leistungsdiagnostischen Grundlagen gegeben. In Abschnitt 1.3 wird anschließend das Herzfrequenzverhalten beim Ausdauersport näher betrachtet und kritische Aspekte werden beleuchtet.

1.1. Allgemeine physiologische Grundlagen

Für einen Überblick über im Ausdauersport relevante physiologische Grundlagen werden zunächst die unterschiedlichen Arten der Energiebereitstellung (Unterabschnitt 1.1.1) sowie die am Sauerstofftransport wesentlich beteiligten Systeme (Unterabschnitt 1.1.2) betrachtet. Dabei steht vor allem das Herz-Kreislauf-System im Fokus. Der Einfluss des Ausdauertrainings auf weitere Komponenten und Systeme des Körpers wird anschließend in Unterabschnitt 1.1.3 betrachtet. In Abschnitt 1.2 werden grundlegende und für die vorliegende Arbeit relevante Aspekte der Leistungsdiagnostik aufgeführt. Auch zentrale physiologische Anpassungsreaktionen, die ein Ausdauertraining zur Folge hat, werden in diesem Kapitel zusammengefasst.

1.1.1. Energiebereitstellung

Als Energiestoffwechsel wird allgemein derjenige Teil des Stoffwechsels von Lebewesen bezeichnet, der zur Energiegewinnung dient. Speziell in der Physiologie wird im Zusammenhang mit dem Energiestoffwechsel der Metabolismus betrachtet, also die Energiebereitstellung innerhalb der Muskelzellen. Im Folgenden wird speziell auf die physiologische Betrachtung der Energiebereitstellung in Muskelzellen eingegangen, da diese für die physiologischen Hintergründe der (Anpassungs-)Reaktionen bei Ausdauerbelastungen von zentraler Bedeutung sind, die ihrerseits eng mit der Energiegewinnung insgesamt zusammenhängen.

Als Energieträger dienen primär energiereiche Phosphate (Adenosintriphosphat (ATP), Kreatinphosphat (CrP)), sekundär aber auch Kohlenhydrate (Glykogen), (intramuskuläre) Triglyzeride und – bei extremer Ausdauerbelastung – sogar körpereigene Aminosäuren. ATP ist mit seinen drei Phosphatgruppen der wichtigste Überträger und Speicher von chemischer Energie auf Zellebene und wird zur Erzeugung jeglicher Bewegungsenergie genutzt: Bei der Übertragung der endständigen Phosphatgruppe auf andere Moleküle wird unter Zuhilfenahme von Wasser Energie freigesetzt. Das so entstehende energieärmere Adenosindiphosphat (ADP) kann in Vorgängen, die Energie liefern, mit Phosphat zu ATP (und Wasser) regeneriert werden. Die Übertragung der Phosphatgruppe auf ein anderes organisches Molekül wird als *Phosphorylierung* bezeichnet. Mit der »primäre Energiebereitstellung« wird die Freisetzung von energiereichem Phosphat durch den Abbau von ATP zu ADP – und manchmal auch Adenosinmonophosphat (AMP) – bezeichnet. Die »sekundäre Energiebereitstellung« bezeichnet die Resynthese von ATP aus ADP über sekundäre Energiequellen. Für diese existieren im Organismus verschiedene Mechanismen in Abhängigkeit von Belastungsdauer und Belastungsintensität. Ein Beispiel hierfür ist die Glykolyse, bei der die Resynthese von ATP mit Hilfe von Zuckermolekülen stattfindet. Hierauf wird später detaillierter eingegangen.

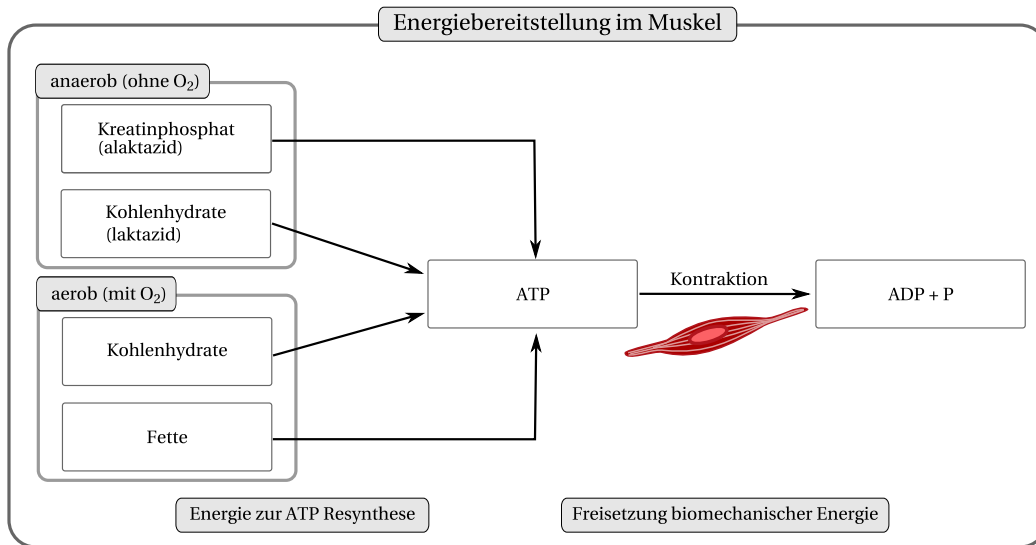


Abbildung 1.1.: Darstellung unterschiedlicher Energieträger zur Energiebereitstellung im Muskel

Es wird zwischen verschiedenen Arten der Energiebereitstellung unterschieden. Bezogen auf den Stoffwechselweg kann Energie anaerob-alaktazid (ohne Sauerstoff und ohne Laktatbildung), anaerob-laktazid (ohne Sauerstoff, unter Bildung von Laktat) oder aerob (mit Sauerstoff) bereitgestellt werden. Abbildung 1.1 zeigt eine Übersicht der unterschiedlichen Energiebereitstellungsarten. Die Eigenschaften und Besonderheiten dieser drei Energierarten sind in Tabelle 1.1 zusammengefasst.

Tabelle 1.1.: Eigenschaften der verschiedenen Wege der Energiebereitstellung

	Anaerob alaktazid	Anaerob laktazid	Aerob	
Dauer	1 - 12 s	12 s - 2 min	2 min - 1,5 h	≥ 1,5 h
Nutzen	Maximalkraft	Kraftausdauer	Ausdauer	Ausdauer
Speicherform	Kreatinphosphat	Glykogen	Glykogen	Fette
Stoffwechselweg	Phosphat-Metabolismus	Glykolyse	Glykolyse	Lipolyse
Stoffwechselort	Cytoplasma	Cytoplasma	Mitochondrien	Mitochondrien
Produkt	ATP (& Kreatin)	ATP (& Laktat)	ATP	ATP
Energieausbeute		2 ATP	36 ATP	129 ATP

Die Prozesse zur Energiebereitstellung laufen teilweise parallel ab. Insbesondere bei häufig wechselnden Belastungsintensitäten oder hohen Intensitäten von 2 – 8 min wird diese als *aerob-anaerob gemischte Ausdauer* bezeichnet. Pro Minute können durch die ATP-Resynthese aus CrP 4,4 kJ gewonnen werden. Der Glykogen-Speicher in den Muskeln liefert 3,0 kJ/min anaerob und 1,0 kJ/min über aerobe Glykolyse. Glykogen aus der Leber sorgt aerob für weitere 0,4 kJ/min, und mittels aerober Lipolyse lassen sich erneut 0,4 kJ/min aus freien Fettsäuren gewinnen. (Hohmann et al., 2020)

Die einzelnen Prozesse zur Energiebereitstellung werden im Folgenden näher beschrieben.

ATP-Speicher Die primäre Energiebereitstellung kommt ausschließlich für Belastungen zum Tragen, die kurzfristig eine hohe Menge an Energie erfordern. Die lokal vorrätige Menge an ATP reicht lediglich für etwa 2 s aus. Bei der *anaerob-alaktaziden* Energiegewinnung durch die Spaltung von ATP entstehen unter Verwendung von Wasser ADP, Phosphat und Energie. Für jede längere Belastung müssen sekundäre Energiequellen genutzt werden, um ATP zu resynthetisieren.

CrP-Speicher Ebenfalls zur kurzfristigen Bereitstellung großer Energiemengen dient die Spaltung von CrP. Die Freisetzung von Phosphaten auf diesem Wege verläuft ebenfalls *anaerob-alaktazid*, d.h. ohne Oxidationsprozess und ohne Freisetzung von Laktat. Dabei reagiert CrP mit ADP zu ATP und Kreatin. Der Vorrat von CrP in der Muskulatur reicht durchschnittlich bei schnellstmöglicher Energiebereitstellung für etwa 10 s. Die Größe des CrP-Speichers ist jedoch durch Training adaptierbar, sodass für Untrainierte eher eine Dauer von ca. 5 – 8 s, und für Hochtrainierte 12 – 20 s angenommen werden. (Dickhuth et al., 2007; Hohmann et al., 2020) Der CrP-Speicher ist bei moderater Belastung nach etwa 5-10 min wieder gefüllt und steht dann erneut für kurze hochintensive Belastungen zur Verfügung.

Glykolyse (anaerob) Der (zeitlich gesehen) zweite Mechanismus der Energiebereitstellung zur ATP-Resynthese ist die *anaerobe Glykolyse*, d.h. die Spaltung von Glykogen. Die anaerobe Glykolyse ist ein Seitenweg der (aeroben) Glykolyse. Bei der Glykolyse entsteht in beiden Fällen Pyruvat. Bei hohen Intensitäten findet die Resynthese von ATP über die Glykolyse jedoch für die geforderten Energiemengen nicht schnell genug statt, sodass über die anaeroben Glykolyse zusätzliche Energie gewonnen werden kann: Hierbei werden nicht alle Pyruvate (über oxidative Decarboxylierung, bei der unter Abspaltung von Kohlenstoff NADH entsteht) in Acetyl-Coenzym-A (Acetyl-CoA) überführt. Acetyl-CoA wird nur bei ausreichender Zufuhr von Sauerstoff (d.h. bei der aeroben Glykolyse) im Zitronensäurezyklus weiter verstoffwechselt. Die nicht decarboxylierten Pyruvate werden mittels NADH und Wasserstoffionen (sowie L-Lactatdehydrogenase als Katalysator) zu Laktat (und NAD^+) reduziert. Bei dieser Reduktion wird Energie freigesetzt und das Laktat sammelt sich in der Muskulatur an. Bei zu hoher Laktatkonzentration im Muskel wird Laktat über die Blutbahn abtransportiert und insbesondere in Herz, Leber und den Nieren abgebaut. Entsteht soviel Laktat, dass der *Laktat-steady-state* überschritten wird, d.h. dass der Abbau in den Organen langsamer verläuft als die Neuproduktion von Laktat, erhöht sich die Laktatkonzentration im Blut deutlich. Insbesondere führen hohe Laktatkonzentrationen zu einer Hemmung des Enzyms Phosphofruktokinase, welches für die Aufrechterhaltung der anaeroben Glykolyse verantwortlich ist. Daher führt ein erhöhter Laktatspiegel zu einem Abfall der Bewegungintensität. Die anaerobe Glykolyse kann für insgesamt etwa 2 min aufrechterhalten werden und erreicht ihren energetischen Höhepunkt nach ca. 45 s. Im Anschluss werden zunehmend ae-

robe Prozesse notwendig, was eine deutliche Reduktion der Belastungsintensität und damit der Leistung zur Folge hat. (Hohmann et al., 2020; Tomasits und Haber, 2011)

Glykolyse (aerob) Der wichtigste Mechanismus bei der aeroben Energiegewinnung ist die (*aerobe*) *Glykolyse*. Das auf diesem Wege gebildete Acetyl-CoA geht vollständig in den Zitronensäurezyklus ein. Zwischenprodukte der Glykolyse wie auch des Zitronensäurezyklus' können schließlich innerhalb der Endoxidation (Atmungskette) zu Kohlendioxid und Wasser verstoffwechselt werden, wobei ATP resynthetisiert wird. Die so verbrauchten Glykogenvorräte in den Muskeln reichen je nach Trainingszustand zwischen 30 min und 60 min bis hin zu 100 min aus. (Heck und Schulz, 2002; Hohmann et al., 2020; Tomasits und Haber, 2011)

Glukoneogenese Erfolgt nach Aufbrauchen der Glykogenvorräte keine Nahrungsaufnahme in Form von Kohlenhydraten und soll die hohe Belastungsintensität weiterhin aufrecht erhalten werden, kann weitere Energie über die *Glukoneogenese* – eine partielle Umkehrreaktion der Glykolyse – bereitgestellt werden. Das aus der Leber gewonnene Glykogen kann ebenfalls über die Glykolyse verstoffwechselt werden. Dieser Prozess verläuft biopositiv, d.h. die durch das Training erfolgenden Adaptationen können zu einer Verbesserung der Leistung führen. (Dickhuth et al., 2007)

Proteinstoffwechsel Genügen hingegen auch die Glykogenvorräte der Leber während eines hochintensiven Trainings nicht aus, um eine ausreichende Energiebereitstellung zu gewährleisten, wird die benötigte Energie bionegativ aus körpereigenen Proteinen gewonnen. Dieser Vorgang kann schnell zu einem Übertraining und einer damit einhergehenden Verschlechterung der Leistung führen. Bei diesem Vorgang wird Stickstoff von den Aminogruppen abgespalten, sodass der Rest der Aminosäure in Abhängigkeit der vorhandenen Kohlenstoff-Atome in den Kohlenhydratstoffwechsel eingeführt wird. Auf diese Weise können verstoffwechselte Aminosäuren als Acetyl-CoA in den Zitronensäurezyklus eingehen oder zu Glukose transformiert werden. Der abgespaltene Stickstoff wird als Harnstoff über die Nieren ausgeschieden und kann zu einer messbaren Erhöhung der Harnstoffkonzentration im Blut führen. (Dickhuth et al., 2007)

Aerobe Lipolyse Um längerfristige Leistung mit niedriger Belastungsintensität zu erbringen, wird Energie aus der *aeroben Lipolyse* gewonnen, d.h. durch die Oxidation von Triglyceriden. Dabei werden freie Fettsäuren abgespalten und langkettige Fettsäuren zu Acyl-CoA aktiviert, welches schließlich schrittweise zu Acetyl-CoA oxidiert wird. Dabei wird das Carrienzym L-Carnitin benötigt. Das so gebildete Acetyl-CoA wird in den Zitronensäurezyklus eingeschleust, sodass die Energiegewinnung wie auch in der Glykolyse durch die Endoxidation stattfindet. Der Prozess ist durch die vorherige Spaltung der Triglyceride allerdings

langsamer als die Energiegewinnung über die Glykolyse. Ein sogenanntes »Fettstoffwechseltraining« findet nur bei weniger intensivem, dafür aber länger andauerndem Training statt, wenn die Energiebereitstellung nicht allein durch die Glykogenvorräte gesichert werden kann. Außerdem wird die Verstoffwechslung der Fette ab einer Blutlaktatkonzentration von etwa 7 mmol/l durch einen Mangel an freiem Carnitin gehemmt. (Dickhuth et al., 2007; Hottenrott und Neumann, 2016)

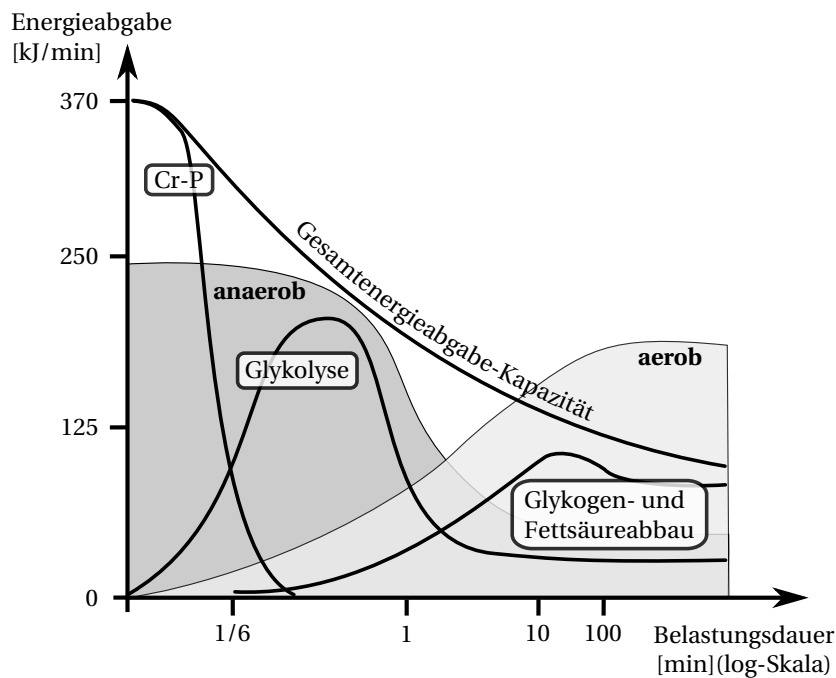


Abbildung 1.2.: Schematische Darstellung der Energiebereitstellung der unterschiedlichen physiologischen Systeme über die Zeit. Unterteilt in anaerobe und aerobe Energiebereitstellung. Adaptiert aus Hohmann et al. (2020), S.59

Abbildung 1.2 stellt schematisch dar, in welchem Umfang die einzelnen Energieträger dem Körper unter Belastung und in Abhängigkeit der Belastungsdauer zur Verfügung stehen: Der (zeitlich) erste Mechanismus zur Resynthese von ATP ist die Spaltung von dem in der Muskulatur vorhandenen CrP. Ist der CrP-Speicher nach 6 bis maximal 20 s durch eine hochintensive Belastung geleert, setzt die anaerobe Glykolyse ein. Die Menge des durch anaerobe Glykolyse resynthetisierten ATPs erreicht nach etwa 45 s ihren Höhepunkt. Nach etwa 2 min anhaltender hochintensiver Belastung lässt die Leistung deutlich nach und die Energiegewinnung wird von den beiden aeroben Mechanismen der ATP-Resynthese dominiert. Sind auch die Glykogenvorräte je nach Trainingszustand nach etwa 30-100 min größtenteils erschöpft, gibt es drei unterschiedliche Optionen:

1. die hochintensive Belastung soll aufrecht erhalten werden und es können Kohlenhydrate von Außen zugeführt werden: Durch die neu aufgenommene Glukose kann ATP weiterhin über die aerobe Glykolyse wiedergewonnen werden;

2. die hochintensive Belastung soll aufrecht erhalten werden und es können keine Kohlenhydrate von Außen zugeführt werden: Die benötigte Energie wird über die Glukoneogenese (biopositiv aus dem Glykogenspeicher der Leber oder bionegativ aus körpereigenen Aminosäuren) gewonnen;
3. die hochintensive Belastung wird nicht aufrechterhalten und spätestens jetzt reduziert: Die Energie wird über die aerobe Lipolyse und somit über den Fettstoffwechsel zur Verfügung gestellt. Dieser Stoffwechselweg dient ebenfalls dann der ATP-Resynthese, wenn die Belastung bereits vor Ausschöpfung der Glykogenvorräte reduziert wird.

Im zuletzt genannten Fall setzt bei niedriger Intensität das sogenannte »Fettstoffwechseltraining« ein, durch das bei einem mindestens 30-minütigen Training die Grundlagenausdauer und die Regenerationsfähigkeit erhöht werden und beispielsweise kardiovaskuläre Risikofaktoren verringert werden können. (Hohmann et al., 2020)

1.1.2. Adaptationen der beteiligten Systeme beim Sauerstofftransport

Wie eingangs erwähnt, sind die zentralen Systeme für den Sauerstofftransport die Lungendiffusionskapazität, das maximale Herzzeitvolumen, die Sauerstofftransportkapazität des Blutes, sowie die Kapillarisation der belasteten Muskulatur. Damit Sauerstoff innerhalb des Körpers zu Organen und in die Muskulatur transportiert werden kann, ist das Zusammenspiel verschiedener Strukturen relevant: Das Blut enthält Hämoglobin als Farbstoff der Erythrozyten, das die Bindung von Sauerstoff oder Kohlenstoffdioxid ermöglicht. Durch die Zusammenarbeit von Herz und Lunge kann das Blut mit Sauerstoff angereichert werden, bevor das Herz mit den Blutgefäßen und Kapillaren dafür sorgt, dass Organe und Muskeln mit Sauerstoff versorgt werden können. Abbildung 1.3 illustriert schematisch den Körperkreislauf zwischen Herz und Lunge (kleiner Körperkreislauf), in dem das Blut mit Sauerstoff angereichert wird, und zwischen Herz und dem restlichen Körper (großer Körperkreislauf), der für den Weitertransport sorgt. Bei einer Beanspruchung der Muskulatur wird innerhalb dieser Muskeln mehr Sauerstoff benötigt, sodass auch die Kapillaren innerhalb der Muskeln eine Rolle spielen.

Entsprechend werden im Folgenden die im Kontext dieser Arbeit relevanten Grundlagen und Adaptationen durch Ausdauerbelastung dargestellt, die das kardiovaskuläre und pulmonale System sowie die Muskulatur betreffen.

Kardiovaskuläres System

Die Versorgung mit Sauerstoff und Nährstoffen aller Körperzellen obliegt in erster Linie dem Herz und dem Gefäßsystem. Dies trifft umgekehrt auch auf den Abtransport entstandener Stoffwechsel- und Stoffwechselabfallprodukte zu. Dabei dient das Herz als »Pumpe« oder

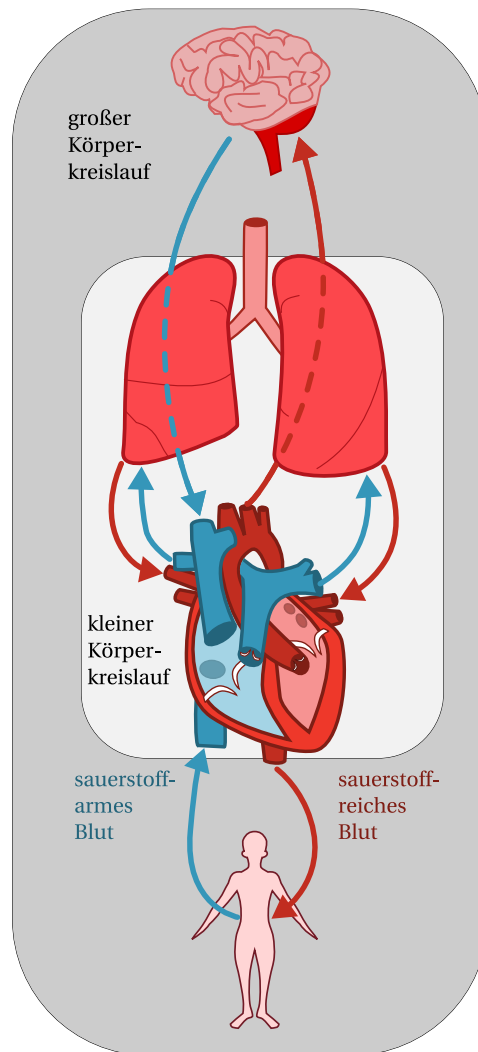


Abbildung 1.3.: Schematische Darstellung des Herz-Lungen-Kreislaufs

»Motor«, um das Blut als Transportmedium durch die Gefäße zu befördern. Das Gefäßsystem selbst reguliert dabei die notwendige Umverteilung des Blutes und sorgt durch Verengung oder Erweiterung einzelner Gefäße dafür, dass die stärker von einer Belastung betroffenen Organe und Muskeln mit ausreichend Blut – und dadurch insbesondere auch mit ausreichend Sauerstoff – versorgt werden. Die das sauerstoffreiche Blut befördernden Arteriolen verzweigen sich an ihren Enden zu Kapillaren (»Haargefäßen«), über die Sauerstoff und Nährstoffe an Zellen abgegeben bzw. über die Stoffwechselprodukte aufgenommen werden. In Ruhe sind etwa zweidrittel der Muskelkapillare undurchlässig und verschlossen, da die meisten Muskeln in Ruhe kaum Sauerstoff und Nährstoffe benötigen.

Wie in Abbildung 1.4 (a) dargestellt, ist das Herz selbst in eine rechte und linke Herzhälfte aufgeteilt, die ihrerseits wiederum in zwei Kammern gegliedert sind: den Vorhof (Atrium) und die Hauptkammer (Ventrikel). Das Blut, das aus dem Körper bzw. aus der Lunge ins Herz gelangt, sammelt sich zunächst in den Vorhöfen. Aus den Vorhöfen fließt das Blut

durch die Segelklappen – die Trikuspidalklappe (rechts) bzw. die Mitralklappe (links) – in die jeweiligen Hauptkammern, von wo es schließlich in die Lungenarterie bzw. die Aorta gepumpt wird. Zwischen den Hauptkammern und den Arterien befinden sich die Taschenklappen – die Pulmonalklappe (rechts) und die Aortenklappe (links). Alle vier Klappen sind Teil der Herzinnenhaut (Endokard), die auch die vier Kammern auskleidet und die innerste von drei Schichten der Herzwand bildet. Die zweite Schicht ist der Herzmuskel (Myokard), die äußere Schicht wird als Herzaußenhaut (Epikard) bezeichnet. Die vier Herzklappen sorgen dafür, dass das Blut nur in eine Richtung fließen kann¹. Sie öffnen bzw. schließen sich während der Herzkontraktion abwechselnd und sichern so den Blutfluss in die richtige Richtung. Die Flussrichtungen des Blutes sind in Abbildung 1.4 (b) illustriert.

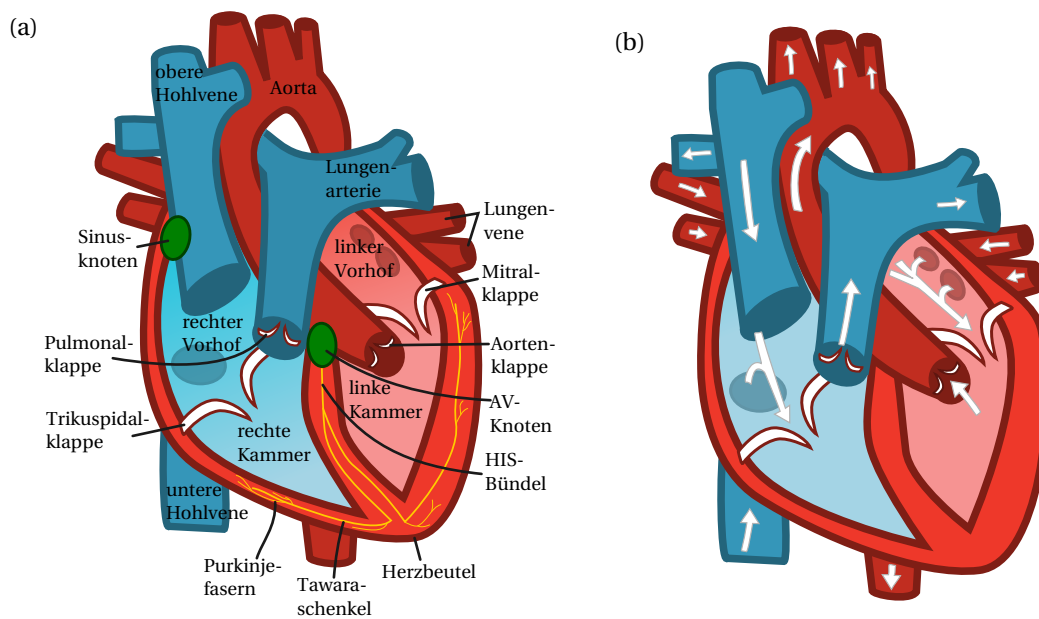


Abbildung 1.4.: Schematische Darstellung vom (a) Aufbau des Herzens, und (b) Blutfluss innerhalb des Herzens und der umgebenden Venen und Arterien.

Innerhalb des Herzbeutels, der das gesamte Herz umschließt, befindet sich das Erregungsleitungssystem des Herzens. Dieses besteht aus dem Sinusknoten, dem Artioventrikular-Knoten (AV-Knoten), dem HIS-Bündel, den beiden Tawara-Schenkeln sowie den Purkinjefasern. Der Sinusknoten wird auch als (primärer) »Schrittmacher« des Herzens bezeichnet. Die im Sinusknoten beginnende Erregung (das Aktionspotenzial) breitet sich über die Herzvorhöfe aus, die dadurch kontrahieren und wird innerhalb von 40 bis 80 ms zum AV-Knoten weitergeleitet. Der AV-Knoten erfüllt dabei zwei zentrale Funktionen: Bei einem gesunden Herzen sorgt er für eine kurze Verzögerung von 60 bis 120 ms bis zur Kontraktion der Hauptkammern des Herzens. Zusätzlich ist der AV-Knoten der sekundäre Schrittmacher des Herzens und sorgt für eine Aufrechterhaltung der (dann etwas verlangsamten) Herzfunktion,

¹Auch bei gesunden Herzen kann es zu kleinen Rückflüssen kommen, da die Klappen teilweise nicht vollständig schließen. Solange der Rückfluss gering ist, stellt dies jedoch kein medizinisches Problem dar.

sofern der Sinusknoten ausfällt. Vom AV-Knoten wird das Aktionspotenzial über das HIS-Bündel, die Tawara-Schenkel und die Purkinje-Fasern zügig in alle Ventrikelbereiche weitergeleitet. In Abhängigkeit der unterschiedlichen Stellungen der Klappen und der Kontraktion des Herzmuskels durchläuft ein Herzschlag vier Aktionsphasen: die Anspannungsphase, die Austreibungsphase, die Entspannungsphase und die Füllungsphase bzw. Phase der Vorhofkontraktion.

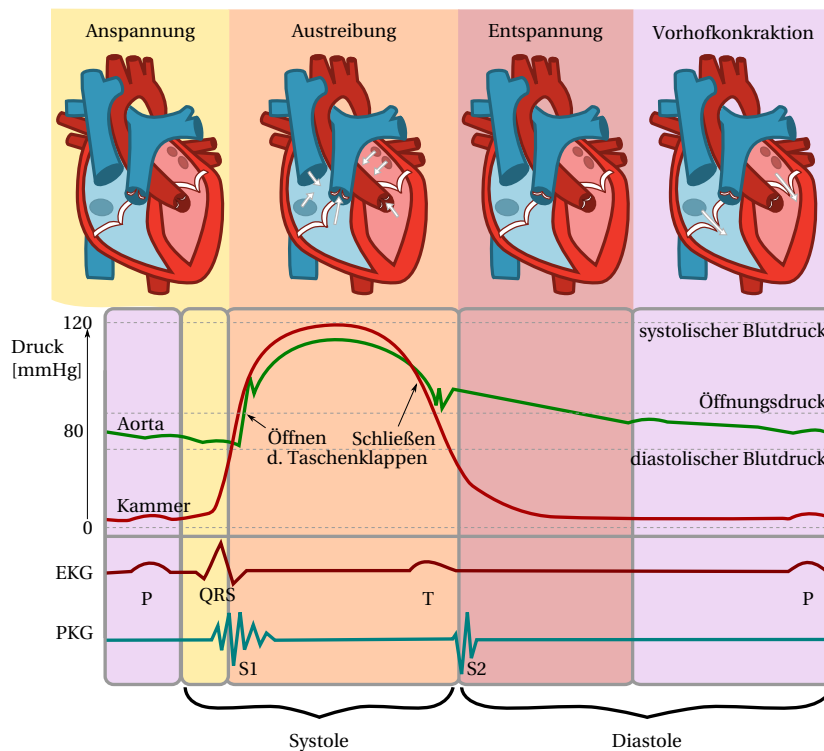


Abbildung 1.5.: Phasen der Herztätigkeit mit unterschiedlichen Druckverhältnissen. Adaptiert von Schwegler und Lucius (2011) (S. 257, Abb. 9.18).

Die unterschiedlichen Phasen der Herztätigkeit sind zusammen mit der Stellung der Klappen und dem Druck in den Kammern in Abbildung 1.5 dargestellt. Die Systole beginnt mit der Anspannungsphase (ca. 60 ms), in der alle Klappen geschlossen sind. Durch eine in der Herzspitze beginnende Kontraktion wird die Ventilebene, d.h. die Ebene der Herzklappen, in Richtung der Herzspitze gezogen. Die Form der Kammern wird dabei nahezu rund, wodurch der erste Herzton erzeugt wird. Gleichzeitig erhöht sich der Druck in den Kammern, bis dieser den Druck in der Aorta bzw. der Lungenarterie übersteigt. Da der Druck innerhalb der Lungenarterie mit 10 mmHg geringer ist als der Druck von 70-80 mmHg in der Aorta, endet die Anspannungsphase der rechten Herzkammer früher. Ist dieser Druck erreicht, beginnt die Austreibungsphase (ca. 200 ms) mit Öffnung der Taschenklappen. Etwa die Hälfte des Blutvolumens wird aus den Kammern in die großen Arterien gepumpt. Die Menge, die dabei aus der linken Herzkammer ausgeworfen wird, wird als »Schlagvolumen« bezeichnet. Die Segelklappen sind währenddessen geschlossen. Da sich das Volumen

in den Ventrikeln während der Auswurfphase verringert, senkt sich die Klappenebene Richtung Herzspitze. Die Herzspitze selbst verschiebt sich etwas zur Herzbasis hin, wodurch die Vorhöfe gedehnt werden und ein Unterdruck entsteht, der Blut aus den Venen in die Vorhöfe saugt. Die nachfolgende Entspannungsphase (ca. 40 ms) ist der Beginn der Diastole. Die Spannung der Ventrikelmuskulatur nimmt ab, der Herzmuskel entspannt sich wieder und der Kammerdruck sinkt unterhalb des Drucks der beiden großen Arterien. Das Schließen der Klappen erzeugt den zweiten Herzton. Sobald die Erregung der Muskulatur vollständig nachgelassen hat, wird der Druck innerhalb der Herzkammern geringer als in den Vorhöfen und die Füllungsphase (mit variabler Dauer) beginnt. Die Segelklappen öffnen sich und das Blut strömt zügig aus den gefüllten Vorhöfen in die Kammern. Gegen Ende der Füllungsphase beginnt die Kontraktion in den Vorhöfen, die dadurch das Blutvolumen innerhalb der Kammern zusätzlich erhöhen.

Das Erregungsleitungssystem und der Sinusknoten werden dabei durch das vegetative Nervensystem reguliert. Während der Sympathikus für eine Erhöhung der Herzfrequenz sorgt, übt der Parasympathikus eine beruhigende Wirkung auf die Erregungsbildung aus und führt zu einer Verlangsamung.

Das Schlagvolumen, das während der Systole aus der linken Herzkammer gepumpt wird, beträgt im Durchschnitt etwa 70 ml, kann jedoch durch Belastung gesteigert werden. Eine Multiplikation des Schlagvolumens mit der Herzfrequenz ergibt das »Herzminutenvolumen« oder auch »Herzzeitvolumen«, also die Menge an Blut, die vom Herzen pro Minute durch den Kreislauf gepumpt wird. In Ruhe beträgt das Herzminutenvolumen in der Regel 4,5-5 l/min. Unter körperlicher Belastung kann sich das Herzminutenvolumen durch die Erhöhung der Herzfrequenz und des Schlagvolumens auf über 25 l/min erhöhen und so die Muskulatur ausreichend mit Sauerstoff versorgen.

Die Frequenz, mit der das Herz schlägt, ist dabei in der Regel nicht gleichmäßig. Vor allem die Füllungsphase des Herzens unterliegt einer variablen Dauer, sodass auch die Zeit des Aufeinanderfolgens zweier Herzschläge variabel ist. Diese zeitliche Variation zwischen den Herzschlägen wird »HRV« genannt. Die HRV ist dabei insbesondere auch von der Atmung abhängig. Die »respiratorische Sinusarrhythmie« sorgt dafür, dass die Herzfrequenz während der Einatmung leicht ansteigt und während der Ausatmung etwas langsamer wird. Die Ursachen für diesen Zusammenhang sind noch nicht vollständig erforscht. Es wird vermutet, dass unter anderem verschiedene Rezeptoren, die den Blutdruck überwachen (hierunter vor allem die Barorezeptoren, die die Dehnung der Gefäßwände der Karotis und der Aorta messen), sowie Sympathikus und Parasympathikus einen Einfluss an der respiratorischen Sinusarrhythmie haben (Karemaker, 2009; Piepoli et al., 1997). Die HRV wird außerdem vom Trainingszustand, Hormonen und dem Stress- bzw. Erholungszustand des Körpers beeinflusst (Billman, 2011; Javorka et al., 2002; Thayer et al., 2012). Eine höhere HRV weist auf einen guten Gesundheitszustand hin – das Herz ist in der Lage, sich flexibel an äußere Umstände und Anforderungen anzupassen.

Respiratorisches System

Während das Herz das Blut in Bewegung hält und somit den Körper mit Nährstoffen, Botenstoffen und insbesondere Sauerstoff versorgt, obliegt es dem pulmonalen System, den Gasaustausch im Blut vorzunehmen und für eine ausreichende Anreicherung mit Sauerstoff zu sorgen.

Diese Versorgung mit Sauerstoff – aber auch mit sonstigen Nährstoffen – erfolgt in fast allen Bereichen des Körpers über das »Kapillarnetz«. *Kapillare* werden die sehr feinen Verzweigungen von Blutgefäßen (und Lymphgefäßen²) des Körpers genannt. Sie haben einen Durchmesser von etwa 5 – 10 μm . Organe wie Gehirn, Herz, Lungenflügel sind ebenso wie die Skelettmuskulatur von zahlreichen Kapillaren durchzogen, da sie im Gegensatz zu beispielsweise Sehnen und Bändern einen besonders hohen Sauerstoffbedarf oder allgemein eine hohe Stoffwechselaktivität aufweisen. Die Länge des gesamten Kapillarnetzes eines Erwachsenen beträgt rund 100 000 Kilometer. Dabei sind unterschiedliche Bereiche von unterschiedlich vielen Kapillaren umgeben. Ausnahmen sind beispielsweise die Gelenkknorpel oder die Augenlinsen, die ihre Nährstoffe über Diffusion aus den umliegenden Gewebestrukturen erhalten und die keine Kapillare besitzen.

Das respiratorische System lässt sich in luftleitende und gasaustauschende Abschnitte unterteilen. Die Anteile des Atmungssystems, die nicht aktiv am Gasaustausch teilnehmen und für den Transport der Luft zuständig sind, bilden den »anatomischen Totraum«. Der anatomische Totraum umfasst dabei Nase, Rachen, Kehlkopf, Luftröhre und die Bronchien. Von den 500 ml Atemvolumen, die ein gesunder erwachsener Mensch durchschnittlich einatmet, gehen 150 – 200 ml im Totraum »verloren«. Die Lungenbläschen (Alveolen) erhalten pro Atemzug die verbleibenden 300 – 350 ml Frischluft für den Gasaustausch. Das individuelle Volumen des Totraums (V_D) kann nach der Bohr-Formel (Gleichung 1.1) berechnet werden:

$$V_D = V_T \cdot \frac{P_{A,\text{CO}_2} - P_{E,\text{CO}_2}}{P_{A,\text{CO}_2}} \quad (1.1)$$

Dabei beschreibt P_{A,CO_2} den Partialdruck von Kohlenstoffdioxid innerhalb der Alveolen, P_{E,CO_2} den Partialdruck von Kohlenstoffdioxid in der ausgeatmeten Luft, und V_T das Expirationsvolumen je Atemzug. Das Expirationsvolumen V_T liegt durchschnittlich bei 0,5 ℓ . Das Volumen des Totraums wird demnach mittels Partialdrücken aus dem Anteil der verbleibenden Luft in den Alveolen relativ zum Gesamtvolumen in Abhängigkeit des Expirationsvolumens berechnet. Da der Partialdruck innerhalb der Alveolen nicht direkt gemessen werden kann, wird dieser seinerseits ebenfalls näherungsweise über die *alveoläre Gasgleichung* anhand des äußeren Luftdrucks, dem Sauerstoffanteil der eingeatmeten Luft, dem respiratorischen Quotienten und weiteren messbaren Komponenten berechnet.

²Üblicherweise sind mit Kapillaren die Blutgefäße gemeint.

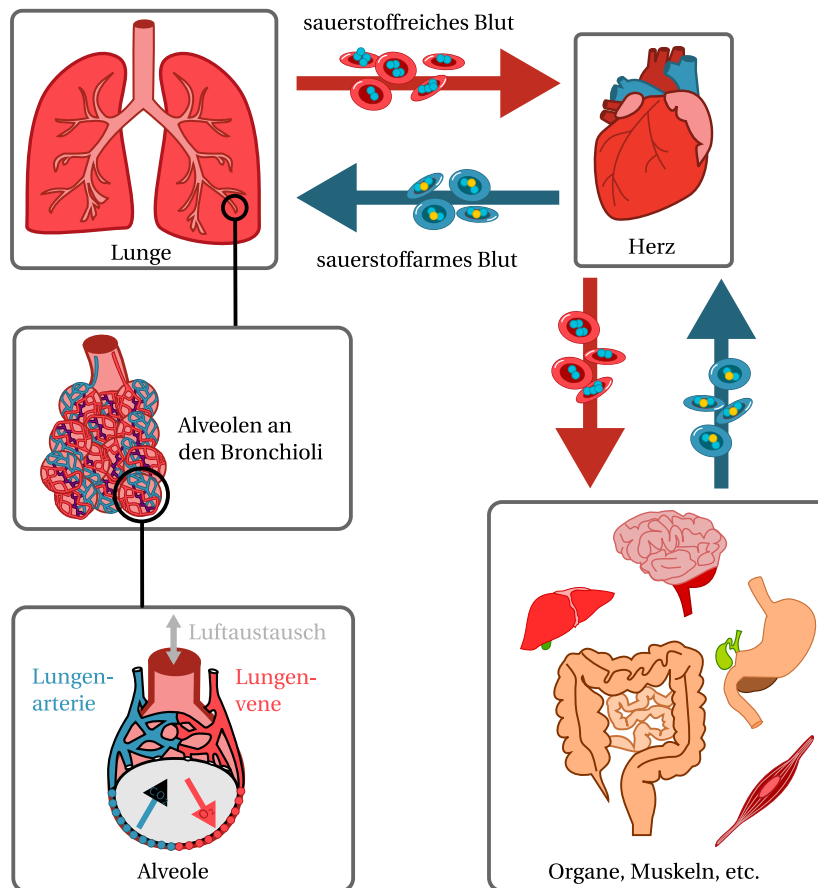


Abbildung 1.6.: Schematische Darstellung des Sauerstofftransports (sauerstoffreiches Blut) von den Lungen über das Herz in die Organe und Kohlenstoffdioxidtransport (sauerstoffarmes Blut) analog zurück. Links ist der schematische Aufbau der Lunge, der Bronchiolen und der Alveolen dargestellt

Der Gasaustausch findet innerhalb der Lunge statt. Dabei werden Gase über die Alveolen ins Blut abgegeben, wo sich Sauerstoffmoleküle an die Erythrozyten binden, um über den Blutfluss in die verschiedenen Organe, die Muskulatur und das Gewebe transportiert zu werden. Die *Lungendiffusionskapazität*, also die Fähigkeit der Lunge zur Aufnahme von Sauerstoff aus der Luft, ist dabei insbesondere abhängig von der diffusionswirksamen Oberfläche der Alveolen und Kapillaren.

Die Atmung als solche wird im menschlichen Körper von unterschiedlichen Muskeln unterstützt. Primär sind dabei das Zwerchfell (»Diaphragma«) und die Interkostalmuskeln an der Atmung beteiligt. Je nach Atemtechnik können zusätzlich die Unterrippenmuskeln und die Atemhilfsmuskulatur (einige Hals-, Brust- und Bauchmuskeln) den Atemvorgang unterstützen.

Für die Ausdauerleistung beim Sport sind in diesem Kontext neben der Lungendiffusionskapazität insbesondere das Atemzug- bzw. -zeitvolumen und die Atemfrequenz von Interesse: Das Atemzugvolumen – also die Menge an Luft, die durchschnittlich unbewusst bei jedem

Atemzug ein- und ausgeatmet wird – beträgt in Ruhe etwa 0,5 ℓ und kann bei großer Belastung auf rund 2,5 ℓ ansteigen. Durch willentliche Beeinflussung und bewusste Aktivierung des Zwerchfells lässt sich das Volumen bis zum Erreichen der maximal nutzbaren Luftmenge bei der Atmung («vollständige Vitalkapazität») weiter steigern. Bei untrainierten Personen liegt diese bei rund 4,5 ℓ. Gut trainierte Personen aus dem Ausdauer-Leistungssport können hingegen eine Vitalkapazität von 7 ℓ und mehr erreichen. Das Atemzugvolumen wird wie die Atemfrequenz üblicherweise durch das vegetative Nervensystem gesteuert, sodass zu jedem Zeitpunkt in Abhängigkeit akuter Belastung eine optimale Sauerstoffversorgung des Körpers sichergestellt ist. Das Atemzeitvolumen (beispielsweise das Atemminutenvolumen) ist das Produkt aus Atemzugvolumen und Atemfrequenz innerhalb einer spezifischen Zeiteinheit.

Während sich das Lungenvolumen eines Erwachsenen auch durch Ausdauersport nicht mehr beeinflussen lässt, kann Ausdauersport dazu beitragen, die Vitalkapazität zu erhöhen und die Atmung insgesamt zu ökonomisieren. Eine Ökonomisierung der Atmung führt dazu, dass auch bei höherer Belastung die Atemfrequenz weniger stark ansteigt und stattdessen eine größere Sauerstoffmenge bei der Inspiration aufgenommen werden kann, also das Atemzugvolumen erhöht wird. Auf diese Weise sind umgekehrt Rückschlüsse über Veränderungen der Ausdauerleistung durch Bestimmung des Atemzugvolumens möglich.

Eine andere Möglichkeit, die Ausdauerleistungsfähigkeit über die Atmung zu ermitteln, ist die maximale Sauerstoffaufnahme ($\dot{V}O_{2\max}$). Dabei wird die relative, maximal mögliche Sauerstoffaufnahme (pro Kilogramm Körpergewicht) bestimmt. Diese gibt an, wie viel Sauerstoff der Körper pro Minute unter Ausbelastung maximal für wenige Sekunden verwenden kann (in Milliliter Sauerstoff pro Minute, $m\ell O_2/min$). Dadurch, dass eine solche Ausbelastung nicht lange aufrechterhalten werden kann, stellt die $\dot{V}O_{2\max}$ automatisch eine obere Grenze der aktuellen Ausdauerleistungsfähigkeit dar. Die absolute $\dot{V}O_{2\max}$ ist vom Körpergewicht unabhängig und lässt sich über das »Ficksche Prinzip« (Gleichung 1.2) ermitteln:

$$\dot{V}O_{2\max} = \text{HMV} \cdot (C_aO_2 - C_vO_2). \quad (1.2)$$

Dabei bezeichnet HMV das Herzminutenvolumen, C_aO_2 den arteriellen Sauerstoffgehalt und C_vO_2 den venösen Sauerstoffgehalt. Da die Bestimmung des Sauerstoffgehalts unter anderem Kenntnis über die Sauerstoffsättigung und den Hämoglobingehalt benötigt, wodurch neben der Atemgasanalyse auch die invasive Bestimmung der Sauerstoffkonzentrationen im arterialisierten und gemischt-venösen³ Blut notwendig ist, gibt es für die Praxis grobe Schätzverfahren zur ungefähren Bestimmung der $\dot{V}O_{2\max}$. Ein Beispiel dafür ist die von Uth et al. (2004) vorgestellte »Uth–Sørensen–Overgaard–Pedersen-Schätzung« (Gleichung 1.3),

³Gemischt-venöses Blut liegt in der Lungenarterie (*Arteria pulmonalis*) vor. Es enthält venöses Blut aus der oberen und unteren Hohlvene (*Vena cava superior* und *inferior*) des großen Körperkreislaufs sowie aus dem venösen Abflusswege des Herzens (*Sinus coronarius*).

bei der die $\dot{V}O_{2\max}$ relativ zur maximalen Herzfrequenz und der Ruheherzfrequenz angenommen wird:

$$\dot{V}O_{2\max} \approx \frac{HF_{\max}}{HF_{\text{Ruhe}}} \cdot 15 \frac{\text{ml}}{\text{kg} \cdot \text{min}}. \quad (1.3)$$

Muskulatur

Damit die Muskulatur gut arbeiten kann, ist eine ausreichende Versorgung mit Sauerstoff essentiell, da nur ein sehr geringer Anteil an Energie anaerob zur Verfügung gestellt werden kann (vgl. Unterabschnitt 1.1.1). Neben dem kardio-vaskulären und dem pulmonalen System tragen die Muskeln selbst durch »Kapillarisation« dazu bei, ausreichend mit Sauerstoff versorgt zu werden. Sowohl Ausdauer- als auch Krafttraining sorgen dafür, dass die Durchblutung der Muskulatur verbessert wird, indem mehr Kapillaren geöffnet werden oder das Lumen – d.h. der innere Durchmesser – der Kapillaren vergrößert wird. Dadurch wird die Austauschkapazität erhöht und der Muskel kann mit mehr Sauerstoff in kürzerer Zeit versorgt werden. Dies führt zu einer Leistungssteigerung des Muskels. Insgesamt kann die Anzahl der Kapillaren relativ zur Masse der Muskulatur durch Ausdauertraining verdoppelt werden. Im Leistungssportbereich ist sogar eine Verdreifachung im Vergleich zu untrainierten Personen möglich (Hohmann et al., 2020).

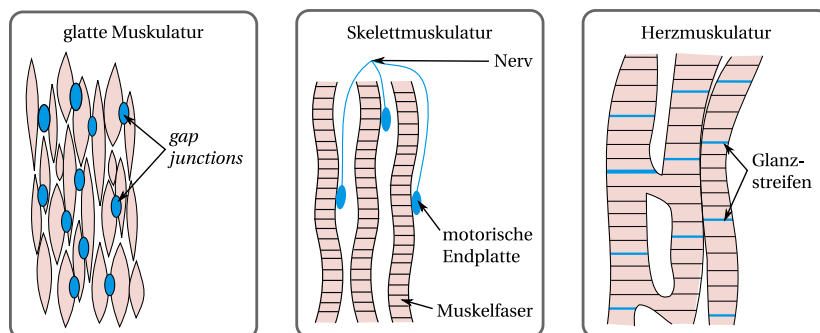


Abbildung 1.7.: Schematische Darstellung der verschiedenen Typen des Muskelgewebes (v.l.n.r.): glatte Muskulatur, quergestreifte Muskulatur und die Herzmuskulatur als Sonderfall der quergestreiften Muskulatur

Grundsätzlich gibt es zwei unterschiedliche Typen von Muskelgewebe im menschlichen Körper: die »glatte« und die »quergestreifte« Muskulatur. Abbildung 1.7 zeigt eine schematische Darstellung. Die glatte Muskulatur befindet sich beispielsweise an den Wänden der Hohlorgane wie Magen und Darm, den Geschlechtsorganen, in Gefäßwänden und den tiefen Atemwegen. Auch der Pupillenmuskel und die Drüsen des Körpers weisen eine glatte Muskulatur auf. Eine besondere Eigenschaft der glatten Muskulatur ist, dass sie nicht ermüdet und nicht willentlich angesteuert werden kann. Ihre Funktionsweise unterliegt damit dem autonomen Nervensystem. Bewegungen der glatten Muskulatur sind hingegen nur vergleichsweise langsam möglich – die Muskelkontraktion einer glatten Muskelzelle kann bis

zu einer Sekunde in Anspruch nehmen. Bei der quergestreiften Muskulatur hingegen beträgt die Kontraktionsgeschwindigkeit wenige Millisekunden. (Schwegler und Lucius, 2011)

Die quergestreifte Muskulatur lässt sich in das Gewebe der Skelettmuskulatur und das des Herzmuskels unterteilen. Denn obwohl der Herzmuskel (Myokard) weder willentlich angesteuert werden kann, noch ermüden darf und sehr ausdauernd arbeiten muss, wäre die Kontraktionsgeschwindigkeit der glatten Muskulatur für den Herzschlag nicht ausreichend. Der Herzmuskel gehört also zur quergestreiften Muskulatur, unterscheidet sich jedoch funktional vom restlichen skelettalen quergestreiften Muskelgewebe: Durch seine autonome Funktionsweise ist der Herzmuskel unabhängig von nervalen Impulsen und hat entsprechend wie die glatte Muskulatur keine motorische Endplatte. Stattdessen verbreiten sich Muskelaktionspotenziale sowohl im Herzmuskel als auch in der glatten Muskulatur beispielsweise mittels Neurotransmittern oder Hormonen über eine Ansammlung von Zell-Zell-Kanälen, den *gap junctions*. Diese verbinden die einzelnen Fasern miteinander. Die Herzmuskelfasern weisen einige Unterschiede zu den Fasern der Skelettmuskulatur auf – für eine detailliertere Übersicht über den genauen Aufbau und die Funktionsweise von Herzmuskulatur und glatter Muskulatur sei an dieser Stelle auf Schwegler und Lucius (2011) verwiesen.

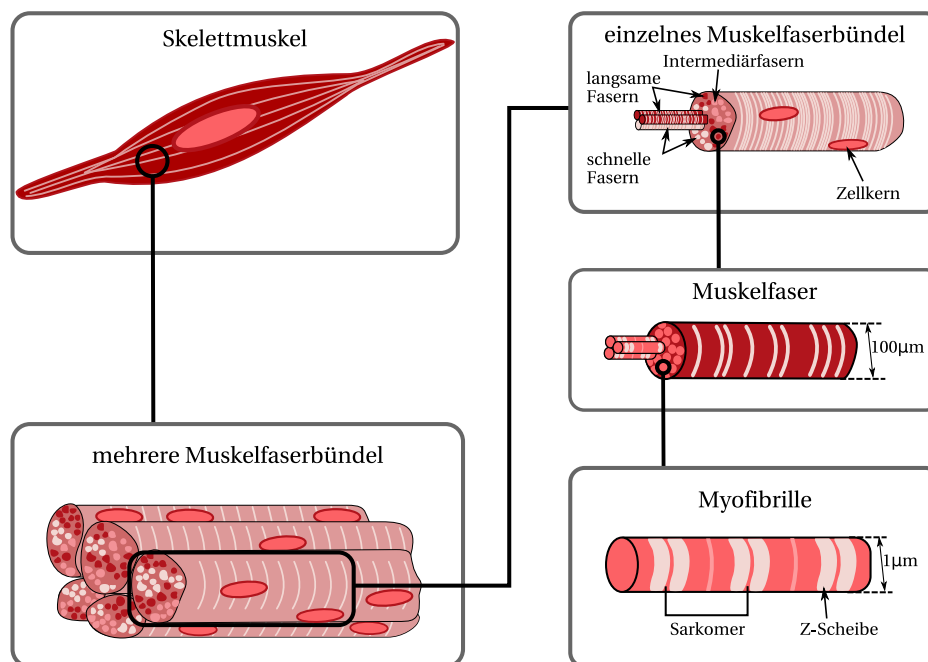


Abbildung 1.8.: Schematische Darstellung des Aufbaus eines Skelettmuskels

Im Folgenden wird speziell der Aufbau des Skelettmuskelgewebes wie in Abbildung 1.8 dargestellt betrachtet und die Veränderungen, die innerhalb dieser Muskulatur durch Ausdauersport verursacht werden. Das Skelettmuskelgewebe macht rund 40-50% des menschlichen Körpergewichtes aus und findet sich im aktiven Bewegungsapparat wieder. Im Gegensatz zur glatten Muskulatur ist die quergestreifte Skelettmuskulatur willkürlich steuerbar. Sie gehört damit zum somatischen, willkürlichen Nervensystem. Ein Skelettmuskel besteht

zunächst aus vielen Muskelfaserbündeln und Blutgefäßen. Ein Muskelfaserbündel ist eine Gruppe von Muskelfasern, die von einer Bindegewebsschicht umgeben ist. Dabei gibt es unterschiedliche Muskelfasertypen, auf die nachfolgend noch näher eingegangen wird. In der Regel kommen innerhalb eines Muskelfaserbündels unterschiedliche Typen von Muskelfasern vor. Muskelfasern bestehen aus mehreren verschmolzenen Zellen. Dadurch enthalten sie auch mehrere Zellkerne und Mitochondrien. Die einzelnen Muskelfasern sind von einer Zellmembran umgeben. Kleine Gruppierungen von Muskelfasern werden von Nerven über eine motorische Endplatte gesteuert. Diese Gruppierungen werden gemeinsam mit ihrem Nerv als *motorische Einheit* bezeichnet. Die Funktionalität der Muskeln obliegt der kleinsten Einheit innerhalb der Muskelfasern: den Myofibrillen. Myofibrillen sind eiweißartige Strukturen, die durch ihren Aufbau in der Lage sind, für Entspannung und Kontraktion im gesamten Muskel zu sorgen. Die Myofibrillen bestehen aus aneinandergereihten Sarkomeren, die ihrerseits Filamente aus den Strukturproteinen Aktin und den Motorproteinen Myosin enthalten. Bei der Muskelkontraktion verschieben sich die Aktin- und Myosinfilamente gegeneinander wie schematisch in Abbildung 1.9 dargestellt. (Haber, 2009; Markworth, 2010; Schwegler und Lucius, 2011)

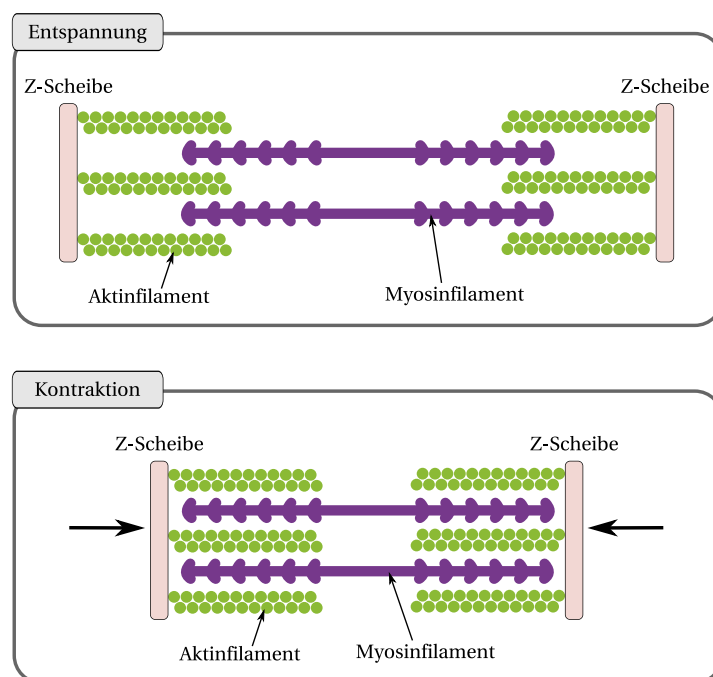


Abbildung 1.9.: Schematische Darstellung von Entspannung und Kontraktion im Sarkomer. Bei der Kontraktion (unten) verschieben sich die Aktin- und Myosinfilamente gegeneinander, sodass sich die Länge zwischen zwei Z-Scheiben und damit das Sarkomer insgesamt verkürzt. Angelehnt an Hottenrott und Neumann (2016) (S. 86)

Je nach Klassifikation lässt sich das Skelettmuskelgewebe unterschiedliche Muskelfasertypen unterteilen. Bei der Klassifikation nach ihrer Farbe gibt es zwei Typen von Muskelfasern: die »langsam zuckenden, roten Muskelfasern« (*Slow Twitch*, ST, Typ I Muskelfasern)

und die »schnell zuckenden, weißen Muskelfasern« (*Fast Twitch*, FT, Typ II Muskelfasern). Innerhalb der FT-Muskelfasern ergeben sich in dieser Klassifikation verschiedene Unterkategorien. In der Regel werden lediglich zwei dieser Unterkategorien näher betrachtet (*Fast Twitch Oxidative* (FTO)-Fasern und *Fast Twitch Glycolytic* (FTG)-Fasern), die sich hauptsächlich im Energieverbrauch unterscheiden. Zusätzlich gibt es die »Intermediärfasern«. Diese werden i.d.R. zwischen den ST- und den FTO-Fasern verortet. Bei einer Klassifikation nach Kontraktionseigenschaften wird üblicherweise zwischen den ST-Fasern, den FT-Fasern und den Intermediärfasern unterschieden.

Die zentralen Unterschiede zwischen den Fasertypen sind neben der Färbung ihre Kontraktionseigenschaften und die Art ihrer Energiebereitstellung: Die langsam zuckenden Muskelfasern erhalten ihre (dunkel-)rote Färbung durch eine große Menge Myoglobin⁴. Durch die große Menge Myoglobin können sie im Vergleich zu den wenig Myoglobin enthaltenden, schnell zuckenden Muskelfasern eine höhere Menge an Sauerstoff speichern. Entsprechend ist die primäre Energiezufuhr der ST-Fasern aerob (langsame aerobe Glykolyse). Die aerobe Energiegewinnung bedingt dabei, dass die ST-Fasern zwar langsamer, dafür aber deutlich ausdauernder kontrahieren. ST-Fasern sind damit besonders bei langandauernden Aktivitäten von Vorteil, die keine schnelle Spitzenleistung erfordern. Im Gegensatz dazu beziehen FT-Fasern ihre Energie primär über die anaerobe Glykolyse. Durch die schnellere Energiebereitstellung können FT-Fasern schneller kontrahieren, sind dafür jedoch weniger ausdauernd und ermüden schneller. FT-Fasern sind damit besonders bei kurzzeitigen Aktivitäten mit hoher Belastung von Vorteil. Innerhalb der FT-Fasern sind FTO-Fasern am ausdauerndsten und ermüden am langsamsten. Die Eigenschaften der hellroten FTG-Fasern und der Intermediärfasern lassen sich zwischen den Eigenschaften der ST- und denen der FTO-Fasern verorten. Ihre Energiebereitstellung verläuft sowohl über die schnelle aerobe Glykolyse wie auch in geringeren Teilen über die anaerobe Glykolyse. Dadurch haben sie trotz ihrer geringeren Menge Myoglobin ähnlich wie die ST-Fasern eine hohe Ermüdungsresistenz und ermöglichen im Vergleich zu FT-Fasern eine länger andauernde Kontraktion. Tabelle 1.2 fasst die zentralen Eigenschaften von ST- und FT-Muskelfasern zusammen und zeigt ihre typischen Funktionsgebiete auf. Die Intermediärfasern sind dabei als Zwischentyp der ST- und FTO-Fasern nicht gesondert aufgeführt.

Für die individuellen Belastungsgrenzen und sportliche Aktivitäten ist vor allem das Verhältnis von ST- zu FT-Muskelfasern von besonderer Bedeutung. Ohne Training ist die Zusammensetzung der Muskulatur hinsichtlich von ST- und FT-Fasern hochgradig individuell. Markworth (2010) zufolge wird allerdings vermutet, dass die ST-Fasern insgesamt leicht überwiegen und das Verhältnis der Fasertypen bei den meisten Menschen nahezu ausgeglichen ist. Dabei hängt der spezifische Anteil an ST- und FT-Fasern eng mit der Aufgabe des je-

⁴Ähnlich wie Hämoglobin ist Myoglobin ein sauerstoffbindendes Protein. Es dient als Sauerstoffspeicher im Muskel und hat – wie Hämoglobin – eine Häm-Gruppe, an deren zentrales Eisen-Ion Sauerstoff reversibel binden kann. Die Häm-Gruppe sorgt auch bei Myoglobin für die rote Färbung.

weiligen Muskels zusammen. So bestehen Muskeln der Haltemuskulatur (beispielsweise der große Rückenstrecker, *Musculus erector spinae*) üblicherweise aus einem deutlich höheren Anteil (bis zu 95%) an ST-Fasern (Markworth, 2010). Auch die Kaumuskulatur und der Zwillingswadenmuskel (*Musculus gastrocnemius*) weisen einen höheren Anteil an ST-Fasern auf, während beispielsweise der Bizeps (*Musculus biceps*) und der Augenmuskel überwiegend FT-Fasern enthalten (Dickhuth et al., 2007).

Tabelle 1.2.: Eigenschaften der einzelnen Muskelfasertypen. Angelehnt an Dickhuth et al. (2007)

	Schnelle Fasern		Langsame Fasern
Kurzbezeichnung	FTG (fast twitch glycolytic)	FTO (fast twitch oxidative)	ST (slow twitch)
Synonyme	Typ IIb-Faser, A-Faser	Typ IIa/c-Faser, C-Faser, Intermediärfasern	Typ I-Faser, B-Faser
Farbe	weiß	rot	dunkelrot
Energiestoffwechsel	anaerob	aerob und anaerob	aerob
Kapillardichte	niedrig	mittel	hoch
Mitochondrien-dichte	niedrig	hoch	(sehr) hoch
Myoglobindichte	niedrig	hoch	hoch
ATP-ase Gehalt	hoch	mäßig	gering
Arbeitsweise	schnelle Kontraktion, geringe Ausdauer	schnelle Kontraktion, mäßige Ausdauer	langsame Kontraktion, hohe Ausdauer
Einsatzbereich	Schnellkraft, Kraftschnellkraft, Maximalkraft	Kraftausdauer, Maximalkraft	Ausdauer, Halte- und Stützfunktion
Trainingsbereich	Intensive Belastung, Kurzzeitige Belastung	Intensive Belastung, Extensive Belastung	Extensive Belastung
Typische Sportarten	Kraftsport, Sprint, Golf	Boxen, Leichtathletik	Marathon, Ausdauersport
Belastungszeit	< 60 s	< 30 min	> 30 min

Die individuelle Zusammensetzung ist dabei den Recherchen von Haber (2009) zufolge zu rund 40% genetisch bedingt und kann zu rund 60% durch Umwelteinflüsse und Trainingsreize modifiziert werden. So lassen sich durch gezieltes Krafttraining FTG- in FTO-Fasern und FTO- in ST-Fasern umwandeln – diese Verschiebung des Muskelfaserspektrums ist jedoch

nur kurzfristig und reversibel. Die umgekehrte Umwandlung von ST- zu FT-Fasern ist nach aktuellem Wissensstand durch langfristiges, gezieltes Ausdauertraining ebenfalls möglich, jedoch nicht reversibel (Hohmann et al., 2020).

1.1.3. Physiologische Reaktionen und Anpassungen durch Ausdauerbelastung

Die Adaptationen, die im menschlichen Körper durch Ausdauerbelastung stattfinden, laufen in verschiedenen molekularen Strukturen und Organen zeitlich unterschiedlich ab. Das Vier-Stufen-Modell von Neumann und Schüler (1994) wird nach heutigem Kenntnisstand als geeignete Modellierung der Trainingsrealität angenommen (Hottenrott und Neumann, 2016) und basiert auf Experimenten und Erfahrungen der Sportpraxis sowie auf tierexperimentellen Ergebnissen.

Tabelle 1.3.: Anpassung einzelner Organe an ein Ausdauertraining (übernommen aus Hohmann et al. (2007))

Organ	Funktionelle und morphologische Anpassungen durch Ausdauertraining
Herz	Senkung von Ruhepuls und Belastungspuls, Vergrößerung des Schlagvolumens, Ökonomisierung der Herzfunktion; Verbesserte Durchblutung des Herzmuskels und Anpassungen in Kammer- und Muskelgröße
Lunge	Ökonomisierung des Gasaustausches, Erhöhung der Sauerstoff-Aufnahmekapazität, Vergrößerung der Vitalkapazität und des Atemminutenvolumens
Blut	Günstiger Einfluss auf die Blutzusammensetzung, Erhöhung der Sauerstoffbindefähigkeit, Erhöhung der Sauerstofftransportfähigkeit
Gefäße	Verringerte Rückbildung der Elastizität von Gefäßen
Muskulatur	Vermehrung der Mitochondrien, verbesserte Kapillarisierung
Immunsystem	Verbesserung des Immunstatus
Hormondrüsen	Geringere Ausschüttung von Stresshormonen unter Belastung

Nach Hottenrott und Neumann (2016) »orientiert sich [das Vier-Stufen-Modell] an der praktischen Gestaltung des Trainings, um optimale Anpassungen in einem biologisch vorgegebenen Zeitraum zu erreichen.« Während »das Ausmaß der Funktionsumstellung (...) von Art, Intensität und Dauer der Trainingsbelastung« abhängt, eignet sich die Herzfrequenz besonders, um die Reaktion des Organismus auf veränderte Trainingsreize zu erkennen. So schreiben Hottenrott und Neumann (2016) weiter, dass die Herzfrequenz »am schnellsten auf Beanspruchung« reagiert und »bereits nach acht Tagen Training signifikant abnehmen« kann. Die einzelnen Stufen beziehen sich dabei auf »Veränderungen im Bewegungsprogramm« (Stufe 1), eine »Vergrößerung der Energiespeicher« (Stufe 2), die »Optimierung geregelter Systeme und Strukturen« (Stufe 3) und die »Koordinierung leistungsbeeinflussender Systeme

me« (Stufe 4). Die gesamte Anpassung dauert für einen Trainingsblock etwa 4-6 Wochen, wobei sich die einzelnen Anpassungsprozesse überschneiden können und nicht vollständig nacheinander stattfinden müssen.

Nachfolgend werden kurz die einzelnen Komponenten betrachtet, die für Adaptationen der Ausdauer besonders relevant sind. Während beispielsweise die flüssige Ausführung eines spezifischen Bewegungsablaufes der Ausdauerleistungsfähigkeit nur indirekt dient, ist wie in der vierten Stufe beschrieben die Synchronisation zwischen der Muskulatur und den leistungsbeeinflussenden Systemen fundamental für den Abschluss der Adaptation an eine spezifische Trainingsbelastung. Zu diesen leistungsbeeinflussenden Systemen gehören das vegetative und zentrale Nervensystem, das kardiopulmonale System, der Elektrolythaushalt, der Energiestoffwechsel, die Muskulatur, das Hormonsystem und das Immunsystem (Hottenrott und Neumann, 2016). Die Anpassung der einzelnen Organe durch Ausdauerbelastung wird von Hohmann et al. (2007) ausführlich beschrieben. Ein Überblick über die leistungsbeeinflussenden Systeme ist in Tabelle 1.3 dargestellt.

Das vegetative Nervensystem Besonders durch die gesamte Stoffwechselökonomisierung stärkt Ausdauersport den Parasympathikus und kann somit die Fähigkeit zur Erholung verbessern. Die Stresshormonausschüttung sinkt und auch andere Bereiche des Parasympathikus werden verbessert (z.B. Regulierung von Magen-Darm-Tätigkeiten). Auch kurzfristige Adaptationen wie beispielsweise ein erhöhter Herzschlag oder die verbesserte Durchblutung beanspruchter Muskulatur während der Belastung werden über das vegetative Nervensystem geregelt.

Das Zentralnervensystem Zimmer et al. (2015) konnten zeigen, dass Ausdauerbelastung die Hirnleistungsfähigkeit steigern und das Risiko für bzw. das Fortschreiten von neurodegenerativen Erkrankungen senken kann.

Das kardiopulmonale System Insbesondere die maximale Sauerstoffaufnahme ist für (aerobe) Ausdauerbelastungen von großer Bedeutung und ein zentraler Parameter in der Bestimmung der (kardiorespiratorischen) Leistungsfähigkeit. Sie gibt an, welche Menge an Sauerstoff (in Milliliter) pro Minute der Körper bei Ausbelastung maximal verwerten kann. Nach dem Fickschen Prinzip ist die $\dot{V}O_{2\max}$ vom Herzminutenvolumen sowie dem Sauerstoff-Gehalt der Venen und Arterien abhängig (vgl. Abschnitt 1.1.2). Das Herzminutenvolumen wiederum wird über das Verhältnis von maximaler Herzfrequenz und Schlagvolumen bestimmt. Während die maximale Herzfrequenz kaum trainierbar ist, lässt sich das Herzminutenvolumen durch das Schlagvolumen verbessern. Das Schlagvolumen lässt sich durch Ausdauertraining (im Vergleich zu untrainierten Personen) etwa verdoppeln, da das Herzvolumen durch Ausdauertraining zunimmt, die Herzzinnenräume erweitert werden, der Herzmuskel kräftiger wird und so mehr Blut pro Schlag und damit mehr Blut pro Minute durch das Herz gepumpt werden kann.

Der Elektrolythaushalt Die Ausdauerleistungsfähigkeit wird bereits von geringfügigem Wasserverlust (2% kurz- bzw. 4% längerfristig) negativ beeinträchtigt.

Der Energiestoffwechsel Je nach Belastungsintensität kann die dominierende Energiebereitstellung angepasst werden. Während bei steigender Belastung die aerobe Fettverbrennung sinkt und die aerobe Glykolyse steigt, geht diese bei weiter ansteigender Belastung in die anaerobe Glykolyse über. Der aktuelle Beanspruchungsgrad einer Sportlerin oder eines Sportlers lässt sich daher gut über die Laktatmenge im Blut bestimmen, da Laktat durch die anaerobe Glykolyse anfällt. Zusätzlich werden durch die Steigerung der Mitochondrienbiogenese je nach Belastung der aeroben Energiestoffwechsel (Glukose- und Fettsäureumsatz) sowie die Kapazität des Kreatinphosphat-Speichers erhöht (Stufe zwei des Modells).

Die Muskulatur Grundsätzlich gibt es drei unterschiedliche Wege, über die sich die Muskulatur an eine andauernde Belastung anpasst: *Synchronisation*, *Hyperplasie* und *Hypertrophie*. Eine Synchronisation findet ohne morphologische Anpassungen der Muskulatur statt. Sie bezieht sich auf die Synchronisierung der motorischen Einheiten, die eine Verbesserung der Kraftentfaltung zur Folge hat. Die Hyperplasie beschreibt eine Vermehrung von Muskelzellen mit entsprechender Oberflächenvergrößerung des Muskels. Diese Form der Adaptation kommt äußerst selten vor (Haber, 2009). Die relevanteste Anpassung der Muskulatur ist die Hypertrophie. Bei dieser nimmt das Muskelvolumen ebenfalls zu – allerdings nicht durch eine Zellvermehrung, sondern durch das Wachstum einzelner Muskelzellen. Durch die muskuläre Beanspruchung werden mit der Zeit neue Myofibrillen gebildet, durch die die einzelnen Muskelzellen und damit schließlich auch der gesamte Muskel im Umfang zunehmen. Gleichzeitig begünstigt Ausdauertraining die muskuläre Durchblutung. Das Lumen der Kapillare im Muskel vergrößert sich und die Zahl der geöffneten Kapillaren steigt an. Auch eine Neubildung von Kapillaren ist bei Sauerstoff- oder Energiemangel des Muskels möglich. Hottenrott und Neumann (2016) gehen davon aus, dass regelmäßiges Ausdauertraining zu einer »Zunahme des Kapillarierungsgrades um 40%« führt. Auch die Stoffwechseleigenschaften der Muskelfasern werden durch das Training beeinflusst: Je nach Intensität und Belastung werden unterschiedliche Enzymaktivitäten erhöht, die die aerobe oder anaerobe Energiegewinnung unterstützen. Letztlich passt sich auch die spezifische Zusammensetzung der Muskelfasern zu einem gewissen Grad durch regelmäßiges Training an die Belastung an: Ein Ausdauertraining mit geringem Krafteinsatz über längere Zeit kann dazu führen, dass in der betreffenden Muskulatur der Anteil an weißen FT-Fasern sinkt, wohingegen die Anzahl an roten ST-Fasern ansteigt (Haber, 2009). Es wird vermutet, dass hauptsächlich die Intermediärfasern durch äußere Einflüsse zu ST- oder FT-Fasern ausgebildet werden können.

Das Hormonsystem (Ausdauer-)Training nimmt auf vielfältige Weise Einfluss auf das Hormonsystem der Körpers. Allgemein wird der Hormonspiegel auf die spezifische Belas-

tung ausgerichtet. Konkret bedeutet das, dass zur Anpassung der muskulären und zellulären Strukturen auf die auftretende Belastung vermehrt Wachstumshormone (Somatotropin) ausgeschüttet werden (Anpassung des Muskel- und Knochenbaus, Muskelhypertrophie); Zur Unterstützung des Knochenwachstums wird der Calciumstoffwechsel angepasst; Die Energiebereitstellung wird verbessert, was mit einer Optimierung des Schilddrüsenstoffwechsels einher geht; Der Organismus kann besser mit Stress umgehen (erhöhte Resilienz und angepasste Ausschüttung von Adrenalin und Noradrenalin); Der Insulinbedarf im Körper sinkt, wodurch der Blutzuckerspiegel besser stabilisiert wird. Gleichzeitig führen individuell zu hohe Belastungen über einen längeren Zeitraum dazu, dass das Hormonsystem zum Schutze des Körpers und der Energiereserven einige Systeme herunterreguliert – beispielsweise die Hormonproduktion in den Eierstöcken bei Frauen.

Das Immunsystem Während einerseits das Immunsystem durch sportliche Aktivitäten gestärkt wird, ist umgekehrt auch ein gesunder körperlicher Zustand zentral, um Leistung erbringen zu können.

Als begrenzende Faktoren bei der Erbringung von Ausdauerleistung heben Hollmann und Mader (2000) besonders die muskuläre Ausdauer hervor.

Im nachfolgenden Kapitel wird zunächst der Begriff der »Leistung« näher betrachtet. Anschließend werden zentrale Aspekte zusammengefasst, die zur qualitativen und quantitativen Bewertung einer (Ausdauer-)Leistung herangezogen werden können.

1.2. Allgemeine Grundlagen der Trainingssteuerung im Ausdauersport

Hohmann et al. (2007) verstehen unter der *trainingswissenschaftlichen Leistungsdiagnostik* die »Strukturierung der sportlichen Leistung und der Leistungsfähigkeit«, und unter der *trainingspraktischen Leistungsdiagnostik* einen »Ist-Sollwert-Vergleich«, um den Trainingserfolg bezüglich eines geplanten Zieles bewerten und kontrollieren zu können. Im Rahmen dieser Arbeit dient die Betrachtung der (trainingspraktischen) Leistungsdiagnostik lediglich einer Bewertung aktueller Leistungszustände und entsprechende empirische Daten als Vergleichswert der Modellierung. Dennoch werden in diesem Kapitel einige wesentliche Grundlagen zusammengefasst. Dabei liegt ein besonderes Augenmerk auf Messgrößen der Diagnostik und gängigen Testverfahren im Ausdauersport.

Zunächst gilt es zu beachten, dass *Ausdauer* nicht gleich *Ausdauer* ist. In der Trainingswissenschaft wird zwischen unterschiedlichen Belastungsbereichen für den Ausdauersport unterschieden, die sich in der Art des Trainings und teilweise in der Art des Stoffwechsels zur Energiebereitstellung unterscheiden. Es gibt Ausdauerbereiche für die Grundlagen-

ausdauer, die Kraftausdauer, sowie den wettkampfspezifischen Ausdauerbereich, der sich wiederum in Kurzzeit-, Mittelzeit- und Langzeitausdauer unterteilt. Zum wettkampfspezifischen Ausdauerbereich gehören auch die Schnelligkeitsausdauer, die Sprintausdauer und die Schnellkraftausdauer. Tabelle 1.4 stellt die unterschiedlichen Ausdauerbereiche ihren Stoffwechselwegen gegenüber.

Tabelle 1.4.: Übersicht über die unterschiedlichen Ausdauerbereiche, ihrer abkürzenden Bezeichnung und der jeweiligen Stoffwechsellaage (übernommen aus Hottenrott und Neumann (2016))

Bereich	Bezeichnung	Stoffwechsellaage
Regeneration und Kompensation	REKOM	aerob
Grundlagenausdauer 1	GA 1	aerob
Kraftausdauer 1	KA 1	aerob
Grundlagenausdauer 2	GA 2	aerob-anaerob
Kraftausdauer 2	KA 2	aerob-anaerob
wettkampfspezifische Ausdauer	WSA	aerob-anaerob bis anaerob
Schnelligkeitsausdauer 1	SA	anaerob

Während ein Training im Bereich der REKOM (»Regeneration und Kompensation«) die Regeneration unterstützen und beschleunigen soll, wird beim GA 1-Training (»Grundlagenausdauer 1«) die allgemeine Grundlagenausdauerfähigkeit entwickelt und stabilisiert. Außerdem dient dieser Trainingsbereich der Vorbereitung auf intensivere Belastungen. Das GA 2-Training (»Grundlagenausdauer 2«) weist im Vergleich zu GA 1 eine höhere Belastungsintensität auf und bereitet auf eine Wettkampfgeschwindigkeit vor. Im WSA-Training (»wettkampfspezifische Ausdauer«) werden entsprechend die wettkampfspezifische Ausdauer und eine Schnelligkeitsausdauer entwickelt. Die Kraftausdauer findet üblicherweise sportartenspezifisch statt. Während KA 1-Training (»Kraftausdauer 1«) die aerobe Kraftausdauer entwickelt, stehen beim KA 2-Training (»Kraftausdauer 2«) höhere Intensitäten und eine maximale Kraftausdauerfähigkeit im Vordergrund. Dabei wird im Vergleich zur Grundlagenausdauer mit höheren Widerständen und geringeren Bewegungsfrequenzen trainiert. (Hottenrott und Neumann, 2016)

Die Energiebereitstellung der WSA ist mit »aerob-anaerob bis anaerob« angegeben. Die Ursache dafür ist, dass sich die WSA ihrerseits nochmal in drei Zeitbereiche unterteilen lässt: die Kurzeitausdauer (KZA), die Mittelzeitausdauer (MZA) und die Langzeitausdauer (LZA). Die LZA kann ebenfalls in Abhängigkeit der Dauer aufgeteilt werden (LZA I-III), da sich bei längerer Belastung die Stoffwechsellaage schrittweise von aerob-anaerob auf ausschließlich aerob umstellt. Abbildung 1.10 gibt einen Überblick über das Verhältnis zwischen anaerober und aerober Energiegewinnung für die Schnelligkeitsausdauer und die unterschiedlichen Bereiche der wettkampfspezifischen Ausdauer. Innerhalb der SA zeigt sich deutlich, dass die Energie bei einem 100 m-Kurz sprint (7 s) fast ausschließlich anaerob bereitgestellt wird,

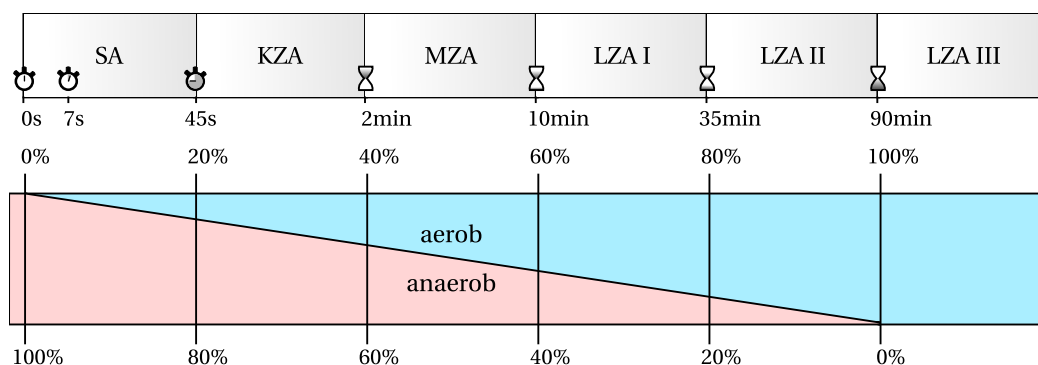


Abbildung 1.10.: Gegenüberstellung der einzelnen Ausdauerbereiche für die Schnelligkeitsausdauer (SA) und die verschiedenen Bereiche der wettkampfspezifische Ausdauer: Kurzzeitausdauer (KZA), Mittelzeitausdauer (MZA) und Langzeitausdauer (LZA). Dargestellt sind die zeitlichen Bereiche der Klassifikation (oben) sowie die angenommene prozentuale Aufteilung der anaeroben und aeroben Energiebereitstellung (unten). (Adaptiert nach Dickhuth et al., 2007, S. 11)

doch schon bei einem 400 m-Sprint (45 s) etwa 20% der Energie aerob benötigt werden. Die Langzeitausdauer (ab etwa 10 min) reicht von 60%-iger aerober Energiebereitstellung bis hin zu einer vollständig aeroben Stoffwechsellage (LZA III, ab etwa 90 min). Wie bereits in Unterabschnitt 1.1.1 (»Energiebereitstellung«) erläutert reichen die Glykogenvorräte in den Muskeln in Abhängigkeit des Trainingszustands bis zu ca. 100 min. Soll die Belastung über diese Zeit hinaus aufrecht erhalten werden, sollten die Speicher durch kohlenhydratreiche Nahrung rechtzeitig aufgefüllt werden. Empfohlen wird die Aufnahme von etwa 30 g Kohlenhydraten pro Stunde.

1.2.1. Belastungs-Beanspruchungs-Konzept

In den empirischen Arbeiten (Kapitel 5-7) werden in den betrachteten Modellen stets zwei unterschiedliche Kenngrößen betrachtet: Belastungs- und Beanspruchungsgrößen. Während es sich – im Kontext des Ausdauersports und ohne Berücksichtigung von psychischen Belastungen – bei der Belastung um äußere Einflussfaktoren und Steuergrößen des Trainings handelt (vgl. nachfolgenden Unterabschnitt »Belastungsnormative«), beschreibt die Beanspruchung die innere Reaktion des Körpers, auf den die Belastung (beispielsweise eine Trainingseinheit) wirkt. Die Beanspruchung macht sich in den unterschiedlichen inneren Funktionssystemen bemerkbar und wirkt auf die Herzfrequenz, die Atmung, die Muskulatur, etc. (vgl. auch »Physiologische Reaktionen und Anpassungen durch Ausdauerbelastung« aus Unterabschnitt 1.1.3). Obwohl eine Belastung im Sinne einer Trainings- und Leistungsanforderung im Sport stets zu einer Beanspruchungsreaktion führt, ist die Stärke der Reaktion von zahlreichen Faktoren abhängig: Neben der tagesaktuellen Leistungsfähigkeit spielen dabei auch die Motivation, sowie kognitive und emotionale Leistungsvoraussetzungen eine Rolle. (Hottenrott und Neumann, 2016) Die körperliche Leistungsfähigkeit setzt sich schließlich aus der Wechselwirkung zwischen Belastung und Beanspruchung zusam-

men und ist auf Grund der Vielzahl an Einflussfaktoren hochgradig individuell. Während einerseits eine Belastung abhängig von der aktuellen Belastbarkeit und Leistungsfähigkeit eine spezifische Beanspruchungsreaktion hervorruft, führt umgekehrt die Beanspruchung der Funktionssysteme zu einer Veränderung von Belastbarkeit und Leistungsfähigkeit, und wirkt damit auf die zukünftig mögliche Belastung. Abbildung 1.11 illustriert die Einflussfaktoren und Wechselwirkungen innerhalb des Belastungs-Beanspruchungs-Konzepts.

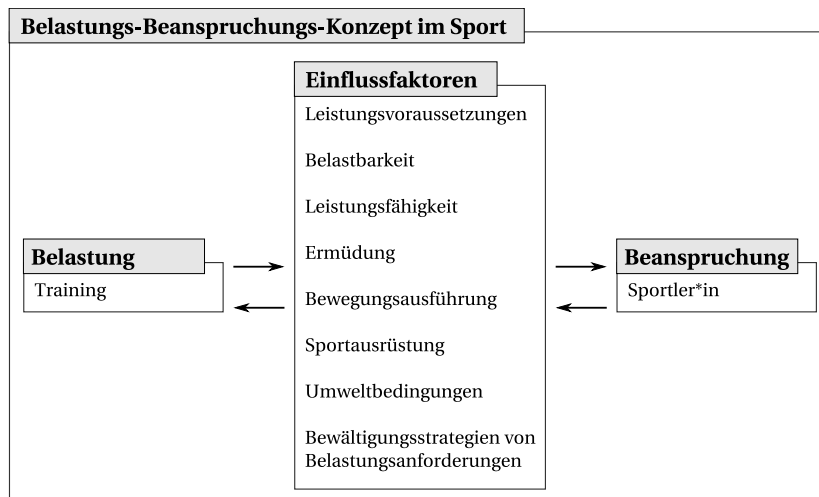


Abbildung 1.11.: Belastungs-Beanspruchungs-Konzept (Adaptiert nach Hottenrott und Neumann, 2016, S. 268-270)

Die *maximale* Leistungsfähigkeit einer Person ist dabei unabhängig von Körperdimensionen der Person, wohingegen die relative Belastung zur Belastung passende Körperdimensionen miteinbezieht, wie beispielsweise die Leistung pro Kilogramm Körpergewicht in Watt. (Graf und Höher, 2009, S. 147) beschreibt die maximale Leistungsfähigkeit als diejenige Leistung, »die eine Person imstande ist zu vollbringen, z.B. auf dem Fahrradergometer«.

Definition 1.2: Leistung.

Leistung lässt sich definieren als »Differentialquotient der Arbeit nach der Zeit ($P = dA/dt$)«, also in etwa der verrichteten Arbeit pro Zeiteinheit. Leistung wird in Watt angegeben. Eine mögliche Berechnung der Leistung ergibt sich damit als Produkt aus Kraft und Geschwindigkeit.

Je nach betrachteter »Kraft« lässt sich Leistung in Gesamtleistung, Muskelleistung, Beschleunigungsleistung und Bremsleistung unterteilen. Eine sportliche Leistung wird oft durch die maximale Leistung eines Muskels (oder einer Muskelgruppe) limitiert.

Basierend auf Arbeiten von Gundlach (1980) und Schnabel et al. (1994) betrachten Hohmann et al. (2020) unterschiedliche Komponenten der sportlichen Leistungsfähigkeit. Ein besonderes Augenmerk wird dabei auf Konstitution und Kondition, Koordination und Technik, sowie Taktik und psychische Eigenschaften gelegt. Da im Kontext dieser Arbeit aus-

schließlich die Ausdauerleistungsfähigkeit beim Radsport betrachtet wird, beschränkt sich dieses Kapitel auf die Betrachtung leistungsdiagnostischer Maßnahmen zur Konstitution und Kondition. Obwohl alle der oben genannten Komponenten sportartenspezifische Bedeutung haben, sind Koordination, Technik, Taktik und psychische Eigenschaften bei den in dieser Arbeit betrachteten Fragestellungen gegenüber der Konstitution und insbesondere der Kondition von untergeordnetem Interesse; auch wenn insbesondere Taktik und Psychologie auch für die Wettkampfleistung hochgradig relevant sind, werden diese jedoch für die Betrachtungen der Ausdauerleistungsfähigkeit nachfolgend nicht berücksichtigt.

1.2.2. Belastungsnormative

Allgemein lassen sich im sportlichen Training Hohmann et al. (2020) zufolge fünf unterschiedliche Normative der Belastung betrachten: Umfang, Dauer, Intensität, Dichte und Häufigkeit, wie in Tabelle 1.5 dargestellt.

Tabelle 1.5.: Übersicht über die fünf Belastungsnormative; angelehnt an Hohmann et al. (2020) (S. 54)

Belastungsnormative	Beschreibung	Merkmale und Messgrößen (exemplarisch)
Umfang	Gesamtumfang der Belastung	Anzahl der Wiederholungen, Distanz, Zeitdauer, Gewicht/Widerstand (»Meter-Kilogramm-Sekunden-System«)
Dauer	Zeit, über die eine Belastungseinwirkung erfolgt	Sekunden, Minuten, Stunden
Intensität	Grad der Anstrengung und Art und Weise der Übungsausführung	z.B. Geschwindigkeit pro Zeiteinheit, Herzfrequenz, Blutlaktatkonzentration, Wattleistung,...; kann gering, submaximal oder maximal sein
Dichte (Pause)	zeitliche Aufeinanderfolge von Belastungen; Wechsel zwischen Belastung und Erholung	Zeitintervalle zwischen Einzelbelastungen, Verhältnis von Belastungs- und Erholungszeit
Häufigkeit	Gesamtanzahl der Trainingseinheiten und Einzelwiederholungen	

Mit diesen Belastungsnormativen ist es möglich, unterschiedliche Trainingsmethoden wie Ausdauer-, Kraft- und Schnelligkeitsmethoden näher zu charakterisieren. Insbesondere kann mit ihrer Hilfe die Trainingsbelastung sowohl im Nachgang an ein Training beurteilt,

als auch im Vorfeld zur Trainingssteuerung charakterisiert und geplant werden. Die einzelnen Belastungsnormative können dabei nicht unabhängig voneinander betrachtet werden und stehen in Wechselwirkung zueinander.

Hottenrott und Neumann (2016) führt zusätzlich noch die Bewegungsausführung als sechste Belastungsnormative an. Nach Haber (2009) lässt sich die Trainingsbelastung abgeleitet von den Belastungsnormativen durch vier wesentliche Maßzahlen quantifizieren: Intensität, Dauer, Häufigkeit und die wöchentliche Netto-Trainingsbelastung. Dabei werden stets nur trainingswirksame Belastungen betrachtet. Als trainingswirksam bezeichnet Haber (2009) eine Belastung, wenn sie eine Mindestintensität für eine gewisse Mindestdauer aufweist und auch eine Häufigkeit pro Woche erreicht wird, die zu dauerhaften organischen Trainingsanpassungen führt. Die minimalen Werte werden dabei sportartenspezifisch betrachtet.

Intensität Die Intensität wird angegeben als prozentuale Leistung während des Trainings im Verhältnis zur maximalen Leistung, also $\text{Intensität} = \frac{\text{Trainingsleistung}}{\text{maximale Leistung}} \cdot 100\%$. Belastungen unterhalb einer spezifischen Mindestintensität haben keine organische Trainingsanpassung für die Leistungsfähigkeit zur Folge und bleiben als nicht-trainingswirksam unberücksichtigt. Die Intensität ist als quantitative Maßzahl qualitativer Natur: Sie kann quantitativ gemessen und angegeben werden, wird jedoch nur ab einer spezifischen Qualität berücksichtigt. Im Ausdauertraining wird als Minimalintensität diejenige Belastung betrachtet, die zur Verbesserung der aeroben Ausdauerfähigkeit und damit zur Entwicklung der $\dot{V}O_{2\max}$ notwendig ist. Ein Bereich von etwa 55 – 70% der $\dot{V}O_{2\max}$ wird nach Haber (2009) als sicher wirksamer Bereich angesehen.

Dauer Als Dauer wird diejenige Zeit betrachtet, in der eine Belastung mit trainingswirksamer Intensität stattfindet. Auch für die Belastungsdauer gilt: Unterhalb einer spezifischen Mindestdauer findet keine organische Trainingsanpassung statt, sodass nur Trainingszeiten oberhalb dieser Dauer berücksichtigt werden. Die Dauer ist als quantitative Maßzahl daher qualitativer Natur. Für den Ausdauersport gibt Haber (2009) hierzu eine Dauer von mindestens 10 Minuten mit ausreichender Trainingsintensität an.

Häufigkeit Als Häufigkeit wird diejenige Anzahl an trainingswirksamen Einheiten pro Woche betrachtet, die sowohl in Bezug auf ihre Intensität als auch in Bezug auf ihre Dauer trainingswirksam stattgefunden haben. Auch für die Häufigkeit gilt: unterhalb einer spezifischen Anzahl stattgefundener (trainingswirksamer) Trainingseinheiten findet keine organische Trainingsanpassung statt, sodass nur Trainingszeiten oberhalb dieser Anzahl berücksichtigt werden. Die Häufigkeit ist als quantitative Maßzahl daher qualitativer Natur. Haber (2009) beschreibt, dass zumindest zwei trainingswirksame Einheiten pro Woche notwendig seien, um im Ausdauersport eine Entwicklung der

$\dot{V}O_{2\max}$ erreichen zu können, da andernfalls die Trainingspause zu lang sei, um noch von der Überkompensation profitieren und damit die Leistung steigern zu können.

Wöchentliche Netto-Trainingsbelastung Die wöchentliche Netto-Trainingsbelastung (WNTB) entspricht der Summe aus allen trainingswirksamen Einheiten, die qualitativ für Intensität, Dauer und Häufigkeit berücksichtigt werden. Unterschreitet nur eine der drei Maßzahlen ihr spezifisches Minimum, wird die Trainingseinheit nicht für die WNTB berücksichtigt. Im Ausdauertraining wird die WNTB in Minuten, Stunden oder als »wöchentliche Netto-Trainingszeit« angegeben. Es handelt sich bei der WNTB um eine rein quantitative Maßzahl. Die WNTB soll dabei klein genug sein, damit eine ausreichende Erholung und Regeneration stattfinden kann. Gleichzeitig soll die WNTB groß genug sein, um die Leistungsfähigkeit steigern zu können. In Abhängigkeit der aktuellen Leistungsfähigkeit einer Person ist daher auch die geplante WNTB individuell und an den aktuellen Trainingszustand anzupassen.

1.2.3. Messgrößen der Beanspruchung in der Diagnostik

Bevor im nächsten Unterkapitel die am häufigsten verwendeten Schwellenkonzepte betrachtet werden, folgt zunächst die Darstellung der drei wichtigsten Messgrößen, die die Beanspruchungsreaktion auf einen Belastungsreiz abbilden und in der professionellen Sportpraxis relativ gut erhoben werden können: die Blutlaktatkonzentration, die Atemgasanalyse und die Herzfrequenz.

Der Verwendung der unterschiedlichen Messgrößen liegt zugrunde, dass die Energiebereitstellung bei körperlicher Belastung in unterschiedlichen Phasen verläuft (vgl. auch Unterabschnitt 1.1.1). Während bei niedriger Belastungsintensität Sauerstoff zur Energiegewinnung genutzt wird (aerobe Phase), findet für moderate und höhere Belastungen auf Grund der geringeren Sauerstoffverfügbarkeit die Energiegewinnung aus Glukose sowohl aerob als auch anaerob statt (aerob-anaerobe Phase). Sehr intensiven Belastungsspitzen werden zunächst über das muskulär vorhandene ATP und CrP (alaktazid-anaerobe Phase) und nach etwa 10 s über die Glykolyse mit Energie versorgt (laktazid-anaerobe Phase), wobei Laktat entsteht (vgl. auch Unterabschnitt 1.1.1). Entsprechend können unterschiedliche Methoden dabei helfen, die belastungsrelevante Stoffwechsellage und insbesondere Zeitpunkte zu detektieren, zu denen Veränderungen der Stoffwechsellage stattfinden.

Blutlaktatkonzentration Die Blutlaktatkonzentration gibt Aufschluss über die individuelle Stoffwechsellage und hilft dabei, den Übergang zwischen unterschiedlichen Phasen der Energiebereitstellung zu bestimmen (metabolische Schwellenkonzepte). Dabei kann die Stoffwechsellage mit einer spezifischen Belastung in Relation gesetzt werden: Bei intensiver Ausdauerbelastung reicht die Sauerstoffversorgung aus der Atmung nicht aus, um die Muskulatur über die aerobe Glykolyse ausreichend mit Energie zu versorgen. Durch die notwendige anaerobe Energiegewinnung entsteht Laktat

(vgl. auch »Allgemeine physiologische Grundlagen«, Abschnitt 1.1). Laktat wird über die Blutbahn aus der Muskulatur abtransportiert und eine gewisse Menge an Laktat kann hinreichend schnell abgebaut werden, um eine Laktat-Ansammlung zu vermeiden. Wird der Laktat-steady-state jedoch überschritten, verläuft der Abbau von Laktat langsamer als die Neubildung zur Energiegewinnung, sodass sich Laktat im Blut anreichert und die Muskulatur übersäuert. Häufig wird die Grenze von etwa 4 mmol/ℓ als maximaler Laktat-steady-state (»maxlass« oder »MLSS«) angenommen, bei dem sich Laktatbildung und Laktatelimination noch gerade so im Gleichgewicht befinden.

Atemgasanalyse Die maximale Sauerstoffaufnahme ($\dot{V}O_{2\max}$) gibt an, wie viel Sauerstoff der Körper pro Minute unter Ausbelastung maximal aufnehmen und insbesondere über die Muskulatur verwerten kann. In der Spirometrie können neben der Sauerstoffaufnahme auch das Abatmen von Kohlendioxid Kohlendioxidabgabe ($\dot{V}CO_2$), das Atemvolumen und die Atemfrequenz gemessen werden. Aus diesen Messgrößen lassen sich je nach Anwendungsfall unterschiedliche weitere Parameter berechnen: Der Respiratorische Quotient beschreibt das Verhältnis zwischen Sauerstoffaufnahme und Kohlendioxidabgabe und ermöglicht Aussagen darüber, wie hoch der Anteil an Fetten und Kohlenhydraten in der Energiebereitstellung ist. Das Atemäquivalent für Sauerstoff oder für Kohlendioxid beschreibt die Menge an Luft (in Litern), die eingeatmet werden muss, damit ein Liter Sauerstoff im Körper genutzt bzw. ein Liter Kohlendioxid abgeatmet werden kann. Durch regelmäßige Beanspruchung wird die Gasaustauschfläche in der Lunge vergrößert, Atmungsabläufe werden optimiert und allgemein wird die Lungenfunktion verbessert. Dadurch sinkt durch regelmäßige Ausdauerbelastung die Atemfrequenz (im Vergleich zum untrainierten Zustand), wohingegen das Atemvolumen steigt. Beide Parameter können somit ebenfalls Indikatoren für Veränderungen des Leistungszustandes sein. Dabei ist zu beachten, dass der gesamte Sauerstoffaustausch nicht nur von der Atmung abhängt, sondern auch vom Sauerstofftransport über das Herz-Kreislaufsystem und der Sauerstoffverwertung in den einzelnen Muskelzellen. Da die aerobe Energiegewinnung wiederum maßgeblich von der Sauerstoffaufnahmekapazität abhängt, bietet die Analyse der Parameter aus der Spirometrie eine weitere Möglichkeit, den Übergang einzelner Phasen der Energiebereitstellung zu bestimmen (ventilatorische Schwellenkonzepte).

Herzfrequenz Die Herzfrequenz lässt sich sowohl zur Steuerung der Belastung als auch zur Kontrolle der erfolgten Beanspruchung nutzen. Je nach Analysezweck lassen sich unterschiedliche Regulationszustände der Herzfrequenz untersuchen: Die Ruhe-HF eignet sich als Kontrollwert der Regeneration, da eine Erhöhung von rund 6-8 bpm bereits auf eine gestörte Regeneration⁵ (z.B. durch beginnende Überlastung oder auch durch Krankheit) hinweisen kann (Karjalainen und Viitasalo, 1986). Zur individuel-

⁵Dabei ist zu beachten, dass die Ruhe-Herzfrequenz bei Frauen durch den Menstruationszyklus zusätzlichen Schwankungen unterliegt.

len Bestimmung der verschiedenen Belastungsbereiche (»Trainingszonen«) sind die maximale Herzfrequenz (HF_{\max}) und die Ruheherzfrequenz (HF_{rest}) erforderlich: Der prozentuale Anteil der aktuellen Herzfrequenz im Vergleich zur Differenz zwischen HF_{\max} und HF_{rest} bestimmt während des Trainings die Trainingszone und wird genutzt, um die Wirksamkeit des Trainings anzugeben. Eine übliche Berechnung der HF_{\max} basiert auf der Formel

$$HF_{\max} = 220 - \text{Alter.}$$

Zur Bestimmung der maximalen Herzfrequenz existieren jedoch zahlreiche Varianten und Verbesserungen der Formel (Nes et al., 2013; Tanaka et al., 2001), die alle heuristischer Natur sind. Insbesondere können sich sowohl Ruhe-HF als auch maximale HF im Laufe der Zeit durch Training verändern (u.a. durch den Trainingszustand und das Alter (Londeree und Moeschberger, 1982)). Auch die Herzfrequenzvariabilität [engl. *Heart Rate Variability*] zeigt charakteristische Veränderungen je nach Belastungsintensität und Erholungszustand. Dadurch kann auch die HRV zur Trainingssteuerung und im Regenerationsmanagement genutzt werden. Auch im allgemeinen Freizeitsport – für den Analysen mit Laktat oder Spirometrie oft zu aufwändig sind – findet die Herzfrequenz immer häufiger Verwendung bei der Kontrolle der eigenen Leistung, wobei vor allem im Bereich der *Wearables* automatisch Trainingszonen berechnet und genutzt werden.

Für spezielle Analysezwecke können auch beispielsweise der Serumharnstoff oder die Kreatinkinase (CK) untersucht werden: Der Serumharnstoff gibt Aufschluss über die Beanspruchung des Proteinkatabolismus, bei dem vermehrt Eiweiße aus der Muskulatur auf Grund von Glukosemangel zur Energiebereitstellung genutzt werden. Kreatinkinase tritt aus den Muskelzellen aus, wenn die muskuläre Belastung durch zu intensives Training oder fehlende Regeneration dazu führt, dass Muskelzellen zerstört werden oder die Energieversorgung nicht ausreicht. Diese beiden Parameter können insbesondere in Trainingslagern hilfreich sein, um die Sportler*innen trotz der mehrmals täglichen hohen Belastungen nicht zu überlasten und gezielte Ruhepausen zu steuern (Hottenrott und Neumann, 2016, S. 327ff).

Trotz ihrer validen Eignung für diverse Anwendungsbereiche weisen allen drei Messgrößen spezifische Besonderheiten auf, die im Folgenden betrachtet werden.

Besonderheiten der Laktatbestimmung Bei der Bestimmung Blutlaktatkonzentration ist die Einhaltung von Hygienemaßnahmen deutlich wichtiger, als bei den anderen beiden Messgrößen. Während die Hersteller i.d.R. eine Messgenauigkeit von ca. 1,5% angeben, können Fehler bei der Messung insbesondere bei den deutlich praktikableren Blutgasanalytoren mit integrierter Laktatmessung je nach Hersteller zu deutlich höheren Abweichungen (rund 10-25%) und damit zu einer Fehleinschätzung der Leistung führen (Medbø et al., 2000). Messungenauigkeiten können dabei insbesondere durch Verunreinigung mit Schweiß entstehen oder durch Verdünnung der Probe mit

Gewebeflüssigkeiten bei zu starkem »Herausmassieren« des Blutes bei unzureichendem Blutfluss. Weitere typische Fehlerquellen sind eine zu langsame Befüllung oder eine Verunreinigung (beispielsweise durch Anfassen) der Messkammern und eine falsche Lagerung der Messstreifen (Röcker und Dickhuth, 2001). Die Vielzahl der möglichen Fehlerquellen zeigt bereits, dass die Laktatmessung sich für die Leistungsdiagnostik auf dem Fahrrad-Ergometer deutlich besser eignet, als für Feldtests, bei denen die Person zur Probenentnahme die Fahrt komplett unterbrechen muss und die Bedingungen weniger standardisiert sind. Weiterhin ist zu beachten, dass stets dieselbe Methodik bei der Laktatbestimmung genutzt werden sollte, da Messungen an Finger und Ohr unterschiedliche Ergebnisse aufweisen und im Sinne einer Leistungskontrolle nicht vergleichbar sind (Feliu et al., 1999). Zusätzlich ist die Messung mit regelmäßigen Kosten für die Laktatstreifen verbunden und die Messwerte liegen stets nur zu einigen wenigen diskreten Zeitpunkten vor, an denen der Laktatspiegel bestimmt wird. Dies sorgt dafür, dass die Blutlaktatkonzentration zwar im professionellen Bereich eine valide Methode zur Leistungsdiagnostik darstellt, für den (ambitionierten) Hobby-Bereich oder eine (kontinuierliche) Modellierung auch kurzfristiger Trainingseffekte nur bedingt geeignet ist.

Besonderheiten der Spirometrie Bei der Bestimmung Atemgaskonzentration ist insbesondere im Vorfeld einer jeden diagnostischen Untersuchung das Material der Spirometrie zu überprüfen. Da sowohl sämtliche Schläuche als auch die Atemmaske wiederverwendet werden, ist hier auf eine besondere Hygiene zu achten, um eine Keimbildung zu verhindern. Scharhag-Rosenberger (2010) haben zusammengefasst, welche Maßnahmen vor und während der spirometrischen Untersuchung durchgeführt werden sollten, um eine gute Datenqualität zu gewährleisten:

- Durchführung einer regelmäßigen Kalibrierung
- je nach Spirometer benötigt das Gerät eine gewisse Warmlaufzeit, i.d.R. etwa 30 min
- bei der Verwendung von Umgebungsluft ist der Raum vor der Kalibrierung des Spirometers zu lüften
- auf eine Passgenauigkeit der Maske ist zu achten, sodass an den Rändern keine Luft entweichen kann
- die Testperson sollte während der Diagnostik nicht sprechen
- während der Ruhemessung und vor Beginn der Belastung sollten einige Parameter auf Plausibilität überprüft werden, insbesondere die Ventilation, die Sauerstoffaufnahme und der Respiratorische Quotient
- regelmäßige Prüfung der Verbrauchsmaterialien auf Verschleiß und Ablagerungen

Zudem empfehlen Scharhag-Rosenberger (2010) eine biologische Qualitätskontrolle in regelmäßigen (etwa monatlichen) Abständen, beispielsweise durch eine spezifische Testperson und ein standardisiertes Belastungsprotokoll. Entsprechend vielfältig sind die möglichen Fehlerquellen – so kann beispielsweise Feuchtigkeit in den Schlauchverbindungen zu einer fehlerhaft niedrigen Messung der Sauerstoffaufnahme ($\dot{V}O_2$) führen (Bischoff, 2002); undichte Stellen (beispielsweise durch Verschleiß oder fehlerhafte Anpassung der Maske) im System haben eine Verunreinigung mit der umgebenden Raumluft und damit eine Veränderung der gemessenen Atemgaskonzentrationen zur Folge; und das Sprechen der Testperson oder auch eine auftretende Hyperventilation kann die Ergebnisse der respiratorischen Kompensation verfälschen. Wird die $\dot{V}O_{2\max}$ mittels Fickscher Formel bestimmt, schließen sich weitere potentielle Fehlerquellen durch die Vielzahl der Parameter (Herzzeitvolumen, Sauerstoffpartialdruck, Sauerstoffsättigung, Hämoglobinkonzentration) an (Bischoff, 2002). Ähnlich wie bei der Laktatdiagnostik sorgt dies dafür, dass auch die Atemgasanalyse nur im professionellen Bereich eine geeignete Methode zur Leistungsdiagnostik darstellt, für den (ambitionierten) Hobby-Bereich oder eine (kontinuierliche) Modellierung auch kurzfristiger Trainingseffekte jedoch durch den hohen Aufwand und die Vielzahl möglicher Fehlerquellen nur bedingt geeignet ist.

Besonderheiten der Herzfrequenzmessung Die Bestimmung Herzfrequenz ist im Vergleich deutlich weniger aufwändig. Grundsätzlich stehen zur Messung sechs verschiedene Sensortechniken mit spezifischen Genauigkeiten und Fehlerquellen zur Verfügung. Die verbreitetste Methode, die auch im sportlichen Kontext sinnvoll eingesetzt werden kann, ist die elektrographische Messung. Andere Messmethoden werden aufgrund ihrer oft unzureichenden Genauigkeit oder ihrer im sportlichen Kontext eingeschränkten Nutzungsmöglichkeit hier nicht betrachtet – eine Übersicht ist bei Ludwig et al. (2018) zu finden. Die elektrographische Messung, wie sie in Elektrokardiographiegeräten (EKGs) Verwendung findet, ist mit 12 Elektroden die bis heute exakteste (nicht invasive) Methode, um die Herzfrequenz zu bestimmen. Fehler in der Messung können durch elektromagnetische Wellen in der Umgebung sowie Bewegungen der Testperson oder eine inkorrekte Platzierung der Elektroden entstehen. Ähnlich wie beim Spirometer ist auch beim EKG auf Verschleiß, Hygiene insbesondere im Bereich der Elektroden und die korrekte und sichere Platzierung der Elektroden zu achten. Die genauesten Resultate werden dabei in Ruhe erzielt. Häufig praktikabler im Bereich der Ausdauerdiagnostik sind Brustgurte, die ebenfalls mit elektrographischen Sensoren ausgestattet sind. Ihre Genauigkeit weist je nach verwendetem Modell im Vergleich zum 12-Kanal-EKG einen r-Wert von 0,85 – 0,99 auf (Weippert et al., 2010). Typische Fehlerquellen sind auch hier ein inkorrekt platzierter Brustgurt und damit der Elektroden, Bewegungsartefakte und andere elektromagnetische Wellen in der Umgebung. Zusätzlich können Fehlmessungen auftreten, wenn der Brustgurt die

kabellose Verbindung zu dem aufzeichnenden Gerät (PC, Smartphone oder Wearable) verliert. Hottenrott und Neumann (2016) betonen in dem Kontext, dass »die Herzfrequenz (HF) bei Belastungen (...) im Sport seit Langem ein zuverlässiger Indikator für die Beurteilung der Trainingswirkung, der Ermüdung oder des Gesundheitszustandes [ist]. (...) Seit Einführung exakt arbeitender HF-Messgeräte (»Pulsuhren«) ist bekannt, dass die HF zu den empfindlichsten Messgrößen gehört, welche die Wirkung des Trainings oder die Veränderung der Belastung (Geschwindigkeit, Streckenprofil) erfassbar macht.« Auch sei inzwischen »gesichert, dass die trainingsbedingten Veränderungen der HF-Regulation bereits nach wenigen Tagen nachweisbar sind«. Heutzutage lässt sich die Herzfrequenz leicht und flexibel auch im alltäglichen Gebrauch messen und bietet auch im Hobbysport-Bereich ohne das Hinzuziehen professioneller Trainer*innen⁶ die Möglichkeit, Trainingsentwicklungen sowohl kurzfristig (akute Belastung) als auch langfristig zu messen und nachzuerfolgen. Dadurch, dass die Herzfrequenz selbst allerdings eine sehr empfindliche Messgröße ist, unterliegt sie einer Vielzahl von Einflüssen, die in Unterabschnitt 1.3.1: »Die Herzfrequenz als Indikator: Kritische Aspekte« (S. 56) genauer betrachtet werden.

Im Kontext der vorliegenden Arbeit wurde auf Basis der Praktikabilität bei grundsätzlich validen Messergebnissen der Fokus auf die Herzfrequenz gelegt. Insbesondere im Bereich der Analyse längerfristiger Leistungsentwicklungen werden die verwendeten Herzfrequenzmessungen nach Möglichkeit geeignet durch Laktatwerte, die $\dot{V}O_{2\max}$ oder vergleichbare diagnostische Messgrößen ergänzt.

Im Folgenden werden nun die gängigsten Schwellenkonzepte im Ausdauersport sowie typische Testverfahren betrachtet, bevor es in Abschnitt 1.3 (»Herzfrequenzverhalten bei körperlicher Belastung«) detaillierter um die Herzfrequenz als Messgröße geht.

1.2.4. Schwellenkonzepte

Die Bestimmung einer ventilatorischen oder metabolischen Schwelle dient der Differenzierung der unterschiedlichen Phasen der Energiebereitstellung. Während die $\dot{V}O_{2\max}$ den zuverlässigsten Parameter zur Bestimmung der individuellen Leistungsfähigkeit darstellt (Åstrand et al. (1986); Hollman und Hettinger (2000) zitiert nach Hottenrott und Neumann (2016)), ist ihre exakte Bestimmung nach dem Fickschen Prinzip mit einem hohen Aufwand verbunden. Nicht zuletzt, da ihre Bestimmung eine maximale Ausbelastung unabdingbar macht. Als Indikator für den aktuellen Trainingszustand muss die $\dot{V}O_{2\max}$ ins Verhältnis zu einer der Dauerleistungsgrenzen wie der anaeroben Schwelle gesetzt werden: Für sich alleine gibt die $\dot{V}O_{2\max}$ zunächst nur Auskunft darüber, welche Leistung eine Person zum aktuellen Zeitpunkt theoretisch maximal erbringen könnte. Sie macht damit jedoch keine Aussage

⁶Dies bezieht sich nur auf die einfache Erhebung der Daten selbst. Für eine geeignete Interpretation und die individuelle Trainingsplanung ist auch hier eine professionelle Beratung und Unterstützung anzuraten.

darüber, in wie weit die mögliche Leistung im Training auch ausgeschöpft wird. Um also die tatsächlich erreichte Leistung abschätzen zu können, muss die $\dot{V}O_{2\max}$ ins Verhältnis mit der Laktatkonzentration gesetzt werden – je später die Laktatkonzentration steigt, desto höher ist der Anteil der tatsächlich genutzten $\dot{V}O_{2\max}$. Je nach Sportart und Trainingszustand beträgt die anaerobe Schwelle ca. 60-85% der $\dot{V}O_{2\max}$, die aerobe Schwelle ca. 40-65% der $\dot{V}O_{2\max}$ (Kindermann, 2004).

Je nach individuellem Leistungszustand unterscheidet sich die maximal mögliche Belastung je Energiebereitstellungsphase stark. Wie in Unterabschnitt 1.1.3 (»Physiologische Reaktionen und Anpassungen durch Ausdauerbelastung«) dargestellt, adaptiert sich der Organismus auf vielfältige Weise an eine vorherrschende Belastung – schon durch eine Vergrößerung der beanspruchten Muskulatur und einer verbesserten Herzleistung kann die aerobe Energiebereitstellung länger aufrecht erhalten werden, sodass die entsprechende Schwellenleistung Aufschluss über den Trainingszustand im Vergleich zu anderen Personen und insbesondere auch im zeitlichen Verlauf gibt. Dadurch sind die Schwellenkonzepte ein wertvolles Hilfsmittel, um das Training zu steuern, die Zielsetzungen des Trainings und die Trainingserfolge zu kontrollieren und ggf. weiter zu optimieren.

Um diese unterschiedlichen Phasen voneinander unterscheiden zu können und so gezieltere Aussagen über die stattfindende Belastungsintensität und damit den Trainingsbereich treffen zu können, werden die Übergänge zwischen der aeroben, der aerob-anaeroben und dem Beginn der anaeroben Phase analysiert. Eine solche Abgrenzung lässt sich während der Durchführung eines Stufentests metabolisch über die Verschiebung der Blutlaktatkonzentration und ventilatorisch über die Veränderung der Atemgaszusammensetzung vornehmen.

Hottenrott und Neumann (2016) nennen die folgenden fünf Schwellenkonzepte, die häufig in der Diagnostik verwendet werden:

Aerobe ventilatorische Schwelle (kurz: AT / VT1) Die AT bezeichnet den ersten Anstieg nicht-metabolischer CO_2 -Produktion, der mittels Spirometrie gemessen werden kann: An der AT fällt vermehrt Kohlendioxid aus der Laktatpufferung an, sodass die Menge der eingeatmeten Luft pro Minute (Ventilation) und die Menge der Kohlendioxidabgabe jeweils im Verhältnis zur Sauerstoffaufnahme ansteigt. Die AT kann über die Veränderung »des Ansteigungsverhaltens des Atemminutenvolumens bestimmt« werden, da die durch die Laktatpufferung entstehende zusätzliche CO_2 -Produktion zur Hyperventilation anregt.

Anaerobe ventilatorische Schwelle (kurz: RCP / VT2) Die VT2 bezeichnet einen überproportionalen Anstieg der Ventilation, der mittels Spirometrie gemessen werden kann. Er entsteht durch vermehrte metabolische Azidose: Oberhalb des »respiratorischen Kompensationspunktes« (oberhalb der aeroben Schwelle) ist eine Pufferung des entstehenden H^+ nicht mehr ausreichend möglich, sodass es zur Übersäuerung (Azido-

se) kommt. Die damit einhergehende Steigerung der Atemfrequenz führt erneut zu einem überproportional starken Anstieg des Atemminutenvolumens.

Individuelle aerobe Laktatschwelle (kurz: LT / LT1) Der LT entspricht dem minimalen Laktat-Leistungs- oder Laktat-Geschwindigkeitsäquivalent im Stufentest. Die aerobe Laktatschwelle wird häufig als der erste Laktatanstieg über den Basis-Laktatwert betrachtet.

Individuelle anaerobe Laktatschwelle (kurz: IAT / IAS / LT2) Die IAT entspricht der Dauerleistungsgrenze. Hierbei wird die maximal kompensierbare Laktatbildung erreicht (oder auch überschritten). Die Bestimmung der IAT ist eines der am weitesten verbreiteten Verfahren. Ausgehend von der individuellen aeroben Laktatschwelle wird für die individuelle anaerobe Laktatschwelle ein Wert von 1,0-1,5 mmol/ℓ auf die LT1 addiert (Hottenrott und Neumann, 2016, S. 312). Je nach Trainingszustand kann die IAT zwischen Werten unterhalb von 2,5 mmol/ℓ und oberhalb von 4 mmol/ℓ Laktat variieren.

Fixe Laktatschwelle (z.B. La2, La3, La4, La6) Die fixe Laktatschwelle stellt einen konkreten Bezug zwischen einer Laktatkonzentration (z.B. 2 mmol/ℓ bei La2) und einem Leistungs- oder Geschwindigkeitswert her.

Bis zu einer bestimmten Belastung kann der Körper eine höhere Menge an Laktat verstoffwechseln. In diesem Fall ist der Laktatspiegel im Blut erhöht, aber konstant. Diese Situation wird als *lactate-steady-state* oder *aerobic-anaerobic threshold* bezeichnet. Die maximale Belastung, die den Laktat-Stationärzustand aufrechterhält, ist die *anaerobe Schwelle*, die auch als *Laktatschwelle* bezeichnet wird. Die höchste Laktatmenge, die zu diesem Zeitpunkt für eine Weile gehalten werden kann, wird als *maximal lactate-steady-state (MLSS)* bezeichnet. Sobald diese Menge überschritten wird, kann sowohl die Pyruvatmenge in den Muskelzellen als auch das Laktat im Blutstrom nicht mehr schnell genug abgebaut werden. Die Muskelzellen geraten in eine Übersäuerung und der Organismus erzwingt drastisch und unmittelbar eine deutliche Reduktion der Belastung. Die Leistung bricht ab und der Sportler fühlt sich erschöpft.

Im Gegensatz zur anaeroben Schwelle wird die *aerobe Schwelle* oft als die niedrigste Belastung definiert, bei der der Laktatspiegel ansteigt und die Muskelzellen ihre Energie nicht nur durch aerobe Prozesse über einen längeren Zeitraum gewinnen können. Diese Schwelle wird häufig mit 2 mmol/ℓ angegeben, während die anaerobe Schwelle mit etwa 4 mmol/ℓ angenommen wird. Die 4 mmol/ℓ-Schwelle wurde als erste fixe anaerobe Schwelle von Mader et al. (1976) definiert. Insbesondere bei Leistungssportler*innen entspricht La4 jedoch nicht den Anforderungen der anaeroben Laktatschwelle. Generell handelt es sich bei fixen Laktatschwellen um verallgemeinerte Näherungswerte, die nur mit Vorsicht und eher als Ausgangspunkt verwendet werden sollten, wenn keine individuelle Schwelle bekannt ist. In der Regel sind sie stark von verschiedenen Faktoren wie der aktuellen körperlichen Verfas-

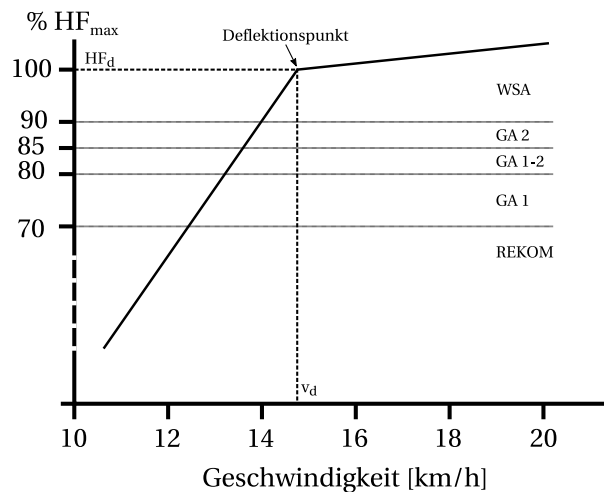


Abbildung 1.12.: Geschwindigkeiten beim Langstreckenlauf mit prozentual vom Deflexionspunkt abgeleiteten Herzfrequenzbereichen für die unterschiedlichen Belastungsbereiche: wettkampfspezifische Ausdauer (WSA), Grundlagenausdauer (GA) sowie Regeneration und Kompensation (REKOM). HF_d bezeichnet die Herzfrequenz am Deflexionspunkt, v_d die Geschwindigkeit am Deflexionspunkt (übernommen aus Hottenrott und Neumann (2016))

sung, der gewählten Sportart und genetischen Veranlagungen abhängig (Tomasits und Haber, 2011). Besonders problematisch bei fixen Laktatschwellen ist jedoch die Varianz der Laktatkonzentration, die von zahlreichen Faktoren abhängt. Dieser Nachteil kann durch die Nutzung individueller Schwellenkonzepte behoben werden.

Ein weiteres Schwellenkonzept, bei dem auf eine Blutabnahme verzichtet werden kann, ist die Bestimmung des Herzfrequenz-Deflexionspunktes wie in Abbildung 1.12 dargestellt. Der Herzfrequenz-Deflexionspunkt nach Conconi et al. (1982) soll ähnlich wie die IAT eine anaerobe Schwellenleistung markieren. Diese Schwelle beruht auf der Beobachtung, dass die Herzfrequenz bei (linear) steigender Belastung zunächst ebenfalls linear ansteigt, die HF-Kurve jedoch bei etwa 80% der maximalen Leistung abflacht. Zur Bestimmung des Deflexionspunktes wird ein spezielles Testprotokoll absolviert, bei dem dieser »Knick« in der Herzfrequenzkurve möglichst deutlich erkennbar ist. Nachteile dieser Methode sind, dass der Deflexionspunkt nicht bei jeder Testdurchführung ermittelt werden kann und dass sich teilweise große Unterschiede zu laktatbasierten Schwellenkonzepten ergeben, sodass die Validität nicht gesichert ist (Bodner und Rhodes, 2000; Jones und Doust, 1995).

Insgesamt existieren heutzutage über 20 verschiedene Konzepte der Laktatschwelle; Dörr (2010) hat in einer umfangreichen Literaturrecherche 27 unterschiedliche Konzepte ausfindig gemacht. Nach Allen und Coggan (2012) gilt die Betrachtung der Laktatschwelle als wichtigster physiologischer Faktor zur Bestimmung der Leistung. Besonders verbreitet sind dabei wie bereits in obiger Übersicht dargestellt vor allem die individuellen Laktatschwellen und fixe Laktatschwellen. Die »4-mmol/l-Schwelle« (La4) geht ursprünglich auf Mader et al. (1976) zurück und wurde in den vergangenen Jahrzehnten umfangreich validiert (Dörr,

2010). Als ein Vorteil der Betrachtung der Laktatkonzentration gegenüber der Bewertung mittels $\dot{V}O_2$, $\dot{V}O_{2\max}$ oder des Herzvolumens gilt die empfindlichere Reaktion auf Änderungen der Ausdauerleistungsfähigkeit und die Unabhängigkeit von beispielsweise Motivation oder Muskelkraft.

Die Betrachtung der Leistung bei einem Anstieg der Laktatkonzentration um $1,5 \text{ mmol/l}$ vom Basalwert wurde von Dickhuth et al. (1991) als eine Möglichkeit zur Bestimmung einer individuellen, anaeroben Laktatschwelle vorgestellt. Für eine umfangreiche Übersicht unterschiedlicher Laktatschwellenkonzepte und ihrer Validierung sei auf die Arbeit von Dörr (2010) verwiesen.

1.2.5. Testverfahren im Ausdauersport (Radsport)

In der Ausdauersport-Diagnostik können zur Bestimmung einer ventilatorischen oder metabolischen Schwelle und somit zur Bestimmung der Leistungsfähigkeit unterschiedliche Tests zum Einsatz kommen. Dabei wird zwischen Feld- und Labortests unterschieden, wobei auch im Labor je nach Möglichkeit und Zielsetzung eine sportartenspezifische Leistung gemessen wird. Im Labor können biologische Messgrößen allerdings deutlich einfacher und durch Standardisierung mit besserer Vergleichbarkeit erhoben werden. Außerdem muss zwischen maximalen und submaximalen Leistungstest unterschieden werden. Während Tests mit maximaler Ausbelastung im Leistungssport verbreitet sind, ist dabei zu beachten, dass die starke körperliche Belastung mit Trainingsplänen oder einer Wettkampfvorbereitung kollidieren kann, sodass je nach Situation auch im Leistungssport auf submaximale Leistungstests zurückgegriffen wird (Lamberts et al., 2011). Übliche Tests in der submaximalen Diagnostik sind Lamberts et al. (2011) zufolge der Cooper-Test, der PWC 170 und der Test von Åstrand. Ergänzend haben Lamberts et al. (2011) selbst einen submaximalen Test vorgestellt, der sich die aktuellen technologischen Fortschritte im Bereich des Herzfrequenz-Monitorings zunutze macht und die *Peak Power* basierend auf submaximaler Leistung, der *Heart Rate Reserve* als einem speziellen Parameter der Herzfrequenz und dem RPE bestimmt. Die *Peak Power* ist die größtmögliche Leistung, die während der ersten 5 s eines Trainings erreicht werden kann. Sie ist ein Indikator für die Energiebereitstellungsfähigkeit von direkt verfügbarem ATP und CrP im Muskel. Die *Peak Power* (PP) ist definiert als: $PP = \frac{\text{Kraft} \cdot \text{Distanz}}{\text{Zeit in Minuten}}$. Die *Heart Rate Reserve* ist definiert als Verhältnis zwischen der Differenz aus der durchschnittlichen HF während des Trainings und der Ruhe-HF sowie der Differenz aus maximaler HF und Ruhe-HF: $\frac{HF_{\text{exercise}} - HF_{\text{rest}}}{HF_{\text{max}} - HF_{\text{rest}}}$. Häufig verwendete Testverfahren mit maximaler (Aus-)Belastung sind Stufentests, Rampentests und Time Trials. Besonders der Laktatstufentest ist eine der Standardmethoden (Clénil, 2019).

Etwas allgemeiner beschreiben Paton und Hopkins (2001) die folgenden sechs unterschiedliche Herangehensweisen für Testprotokolle für den Radsport:

Konstante Arbeit Bei dieser Testart wird eine vordefinierte Arbeit in kürzest möglicher Zeit verrichtet. Alternativ kann eine vordefinierte Strecke in kürzestmöglicher Zeit zurückgelegt werden. Ein typisches Beispiel für diese Art von Test sind *Time Trials*.

Konstante Zeit Bei dieser Testart wird für eine vordefinierte Dauer so viel Arbeit wie möglich verrichtet. Alternativ kann in der vorgegebenen Dauer so viel Strecke wie möglich zurückgelegt werden.

Konstante Leistung Bei dieser Testart wird eine vorgegebene Leistung bis zur Ausbelastung aufrechterhalten.

Inkrementell: Peak Power oder $\dot{V}O_{2\max}$ Bei dieser Testart wird die Leistung bis zur maximal möglichen Belastung schrittweise zügig erhöht. Eine Testdauer von etwa 8–12 Minuten wird dafür empfohlen. Für die Peak Power sind Sprinttests ein typisches Beispiel. Für die Bestimmung der $\dot{V}O_{2\max}$ können stattdessen Rampentests mit einer Steigerung alle 30 s oder Stufenprotokolle mit einer Steigerung alle 3–4 min verwendet werden. Es wird eine Dauer von 7–26 Minuten bis zur Ausbelastung empfohlen (Scharhag-Rosenberger und Schommer, 2013).

Inkrementell: anaerobe Schwelle Bei dieser Testart wird die Belastung für eine vorgegebene Dauer im Wechsel konstant gehalten und schrittweise erhöht. Ein solcher Test findet besonders zur Bestimmung der anaeroben Schwelle mittels Blutlaktatkonzentration oder Atemgasanalyse Verwendung. Ein typisches Beispiel für diese Art von Test sind Stufentests.

Critical Power (CP) Bei dieser Testart wird eine konstante Arbeit verrichtet, eine konstante Zeit gefahren oder eine konstante Leistung erbracht. Dabei werden jeweils unterschiedliche Intensitäten eingestellt. Die jeweilige Dauer und die mittlere Arbeit eines jeden Tests werden anschließend für die Berechnung der CP verwendet. Auf diese Weise werden die anaerobe Leistungsgrenze und die maximale aerobe Leistung geschätzt.

Für eine umfangreiche Übersicht unterschiedlicher konkreter Belastungstests, zu ihrer Durchführung und ihrem jeweiligen Anwendungszweck sei auf die Arbeit von Mackenzie (2005) verwiesen, in der insgesamt 101 Leistungstests für unterschiedliche Belastungsarten (Ausdauer, Mobilität, Kraft und andere) präsentiert werden.

Im Kontext dieser Arbeit werden Studien analysiert, in denen die anaerobe Schwellenleistung im Vordergrund steht, sodass für die Betrachtung der Leistungsmodellierung ausschließlich Stufentests verwendet wurden (Kapitel 7). Zur Analyse der Herzfrequenz-Vorhersage wurden Tests mit unterschiedlichen Herzfrequenz-Dynamiken betrachtet, sodass in diesem Kontext neben Stufentests auch Rampentests, Sprints und eine künstliche Outdoor-Einheit sowie reguläre »Outdoor«-Trainingseinheiten betrachtet werden (Kapitel 5 und Kapitel 6).

1.3. Herzfrequenzverhalten bei körperlicher Belastung

Wie bereits erwähnt gilt nach Hottenrott und Neumann (2016) »die Herzfrequenz (HF) bei Belastungen (...) im Sport seit Langem [als] ein zuverlässiger Indikator für die Beurteilung der Trainingswirkung, der Ermüdung oder des Gesundheitszustandes«. Durch Training bedingte Veränderungen der HF seien bereits nach wenigen Tagen erkennbar und auch während eines Trainings reagiert die HF auf die akute Beanspruchung.

Nach einer kurzen Rekapitulation der akuten und chronischen Reaktionen der Herzfrequenz auf körperliche Belastung werden kritische Aspekte betrachtet, die bei der Nutzung der HF im Kontext von Trainingsmonitoring und Trainingsplanung beachtet werden sollten. Ein Überblick über die Gemeinsamkeiten akuter und chronischer Reaktionen schließt das Kapitel ab.

Bei einer akuten Belastung reagiert die HF mit als erste auf die körperliche Beanspruchung. Åstrand et al. (1986) zufolge findet diese Reaktion der HF in zwei Phasen statt: einer »onset«-Phase und einer »steady state«-Phase der HF. Bunc et al. (1988) bezeichnet die Reaktion bei submaximaler Belastung als schnelle und langsame Phase der HF, wobei die schnelle Phase dem »onset« und die langsame Phase dem »steady state« entspricht: Während einer submaximalen Belastung ist das Verhältnis zwischen Herzfrequenz und Belastungsintensität in der schnellen Phase linear. Ab einer spezifischen Intensität steigt die HF im Verhältnis zur Belastung langsamer an (langsame Phase). Der Punkt zwischen den beiden Phasen entspricht Conconi et al. (1982) und Bunc et al. (1988) zufolge der anaeroben Schwelle und wird von Conconi et al. (1982) als Deflexionspunkt beschrieben (vgl. Unterabschnitt 1.2.4, »Schwellenkonzepte«). In ihrer Studie überprüften und bestätigten Bunc et al. (1988) die beiden Thesen, dass 1. die HF-Reaktion mit ihren zwei unterschiedlichen Phasen genutzt werden kann, um den Trainingszustand einer Person zu beschreiben und dass 2. der Deflexionspunkt bei inkrementell steigender Belastung mit der ventilatorischen Schwelle korreliert.

Langfristig betrachtet wird die Herzleistung durch Ausdauertraining deutlich verbessert, wie bereits in Unterabschnitt 1.1.3 (»Physiologische Reaktionen und Anpassungen durch Ausdauerbelastung«) dargestellt. Durch regelmäßiges Ausdauertraining werden die Herzinräume erweitert und der Herzmuskel gekräftigt. Dadurch kann eine größere Menge Blut pro Minute durch das Herz gepumpt werden und das Schlagvolumen nimmt zu. Insgesamt kann das Schlagvolumen durch regelmäßiges Training nahezu verdoppelt werden, sodass das Herz eines langjährigen Ausdauerathleten 19 ml/kg Körpergewicht Blut pro Minute pumpen kann (bzw. 18 ml/kg bei Ausdauerathletinnen) (Hottenrott und Neumann, 2016). Bei einer untrainierten männlichen Person pumpt das Herz rund 13 ml/kg Körpergewicht Blut pro Minute durch den Körper, bei weiblichen Personen sind es rund 12 ml/kg . Infolge des höheren Schlagvolumens sinkt die HF in Ruhe und auch unter Belastung – die gesamte Herzfunktion wird ökonomisiert und die Durchblutung insgesamt verbessert.

1.3.1. Die Herzfrequenz als Indikator: Kritische Aspekte

Die exklusive Messung der HF als physiologische Reaktion des Körpers auf Bewegung bzw. Belastung findet in unterschiedlichen Bereichen Anwendung. Zum Beispiel wird die HF genutzt,

- um die Erschöpfung oder den Grad der Müdigkeit einer Person zu schätzen (She et al., 2013; Vautier et al., 1994);
- um die individuelle Herz-Kreislauf-Funktion anzuzeigen (Borresen und Lambert, 2008; Carter et al., 2003);
- um Trainingsparameter wie die Intensität oder Trainingsdauer von Einzelpersonen oder ganzen Gruppen zu überwachen (Lee et al., 2015; Sornanathan und Khalil, 2010);
- oder schlicht zur Steuerung des individuellen Trainings (Hunt und Hunt, 2016; Weghorn, 2013).

Ein wesentlicher Grund bei der Nutzung der HF zur Trainingssteuerung und -überwachung ist sicherlich die verhältnismäßig einfache und kostengünstige Messung auch während einer Belastung. Trotz oder gerade wegen der breiten Verwendung der HF als Beanspruchungsparameter müssen auch kritische Aspekte berücksichtigt werden.

So stellen mehrere neuere Studien dar, dass die Genauigkeit und Präzision der HF-Messung stark von dem verwendeten tragbaren Gerät, der genutzten Technik der HF-Messung und der durchgeführten Aktivität abhängt: El-Amrawy und Nounou (2015) verglichen insgesamt neun *Smartwatches* und acht Fitness-Tracker – die Genauigkeit der HF-Messung (im Vergleich zum EKG-Referenzpulssignal) lag zwischen 92,8% und 99,9% je nach Gerät und die Präzision zwischen 5,9% und 20,6%. Weghorn (2016) fand heraus, dass die Verwendung eines optischen Reflexionssystems (Puls-Plethysmographie, PPG) anstelle eines Brustgurtsensors immer dann versagt, wenn Bewegung und schnellere Impulsschwankungen auftreten; beispielsweise fanden sie Messungen von 118 bpm, während die EKG-Referenzmessung bei 65 bpm lag. Ähnliche Ergebnisse wurden von Gillinov et al. (2017) dargestellt, während Parak und Korhonen (2014) zwei PPG-basierte HR-Monitore evaluierte, wobei die HF-Messung in einem 10 bpm-Intervall in etwa 87% der Zeit im Vergleich zum EKG-Referenzpuls lag.

Auch unabhängig von möglichen Fehlern bei der Messung unterliegt die HF unterschiedlichen Schwankungen und äußeren Einflüssen. Selbst unter kontrollierten und standardisierten Bedingungen kommt es zu Schwankungen von Tag zu Tag – sowohl in Bezug auf die maximale HF, als auch bezüglich durchschnittlicher HF-Werte in standardisierten Belastungstests (Becque et al., 1993; Brisswalter und Legros, 1994; Lamberts und Lambert, 2009). In ihrem Review-Bitrag zum Herzfrequenz-Monitoring haben Achten und Jeukendrup (2003) eine Reihe von Studien zu unterschiedlichen Einflüssen auf die Herzfrequenz zusammengetragen. Dabei unterscheiden sie zwischen allgemeinen Faktoren, die während

des Trainings auftreten (Varianz der HF von Tag zu Tag), physiologischen Einflüssen (Kardiovaskulärer Drift, Flüssigkeitshaushalt) und Umwelteinflüssen (Umgebungstemperatur, Höhenlage). Zusätzlich wird die Herzfrequenz durch die respiratorische Sinusarythmie von der Atmung beeinflusst (Hirsch und Bishop, 1981) und reagiert auch auf Stress (Jeukendrup und Diemen, 1998; Lambert et al., 1998; Taelman et al., 2009), hormonelle Schwankungen (Leicht et al., 2003; Vargas-Uricoechea et al., 2014) und allgemein Veränderungen im Gesundheitszustand (Buchheit et al., 2013; Buchman et al., 2002; Griffin et al., 2003; Hartmann et al., 2019; Vargas-Uricoechea et al., 2014). Insbesondere den Kardiovaskulären Drift hebt auch Gilman (1996) als Faktor hervor, der besonders zu berücksichtigen sei. Speziell beim Radsport kommt hinzu, dass auch die Sitzposition auf dem Rad die Herzfrequenz beeinflussen kann (Gnehm et al., 1997), das Fahren im Windschatten zu einer Abnahme der HF um 20 – 30 % führt (Hottenrott und Neumann, 2016) und umgekehrt eine höhere Trittfrequenz bei gleicher Leistung eine höhere HF zur Folge hat (Hottenrott und Neumann, 2016).

Während die HF bei submaximaler Belastung unter Laborbedingungen oft stark mit $\dot{V}O_2$ korreliert und genutzt werden kann, um $\dot{V}O_2$ und Laktat zu schätzen (Achten und Jeukendrup, 2003), haben Lamberts und Lambert (2009) je nach Belastungsintensität die Varianz der HF im Vergleich unterschiedlicher Tage genauer untersucht. Die Testpersonen ihrer Studie wurden in vier Gruppen mit jeweils unterschiedlicher Trainingsintensität (Prozentual zur maximalen HF) eingeteilt. Dabei traten je nach untersuchter Gruppe und Messzeitpunkt Abweichungen der Erholungsherzfrequenz von über 6 bpm auf, für submaximale Herzfrequenzen betrug die Abweichung in einigen Gruppen im Mittel rund 3 bpm. Lamberts und Lambert (2009) kommen daher zu dem Schluss, dass diese Abweichungen im Rahmen einer kontrollierten Testumgebung als hoch bewertet werden müssen, da eine Abweichung von 3 – 6 bpm in der HF bereits als Indikator für ein Übertraining oder eine beginnende Erschöpfung interpretiert werden kann. Für Analysen dieser Art sei demnach Vorsicht in der Interpretation geboten. Lucía et al. (2000) haben in ihrer Studie hingegen festgestellt, dass die HF als Parameter zur Leistungskorrelation auch im Vergleich zur Laktatschwelle sowie den beiden ventilatorischen Schwellwerten über unterschiedliche Trainingsperioden (Erholungsperiode, vor der Wettkampfperiode, während der Wettkampfperiode) stabil bleibt und geringe Schwankungen von maximal ± 3 bpm aufweist. Obwohl die Schwankungsbreite der HF in einem ähnlichen Bereich liegt, bewerten sie die HF damit als stabilen Parameter mit geringer Varianz. Dies verdeutlicht, dass die Einordnung, welche Abweichungen der HF toleriert werden können, stark von der aktuellen Analyse abhängt und nicht verallgemeinert werden sollte. Während die Schwankungsbreite für Aussagen bezüglich eines Übertrainings bereits grenzwertig oder zu hoch sein kann, scheint die HF für allgemeinere Einschätzungen der Leistungsfähigkeit und im Vergleich zu anderen Messmethoden zur Bestimmung metabolischer oder respiratorischer Schwellen stabil genug zu sein.

1.3.2. Gemeinsamkeiten akuter und chronischer Reaktionen auf Ausdauerbelastung

Die akuten und chronischen Reaktionen eines Organismus auf eine Belastung können nicht unabhängig voneinander betrachtet werden. Es besteht eine beidseitige Wechselwirkung: Während sich chronische Anpassungsvorgänge auf die akute Reaktion auswirken, beeinflusst umgekehrt jedes Training mit seinen kurzfristigen Anpassungsreaktionen auch die langfristigen Trainingseffekte.

Die Herzfrequenz oder bestimmte Eigenschaften der Herzfrequenz werden seit einigen Jahren immer wieder auf ihre Zusammenhänge zu den unterschiedlichsten Leistungsparametern hin untersucht. In zahlreichen Studien (beispielsweise von Davis und Convertino (1975), Gilman und Wells (1993) und Lucía et al. (2000)) konnte ein Zusammenhang zwischen der HF und der Trainingsintensität nachgewiesen werden.

Allerdings sind derartige Zusammenhänge nicht in jedem Studiendesign nachweisbar. Auch bezogen auf die Ausdauerleistung legen die Analysen von ten Haaf et al. (2019) zunächst nahe, dass ein solcher Zusammenhang nicht immer gegeben ist: In ihrer Untersuchung von 30 männlichen Probanden wurden über einen Untersuchungszeitraum von etwa zwei Monaten inklusive eines Amateur-Radrennens signifikante Veränderungen der Herzfrequenz festgestellt. Bei der Betrachtung der $\dot{V}O_2$ und der ventilatorischen Schwelle als Leistungsindikatoren konnten hingegen keine signifikanten Veränderungen über den betrachteten Zeitraum beobachtet werden, sodass die submaximale Herzfrequenz nicht mit der Leistungsfähigkeit der Probanden korreliert werden konnte. Sie schlussfolgern, dass die HF nicht geeignet scheint, um die Leistung in (intensivem) Ausdauertraining zu kontrollieren oder vorherzusagen.

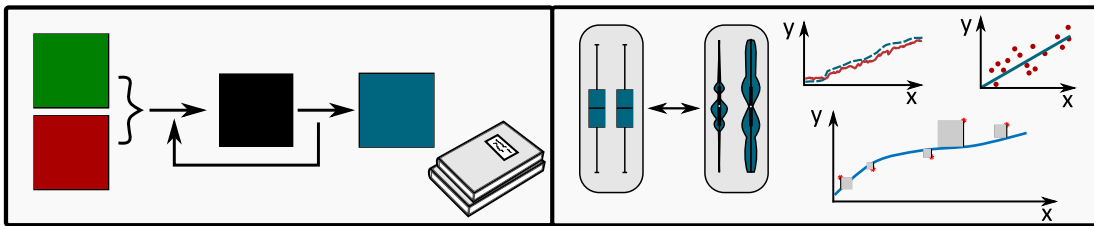
Umgekehrt haben Daanen et al. (2012) in ihrem Review-Artikel festgestellt, dass die Erholungsherzfrequenz [engl. *Heart Rate Recovery*] (HRR) mit Änderungen des Trainings- bzw. Leistungsstatus' korreliert. Bellenger et al. (2016) führten eine Meta-Analyse zur Evaluation der HF-Regulation als Indikator der Trainingsanpassung durch. Dabei haben sie eine Gesamtanzahl von 5377 Papern berücksichtigt. Sie kommen zu dem Schluss, dass eine Steigerung der Leistung insgesamt besonders gut über verschiedene Parameter der der HF (konkret: die Ruhe- und Post-Trainings-HRV, die Post-Trainings-HRR und die Herzfrequenzbeschleunigung) feststellbar sei. Eine trainingsinduzierte Ermüdung lasse sich ebenfalls gut über die Herzfrequenzbeschleunigung feststellen. Allein die Vielzahl der betrachteten Publikationen zeigt, dass Zusammenhänge gewisser Parameter der HF und der Leistung kein Einzelfall sind. Konkret haben auch Lucía et al. (2000) die Korrelation zwischen einer Ziel-HF und drei verschiedenen Leistungsmarkern untersucht: Sie konnten gute Korrelationen für die Laktat-Schwelle sowie die erste und zweite ventilatorische Schwelle feststellen und kommen zu dem Schluss, dass ein Test zu Beginn der Trainingsphase ausreicht, um die Trai-

ningsbelastung basierend auf HF-Daten zu definieren. In ihrem Fazit heben sie die HF als besonders geeignet zur Bestimmung der Trainingsintensität hervor:

»To date, HR is the most useful parameter for evaluating the level of intensity attained during training sessions and competition in professional cycling« (Lucía et al., 2000).

Wenngleich die Herzfrequenz sich nicht in jedem Szenario und für jede Art der Analyse als hinreichend zuverlässig erwiesen hat, verfügt sie in vielen Studien hinsichtlich spezifischer Fragestellungen über eine gute Stabilität und Zuverlässigkeit. Besonders ihre Sensitivität bei akuten wie auch bei chronischen Anpassungen macht sie zu einem geeigneten Primärparameter für die Fragestellungen dieser Arbeit.

Grundlagen zur Modellierung des Herzfrequenzverhaltens im Ausdauersport



Abstract.

Zur Analyse, Simulation und Vorhersage der Herzfrequenz- und Leistungsentwicklung beim Ausdauersport sind sowohl Modelle sowie mathematische und statistische Grundlagen notwendig. Während über die in der Modellierung genutzten Begrifflichkeiten Klarheit herrschen sollte, existieren gerade im Bereich der Analyse eine Vielzahl an möglichen Methoden, aus denen es stets gilt, die geeigneten Werkzeuge auszuwählen. Sowohl bei statistischen Verfahren als auch bei visuellen Darstellungen sind dabei zentrale Voraussetzungen wie auch die tatsächliche Eignung für eine vorliegende Fragestellung zu überprüfen.

In diesem Kapitel werden die zentralen Aspekte der modelltheoretischen und statistischen Grundlagen sowie geeignete Visualisierungsmethoden betrachtet und kritisch hinterfragt. Anschließend werden relevante Modellierungsaspekte für die Herzfrequenzmodellierung aufgezeigt und es werden typische Modelle aus der Literatur vorgestellt.

Die Begriffe *Modell* und *Modellierung* werden in unterschiedlichen wissenschaftlichen Disziplinen oft unterschiedlich genutzt, dienen aber dennoch stets einem ähnlichen Zweck: Es soll von einem Problem der realen Welt abstrahiert werden, um damit das Verständnis eines Phänomens zu erhöhen oder sich der Lösung eines Problems (zumindest teilweise) nähern zu können. Ortlieb et al. (2013) stellt die Modellierung als Prozess (Abbildung 2.1) dar, bei dem ein reales Problem zunächst in ein mathematisches Problem übersetzt wird. Für das mathematische Problem kann über Simulation und/oder Analysen eine Lösung eruiert werden, die ihrerseits wieder einer Rück-Übersetzung in eine reale Lösung benötigt. Die Rück-Übersetzung ist in der Regel mehr eine Interpretation und weniger eine Übersetzung im klassischen Sinne. Sie ist häufig mit unterschiedlichen Annahmen verbunden, die während der Analyse zur (mathematischen) Lösungsfindung getroffen werden mussten. Schließlich ist auch die aus der Interpretation entstandene reale Lösung erneut mit dem ursprünglichen Problem in Beziehung zu setzen. Es sollte stets nochmal geprüft werden, ob die vermeintliche Lösung tatsächlich eine realistische und plausible Lösung für das ursprüngliche Problem darstellt, oder ob die Modellierung ggf. angepasst und erneut betrachtet werden muss.

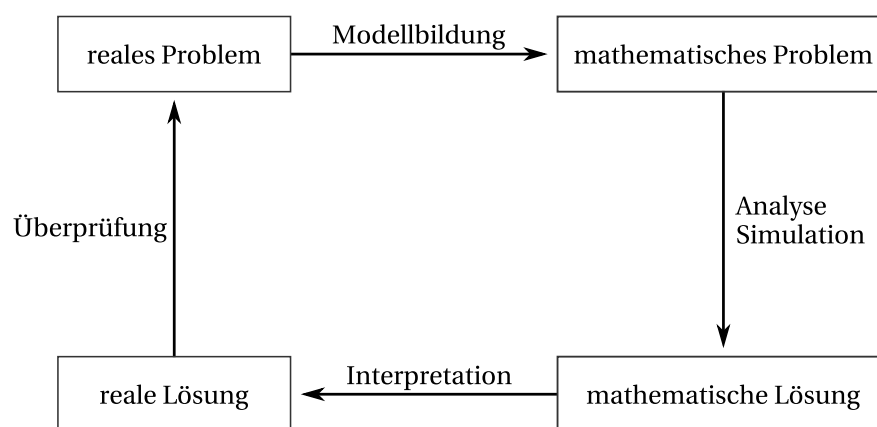


Abbildung 2.1.: Schematische Darstellung der Modellbildung, entnommen aus Ortlieb et al. (2013)

In der Modellierung selbst werden häufig unterschiedliche Begriffe genutzt (Simulation, Vorhersage und ähnliche), die nachfolgend in »Modellierung, Simulation und Vorhersage« (Abschnitt 2.1) näher erläutert werden. Ein weiterer wichtiger Aspekt bei jeder Modellierung ist die Güte der Ergebnisse in Form unterschiedlicher Maßzahlen (z.B. mittlerer Fehler, mittlere prozentuale Abweichung, u.v.a.). Im Bereich der Statistik ist dabei zusätzlich die Fragestellung relevant, mit *welcher Wahrscheinlichkeit* Ergebnisse unterschiedlicher Experimente zufällig oder systematisch vergleichbar zueinander sind. Diese werden in Unterabschnitt 2.1.1, »Zeitreihenanalyse und statistische Parameter« aufgeführt. Speziell bei der Modellierung und Vorhersage von Herzfrequenzen kommen spezifische Methoden zum Einsatz — ein Überblick als Querschnitt über die in der Literatur bekannten Modelle gibt Abschnitt 2.2 »Modellierung der Herzfrequenz als akute Anpassung«.

2.1. Modellierung, Simulation und Vorhersage

Biologische Fragestellungen zeichnen sich oft durch zahlreiche Gemeinsamkeiten und Charakteristiken aus. Dadurch können Modelle entwickelt werden, die auf diese spezifischen Charakteristiken ausgelegt sind und in allen Fragestellungen mit diesen Charakteristiken angewendet werden können. Beispielsweise lassen sich im Bereich von Herzfrequenzverläufen Modelle entwickeln, die auf die grundlegenden Charakteristika des Herzfrequenzverhaltens angepasst sind. Ein Vorteil der Entwicklung von Modellen anhand spezifischer Charakteristika ist, dass diese damit nicht ausschließlich für einen einzigen Spezialfall nutzbar sind und nicht für jede neue Menge an Daten ein eigenes Modell kreiert werden muss. Gleichzeitig sollen jedoch auch individuelle Unterschiede modelliert werden, da gerade biologische Prozesse trotz zahlreicher gemeinsamer Eigenschaften im Detail sehr individuell ablaufen. Während beispielsweise bei der Herzfrequenz zugrunde gelegt werden kann, dass diese (bei herzgesunden Menschen) bei einer zunehmenden Belastung ansteigt und bei abnehmender oder ausbleibender Belastung wieder absinkt, hängen insbesondere Geschwindigkeit und Stärke des Anstiegs bzw. Abfalls von unterschiedlichen individuellen Faktoren ab. Dabei spielen beispielsweise der Trainingszustand und das Alter eine entscheidende Rolle (vgl. auch »Herzfrequenzverhalten bei körperlicher Belastung«, Abschnitt 1.3). Die Einstellung von solchen individuellen Charakteristika werden in vielen Modellen über »Parameter« vorgenommen.

Abbildung 2.2 veranschaulicht den Prozess der Modellierung bei gegebenem Modell und gegebenen Daten, wie er in der Herzfrequenzmodellierung mit mathematischen Modellen in der Regel stattfindet¹, um externe Belastungen mit der internen Beanspruchungsreaktion zu korrelieren.

Damit die Parameter des Modells auf die individuelle Situation angepasst werden können, werden Messwerte aus der individuellen Problemstellung verwendet. Diese Messwerte werden für die Modellierung häufig als »Daten« oder als »Datensatz« bezeichnet. Im Falle der Herzfrequenz- und Leistungsmodellierung liegen hierbei Messwerte zur Belastung einer Person vor, sowie Messwerte der individuellen Beanspruchungsreaktion des Körpers. Erhält das Modell diese Daten als Eingabe, kann es über eine geeignete »Parameteroptimierung« auf das vorliegende Problem angepasst werden. Dieser Prozess wird oft als »Fitting« oder »Kalibrierung« bezeichnet. Bei der Parameteroptimierung findet im Wechsel eine Anpassung der Parameter und ein Abgleich der Ergebnisse des Modells mit den Eingangsdaten statt: Solange die Abweichung zwischen den realen Messwerten und dem Ergebnis der Modellierung einen spezifischen (vorgegebenen) Fehlerwert überschreitet, werden die Parameter weiter verändert. Der Abgleich zwischen den modellierten und den empirischen (gemessenen) Daten liefert währenddessen die aktuelle »Anpassungsgüte« des Modells. Nach der

¹Modelle aus der künstlichen Intelligenz und dem maschinellen Lernen funktionieren grundlegend etwas anders und werden im Rahmen dieser Arbeit nicht im Detail betrachtet.

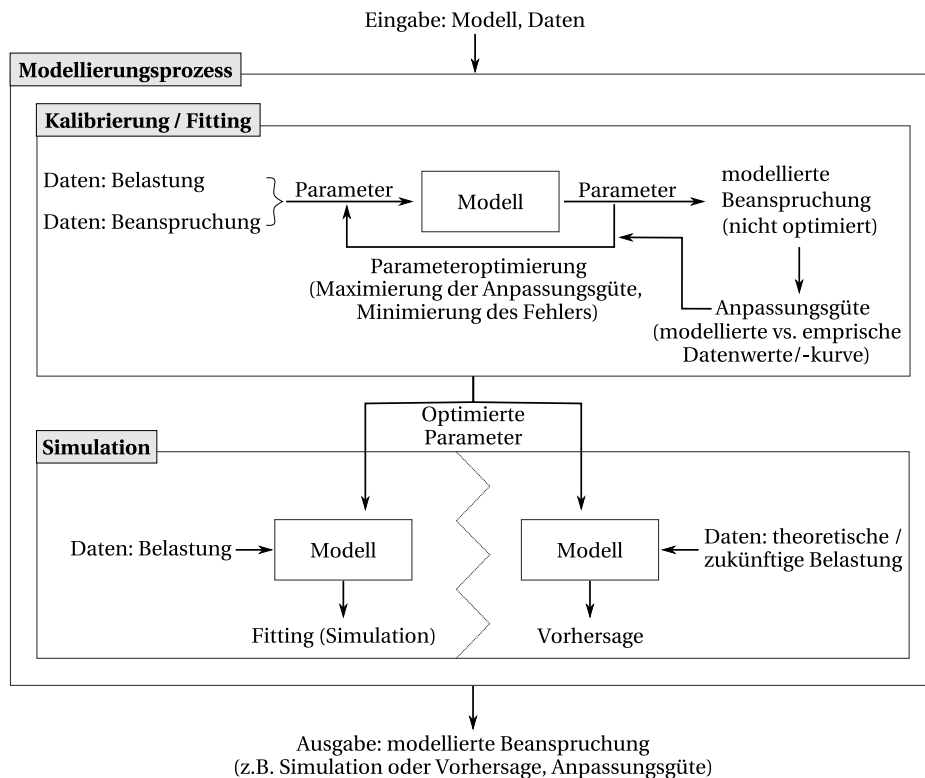


Abbildung 2.2.: Schematische Darstellung eines Modellierungsprozesses in zwei Phasen: Kalibrierung und Simulation

Kalibrierungsphase kann das hinsichtlich eines vorgegebenen Gütekriteriums parameteroptimierte Modell zur »Simulation« verwendet werden. Dabei lassen sich zwei verschiedene Arten der Simulation betrachten: eine »Fitting-Simulation« und eine »Vorhersage« bzw. »Vorhersage-Simulation«. Werden explizit die Modellergebnisse von Daten betrachtet, die bereits für die Parameteroptimierung genutzt wurden, handelt es sich um die Simulation des Fittings; bei der Betrachtung von Daten, die für das Modell neu sind und nicht zur Kalibrierung genutzt wurden, handelt es sich um eine Vorhersage. Das Ergebnis der Simulation ist im Kontext der Herzfrequenz- oder Leistungsmodellierung eine modellierte Beanspruchungsreaktion bei gegebener Belastung. Diese kann auf unterschiedliche Weise dargestellt und evaluiert werden — beispielsweise mit Hilfe statistischer Kennzahlen zur Berechnung der Modellgenauigkeit, oder auch in Form einer geeigneten graphischen Darstellung zum visuellen Vergleich von simulierten und empirischen Daten. Die Parameteroptimierung kann beispielsweise mit der »Methode der kleinsten Quadrate« durchgeführt werden. Diese wird im nachfolgenden Kapitel näher beschrieben.

Im Folgenden werden die im Kontext der Herzfrequenzmodellierung häufig verwendeten statistischen Parameter und Tests (Unterabschnitt 2.1.1, »Zeitreihenanalyse und statistische Parameter«) sowie gängige bzw. geeignete Formen der Visualisierung dargestellt (Unterabschnitt 2.1.2, »Visualisierung«). Anschließend werden diese hinsichtlich ihrer zielführenden

Eignung kritisch diskutiert (Unterabschnitt 2.1.3, »Kritische Betrachtung statistischer Parameter«), bevor die gängigen Modell für die Herzfrequenzmodellierung (Abschnitt 2.2, »Modellierung der Herzfrequenz als akute Anpassung«) präsentiert werden.

2.1.1. Zeitreihenanalyse und statistische Parameter

Für die Modellierung von Herzfrequenzen oder Leistungsdaten müssen in der Regel zeitdiskrete Signale vorliegen. In der Herzfrequenz- oder Wattleistungsmessung entstehen die Aufnahmen solcher zeitdiskreter Signale durch die Abtastung zu meist äquidistanten Zeitpunkten des jeweiligen Signals. Während die Abtastrate bei der Modellierung akuter Reaktionen auf körperliche Belastung im Millisekunden- oder Sekundenbereich liegt, beträgt die Abtastrate bei der Untersuchung chronischer Belastungsreaktionen oft einzelne oder mehrere Tage bis hin zu wenigen Wochen. Da die Daten zusätzlich von der zur Abtastrate gehörenden Zeitkomponente abhängen, werden solche Daten *Zeitreihen* genannt.

Die *Zeitreihenanalyse* gehört zum Gebiet der Regressionsanalysen und wird häufig zur Prognose künftiger Entwicklungen auf Basis vergangener Strukturen in der Zeitreihe genutzt. Unabhängig von der spezifischen Analyse sind häufig eine Vielzahl statistischer Parameter hilfreich, um die Qualität einer Modellierung einschätzen zu können.

Die Güte eines Modells beziehungsweise die Güte der Simulation oder Vorhersage mit einem bestimmten Modell kann qualitativ und quantitativ bewertet werden. Während eine qualitative Betrachtungsweise Einzelfallanalysen voraussetzt und schwierig zu vergleichen ist, liefert eine quantitative Analyse zunächst besser vergleichbare Kriterien für die Modellgüte. Hierzu stehen unterschiedliche statistische Parameter zur Verfügung. Im trainingswissenschaftlichen Kontext werden dazu häufig die mittlere quadratische Abweichung (*mean-square error* (MSE)) oder die Wurzel der mittleren quadratischen Abweichung (*root-mean-square error* (RMSE)), der mittlere prozentuale Fehler (*mean absolute percentage error* (MAPE)) bzw. für kleinere Datensätze der gewichtete mittlere prozentuale Fehler (*weighted average percentage error* (WAPE)), sowie unterschiedliche Korrelationskoeffizienten (Pearson's r , Intra-Klassen-Korrelation (ICC)) oder der Determinationskoeffizient (R^2) herangezogen.

Definition 2.1: Fehlermaße und Korrelationskoeffizienten.

Seien n die Anzahl der Datenpunkte, p die Anzahl der Modellparameter $\alpha = a_1, a_2, \dots, a_p$, x_i, y_i empirische Datenpunkte mit Mittelwert \bar{x} bzw. \bar{y} und $z_i := f(x_i, \alpha)$ die simulierten Datenpunkte mit Mittelwert \bar{z} , $i \in \{1, 2, \dots, n\}$. Dabei entspricht z_i auf Basis von x_i der Modellierung bzw. Schätzung für y_i .

SSE Der quadratische Fehler SSE (*sum of squares error*) ist definiert als
$$SSE = \sum_{i=1}^n (y_i - z_i)^2.$$

SST Das Maß der totalen Abweichung SST (*total sum of squares*) ist definiert als

$$SST = \sum_{i=1}^n (z_i - \bar{z})^2.$$

MSE Die *mittlere quadratische Abweichung* (engl. *mean-square error*) ist ein Maß für die Modellgüte und ist definiert als $MSE = \frac{SSE}{n-p}$.

MAE Der *mittlere absolute Fehler* (engl. *mean absolute error*) ist definiert als

$$MAE = \frac{1}{n} \cdot \sum_{i=1}^n |y_i - z_i|.$$

RMSE Die *Wurzel aus der mittleren quadratischen Abweichung* (engl. *root-mean-square error*) ist definiert als Wurzel aus dem MSE, $RMSE = \sqrt{MSE}$.

MAPE Der mittlere prozentuale Fehler (engl. *mean absolute percentage error*) dient dazu, die prozentuale Abweichung zwischen einer Messreihe und ihrer Simulation zu berechnen und ist definiert als $MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - z_i|}{y_i}$.

WAPE Der gewichtete mittlere prozentuale Fehler (engl. *weighted average percentage error*) bringt eine andere Gewichtung in den MAPE. Er ist definiert als $WAPE = \frac{\sum_{i=1}^n |y_i - z_i|}{\sum_{i=1}^n |y_i|}$. Der WAPE ist eine Abwandlung des MAPE, bei dem die Fehlerwerte durch die empirischen Messwerte gewichtet werden. Insbesondere wenn nur wenige Datenpunkte zur Verfügung stehen (d.h. n ist sehr klein), ist der WAPE dem MAPE vorzuziehen.

Pearson's r Der empirische Korrelationskoeffizient nach Pearson ist eine Maßzahl für die Stärke der Korrelation zweier Merkmale (z.B. empirische Messung vs. Simulation) und ist definiert als

$$r = r_{XY} = \frac{\sum_{i=1}^n (y_i - \bar{y}) \cdot (z_i - \bar{z})}{\sqrt{(\sum_{i=1}^n (y_i - \bar{y})^2) \cdot (\sum_{i=1}^n (z_i - \bar{z})^2)}}$$

d.h. die empirische Kovarianz wird dividiert durch die Wurzel des Produkts aus den empirischen Varianzen. Es gilt $-1 < r < 1$, wobei -1 eine negative und 1 eine perfekte Korrelation beschreibt. In der Interpretation deuten Werte im Bereich $0,0 \leq r \leq 0,2$ auf keinen bis einen sehr geringen linearen Zusammenhang hin, Werte im Bereich $0,2 < r \leq 0,5$ auf einen schwachen bis mäßigen linearen Zusammenhang, Werte im Bereich $0,5 < r \leq 0,8$ auf einen deutlichen linearen Zusammenhang und Werte im Bereich $0,8 < r \leq 1$ auf einen hohen bis perfekten linearen Zusammenhang. Für negative Werte von r gelten die Bereiche analog für einen negativen linearen Zusammenhang.

ICC Der Intraklassen-Korrelationskoeffizient ist eine Maßzahl zur Quantifizierung der Interrater-Reliabilität. Seine Verwendung eignet sich unter anderem dann, wenn mehrere Beobachtungszeitpunkte vorhanden sind. Je nach Situation können unterschiedliche Varianten des ICC betrachtet werden. Im Kontext dieser Arbeit ist lediglich der ICC(3,1) von Bedeutung; für eine detaillierte Beschreibung wird auf gängige Statistikbücher verwiesen. Die Werte des ICC liegen im Intervall $[-1; 1]$;

da negative Werte für die Reliabilität keine Bedeutung haben, wird diesen eine Reliabilität von Null zugewiesen. Werte größer als 0,7 werden im Allgemeinen als Indikator für gute Übereinstimmung angesehen.

R² Der *Determinationskoeffizient* gibt an, wie dicht die empirischen Daten und eine angepasste Regressionsgerade zueinander liegen. Dieser ist definiert durch $R^2 = 1 - \frac{SSE}{SST}$. Es gilt $0 < R^2 < 1$, wobei bei $R^2 = 1$ ein vollständiger linearer Zusammenhang zwischen der Regression und den empirischen Werten vorherrscht (Bonate, 2011). Hierbei ist zu beachten, dass sich R^2 nicht eignet, um Messdaten mit der Simulation eines beliebigen Modells zu vergleichen. Der Determinationskoeffizient sollte ausschließlich in der Regressionsanalyse verwendet werden. Die simulierten Datenpunkte z_i innerhalb von SSE und SST müssen in diesem Fall entsprechend aus einem linearen Regressionsmodell stammen.

Dabei ist zu beachten, dass der MAPE die Werte der Simulation eher überschätzt, d.h. höher annimmt, als die Messwerte tatsächlich sind. Der MAPE sollte daher bei einer Evaluation nicht als alleiniges Maß, sondern stets gemeinsam mit und in Relation zu anderen statistischen Maßen (wie beispielsweise dem RMSE) oder einer geeigneten Visualisierung betrachtet werden.

Statistische Tests

Beim Vergleich eines Testergebnisses mit der (zumeist unbekannt)en Wahrheit lassen sich die in Abbildung 2.3 dargestellten vier Fälle unterscheiden: Sofern das Testergebnis mit der Realität der Grundgesamtheit übereinstimmt – also H_1 zurecht verworfen oder nicht verworfen wird – ist der Test korrekt, andernfalls inkorrekt. Die Fehler werden dabei in einen »Fehler erster Art« und einen »Fehler zweiter Art« unterteilt. Ein Fehler erster Art liegt vor, wenn die Nullhypothese zutreffend wäre, der Test jedoch signifikant für die Alternativhypothese ausgefallen ist. Dieser Fehler wird auch als α -Fehler bezeichnet. Umgekehrt liegt ein Fehler zweiter Art vor, wenn die Nullhypothese falsch ist, der Test jedoch nicht signifikant für die Alternativhypothese ausgefallen ist und diese verworfen wurde. Dieser Fehler wird auch als β -Fehler bezeichnet.

Entsprechend dieser beiden Fehlerarten geben der α -Wert die Wahrscheinlichkeit für einen Fehler erster Art und der β -Wert die Wahrscheinlichkeit für einen Fehler zweiter Art an. Die Wahrscheinlichkeit $1 - \alpha$ dafür, dass bei wahren H_0 die richtige Entscheidung (für H_0) getroffen wird, ist die »Spezifität« des Tests. Die Wahrscheinlichkeit $1 - \beta$ dafür, dass bei falschem H_0 die richtige Entscheidung (für H_1) getroffen wird, ist die »Sensitivität« oder auch »Trennschärfe« oder »(statistische) Power« des Tests.

		Testentscheidung	
		für H_0	für H_1
Realität / Wahrheit	H_0 : wahr	richtig positiv A	falsch positiv <i>Fehler 1. Art</i> B
	H_1 : wahr	falsch negativ <i>Fehler 2. Art</i> C	richtig negativ D

H_0 ist wahr: $A+B$
 Spezifität: $\frac{A}{A+B}$

H_1 ist wahr: $C+D$
 Sensitivität: $\frac{D}{C+D}$

Abbildung 2.3.: Übersicht über die möglichen Ergebnisse eines statistischen Tests im Vergleich zur Realität

Ist eine »gleiche Grundgesamtheit« zwischen den betrachteten Gruppen (beispielsweise der Gruppe der Messwerte und der Gruppe der Simulationswerte) gefordert, müssen drei Bedingungen erfüllt sein:

1. Die Mittelwerte bzw. Mediane beider Gruppen sind (in etwa) gleich;
2. Die Streuung (Varianz) beider Gruppen ist (in etwa) gleich;
3. Die Verteilung beider Gruppen hat die »gleiche Form«.

Ist also beispielsweise eine der betrachteten Gruppen gleichverteilt, die zweite Gruppe jedoch normalverteilt, kann nicht vom Vorliegen einer gleichen Grundgesamtheit ausgegangen werden. Bei schiefen Verteilungen sollten beide ähnlich schief sein, d.h. die Verteilung beider Gruppen muss rechtsschief *oder* linksschief sein.

Auch ohne die Forderung nach einer gleichen Grundgesamtheit müssen die Daten jeder Teilstichprobe in sich unabhängig und identisch verteilt sein: Unabhängig bedeutet dabei, dass die einzelnen Werte der Stichprobe sich nicht gegenseitig beeinflussen. Die identische Verteilung sorgt dafür, dass die einzelnen Werte der Stichproben mit gleicher Wahrscheinlichkeit gleiche Werte annehmen und das Experiment damit wiederholt durchgeführt werden kann, ohne andere Ergebnisse zu erzielen. Zwei oder mehr Stichproben sind (im Sinne von Zufallsvariablen) unabhängig und identisch verteilt, wenn jede Stichprobe der gleichen Wahrscheinlichkeitsverteilung unterliegt und wenn alle Stichproben voneinander unabhängig sind, sich also nicht gegenseitig beeinflussen. Andernfalls ist die Anwendung eines statistischen Testverfahrens geeignet. Ob die Daten unterschiedlicher Gruppen der gleichen Verteilung unterliegen (die Verteilungen also die gleiche Form aufweisen) und ob die Werte innerhalb einer Gruppe unabhängig voneinander sind, muss dabei ggf. im Vorfeld separat überprüft werden.

Die Wahl des richtigen Tests unterliegt neben konkreten Voraussetzungen zunächst der grundlegenden Entscheidung, wie viele Stichproben gemeinsam analysiert werden und in welchem Verhältnis diese zueinander stehen: Neben Tests für eine Stichprobe (wie dem

Einstichproben-t-Test oder dem Vorzeichentest) sind bei der Analyse eines Zusammenhangs zweier oder mehrerer unterschiedlicher Stichproben statistische Tests für zwei (oder mehr) »abhängige« oder »unabhängige« Stichproben notwendig.

»Abhängigen« Stichproben liegen Daten zugrunde, die vom gleichen Probandenkollektiv oder der gleichen einzelnen Person stammen. Werden unterschiedliche Interventionen der gleichen Person(en) untersucht und geht es dabei um den Effekt der Intervention, sind die Stichproben abhängig voneinander. Wird beispielsweise ein spezifischer physiologischer Parameter in einer Probandengruppe vor und nach einer Trainingseinheit gemessen, sind die Stichproben abhängig. Andere Bezeichnungen für abhängige Stichproben sind *verbundene* oder *gepaarte* Stichproben. »Unabhängige« Stichproben basieren auf Daten, bei denen unterschiedliche Untersuchungsgruppen miteinander verglichen werden. Wird beispielsweise der Effekt einer Intervention auf Männer mit dem Effekt der gleichen Intervention auf Frauen verglichen, lassen sich die Daten in zwei unterschiedliche und voneinander unabhängige Stichproben (Männer vs. Frauen) unterteilen. Andere Bezeichnungen für unabhängige Stichproben sind *unverbundene* oder *ungepaarte* Stichproben.

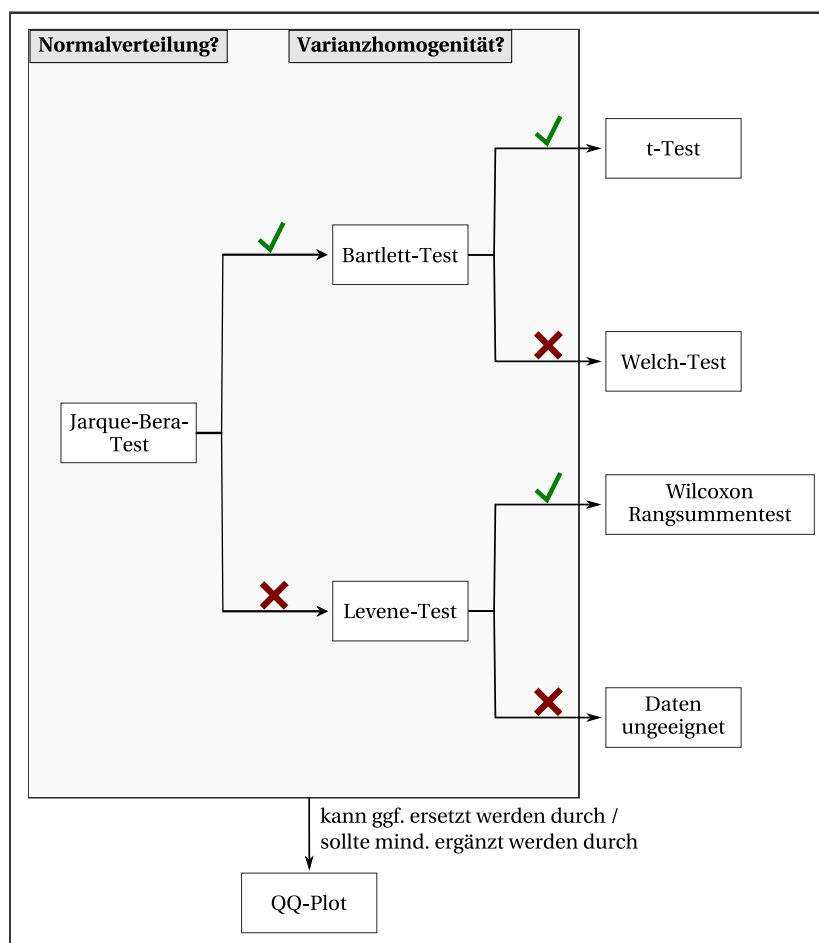


Abbildung 2.4.: Übliches Schema zur Auswahl eines statistischen Tests bei verbundenen Stichproben

Bei den durchgeführten Analysen innerhalb dieser Arbeit werden ausschließlich verbundene Stichproben betrachtet: Es handelt sich innerhalb einer Studie stets um Ergebnisse innerhalb des selben Probandenkollektivs. Entsprechend kommen für die statistischen Auswertungen ausschließlich Tests für verbundene Stichproben in Frage. Die Standardvorgehensweise für eine Testauswahl ist schematisch in Abbildung 2.4 dargestellt: Mit einem Test wie dem dem Jarque-Bera-Test lässt sich überprüfen, ob die Daten der Stichprobe normalverteilt sind. Anschließend sind die Daten auf Varianzhomogenität zu überprüfen. Dazu kann beispielsweise der Bartlett-Test (für normalverteilte Daten) oder der Levene-Test (für Daten ohne Annahme der Normalverteilung) verwendet werden. Je nach erfüllten Voraussetzungen kommen der t-Test, der Welch-Test oder der verteilungsfreie Wilcoxon-Rangsummentest für die weiteren Analysen in Frage. Diese Tests werden zum besseren Verständnis im Folgenden kurz vorgestellt, wobei auf die detaillierte Beschreibung der Teststatistiken weitgehend verzichtet wird. Dies jeweilige Teststatistik kann in allen gängigen Statistik-Lehrbüchern nachgeschlagen werden.

Jarque-Bera-Test Der Jarque-Bera-Test gehört zur Gruppe der Anpassungstests (engl. »goodness-of-fit test«) und wird verwendet, um das Vorliegen einer Normalverteilung zu prüfen. Dabei werden die Schiefe und die Wölbung einer Stichprobe aus der vorliegenden Verteilung in der Berechnung der Statistik berücksichtigt. Konkret besteht beim Jarque-Bera-Test das Hypothesenpaar aus der Hypothese H_0 : »Die Stichprobe ist normalverteilt«, und der Hypothese H_1 : »Die Stichprobe ist nicht normalverteilt«. In MATLAB erfolgt die Berechnung des Jarque-Bera-Tests über die Funktion *jbtest*.

Bartlett-Test Der Bartlett-Test wird verwendet, um das Vorliegen einer Varianzhomogenität zweier oder mehrerer Stichproben zu prüfen, wobei die Daten der einzelnen Stichproben normalverteilt sein müssen. Das Hypothesenpaar besteht beim Bartlett-Test aus der Hypothese H_0 : »Die Streuung innerhalb der Stichproben ist homogen; alle Varianzen sind gleich«, und der Hypothese H_1 : »Die Streuung innerhalb der Stichproben ist heterogen; nicht alle Varianzen sind gleich«. In MATLAB erfolgt die Berechnung des Bartlett-Tests über die Funktion *barttest*.

Levene-Test Der Levene-Test wird verwendet, um das Vorliegen der Varianzhomogenität zweier oder mehrerer Stichproben zu prüfen, wobei keine Normalverteilung vorliegen muss. Die Trennschärfe des Levene-Tests ist geringer als die des Bartlett-Tests. Liegen normalverteilte Daten vor, sollte daher stets der Bartlett-Test zur Überprüfung der Varianzhomogenität bevorzugt werden. Wie auch beim Bartlett-Test besteht das Hypothesenpaar beim Levene-Test aus der Hypothese H_0 : »Die Streuung innerhalb der Stichproben ist homogen«, und der Hypothese H_1 : »Die Streuung innerhalb der Stichproben ist heterogen«. In MATLAB erfolgt die Berechnung des Levene-Tests über die Funktion *vartestn* (mit Parameter 'TestType': 'LeveneAbsolute').

t-Test Der t-Test dient dazu, anhand von Stichproben Hypothesen zu überprüfen, die den Mittelwert (oder bei mehreren Gruppen: die Mittelwerte) der Grundgesamtheit betrifft. Voraussetzung dafür ist, dass der Grundgesamtheit eine Normalverteilung unterstellt werden kann. Die Verteilung der Varianz darf unbekannt sein, allerdings darf die Streuung nicht zu groß sein. Es wird eine *Varianzhomogenität* vorausgesetzt. Sofern die Verteilung der Varianz ebenfalls bekannt ist, eignet sich der Gauß-Test (auch als z-Test bezeichnet) besser. Beim t-Test gibt es drei unterschiedliche Teststatistiken, je nach dem, ob eine einzelne Stichprobe, zwei verbundene Stichproben oder zwei unverbundene Stichproben analysiert werden. Bei zwei verbundenen Stichproben wird ein *gepaarter t-Test* verwendet. Das Hypothesenpaar besteht beim gepaarten t-Test aus der Hypothese H_0 : »Die mittlere Differenz zwischen den Messungspaaren ist gleich Null«, und der Hypothese H_1 : »Die mittlere Differenz zwischen den Messungspaaren ist ungleich Null«. Eine andere Formulierung des Hypothesenpaares ist H_0 : »Die Mittelwerte der beiden Stichproben sind gleich« und H_1 : »Die Mittelwerte der beiden Stichproben sind ungleich«. In MATLAB erfolgt die Berechnung des t-Tests für einzelne Stichproben oder zwei verbundene Stichproben über die Funktion *ttest* und für zwei unverbundene Stichproben über *ttest2*.

Welch-Test Der Welch-Test ist eine Alternative zum t-Test mit gleichem Hypothesenpaar wie beim t-Test, d.h. H_0 : »Die Mittelwerte der beiden Stichproben sind gleich« und H_1 : »Die Mittelwerte der beiden Stichproben sind ungleich«. Die Daten der Stichprobe müssen nicht die gleiche Streuung haben (Varianzheterogenität), sollten aber möglichst normalverteilt sein. In MATLAB erfolgt die Berechnung des Welch-Tests für zwei verbundene Stichproben ebenso wie beim gepaarten t-Test über die Funktion *ttest2* mit zusätzlichem Parameter 'Vartype': 'unequal'.

Wilcoxon-Rangsummentest Der Wilcoxon-Rangsummentest ist eine »verteilungsfreie« Alternative zum t-Test mit gleichem Hypothesenpaar wie beim t-Test, d.h. H_0 : »Die Mittelwerte der beiden Stichproben sind gleich« und H_1 : »Die Mittelwerte der beiden Stichproben sind ungleich«. Die beiden Gruppen müssen nicht normalverteilt sein. Damit der Wilcoxon-Rangsummentest ein verlässliches Ergebnis liefert, müssen allerdings auch hier beide Gruppen die gleiche Streuung haben (Varianzhomogenität). In MATLAB erfolgt die Berechnung des Wilcoxon-Rangsummentests über die Funktion *ranksum*.

Sind die jeweiligen Voraussetzungen nicht erfüllt und ein Test wird dennoch verwendet, ist die Wahrscheinlichkeit für ein falsch-positives Ergebnis erhöht (Fehler 1. Art, » α -Risiko«). Bei einem Test, mit dem die Daten auf beispielsweise Mittelwertunterschiede geprüft werden sollen, zeigt das Testergebnis mit höherer Wahrscheinlichkeit eine Signifikanz an, obwohl keine Mittelwertunterschiede vorliegen. Die Nullhypothese wird zurückgewiesen, obwohl sie wahr ist.

Parameteroptimierung

Eine klassische Möglichkeit zur Parameteroptimierung ist die »Methode der kleinsten Quadrate«; mit dieser werden auch alle Kalibrierungen der Parameter in den späteren empirischen Arbeiten durchgeführt. Diese Methode beschreibt eine Kurvenanpassungs- bzw. Regressionsanalysetechnik, die eine Kurve – bzw. die Parameter eines Modells, das eine entsprechende Kurve erzeugt – mit minimalem euklidischen Abstand zu gegebenen (empirischen) Datenpunkten berechnet: Sei (X, Y) eine Menge von n Datenpaaren $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ und sei $f(X, \alpha)$ die Modellfunktion mit m Parametern $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m)$, die die Daten $y_i \in Y$ schätzen soll.

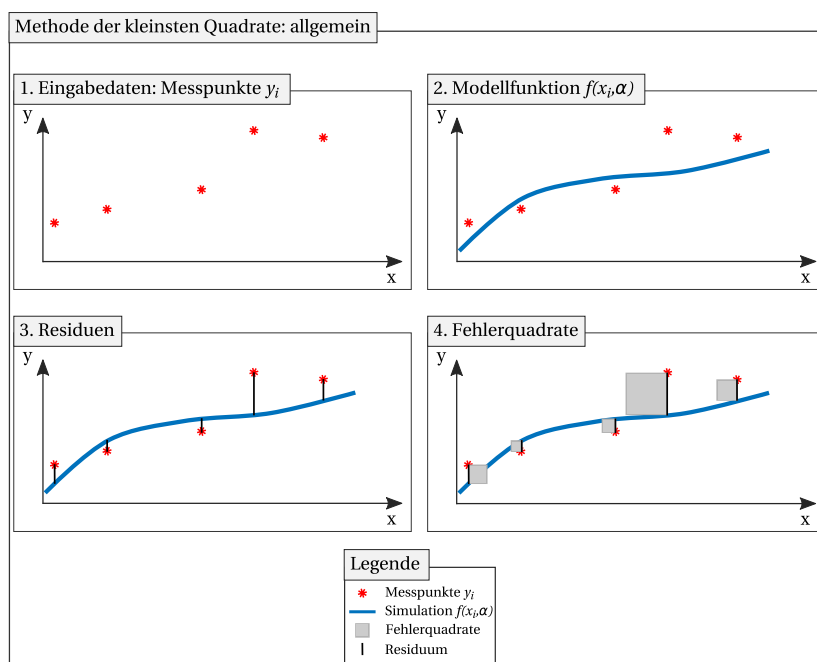


Abbildung 2.5.: Schematische Darstellung zur Methode der kleinsten Quadrate: (1) Zunächst liegen Daten (X, Y) vor. Zur besseren Visualisierung der Methodik sind nur die zu approximierenden Daten Y dargestellt. (2) Mit dem Modell f , dessen Parameter α optimiert werden sollen, wird für ein vorgegebenes α eine Simulation $f(X, \alpha)$ durchgeführt. (3) Die Methode der kleinsten Quadrate bestimmt die lotrechten Abstände (Residuen) zwischen den empirischen Eingabewerten y_i und den zugehörigen Simulationen $f(x_i, \alpha)$. (4) Basierend auf den Residuen werden Quadrate gebildet, deren Summe für ein optimiertes α minimiert wird

Die $x_i \in X$ dienen also als Eingabe für das Modell, um mit $f(X, \alpha)$ die Datenwerte $y_i \in Y$ anzunähern. Im Kontext der betrachteten Zeitreihenanalysen sind die $x_i \in X$ voneinander unabhängig, während die Modellannahme darauf basiert, dass die $y_i \in Y$ von den $x_i \in X$ abhängen. Das Ziel der Methode der kleinsten Quadrate ist es, α derart zu bestimmen, dass für die Summe der Quadrate S gilt: $S = \operatorname{argmin}_{\alpha} \sum_{i=1}^n (y_i - f(x_i, \alpha))^2$, also die Summe der Quadrate des Fehlers zwischen vorgegebenen und geschätzten Datenpunkten minimiert wird. Abbildung 2.5 veranschaulicht diese Vorgehensweise. Liegen – wie bei Zeitreihenanalysen

besonders im Bereich der Trainingswirkungsanalyse häufig der Fall – deutlich mehr Eingabewerte für X vor und vergleichsweise wenige diskrete Datenpunkte für Y , findet eine geeignete Umnummerierung statt. Während das Modell für jeden Eingabewert von X einen Simulationswert $f(X, \alpha)$ bestimmt, findet die Auswertung der kleinsten Quadrate nur dort statt, wo ein Wertepaar (x_i, y_i) existiert.

Die Methode der kleinsten Quadrate eignet sich aus zwei Gründen besonders gut zur Parameteroptimierung: Die Quadrierung sorgt dafür, dass der Fehler unabhängig davon ist, ob die empirischen Werte über- oder unterschätzt werden. Für den Fehlerwert macht es damit keinen Unterschied, ob die Simulationswerte die empirischen Daten chronisch unterschätzen, chronisch überschätzen oder im Wechsel um die empirischen Daten oszillieren. Gleichzeitig führt die Quadrierung zu einer geeigneten Gewichtung der Fehler, indem wenige große Abweichungen »stärker bestraft« werden als viele sehr kleine Abweichungen. Auf fehlerhafte Ausreißer in den Daten, wie sie beispielsweise durch Messfehler entstehen können, sollte daher im Vorfeld unbedingt geachtet werden.

2.1.2. Visualisierung

Die Visualisierung von Daten und Simulationsergebnissen ist oftmals eine sinnvolle und teilweise notwendige Ergänzung der statistischen Auswertung. Sie bietet nicht nur die Möglichkeit einer qualitativen Bewertung, sondern kann insbesondere die Interpretation der (nicht nur statistischen) Ergebnisse erleichtern und vor Fehlinterpretationen bewahren. Dies wird im nachfolgenden Abschnitt über die »Kritische Betrachtung statistischer Parameter« (Unterabschnitt 2.1.3) diskutiert.

Boxplots

Zur Visualisierung ordinalskaliertter Merkmale werden häufig Box-Whisker-Plots (kurz: Boxplots) verwendet. Der Boxplot – wie in Abbildung 2.6 dargestellt – stellt dabei eine übersichtliche Zusammenfassung einiger statistischer Kenngrößen (vgl. Definition 8.1) des betrachteten Merkmals dar: eine Art »Minimum« und »Maximum« der Daten (kleinster bzw. größter Wert ohne Ausreißer und Extremwerte), den Median, das obere und untere Quartil sowie mögliche Ausreißer und Extremwerte. Dabei entspricht der Bereich zwischen dem größten und dem kleinsten Wert ohne Ausreißer dem vierfachen des Interquartilsabstand. Werte, die mehr als der 1,5-fache Interquartilsabstand (IQA) von Q_1 bzw. Q_3 entfernt sind, sind Ausreißer, alle darüber hinausgehenden Werte werden als Extremwerte klassifiziert. Die Box innerhalb des Boxplots entspricht dem IQA und enthält entsprechend 50% der Daten.

Boxplots eignen sich, um einen ersten (groben) Überblick über die Datenlage zu erhalten. Allerdings gehen Informationen über die Verteilung der Daten bei der Betrachtung von Boxplots vollständig verloren. Matejka und Fitzmaurice (2017) stellen in ihrer Publikation an-

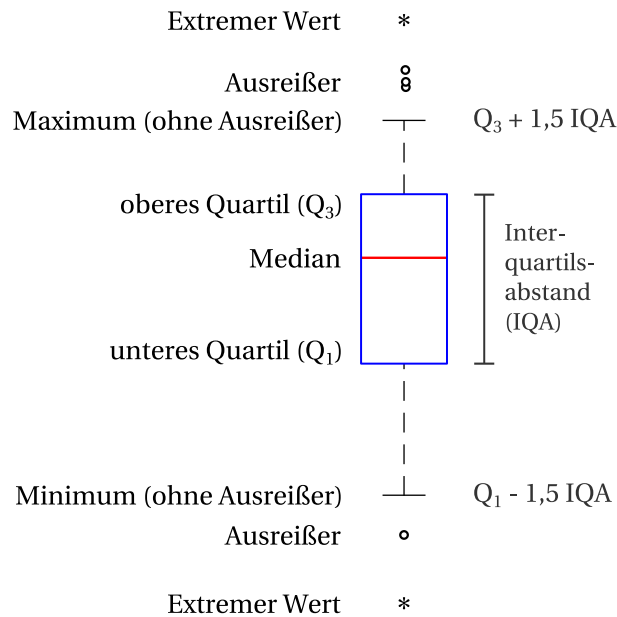


Abbildung 2.6.: Schematische Darstellung eines Boxplots

schaulich dar, wie viele Informationen durch die Abstrahierung auf statistische Kennzahlen verloren gehen. In Bezug auf die Boxplot-Visualisierung ist eines ihrer Beispiele in Abbildung 2.7 abgebildet: Die dargestellten sechs Verteilungen der Datenpunkte und Histogramme (A-E) weisen alle identische Quartile, den gleichen Median und das gleiche Minimum bzw. Maximum (im Sinne des 1,5 IQA-Wertes) auf und resultieren trotz ihrer Unterschiede folglich im gleichen Boxplot.

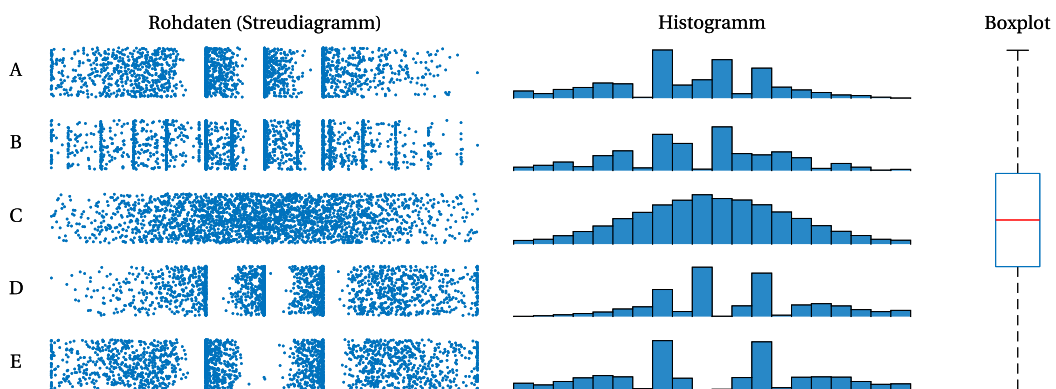


Abbildung 2.7.: Unterschiedlich verteilte Daten (A-E) dargestellt als Streudiagramm, Histogramm, und dem zu A-E zugehörigen Boxplot. Adaptiert von und mit Daten aus Matejka und Fitzmaurice (2017)

Es gibt zahlreiche Varianten des klassischen Boxplots oder alternative Darstellungen, die dieselben statistischen Kennzahlen visualisieren, jedoch unterschiedliche weitere Informationen darstellen. Boxplots mit variabler Breite beispielsweise eignen sich, um zumindest die Anzahl der dem Plot zugrunde liegenden Datenpunkte zusätzlich berücksichtigen zu können. Der Verlust der Information über die Verteilung der Daten bleibt allerdings wie auch

bei vielen anderen Varianten des Boxplots bestehen – es entsteht schnell der Eindruck, die Daten lägen normalverteilt vor.

Ein Histogramm hat umkehrt den Nachteil, dass einerseits ein quantitativer Vergleich unterschiedlicher Daten oft mehr Zeit in Anspruch nimmt und eine genauere Betrachtung benötigt und andererseits statistische Kenngrößen (wie Median und IQA) nicht visuell dargestellt sind. Diese müssen dabei separat berechnet und nominell zusätzlich angegeben werden. Ein ähnlicher Nachteil besteht, wenn die Gesamtheit der Daten in Form einer Dichteverteilung dargestellt wird – auch hier werden die Daten stark vereinfacht, wobei Kenngrößen nicht direkt ablesbar sind.

Eine Möglichkeit, die Vorteile der Darstellung des Boxplot und eines Histogramms zu vereinen und gleichzeitig einen übersichtlichen Eindruck zu vermitteln, bietet der nachfolgend vorgestellte Violinenplot.

Violinenplots

Der Violinenplot eignet sich insbesondere für die qualitative und quantitative Visualisierung von nicht-normalverteilten Daten und bietet neben einem visuellen Eindruck statistischer Kennzahlen auch die Möglichkeit zur Betrachtung der Verteilung der Daten. Die Verteilung enthält zusätzliche, oft hilfreiche Informationen zur Interpretation der Daten. Die Vorgehensweise, wie ein solcher Plot entsteht, ist in Abbildung 2.8 veranschaulicht: Bei dieser Art der Visualisierung wird zunächst die Dichteverteilung der Daten anhand eines Histogramms berechnet und zur besseren Darstellung geglättet. Anschließend wird zusätzlich ein Boxplot ergänzt. Um einen klassischen Violinenplot wie von Hintze und Nelson (1998) vorgestellt zu erhalten, wird die Dichteverteilung zusätzlich gedreht und gespiegelt. Der Boxplot befindet sich vertikal in der Mitte. Die Häufigkeit, mit der eine Merkmalsausprägung auftritt, ist visuell über die Breite der Dichteverteilung mit der Häufigkeit anderer Merkmalsausprägungen vergleichbar.

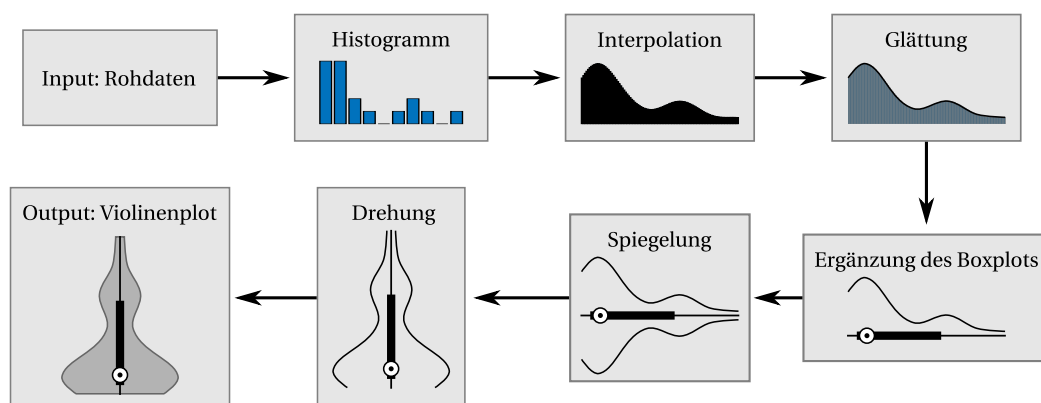


Abbildung 2.8.: Vorgehensweise zur Erstellung eines Violinenplots

Die zuvor in Abbildung 2.7 dargestellten Daten sind in Abbildung 2.9 zum Vergleich als Violinenplots gezeigt.

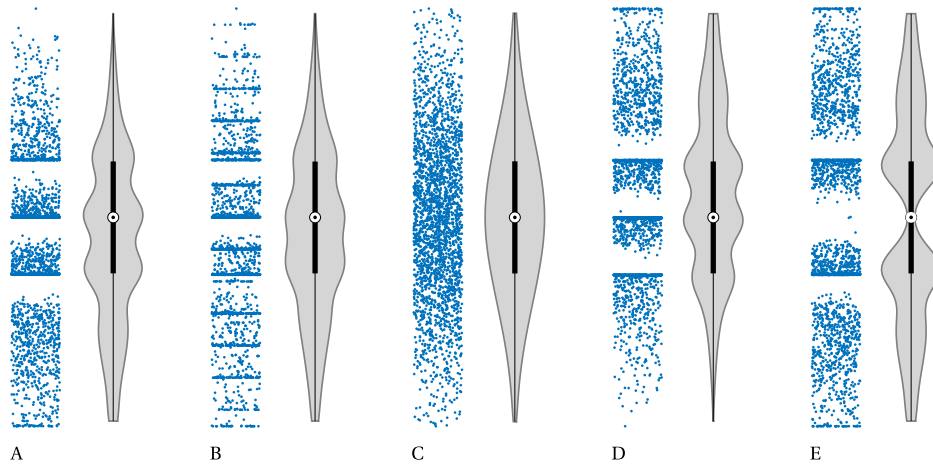


Abbildung 2.9.: Violinenplots zu unterschiedlich verteilten Datensätzen A-E. Die zugehörigen Streudiagramme der Rohdaten sind zur besseren Übersicht erneut dargestellt. Adaptiert von und mit Daten aus Matejka und Fitzmaurice (2017)

Wie auch beim Boxplot sollte auf die Darstellung mittels Violinenplot verzichtet werden, wenn die Anzahl der betrachteten Daten zu gering ist – durch die Reduktion der Daten auf fünf Kenngrößen sollten zumindest 10, besser 20 oder mehr Datenpunkte vorliegen (auch wenn dies rein technisch bei automatisierten Berechnungen wie über die *boxplot*-Funktion in MATLAB nicht erforderlich ist). Zudem ist zu beachten, dass die Berechnung der Dichteverteilung bei kleinen Datenmengen zu einer starken Glättung führt.

Auf Grund der genannten Schwächen von Boxplots, auf die auch Hintze und Nelson (1998) hinsichtlich der Aussagekraft von Violinenplots gegenüber derer aus Boxplots hinweisen, wird in dieser Arbeit auf ihre Verwendung verzichtet; bei Datenmengen, die eine entsprechende Visualisierung rechtfertigen, wird stattdessen der Violinenplot verwendet.

Qualitative Visualisierungen

Zur qualitativen Analyse eignen sich Graphiken, in denen das Ergebnis einer Simulation unmittelbar mit den empirischen Daten verglichen werden können. Abbildung 2.10 stellt exemplarisch eine Möglichkeit für eine solche Visualisierung dar. In der Darstellung kann es hilfreich sein, auf die Strichdicke zu achten: Während zu feine Linien schlecht erkennbar sind, können zu dicke Linien den Vergleich optisch erschweren und für die Interpretation interessante Stellen verbergen. Wenngleich die Visualisierung keinen Einfluss auf die statistische Auswertung hat, kann es sich in einigen Fällen lohnen, unterschiedliche Einstellungen für die Graphik zu betrachten. Je nach verwendeten Daten sind dabei andere graphische Elemente anstelle von Linien geeignet: Für einzelne Belastungswerte über mehrere Tage, wie dies in der Leistungsmodellierung häufig auftritt, eignen sich diskrete, unzusam-

menhängende Blöcke eher. Auch bei einzeln gemessenen Leistungswerten ist es sinnvoll, den angenommenen Verlauf für die Zeit zwischen zwei Messungen anders darzustellen als die tatsächlichen Messwerte. Andernfalls wäre in der Visualisierung nicht erkennbar, welche Daten tatsächlich empirisch ermittelt wurden und welche Werte auf Annahmen oder Simulationen beruhen.

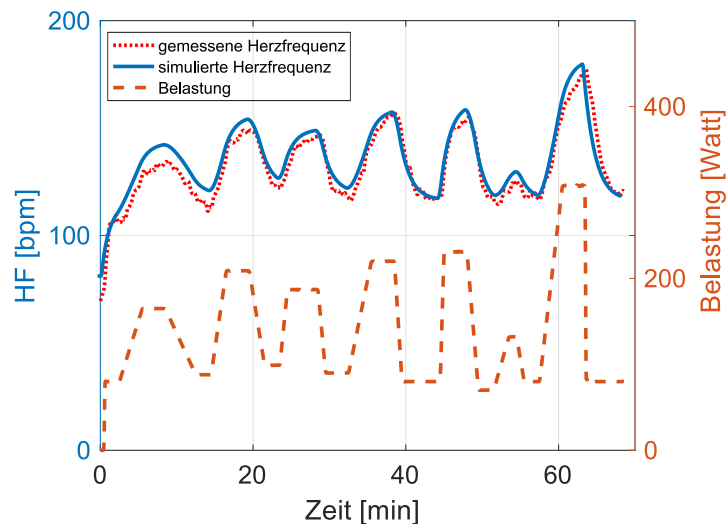


Abbildung 2.10.: Visualisierung eines Vergleichs zwischen modellierten (blau) und empirischen (rot) Herzfrequenz-Daten in Schlägen pro Minute im zeitlichen Verlauf einer Trainingseinheit. Die Belastungswerte (orange) sind in Watt angegeben

QQ-Plots

Eine Möglichkeit, um die Verteilung zweier Stichproben auch visuell miteinander zu vergleichen, ist das Quantil-Quantil-Diagramm (QQ-Plot, in MATLAB über *qqplot*): Dabei werden zunächst die Werte beider Stichproben separat der Größe nach sortiert. Anschließend werden Wertepaare gebildet und die jeweils i -ten Werte der Sortierung aus beiden Verteilungen zu einem Paar zusammengefasst. Diese Wertepaare werden in ein Koordinatensystem übertragen, wodurch die Quantile der beiden Verteilungen gegeneinander aufgetragen werden. Je ähnlicher sich die Verteilungen und Streuungen der beiden Gruppen sind, desto näher liegen die Punkte beim QQ-Plot an der Diagonalen. Umgekehrt zeigt eine große Abweichungen von der Diagonalen, dass die beiden Stichproben tendenziell unterschiedliche Verteilungen besitzen.

Der QQ-Plot kann außerdem genutzt werden, um die Verteilung einer Stichprobe visuell mit einer vorgegebenen Verteilung (häufig: der Standardnormalverteilung) zu vergleichen. Die Quantile werden dabei in Abhängigkeit der Anzahl an Datenpunkten in der Stichprobe gewählt: Für eine Stichprobe mit n Datenpunkten werden n Quantile gebildet und der i -te Wert (der sortierten Stichprobe) wird gegen Quantil $\frac{i-0,5}{n}$ der gewählten Verteilung aufgetragen.

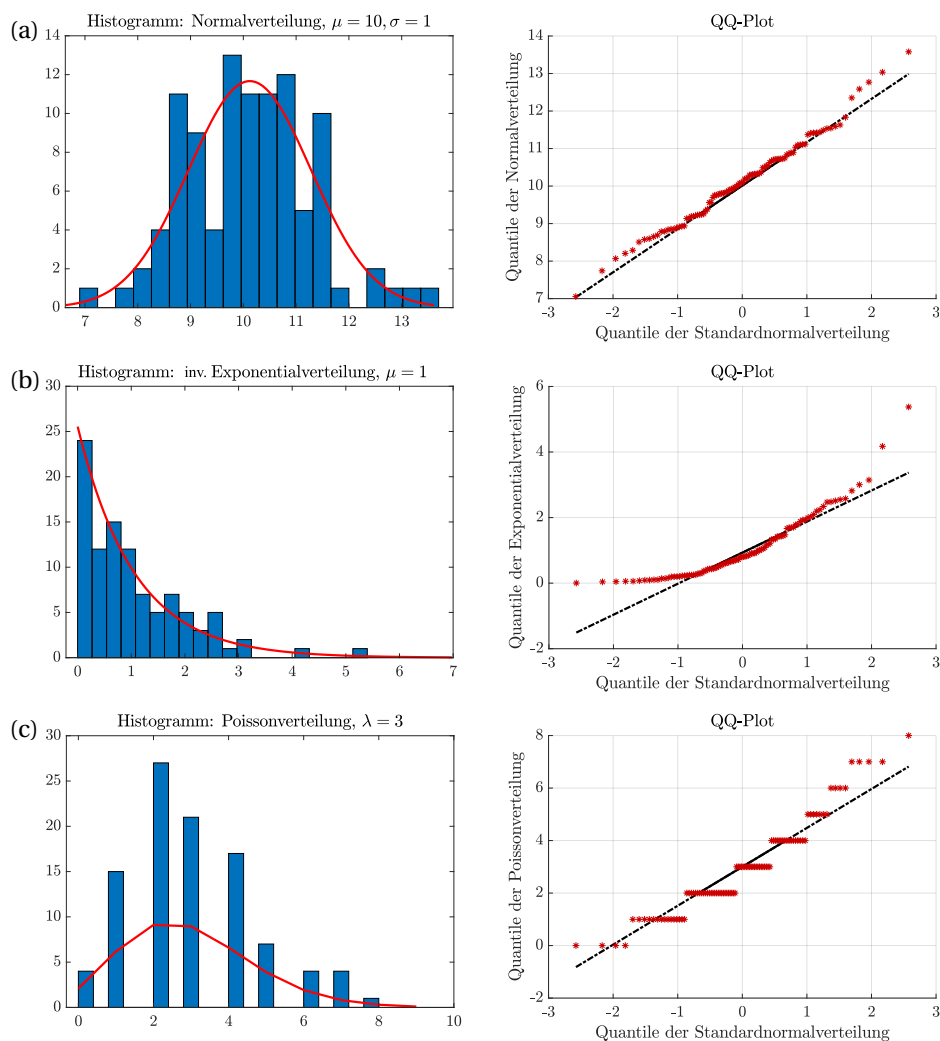


Abbildung 2.11.: Histogramme mit Dichtefunktion und QQ-Plots im Vergleich zur Standardnormalverteilung für Zufallszahlen der **(a)** Normalverteilung ($\mu = 10, \sigma = 1$), **(b)** inverse Exponentialverteilung ($\mu = 3$), **(c)** Poissonverteilung ($\lambda = 3$)

Drei Beispiele für unterschiedliche QQ-Plots im Vergleich zur Standardnormalverteilung sind in Abbildung 2.11 dargestellt: Die zugrundeliegenden Daten sind entsprechend der betrachteten Verteilung generierte Zufallszahlen. Dabei ist jeweils links das Histogramm mit zugehöriger Kurve der Dichtefunktion für die zugrunde liegende Verteilung abgebildet. Auf der rechten Seite ist jeweils der zugehörige QQ-Plot im Vergleich zur Standardnormalverteilung (mit Mittelwert $\mu = 0$ und Standardabweichung $\sigma = 1$) dargestellt. Die Daten in Abbildung 2.11 **(a)** unterliegen einer Normalverteilung mit Mittelwert $\mu = 10$ und Standardabweichung $\sigma = 1$. Die Daten der Quantile im QQ-Plot liegen alle annähernd auf der Diagonalen – die Normalverteilung ist der Standardnormalverteilung sehr ähnlich. Die Daten in Abbildung 2.11 **(b)** unterliegen einer inversen Exponentialverteilung mit Mittelwert $\mu = 3$. Die Daten der Quantile im QQ-Plot liegen im mittleren Teil annähernd auf der Diagonalen. Da die inverse Exponentialfunktion wie auch im Histogramm zu sehen keine negativen Wer-

te annimmt und besonders viele Datenpunkte nahe Null hat, beginnen auch im QQ-Plot die Datenpunkte der Exponentialfunktion bei Null. Im Bereich von etwa -3 bis -1 auf der x-Achse gibt es daher größere Abweichungen von der Diagonalen. Ähnlich sieht es im hinteren Abschnitt der Graphik (im Bereich $2 - 3$ auf der x-Achse) aus: Bedingt durch einige wenige höhere Datenpunkte der Verteilung entsteht auch in diesem Bereich eine deutliche Abweichung zur Standardnormalverteilung. Die Daten in Abbildung 2.11 (c) unterliegen einer Poissonverteilung mit Parameter $\lambda = 3$, der sowohl den Mittelwert als auch die Standardabweichung beschreibt (d.h. $\lambda = \mu = \sigma^2$). Die Verteilung ist für kleine Werte von λ linksschief und weicht in ihrer Struktur und Dichte deutlich von der Standardnormalverteilung ab (für große Werte von λ ab etwa $\lambda \geq 30$ entspricht die Poissonverteilung annähernd einer Normalverteilung mit $\mu = \sigma^2 = \lambda$). Die Poissonverteilung gehört zu den diskreten Wahrscheinlichkeitsverteilungen². Sie ist ausschließlich für natürliche Zahlen definiert. Die Daten der Quantile im QQ-Plot zeigen das deutlich in Form stückweise linearer Ansammlungen der Datenpunkte.

Heatmaps

Heatmaps bieten eine Möglichkeit, dreidimensionale Daten in einem zweidimensionalen Diagramm darzustellen. Die dritte Dimension wird dabei über verschiedene Farben kodiert und die Daten sollten entsprechend ordinalskaliert oder kardinalskaliert sein. Diese Form der Visualisierung eignet sich vor allem, um Muster, Zusammenhänge oder Veränderungen (beispielsweise aber nicht notwendigerweise im zeitlichen Verlauf) zu analysieren.

Zwei der betrachteten Variablen werden auf der x- und y-Achse aufgetragen. Werden zeitliche Veränderungen betrachtet, wird klassischer Weise das Zeitmaß der x-Achse zugeordnet. Das Innere der Heatmap besteht aus gleichgroßen Rechtecken, die entsprechend der dritten Variablen und einer Farbskala eingefärbt sind.

Abbildung 2.12 zeigt zwei Beispiele für eine Heatmap, in der MAPE-Fehlerwerte für drei Probanden über vier Tests aufgeführt sind. Spezifikationen der Probanden, die Art der Tests und wie die Fehlerwerte ermittelt wurden sind für dieses Beispiel nicht von Belang. Für jede Kombination aus einem Probanden und einem Test liegt ein Fehlerwert vor. In der Heatmap gibt es entsprechend für jede Kombination aus Proband und Test ein Rechteck, in dem über die Farbkodierung der MAPE-Wert eingetragen ist³. In diesem Fall sind die zugrundeliegenden MAPE-Werte ebenfalls in das jeweilige Rechteck eingetragen. Als Farbkodierung sind hier Graustufen gewählt, wobei dunklere Graustufen bis hin zu Schwarz niedrigere Werte kodieren und hellere Graustufen bis hin zu weiß höhere Werte. Grundsätzlich sind beliebige

²Im Gegensatz dazu sind die Standardnormalverteilung, die Normalverteilung und die Exponentialverteilung stetige Wahrscheinlichkeitsverteilungen.

³Liegen pro Proband und Test mehrere Fehlerwerte vor, kann beispielsweise auch der Median, der Mittelwert, das Minimum oder eine andere Spezifikation der Fehlerwerte betrachtet werden.

ge Farbskalen möglich. In der ersten Heatmap (links, **(a)**) liefert die Heatmap lediglich einen visuellen Überblick über die MAPE-Werte, ein spezifisches Muster ist nicht eindeutig zu erkennen. In der zweiten Heatmap (rechts, **(b)**) lässt sich ablesen, dass Proband 1 (unterste Zeile der Heatmap) von allen drei Probanden die niedrigsten MAPE-Werte in jedem Test erzielt hat, und dass für Test 3 (dritte Spalte der Heatmap) von allen vier Tests die höchsten MAPE-Werte vorliegen.

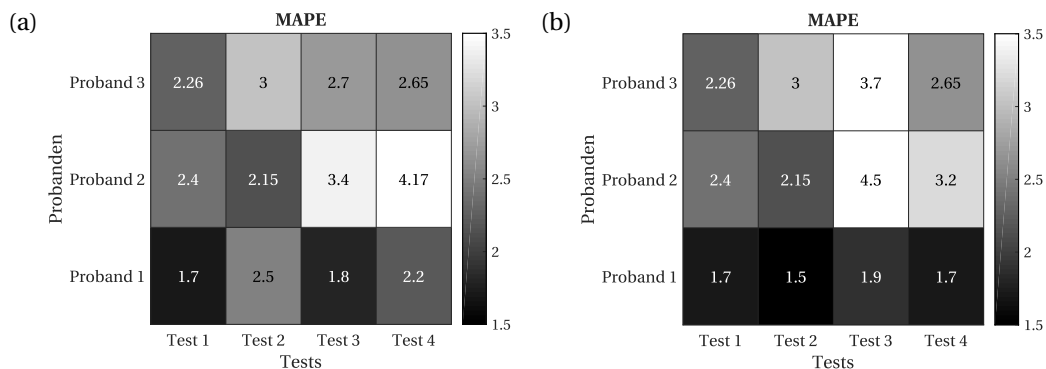


Abbildung 2.12.: Beispiel von zwei Heatmaps, in denen jeweils die MAPE-Werte über eine Graustufen-Farbskala für jede Kombination aus drei Probanden und vier Tests dargestellt sind

2.1.3. Kritische Betrachtung statistischer Parameter

In vielen Bereichen hat sich die Verwendung und Angabe gewisser statistischer Parameter etabliert. Jedoch ist nicht jedes Verfahren und nicht jeder statistischer Parameter für jede Problemstellung gleich gut – oder überhaupt – geeignet. Im Folgenden werden daher die für diese Arbeit relevanten Kenngrößen und Visualisierungen kritisch hinterfragt und bezüglich einer geeigneten Nutzbarkeit bewertet.

Statistische Maße

Statistische Maße sind für sich allein betrachtet nur begrenzt aussagekräftig. Ein anschauliches Beispiel liefert das von Anscombe (1973) beschriebene »Anscombe-Quartett«: Das Anscombe-Quartett umfasst vier Mengen von Datenpunkten. Die einzelnen Werte sind in Tabelle 2.1 dargestellt.

Tabelle 2.2 zeigt verschiedene statistische »Standard-Parameter« wie Mittelwert und Varianz. Auch die Korrelation zwischen den jeweiligen x - und y -Werten und die zugehörigen linearen Regression dieser vier Mengen sind dort erfasst. Es wird deutlich, dass unter Beachtung dieser statistischen Merkmale die vier Mengen bis zur zweiten und zum Teil bis zur dritten Nachkommastelle nahezu identisch sind. Bei alleiniger Betrachtung dieser statistischen Werte ist völlig unklar, wie ähnlich oder verschieden die Mengen untereinander

Tabelle 2.1.: Daten der vier Mengen aus dem Anscombe-Quartett von Anscombe (1973)

Menge 1		Menge 2		Menge 3		Menge 4	
x	y	x	y	x	y	x	y
4,0	4,26	4,0	3,10	4,0	5,39	8,0	5,25
5,0	5,68	5,0	4,74	5,0	5,73	8,0	5,56
6,0	7,24	6,0	6,13	6,0	6,08	8,0	5,76
7,0	4,82	7,0	7,26	7,0	6,42	8,0	6,58
8,0	6,95	8,0	8,14	8,0	6,77	8,0	6,89
9,0	8,81	9,0	8,77	9,0	7,11	8,0	7,04
10,0	8,04	10,0	9,14	10,0	7,46	8,0	7,71
11,0	8,33	11,0	9,26	11,0	7,81	8,0	7,91
12,0	10,84	12,0	9,13	12,0	8,15	8,0	8,47
13,0	7,58	13,0	8,74	13,0	12,74	8,0	8,84
14,0	9,96	14,0	8,10	14,0	8,84	19,0	12,50

sind, sodass ergänzend stets geeignete weitere Methoden zur Analyse (wie beispielsweise Visualisierungen) genutzt werden sollten.

Tabelle 2.2.: Statistische Merkmale der vier Mengen von Anscombe (1973)

Statistische Eigenschaft	Menge 1	Menge 2	Menge 3	Menge 4
Mittelwert von x	für jede der Mengen exakt: 9			
Varianz von x	für jede der Mengen exakt: 11			
Mittelwert von y	7,5009	7,5009	7,5000	7,5009
Varianz von y	4,127	4,128	4,123	4,123
Korrelation zw. x und y	0,816	0,816	0,816	0,817
Lin. Regression	für jede der Mengen: $y = 3,00 + 0,500x$			

Erst die Rohdaten aus Tabelle 2.1 oder eine Visualisierung wie in Abbildung 2.13 verdeutlicht die Unterschiede und zeigt, dass die vier Mengen einander keineswegs so ähnlich sind, wie die betrachteten statistischen Parameter suggerieren. Insbesondere zeigt sich, dass beispielsweise die Exklusion des vorletzten Datenpunktes aus Menge 3 oder des letzten Datenpunktes aus Menge 4 eine völlig andere Regressionsgleichungen ergäben, wohingegen die Entfernung eines einzelnen Punktes in den Mengen 1 und 2 keinen großen Unterschied herbeiführt.

Infolgedessen sollten bei Analysen neben den statistischen Parametern auch stets andere Methoden wie geeignete Visualisierungen zur Bewertung und Interpretation herangezogen werden. In der vorliegenden Arbeit werden bei statistischen Analysen daher neben der Angabe statistischer Parameter ergänzend Möglichkeiten einer geeigneten Visualisierung genutzt, um eine valide Interpretation der Ergebnisse zu ermöglichen.

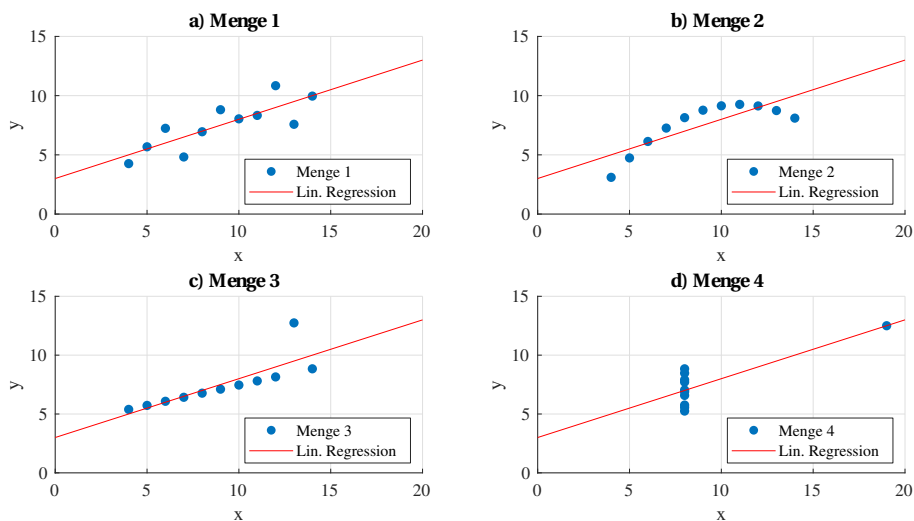


Abbildung 2.13.: Graphische Visualisierung der vier Mengen aus dem Anscombe-Quartett mit ihrer jeweiligen linearen Regression

Signifikanz und p-Wert

Bei statistischen Tests ist es nicht nur wichtig, ein für die Problemstellung geeignetes Testverfahren auszuwählen und die zugehörigen Voraussetzungen sorgsam zu überprüfen – auch die Bewertung der Signifikanz anhand des »p-Wertes« führt gerade in der Interpretation eine eigene Fehlerquelle mit sich.

Statistische Tests können genutzt werden, um zu überprüfen, ob zwei Gruppen als »gleichwertig« betrachtet werden können. Häufig wird dies jedoch in der Interpretation und Ergebnisdiskussion gleichgesetzt mit der Aussage, dass es »keine signifikanten Unterschiede zwischen den Gruppen« gibt. Amrhein et al. (2019) haben in einer Analyse von 791 Artikeln aus unterschiedlichen Journals festgestellt, dass etwa bei der Hälfte (51%) der dort interpretierten Testverfahren aus einem »nicht signifikanten« Test geschlossen wurde, dass kein Effekt vorhanden sei, also die Nullhypothese als bestätigt betrachtet werden könne. Dem liegt der Irrglaube zugrunde, dass die Nullhypothese als bewiesen gelte, wenn der Test die Nullhypothese nicht verwirft, das Testergebnis also »nicht signifikant« zum gegebenen α -Wert (d.h. $p \geq \alpha$) ist. Allerdings kann die Nullhypothese über einen statistischen Test *grundsätzlich nicht* bewiesen werden. Lediglich signifikante Testergebnisse liefern eine Aussage, die jedoch auch keinen festen Beweis für einen Effekt darstellt. Die einzige Aussage, die ein signifikantes statistisches Testverfahren liefert, ist, dass die Nullhypothese mit der zuvor festgelegten *Wahrscheinlichkeit* $1 - \alpha$ nicht gilt bzw. dass die Alternativhypothese mit Wahrscheinlichkeit $1 - \alpha$ gültig ist. Je kleiner der berechnete p-Wert ist, desto eher liefert er demnach ein gutes *Indiz* für die Gültigkeit der Alternativhypothese. Jedoch sagt auch ein kleiner p-Wert nicht aus, welche Hypothese nun tatsächlich wahr ist: Mit $p < \alpha$ gilt *entwe-*

der, dass tatsächlich die Nullhypothese zu verwerfen ist und die Alternativhypothese gilt, oder dass zwar die Nullhypothese gilt, in der Stichprobe des Tests jedoch zufällig eher untypische Daten verwendet wurden (mit Irrtumswahrscheinlichkeit α). Bei einer fehlenden Signifikanz ($p \geq \alpha$) ist keine sinnvolle Aussage möglich. Ein Test, der zu einem gegebenen α -Wert anhand des berechneten p-Wertes als nicht signifikant gewertet wird, sagt lediglich aus, dass es anhand der statistischen Kriterien und unter Verwendung der Daten dieser zufälligen Stichprobe *nicht* möglich ist, die Nullhypothese mit Wahrscheinlichkeit $1 - \alpha$ zu *verwerfen*. Eine Aussage über die Gültigkeit der Nullhypothese kann damit jedoch nicht getroffen werden. Insbesondere ist dabei zu beachten, dass der p-Wert bei wahrer Nullhypothese auf dem Intervall $[0,1]$ gleichverteilt ist (Colquhoun, 2014; Hung et al., 1997), also mit gleicher Wahrscheinlichkeit jeden Wert dieses Intervalls annehmen kann. Bei einer wahren Nullhypothese ist ein großer p-Wert als Testergebnis entsprechend ebenso wahrscheinlich wie ein kleiner p-Wert.

Amrhein et al. (2017) haben in ihrem Review dargestellt, warum die bloße Unterscheidung zwischen »signifikant« und »nicht signifikant« anhand des p-Wertes problematisch ist. Unter anderem führen sie mehrere Studien an, die gezeigt haben, dass Testwiederholungen zu völlig unterschiedlichen Ergebnissen führen können – unabhängig davon, welche Hypothese wahr ist (Boos und Stefanski, 2011). Beispielsweise konnte Cumming (2008) zeigen, dass eine Studie bei vorgegebener Irrtumswahrscheinlichkeit $\alpha = 0,2$ mit erstem Testergebnis $p = 0,05$ bei Wiederholung mit 80%iger Wahrscheinlichkeit einen p-Wert zwischen 0,00008 und 0,44 liefert und mit 10%iger Wahrscheinlichkeit sogar $p > 0,44$ liefert. Lazzeroni et al. (2014) untersuchten eine Studie, die bei Irrtumswahrscheinlichkeit $\alpha = 0,05$ zunächst einen p-Wert von $p = 0,5$ lieferte. Eine mehrfache Wiederholung des Tests führte in 95% der Fällen zu einem p-Wert zwischen 0,003 und 0,997. Diesen Effekt hat Cumming (2013) als »Tanz der p-Werte« beschrieben. Erklären lässt sich der Effekt damit, dass der p-Wert basierend auf Zufallsvariablen berechnet wird und damit selbst dem Zufall unterliegt.

Ein weiterer zentraler Aspekt bei der Verwendung statistischer Testverfahren ist die Wahl von α für die Wahrscheinlichkeit eines Fehlers 1. Art (falsche Testentscheidung für die Alternativhypothese H_1 , obwohl H_0 wahr wäre) und damit die Festlegung einer Grenze für den p-Wert. In der Realität liegt häufig kein großer Unterschied zwischen beispielsweise einem Testergebnis mit $p = 0,048$ und einem mit $p = 0,053$ vor, wenngleich dies durch die Signifikanz-Aussage suggeriert wird, die in diesem Fall bei $\alpha = 0,05$ unterschiedlich ausfallen würde. Das verdeutlicht, dass der p-Wert alleine und eine aus dem p-Wert gefällte Entscheidung für »signifikant« oder »nicht signifikant« eine geringe Aussagekraft hat und mit vielen Unsicherheiten behaftet ist. Eine Ergänzung der Effektgröße und eines Unsicherheits- bzw. Konfidenzintervalls könnten Korner-Nievergelt und Hüppop (2016) und Ioannidis (2019b) zufolge zu einem besseren Verständnis des Testergebnisses und damit zu sinnvolleren Interpretationen der Ergebnisse beitragen. Auch Knottnerus und Tugwell (2020) betonen, dass neben der Signifikanz-Entscheidung unbedingt das Konfidenzintervall sowie die

Größe der in der Studie untersuchten Population und die »statistische power« (die Sensitivität des Tests, also die Wahrscheinlichkeit, dass ein existierender Effekt auch entdeckt wird: $1 - \beta$) zur Interpretation der Ergebnisse herangezogen werden sollten.

Ioannidis (2019a) betont jedoch auch, dass dies nicht bedeutet, dass statistische Tests und die Verwendung des p-Werts keinen Nutzen haben – wichtig sei jedoch, die statistischen Ergebnisse als nur einen von mehreren Bausteinen für die Interpretation und ggf. darauf basierende Entscheidungen zu betrachten. Auch mögliche Alternativen zu statistischen Tests müssten zunächst geeignet analysiert und evaluiert werden.

Im Rahmen der hier vorliegenden Arbeit werden bei statistischen Analysen daher neben der häufig verwendeten zweiteiligen Bewertung »signifikant« vs. »nicht signifikant« auch qualitative Visualisierungen und numerische Ergebnisse sowie bei für das Ergebnis zentralen Tests nach Möglichkeit entsprechende Konfidenzintervalle oder die statistische Power zur Interpretation hinzugezogen.

Über die Wahl des statistischen Tests basierend auf Tests zur Überprüfung der Voraussetzungen

Bei der im Rahmen der Vorstellung gängiger statistischer Testverfahren beschriebenen Vorgehensweise (S. 69) werden üblicherweise mehrere statistische Tests hintereinander auf den gleichen Daten durchgeführt: Bevor eine Entscheidung für den t-Test, den Welch-Test oder den Wilcoxon-Rangsummentest getroffen werden kann, werden mithilfe zweier Tests die Daten hinsichtlich ihrer Normalverteilung und ihrer Varianzhomogenität überprüft. Die Kombination aus einem Test (oder mehreren Tests) zur Überprüfung der Voraussetzung und nachfolgend einem statistischen Test zur Überprüfung der eigentlich relevanten Hypothese führt jedoch zu eigenen Herausforderungen. Ruxton (2006) bemerkt, dass der Fehler 1. Art bei der Kombination eines Vorabtests mit dem t-Test weniger gut kontrolliert werden kann, als bei der alleinigen Durchführung des Welch-Tests. Dazu verweist er unter anderem auf die Analysen von Gans (1981) und Moser und Stevens (1992): Gans (1981) hat gezeigt, dass es bei der Verwendung des t-Tests insbesondere dann zu höheren Fehlern 1. Art kommt, wenn die Stichprobengröße der beiden Gruppen unterschiedlich ist. Dieser Unterschied kann ihren Analysen zufolge auch nicht ausreichend durch einen der üblichen Vorabtests ausgeglichen werden. Die Ergebnisse mit dem Welch-Test hingegen zeigten sich auch in diesem Szenario robuster, sodass der Welch-Test besonders bei ungleichen Stichprobengrößen anstelle der Kombination aus Vorabtest und t-Test von Gans (1981) empfohlen wird. Da der Welch-Test ihren Untersuchungen zufolge jedoch auch bei gleichen Stichprobengrößen nicht unzuverlässiger ist als die Kombination aus Vorab-Test und t-Test, plädiert er dafür, grundsätzlich den Welch-Test (ohne vorherige Testung auf Normalverteilung oder Varianzhomogenität) zu nutzen. Moser und Stevens (1992) beschäftigten sich konkret mit der Frage, inwieweit ein vorheriger Test auf Varianzhomogenität vor der Durchführung eines t-Tests ge-

eignet und sinnvoll ist. Dabei kommen sie zu dem Ergebnis, dass diese Vorgehensweise (erst ein Test auf Varianzhomogenität und basierend auf dem Ergebnis ein entsprechender nachfolgender Test) lediglich einen zusätzlichen Analyseschritt ohne statistischen Mehrwert beinhaltet. Basierend auf dem Vergleich unterschiedlicher Testverfahren führen sie eine Fallunterscheidung durch: Weichen die Stichprobengrößen nicht voneinander ab, haben der t-Test und u.a. der Welch-Test die gleiche Sensitivität und die gleiche Spezifität. Lediglich bei ungleichen Stichprobengrößen zeigt sich ein Unterschied – zugunsten des Welch-Test. Während der t-Test bei ungleichen Stichprobengrößen und einer vorherrschenden Varianzheterogenität (d.h. die Differenz der Varianzen weicht um mehr als 1 voneinander ab) eine hohe Wahrscheinlichkeit für einen Fehler 1. Art und entsprechend eine schlechte Sensitivität (»power«) liefern kann, zeigt sich der Welch-Test auch in diesem Szenario robust. Sie ziehen das Fazit, dass ein vorheriger Test auf Varianzhomogenität grundsätzlich nicht angebracht ist und es sinnvoller ist, von vorneherein den Welch-Test anstelle des t-Tests zu verwenden. Auch Zimmerman (2004) kommt zu dem Ergebnis, dass der Welch-Test der Kombination aus Levene-Test und t-Test stets zu bevorzugen ist. Selbst bei Daten, die nicht normalverteilt sind, ist der Welch-Test im Vergleich zum Wilcoxon-Rangsummentest keine Nachteile hat, wie Zimmerman und Zumbo (1993) in ihrer Studie darlegen. Insbesondere können mit dem Welch-Test ausgeführt auf Rängen (wie beim Wilcoxon-Rangsummentest) zusätzlich valide Testergebnisse erzielt werden, wenn die Daten weder normalverteilt noch varianzhomogen sind.

Ruxton (2006) empfiehlt auf Basis dieser Resultate und eigenen Analysen, grundsätzlich den Welch-Test (»unequal variance t-test«) anstatt des t-Tests (»Student's t-test«) oder des Wilcoxon-Rangsummentests (»Mann-Whitney U test«) zu verwenden. Dabei schlägt er eine zweistufige Vorgehensweise vor: In einem ersten Schritt sollen die Verteilungen der beiden Stichproben visuell überprüft werden. Im zweiten Schritt kann entweder unmittelbar der Welch-Test durchgeführt werden, sofern die Daten visuell betrachtet normalverteilt aussehen. Kann nach graphischer Durchsicht bei mindestens einer der beiden Stichproben keine Normalverteilung angenommen werden, empfiehlt Ruxton (2006), die Daten zunächst wie beim Wilcoxon-Rangsummentest in Ränge zu sortieren und anschließend auf den sortierten Daten den Welch-Test durchzuführen. Ergänzend weist er darauf hin, dass für eine geeignete Interpretation des Testergebnisses neben dem p-Wert zusätzlich der Mittelwert, die Varianz, die Größe der Stichprobe in jeder Gruppe, der berechnete t-Wert und der berechnete Freiheitsgrad betrachtet werden sollten.

Auch Rasch et al. (2011) schließen sich diesen Ansichten an und empfehlen, gänzlich auf vorherige Tests zur Überprüfung der Voraussetzung und der Verwendung des t-Tests oder des Wilcoxon-Rangsummentests zu verzichten. Stattdessen raten auch sie dazu, für die Überprüfung von Hypothesen bezüglich eines gleichen Mittelwertes von vorneherein den Welch-Test zu verwenden.

In der vorliegenden Arbeit wird daher ebenfalls dem Welch-Test der Vorzug gegenüber einer Kombination mehrerer Tests gegeben. Entsprechend finden weder der t-Test noch der Wilcoxon-Rangsummentest oder Tests zur Überprüfung der Normalverteilung und der Varianzhomogenität in den durchgeführten Studien Verwendung. Je nach Analyse wird der Welch-Test durch weitere statistische Kennzahlen oder graphische Darstellungen der Stichproben anhand von QQ-Plots und insbesondere der exemplarischen Visualisierung der Simulationsergebnisse ergänzt.

Korrelationskoeffizienten

Korrelationskoeffizienten wie Pearson's r , der Intraklassen-Koeffizient ICC oder auch der Determinationskoeffizient R^2 werden häufig genutzt, um die Ähnlichkeit zweier Messreihen in Bezug auf ihren Verlauf darzustellen. Dabei dürfen die Messreihen im Sinne eines RMSE-Wertes oder eines anderen Abstandsmaßes durchaus stark voneinander abweichen; relevant ist lediglich der inhärente Zusammenhang der einzelnen Datenpunkte innerhalb der einzelnen Messreihen und wie sich dieser im Vergleich zur anderen Messreihe verhält. In zahlreichen Publikationen (beispielsweise bei Bland und Altman (1986); Borresen und Lambert (2009); Borszcz et al. (2018); Buchheit et al. (2010); Tanaka et al. (1979) u.v.m.) werden Korrelationskoeffizienten als Gütemaß einer akkuraten Simulation oder Vorhersage auch bei Zeitreihen betrachtet. Allerdings ist zur Bestimmung der Korrelation zweier Zeitreihen eine vorhergehende *Trendbereinigung* (engl. *detrend*) unabdingbar, die systematische Änderungen ausblendet. Ohne eine solche Trendbereinigung ist ein angegebener Korrelationskoeffizient als Gütekriterium der Simulations- oder Vorhersagegenauigkeit ungeeignet und kann lediglich eine Scheinkorrelation beschreiben, jedoch keine echte Korrelation quantifizieren. Das liegt daran, dass bei Zeitreihen stets eine systematische Abhängigkeit der Messwerte von der Zeit vorhanden ist. Eine berechnete Korrelation zweier (nicht-trendbereinigter) Zeitreihen stellt somit fest, wie stark beide Zeitreihen *mit der Zeit* korrelieren, sagt jedoch nichts über eine mögliche Korrelation der Zeitreihen *zueinander* aus. Dieser Effekt wird nachfolgend anhand eines Beispiels detailliert erläutert.

Zur Durchführung einer linearen Trendbereinigung einer Zeitreihe wird die Regressionsgerade über die »Methode der kleinsten Quadrate« bestimmt. Anschließend wird die entstandene Regressionsgerade von den Simulationsdaten subtrahiert. Ob vor der Berechnung der Korrelationskoeffizienten tatsächlich eine Trendbereinigung durchgeführt wurde, wird häufig jedoch nicht in den entsprechenden Publikationen erwähnt. Es ist daher bei vielen dieser Publikationen unklar, ob die Daten trendbereinigt wurden oder nicht.

Dieser Effekt lässt sich mit zwei zufälligen Zeitreihen veranschaulichen: Werden zwei völlig zufällige Zahlenreihen Y_1 und Y_2 mit je 100 Werten im Bereich 0 – 1 erstellt, ist keine Korrelation zu erwarten. Zwei solche Reihen sind in Abbildung 2.14 dargestellt, der Korrela-

tionskoeffizient zwischen Y_1 und Y_2 in diesem Beispiel beträgt $-0,1776$, es besteht keine Korrelation.

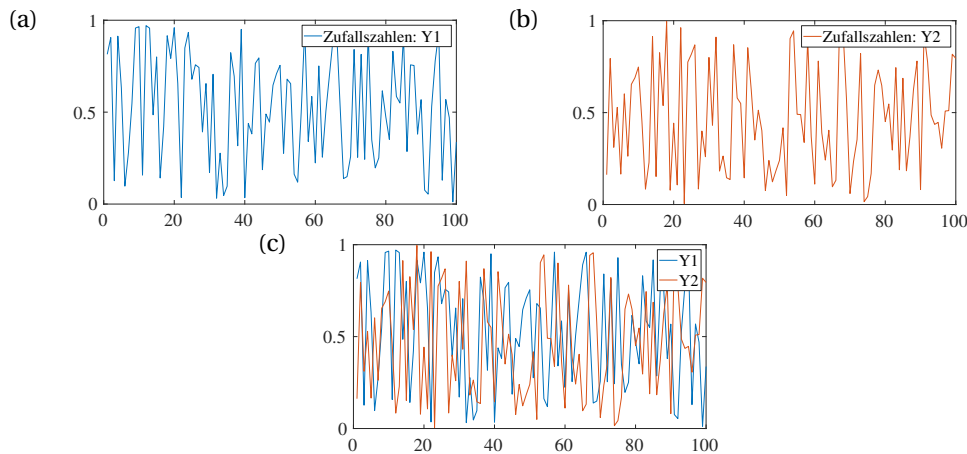


Abbildung 2.14.: Plots zu (a) den Zufallszahlen Y_1 , (b) den Zufallszahlen Y_2 , (c) den Zufallszahlen Y_1 und Y_2 übereinandergeplottet

Die Analyse ändert sich hingegen deutlich, wenn beiden Reihen künstlich der selbe »Trend« hinzugefügt wird: Hierzu kann beispielsweise eine Gerade mit einem geringen Anstieg erstellt werden, anhand derer die beiden Reihen ausgerichtet werden. Die Gerade ℓ in diesem Beispiel weist über die 100 Werte einen geringen Anstieg von 6 (im Bereich von -3 bis 3) auf (Abbildung 2.15 (a)). Für beide Reihen Y_1, Y_2 wurden die einzelnen Werte zum jeweiligen korrespondierenden x -Wert (entsprechend eines Zeitpunktes in einer Zeitreihe) der Geraden ℓ addiert. Die entstandenen Reihen Y_{1_t}, Y_{2_t} sind mit einem Trend behaftet: $Y_{1_t}(x) = Y_1(x) + \ell(x)$, $Y_{2_t}(x) = Y_2(x) + \ell(x)$. Diese neuen Reihen mit Trend sind in Abbildung 2.15 (b) dargestellt. Allein durch die Ergänzung der Reihen um einen Trend steigt der Korrelationskoeffizient auf $0,9679$ an, was einer nahezu perfekten Korrelation entspricht.

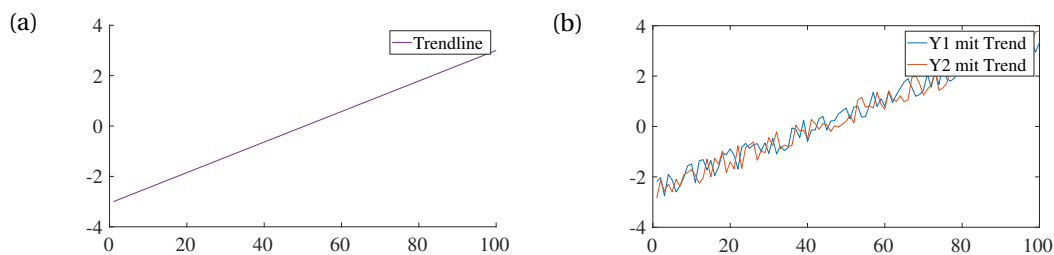


Abbildung 2.15.: (a) Die im Beispiel genutzte Trendlinie; (b) Die Zufallszahlen mit künstlich ergänztem Trend Y_{1_t} und Y_{2_t}

Zur Verdeutlichung sind in Abbildung 2.16 vier weitere Beispiele mit unterschiedlichen Trends dargestellt:

- (a) Reihe Y_{1_t} ist wie zuvor mit dem Trend aus Gerade ℓ behaftet. Für Reihe Y_{2_t} ist ein Trend gewählt, bei dem die zugrundeliegende Gerade ℓ_2 über die 100 Werte einen stär-

keren Anstieg von 40 (im Bereich von -20 bis 20) aufweist. Der Korrelationskoeffizient beträgt 0.9846 .

- (b) Reihe $Y1_t$ ist wie zuvor mit dem Trend aus Gerade ℓ behaftet. Für Reihe $Y2_{t3}$ ist ein Trend gewählt, bei dem die zugrundeliegende Gerade ℓ_2 über die ersten 70 Werte einen stärkeren Anstieg von 40 (im Bereich von -20 bis 20) aufweist. Für die hinteren 30 Werte wurde ein negativer Trend ergänzt. Der Korrelationskoeffizient beträgt 0.7749 .
- (c) Reihe $Y1$ ist ohne Trend. Für Reihe $Y2_{t4}$ ist ein sehr geringer Trend gewählt, bei dem die zugrundeliegende Gerade ℓ_4 über die 100 Werte einen minimalen Anstieg von 1 (im Bereich von -0.5 bis 0.5) aufweist. Der Korrelationskoeffizient beträgt -0.2361 .
- (d) Reihe $Y1_{t5}$ ist ein geringer Trend gewählt, bei dem die zugrundeliegende Gerade ℓ_5 über die 100 Werte einen minimalen Anstieg von 2 (im Bereich von -1 bis 1) aufweist. Für Reihe $Y2_{t4}$ ist ein ähnlich geringer, aber abweichender Trend (wie aus Beispiel c) gewählt, bei dem die zugrundeliegende Gerade ℓ_4 über die 100 Werte einen minimalen Anstieg von 1 (im Bereich von -0.5 bis 0.5) aufweist. Der Korrelationskoeffizient beträgt 0.578 .

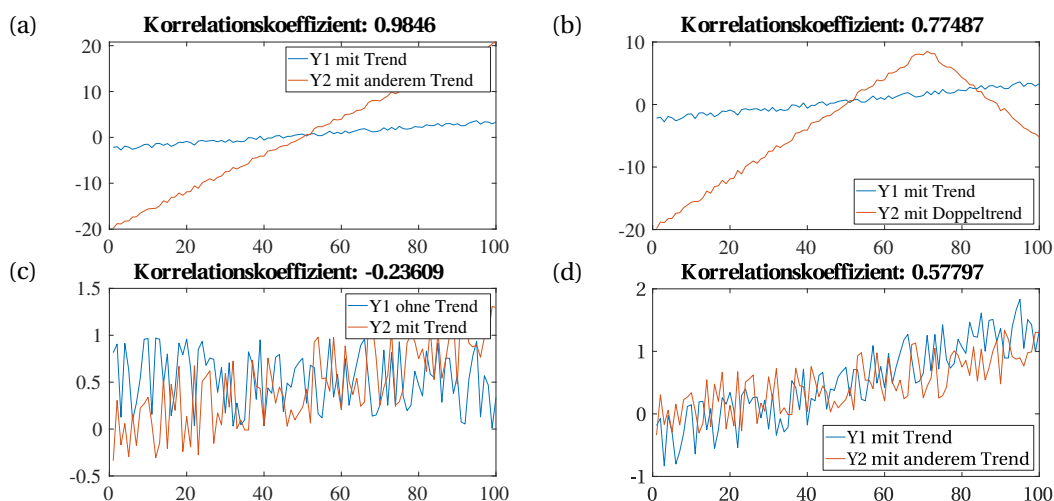


Abbildung 2.16.: (a) Zwei unterschiedliche Trends für $Y1_t$ und $Y2_{t2}$; (b) Zwei unterschiedliche Trends für $Y1_t$ und $Y2_{t3}$, wobei $Y2_{t3}$ einen Doppeltrend aufweist; (c) $Y1_t$ ist ohne Trend, $Y2_{t4}$ hat einen geringen Trend; (d) $Y1_{t5}$ und $Y2_{t4}$ haben beide einen geringen, voneinander abweichenden Trend

Obwohl die beiden Reihen in Abbildung 2.16 (a) und (b) optisch durch die unterschiedlichen Trends keine Gemeinsamkeit haben, werden in beiden Fällen hohe oder gar sehr hohe Korrelationen detektiert. Der Vergleich von Abbildung 2.16 (c) und (d) verdeutlicht, dass bereits ein geringer, unterschiedlicher Trend in beiden Reihen zu einer deutlich höheren berechneten Korrelation führt. Sobald eine Reihe nicht trendbehaftet ist, kann auch auf Basis des Korrelationskoeffizienten keine Korrelation mehr festgestellt werden. Insbesondere zeigt der Vergleich von Abbildung 2.16 (a), Abbildung 2.16 (d), dass ein stärkerer Trend einer

der Reihen eine höhere Korrelation suggeriert, da hier die tatsächlichen Werte der Zufallszahlen weniger stark ins Gewicht fallen, wie die Existenz eines Trends.

Die einzig detektierbare Korrelation, die den unterschiedlichen Reihen von Zufallszahlen tatsächlich zugrunde liegt, betrifft jedoch den (hier künstlich erzeugten) Trend. Eine vergleichbare Abhängigkeit existiert bei Zeitreihen systemimmanent, da Zeitreihen per definitionem von der Zeit abhängen. Eine Korrelationsanalyse ohne vorherige Untersuchung und ggf. Bereinigung eines Trends ist bei Zeitreihen entsprechend kritisch zu betrachten. Bei den in dieser Arbeit durchgeführten Analysen werden Korrelationen ausschließlich bei geeigneten Fragestellungen und nach vorheriger Trendbereinigung berechnet.

Besonderheit: Korrelationskoeffizienten in der Analyse der Leistungsentwicklung

Modelle der Trainingswirkungsanalyse befassen sich mit der Trainingswirkung über einen längeren – in der Regel mehrwöchigen – Zeitraum und der damit zusammenhängenden Veränderung der Leistungsfähigkeit, also der Leistungsentwicklung. Pfeiffer (2020) beschreibt Modelle der Trainingswirkungsanalyse als »Modelle zur prozessualen Abbildung der Relation von Training und Leistung«. Dabei ist das Verhältnis von externer Trainingsbelastung bzw. interner Trainingsbeanspruchung und der sportlichen Leistungsfähigkeit nicht linear; es wird durch unterschiedliche Anpassungsprozesse beeinflusst. Pfeiffer (2020) benennt in diesem Kontext insbesondere drei zentrale Aspekte:

- Die Dynamik der Anpassungsprozesse: Eine Belastungsreiz führt erst mit einer zeitlichen Verzögerung zu Adaptationsprozessen im Körper (vgl. »Physiologische Reaktionen und Anpassungen durch Ausdauerbelastung«, Unterabschnitt 1.1.3). Zusätzlich beeinflusst der tagesaktuelle Leistungszustand einer Person die Reaktion auf einen Trainingsreiz, sodass auch identische Trainingsreize zu unterschiedlichen Zeitpunkten verschiedene Beanspruchungsreaktionen und in Folge dessen verschieden starke Anpassungsprozesse hervorrufen können;
- Die hohe Individualität: Die körpereigenen Anpassungsprozesse sind in ihrer zeitlichen, qualitativen und quantitativen Ausprägung hochgradig individuell;
- Die Beeinflussung des Gesamtsystems im Training: Auch gezieltes und systematisches Training führt i.d.R. nicht dazu, dass ein einzelner physiologischer Prozess ausgelöst werden kann. Jedes Training wirkt sich fast immer auf das »Gesamtsystem« aus.

In der Trainingswirkungsanalyse dienen die über mehrere Wochen hinweg erfassten Trainings- und Belastungsdaten als Eingabewerte für ein parametrisiertes Modell. Nach Individualisierung der Parameter über ein Kalibrierungsverfahren kann so die Leistungsfähigkeit approximiert werden. Ist das Modell an den Probanden angepasst, kann es optimalerweise schließlich auch zur Trainingsplanung genutzt werden, indem der gewünschte Leistungsverlauf vorgegeben und umgekehrt die dazu notwendige Belastung geschätzt wird.

Insbesondere lässt sich auf diese Weise ein Trainingsverlauf planen, der eine Überlastung der Probandin oder des Probanden vermeidet und die benötigte Belastung minimiert.

Die Daten, die einer Analyse der Leistungsentwicklung zugrunde liegen, gehören ebenfalls zu den Zeitreihen. Dabei unterliegen sowohl empirisch gemessene wie auch simulierte Leistungswerte tagesformabhängigen Schwankungen. Werden beispielsweise Laktatwerte zur empirischen Leistungsüberprüfung gemessen, ist zu berücksichtigen, dass die tagesaktuellen Laktatwerte unter anderem durch die Ernährung, durch sportliche Aktivitäten, die Außentemperatur oder bei Frauen durch den Zeitpunkt im Menstruationszyklus beeinflusst werden (Berend et al., 1994; No und Kwak, 2016). Eine vollständige Ausbelastung innerhalb der Leistungsdiagnostik wiederum ist gerade im Leistungssportbereich anfällig für Schwankungen in der Motivation und dem allgemeinen Befinden während der Untersuchung (Andreacci et al., 2002; Wilmore, 1968). Auch die Herzfrequenz unterliegt zahlreichen Schwankungen wie sie bereits in Unterabschnitt 1.3.1: »Die Herzfrequenz als Indikator: Kritische Aspekte« diskutiert wurden.

Für die Überprüfung der Leistungsentwicklung sind tagesformabhängige Schwankungen hingegen selten⁴ von Interesse. Gerade bei austrainierten Profi-Athlet*innen im Leistungssport-Bereich sind über einen längeren Zeitraum kaum Schwankungen in der Leistungsentwicklung festzustellen, sodass kleine Schwankungen auf Basis der Tagesform während der regulären Trainingsphasen keine wichtige Rolle spielen. Umgekehrt können gerade diese Schwankungen innerhalb der Modellierung je nach Skalierung der Daten groß erscheinen, obwohl sie innerhalb der Leistungsentwicklung irrelevant sind. Wichtiger zur Analyse der Leistungsentwicklung ist daher die *Trendentwicklung* über einen größeren Zeitraum. Je nach Trainingszustand der Athletin oder des Athleten entspricht der Trend häufig einer Nulllinie. Lediglich nach einer Trainingspause oder innerhalb einer spezifischen Wettkampfvorbereitung stellt sich teilweise eine leicht positive Trendentwicklung ein.

Innerhalb der Modellierung ist daher die Analyse des Trends selbst anstelle von tagesformabhängigen Schwankungen von Interesse, sodass in diesem Fall auf eine Trendbereinigung der Daten verzichtet werden sollte. Im Gegenteil: Je nach Untersuchungsziel kann es sinnvoll sein, explizit den Trend zu analysieren. Bei den später durchgeführten Analysen zur Feststellung, ob die Parameterveränderung bei der Simulation mit dem Faltungsmodell der Veränderung der empirischen Leistung entspricht, wird daher einerseits auf eine Trendbereinigung verzichtet, andererseits ergänzend spezifisch der Trend beider Zeitreihen betrachtet.

⁴Eine Ausnahme hiervon bildet beispielsweise die Vorhersage der Leistung für einen Wettkampf

2.2. Modellierung der Herzfrequenz als akute Anpassung

Zur gezielten Modellierung der akuten Reaktion des menschlichen Körpers auf physikalische Belastung und Aktivität können prinzipiell unterschiedliche Arten von Modellen betrachtet werden. Im Kontext dieser Arbeit werden dazu speziell Modelle zur Simulation der Herzfrequenz betrachtet. Herzfrequenzmodelle werden üblicherweise dazu genutzt, die Herzfrequenz während des Trainings zu kontrollieren und so das eigene Training gezielt zu steuern (Su et al., 2010a). Auch die Belastung auf dem Fahrrad- oder Laufbandergometer kann abhängig von der Herzfrequenz reguliert werden. Für derartige Anwendungen genügt es im Allgemeinen, die Herzfrequenzen der nächsten 2-30 Sekunden vorherzusagen. Hierzu existieren zahlreiche Modellvarianten. Viele der Modelle basieren dabei auf Differentialgleichungsmodelln in unterschiedlichen Varianten, wie beispielsweise Linear-Zeitinvariante-Modelle (LTI-Modelle, engl. *linear time-invariant models*) Hammerstein-Modelle, Wiener-Modelle und Hammerstein-Wiener-Modelle (Baig et al., 2010; Cheng et al., 2007a; Mohamad et al., 2011a; Su et al., 2010a)

Die nachfolgenden beiden Exkurse bieten eine kurze Übersicht über die für die HF-Modellierung zentralen Charakteriska von Differentialgleichungsmodelle (DGL-Modelle) und Hammerstein-Wiener-Modellen, da insbesondere Hammerstein-Wiener-Modelle in den späteren Analysen (»Analyse des Faltungsmodells: Modellvergleich«, Abschnitt 5.2) von besonderem Interesse sind.

Exkurs: Differentialgleichungen

Grundlegend gibt es zunächst zwei unterschiedliche Klassen von Differentialgleichung: gewöhnliche und partielle Differentialgleichungen. Während gewöhnliche Differentialgleichungen (DGLen) von einer einzigen Variable abhängig sind (beispielsweise von der Zeit), sind partielle DGL häufig von mehreren Variablen abhängig und können zudem partielle Ableitungen enthalten. Da partielle Differentialgleichungen für die innerhalb dieser Arbeit betrachteten Modelle keine Rolle spielen, werden diese hier nicht betrachtet. Weiterhin werden DGLen in lineare und nichtlineare DGLen unterteilt. In linearen DGLen treten die gesuchte Funktion und ihre Ableitungen ausschließlich linear auf. Bei nichtlinearen DGLen können diese als Argument einer nichtlinearen Funktion wie beispielsweise einer Sinus-Funktion auftreten.

Ist im Rahmen dieser Arbeit allgemein von Differentialgleichungen ohne nähere Spezifikation die Rede, sind stets *gewöhnliche, lineare Differentialgleichungen* gemeint.

Für eine geeignete Interpretation werden (gewöhnliche) Differentialgleichungen und zwei spezielle Typen von Differentialgleichungen – die Anfangswertprobleme und lineare Differentialgleichungssysteme – im Folgenden kurz motiviert.

Allgemein betrachtet beinhalten Differentialgleichungen eine Funktion mit einer unabhängigen Variablen und deren Ableitung(en), d.h. $f : X \rightarrow Y$ ist eine Differentialgleichung, wenn f und mindestens eine Ableitung dieser Funktion in der Gleichung vorkommen. Im Kontext der Herzfrequenz- oder Leistungsmodellierung, die im Rahmen dieser Arbeit stets als Zeitreihen betrachtet werden, ergeben sich für das Aufstellen von Differentialgleichungen drei wesentliche Aspekte:

1. Es ist insbesondere die Ableitung nach der Zeit von Interesse;
2. Die Bildung von mindestens der ersten Ableitung ist möglich, da solche Zeitreihen i.d.R. stetig und differenzierbar sind;
3. Die äußere Belastung hat im Kontext eines einzelnen, abgeschlossenen Trainings keine Vergangenheit, sodass alle Werte für Zeiten $t \leq 0$ gleich Null gesetzt werden können.

Die erste Ableitung einer Funktion nach der Zeit lässt sich als *Veränderung über die Zeit* interpretieren. Bildet die Funktion also beispielsweise den Verlauf einer Herzfrequenz (über die Zeit) ab, gibt ihre Ableitung an, wie sich die Herzfrequenzen im Zeitverlauf verändern, also ob die Werte der Herzfrequenzfunktion steigen oder fallen. Gerade für Herzfrequenzverläufe ergibt daher durchaus Sinn, genau diejenige Änderungsrate zu betrachten, die beschreibt, wie sich die Herzfrequenz im Vergleich zu ihren vorherigen Werten entwickelt hat – also die erste Ableitung zu betrachten. Die zweite Ableitung entspricht einer »Veränderung dieser Veränderungen« – also beispielsweise eine durch den kardialen Drift induzierte Abweichung der ansonsten üblichen Veränderung der Herzfrequenz über die Zeit.

Ein einfaches Beispiel für eine DGL ist das Wachstum einer Population: Die Differentialgleichung $\dot{f}(t) = a \cdot f(t)$ beschreibt, wie das Wachstum (der Population) und die Funktion (die die Population z.B. in Abhängigkeit der Zeit t beschreibt) zusammenhängen. Da diese Gleichung nach der höchsten Ableitung (hier: \dot{f}) aufgelöst dargestellt ist, handelt es sich um eine *explizite* Darstellung der DGL. Dieselbe DGL kann *implizit* dargestellt werden als $\dot{f}(t) - a \cdot f(t) = 0$, d.h. die Gleichung wird nach Null aufgelöst angegeben.

Werden mehrere Differentialgleichungen simultan betrachtet, handelt es sich um ein *Differentialgleichungssystem*.

Hat eine gewöhnliche Differentialgleichung $\dot{y}(t) = f(t, y(t))$ als Nebenbedingung, dass ein vorgegebener Wert einen spezifischen Funktionswert hat, z.B. $y(t_0) = y_0$, liegt ein *Anfangswertproblem* vor. Dabei ist y eine Lösung für das Anfangswertproblem, wenn y beide Gleichungen – also $\dot{y}(t) = f(t, y(t))$ und zugleich $y(t_0) = y_0$ – erfüllt.

Exkurs:

Hammerstein- und Wiener-Modelle bestehen aus einem linearen und einem nichtlinearen Anteil, wobei der lineare Anteil auch als lineare DGL dargestellt werden kann.

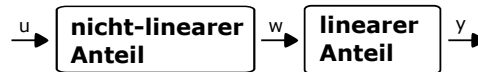


Abbildung 2.17.: Box-Diagramm für Hammerstein-Modelle mit Eingangssignal u , einem durch eine nicht-lineare Funktion transformierten Signal w , und dem durch eine zusätzliche lineare Funktion transformierten Ausgangssignal y

Ein Hammerstein-Modell (Eskinat et al., 1991; Hunter und Korenberg, 1986; Narendra und Gallman, 1966) wie in Abbildung 2.17 veranschaulicht, wird verwendet, um nichtlineare Systeme zu modellieren, bei denen die Nichtlinearität direkt am Modell-Input vorhanden ist. Damit setzt sich dieser Modelltyp zusammen aus einer nichtlinearen Funktion des Eingangssignals u zum Zeitpunkt t , z.B. einer q -Grad Polynomfunktion

$$w(t) = f(u(t)) = a_0 + a_1 u(t) + a_2 u(t)^2 + \dots + a_q u(t)^q, \quad (2.1)$$

gefolgt von einem linearen Teil, z.B. einem LTI-Modell. Im Allgemeinen ist die Wirkung eines LTI-Systems auf ein Eingangssignal durch eine Faltung des Eingangssignals mit der Impulsantwortfunktion des Systems gegeben, die die Dynamik des Systems charakterisiert (Chinarro, 2014). So kann das LTI-Modell durch eine Faltung dargestellt werden. Das resultierende Hammerstein-Modell mit Ausgangssignal y kann dann wie in Gleichung 2.2 beschrieben werden:

$$y(t) = (h * w)(t) = \sum_{\tau} h(\tau) w(t - \tau), \quad (2.2)$$

wobei die Funktion h die entsprechende Impulsantwortfunktion des Systems ist.

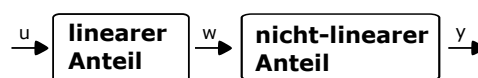


Abbildung 2.18.: Box-Diagramm für Wiener-Modelle mit Eingangssignal u , einem durch eine lineare Funktion transformierten Signal w , und dem durch eine zusätzliche nicht-lineare Funktion transformierten Ausgangssignal y

Umgekehrt wird ein Wiener-Modell (Billings und Fakhouri, 1977; Hunter und Korenberg, 1986) wie in Abbildung 2.18 dargestellt, verwendet, um nichtlineare Systeme zu modellieren, bei denen die Nichtlinearität am Ausgang des Systems auftritt. Der erste Teil eines Wiener Modells ist daher ein lineares System, das erneut unter Verwendung der entsprechenden Impulsantwort h als Faltung ausgedrückt werden kann,

$$w(t) = f(u(t)) = (h * u)(t) = \sum_{\tau} h(\tau) u(t - \tau), \quad (2.3)$$

gefolgt von einem nichtlinearen Teil, z.B. einer q -Grad-Polynomfunktion wie in Gleichung 2.1, sodass das Wiener-Modell mit dem Ausgangssignal y wie in Gleichung 2.4 beschrieben werden kann als

$$y(w(t)) = a_0 + a_1 w(t) + a_2 w^2(t) + \dots + a_q w^q(t). \quad (2.4)$$

Eine Kombination der beiden Modelltypen zu einem Hammerstein-Wiener- oder einem Wiener-Hammerstein-Modell entspricht der Verkettung beider Systeme. Bei einem Hammerstein-Wiener-Modell wird also das Eingangssignal zunächst nichtlinear transformiert, gefolgt von einer linearen Transformation (Hammerstein-Modell), und erneut gefolgt von einem nichtlinearen Anteil (Wiener-Modell). Das Wiener-Hammerstein-Modell ist analog umgekehrt verkettet.

In Unterabschnitt 2.2.2, »Herzfrequenzmodellierung: Stand der Forschung«, werden gängige Herzfrequenzmodelle aus der Literatur für einen Überblick des Forschungsstandes zusammengefasst und ihre publizierten Modelleigenschaften werden näher beleuchtet. Wenn gleich sich die bekannten Modelle gut zur Vorhersage der Herzfrequenz innerhalb kurzer Zeiträume (üblicherweise im Bereich 2-30 Sekunden) eignen, scheinen kaum geeignete Modelle zu existieren, die die Herzfrequenz für eine komplette Trainingseinheit vorhersagen können. Gonzalez et al. (2016) haben ein solches Modell zur Vorhersage der Sauerstoffaufnahme entwickelt. Das Modell entstammt einer automatisierten Modellidentifikation einer MATLAB-Toolbox und wird weder physiologisch, noch phänomenologisch erklärt.

Zum Zweck von Vorhersagen vollständiger Trainingseinheiten wurde daher ein eigenes Modell (*Faltungsmodell*) entwickelt. Das Faltungsmodell ist für die Vorhersage der Herzfrequenz kompletter Trainingseinheiten gut geeignet und weist dabei geringere Fehler auf, als unterschiedliche Modellvarianten aus der Literatur (Ludwig et al., 2016). Teile dieses Kapitels entsprechen den Inhalten aus Ludwig et al. (2018).

2.2.1. Anforderungen an Modelle zur Herzfrequenzmodellierung

Die Modellierung phänomenologischer Aspekte setzt die Kenntnis darüber voraus, welche Phänomene ein Modell abbilden soll und wie relevant welches Phänomen in der Modellierung ist. Die Beanspruchungsreaktion der HF auf Belastung mit unterschiedlicher Intensität und Dauer beinhaltet insbesondere im Ausdauersport vier zentrale Phänomene:

1. der verzögerten exponentiell abgeschwächten HF-Reaktion bei ansteigender oder abschwächender Belastung, z.B. durch variierende Geschwindigkeits-, Steigungs- oder Gefällebedingungen auf einer Strecke (Boucsein, 2000; Bunc et al., 1988);
2. dem S-förmiges HR-Verhalten bei kontinuierlicher inkrementeller Belastung (Brooke und Hamley, 1972);

3. dem kardiovaskulären Drift (engl. *cardiac drift*) bei länger andauernder Aktivität (Heaps et al., 1994);
4. der Erschöpfungsreaktion, die auch als »Hitting the wall« bezeichnet wird; dies wird als der Moment beschrieben, in dem die Glykogenvorräte erschöpft sind und Energie vollständig aus Fetten erzeugt werden muss (Stevinson und Biddle, 1998).

Darüber hinaus können andere Phänomene wie eine Vorbelastung der HF oder die maximale HF einer Person direkt oder implizit in einem Modell berücksichtigt werden. Im folgenden Unterkapitel werden verschiedene Arten von Modellen vorgestellt und in Kürze die Ergebnisse der entsprechenden Studien zusammengefasst. Der Abschnitt endet mit einer Diskussion über die Verwendung der vorgestellten Modelle im Hinblick auf modellierte physiologische Phänomene.

2.2.2. Herzfrequenzmodellierung: Stand der Forschung

In den letzten Jahrzehnten wurden zahlreiche Modelle zur HF-Modellierung vorgestellt und diskutiert. Physiologische Modelle werden in der Regel konzipiert, um ein bestimmtes Verhalten eines biologischen Systems mit hoher Genauigkeit zu simulieren. Diese Simulationen des menschlichen Herz-Kreislauf-Systems umfassen eine Vielzahl unterschiedlicher Zwecke und decken große Komplexitätsunterschiede ab.

Zum Beispiel beschreibt Grodins (1959) das Herz-Kreislauf-System als »Feedback-Regulator« und betont die Bedeutung der Identifizierung der relevanten Komponenten in einem System mit Ein- und Ausgängen, sowie der Verbindungen zwischen beiden. Daher identifiziert er Ein- und Ausgangsparameter für die rechte und linke Herzkammer und beschäftigt sich mit dem offenen Lungenkreislauf, auf Basis dessen er das Herz-Kreislauf-System formalisiert und modelliert.

Ähnliche Arten von Modellen mit Fokus auf spezielle Aspekte des Herz-Kreislauf-Systems im Allgemeinen finden sich z.B. bei Asteroth (2000); Hotehama et al. (2003); McLeod (1966); Ursino (1998); Whittam et al. (1998). Eine detaillierte Übersicht mit Fokus auf die Dynamik des Herz-Kreislauf-Systems und physiologische Modelle findet sich bei Lim et al. (2012). Solche physiologischen Modelle werden häufig konzipiert, um bestimmte Teile der physiologischen Funktionen des menschlichen Körpers darzustellen und wissenschaftlich erklären zu können. Darüber hinaus gibt es viele Techniken, die phänomenologische Beobachtungen modellieren.

Phänomenologische Modelle und Modelle aus dem Bereich des maschinellen Lernens sind in Bezug auf die Simulation und Vorhersage der HF unter Belastung oft besser geeignet als physiologische Modelle, auch wenn sie nicht alle Effekte des Körpers exakt widerspiegeln können. Mit ihnen lassen sich jedoch wesentliche Effekte während des Trainingsprozesses performant und vergleichsweise einfach beobachten und modellieren. Insbesondere durch

die Einschränkungen von physiologischen Messgrößen während des Trainings ist ein zu exaktes Modell mit vielen verschiedenen physiologischen Aspekten häufig nicht anwendbar – in der Regel aber auch nicht notwendig.

Die HF-Reaktion auf eine Belastung kann in vier verschiedene Zeithorizonte unterteilt werden: ultrakurzfristig, kurzfristig, mittelfristig und langfristig. Während sich die Langzeitmodellierung mit Trainingseffekten über mehrere Wochen oder Monate beschäftigt, spiegeln die ultrakurzfristigen Reaktionen die HRV wider. Letztere wird im Kontext dieser Arbeit nicht betrachtet. Eine Betrachtung eines möglichen Zusammenhangs zwischen der modellierten HF und der Leistungsentwicklung erfolgt später in Kapitel 7, »Nutzen des parameterreduzierten Faltungsmodells bei der Leistungsbeurteilung«. In den nachfolgenden Abschnitten liegt der Fokus auf der kurz- und mittelfristigen Modellierung der menschlichen HF unter Belastung, wie sie in zahlreichen phänomenologischen Modelle betrachtet wird.

Im Folgenden werden alle betrachteten Modelle in Unterabschnitten entsprechend der zugrunde liegenden Modellart beschrieben. Die jeweils genannten Ergebnisse der Studien beziehen sich in der Regel entweder auf die Kalibrierungssimulation (d.h. die Auswertung des Trainingsdatensatzes, mit dem das Modell kalibriert wurde) oder eine Vorhersage von wenigen Sekunden.

Künstliche Neuronale Netze

Yuchi und Jo (2008) implementieren ein feedforward künstliches neuronales Netzwerk (KNN)⁵, um die HF für die nächste Sekunde basierend auf körperlicher Aktivität (erhalten als 3D-Beschleunigungssignale) vorherzusagen, während Mutijarsa et al. (2016) das Gleiche basierend auf der Trittfrequenz beim Radfahren analysiert. In beiden Netzen werden die aktuelle HF und der jeweilige Stresswert (körperliche Aktivität bzw. Kadenz) als Eingangsgrößen (Input) verwendet. Die HF für die unmittelbar folgende Sekunde wird als Ausgangssignal (Output) definiert. Yuchi und Jo (2008) haben einen mittleren absoluten Fehler von 3,31 bpm in ihrem Testset ermittelt; in ihren Untersuchungen haben sie eine Anzahl von 50 Neuronen in der versteckten Schicht als geeignet identifiziert. Mutijarsa et al. (2016) haben einen mittleren absoluten Fehler von 3,02 bpm in ihrem Testset ermittelt und eine Anzahl von 333 Neuronen in der versteckten Schicht in ihren Analysen identifiziert. Das Testset wird als 30 sec. Vorhersageintervall angegeben.

Xiao et al. (2009), Xiao et al. (2010) und Xiao et al. (2011) präsentieren verschiedene Optimierungsmethoden, die auf evolutionären Algorithmen basieren. Damit sollen neuronale Netze für die HF-Vorhersage bei körperlicher Aktivität basierend auf dem von Yuchi und Jo (2008)

⁵Künstliche Neuronale Netze werden in der vorliegenden Arbeit nicht verwendet; die Modelle werden nur der Vollständigkeit halber erwähnt. Für eine detaillierte Übersicht über die Funktionsweise von künstlichen neuronalen Netzen sowie spezifische Unterschiede und Eigenschaften verschiedener Netze sei beispielsweise auf Rojas (2013) verwiesen.

beschriebenen Netz trainiert werden. Die HF-Werte werden alle 30 Sekunden für eine einzelne Testperson vorhergesagt. Der mittlere absolute Fehler beträgt bei einer kurzfristigen Vorhersagedauer 4,38 bpm im Testdatensatz.

Differentialgleichungen

DGLen werden häufig verwendet, um Änderungsraten innerhalb einer mathematischen Gleichung auszudrücken und sind daher ein gängiger Modelltyp, um Änderungen und Abweichungen der HF während des Trainings zu modellieren.

Um die Unterschiede innerhalb der nachfolgend beschriebenen drei DGLen genauer untersuchen zu können und um sie exemplarisch in einer späteren Betrachtung innerhalb von »Das Faltungsmodell als Differentialgleichungsmodell« (Unterabschnitt 5.3.1) mit der Struktur eines eigenen Modells (»Faltungsmodell«, Abschnitt 5.1) vergleichen zu können, wird in Gleichung 2.5 zunächst eine Struktur betrachtet, die allen drei Modellen zugrunde liegt:

$$\begin{aligned}\dot{x}_1(t) &= -a_1 \cdot x_1(t) + a_2 \cdot x_2(t) + h(u(t)) \\ \dot{x}_2(t) &= -a_4 \cdot x_2(t) + g(x_1(t), x_2(t)) \\ y(t) &= x_1(t)\end{aligned}\tag{2.5}$$

Dabei sind $a_i \in \mathbb{R}$ mit $i \in \mathbb{N}^+$ freie Parameter, u dient als Modell-Input (Belastung), und y dient als Modell-Output (berechnete HF). Die Funktionen h und g werden in der jeweiligen Modellbeschreibung angegeben, um Unterschiede in den Modellen zu verdeutlichen.

Cheng et al. (2007b) stellen eine DGL vor, die ursprünglich für das Training auf dem Laufband verwendet wird. Mit dem Modell soll das HF-Verhalten während länger andauernder Belastung sowie für die Erholungsphase beschrieben werden. Ein Jahr später veröffentlichte Cheng et al. (2008) eine etwas andere DGL, die zur Steuerung der Geschwindigkeit eines Laufbandes verwendet wird: Beim Gehen mit unterschiedlichen Geschwindigkeiten soll das Laufband seine Geschwindigkeit automatisch anhand der Herzfrequenz regulieren. In beiden Modellen formulieren die Autoren zwei kurzfristige Komponenten für unterschiedliche Reaktionen von HF-Veränderungen: Eine Komponente (x_1) wird angegeben, um Veränderungen in der HF basierend auf parasymphatischen und sympathischen neuronalen Effekten als zentrale Antwort auf eine Trainingsbelastung zu beschreiben. Die zweite Komponente (x_2), um Veränderungen in der HF basierend auf Effekten des Hormonsystems, der Erhöhung der Körpertemperatur oder anderen langsamer wirkenden Einflussfaktoren aus dem peripheren lokalen Stoffwechsel zu beschreiben. Die Ausgabe $y(t)$ in beiden Modellen beschreibt die Veränderungen in der HF gegenüber der Ruhe-HF, während das Inputsignal $u(t)$ während des Trainings auf die Gehgeschwindigkeit eingestellt wird (und auf 0 für die Erholung). Die Geschwindigkeit wird mit einem quadratischen Einfluss auf Veränderun-

gen der HF in beiden Modellen modelliert: Bezüglich Gleichung 2.5 definieren Cheng et al. (2007b) die Funktion h als

$$h(u(t)) = \frac{a_2 \cdot u^2(t)}{1 + \exp(-u(t) + a_3)}, a_2 = 1,$$

wobei die Exponentialfunktion genutzt wird, um weitere nicht-lineare Einflüsse auf die HF zu beschreiben; Cheng et al. (2008) haben diesen Teils des Modells reduziert zu

$$h(u(t)) = a_2 \cdot u^2(t).$$

Die langsame Erholung der HF nach dem Training modellieren Cheng et al. (2007b) ebenfalls mit einer hyperbolischen Tangensfunktion:

$$g(x_1(t), x_2(t)) = a_4 \cdot \tanh(x_2(t)) + a_5 \cdot x_1(t).$$

Nur Änderungen der ersten Komponente sind innerhalb einer Sigmoidalfunktion von der Eingangsgeschwindigkeit abhängig. Die fünf in diesem Modell verwendeten Parameter werden mit Levenberg-Marquardt angepasst. Eine Analyse der Approximationsgenauigkeit erfolgt nur visuell.

Das in Cheng et al. (2008) präsentierte Modell hat keine explizite Komponente, die eine langsame Erholung modelliert. Während die Eingangsgeschwindigkeit in diesem Modell lediglich Änderungen der ersten Komponente bewirkt, deckt die Sigmoidalfunktion Änderungen der zweiten Komponente ab; sie ist dabei abhängig von der ersten Komponente mit

$$g(x_1(t), x_2(t)) = \frac{a_4 \cdot x_1(t)}{1 + \exp(-(x_1(t) - a_5))}.$$

Obwohl das Modell prinzipiell anhand der fünf freien Parameter individualisiert werden kann, werden die Parameter in der Studie von Cheng et al. (2008) basierend auf den Daten aller Probanden fixiert. Damit haben sie ein Modell ohne freie Parameter für ihr Steuerungsdesign identifizieren können. Auch hier erfolgt eine Analyse der Approximationsgenauigkeit nur visuell, da der Schwerpunkt der vorgestellten Arbeit auf dem Reglerdesign und der Parameterstabilität liegt.

Während Scalzi et al. (2012) das Modell von Cheng et al. (2008) zur Beschreibung eines neuen Steuerungsdesigns verwenden, passen Paradiso et al. (2013) dieses Modell leicht für die Verwendung auf einem Fahrrad-Ergometer an. Im Vergleich zum Originalmodell verwenden sie einen neuen Skalierungsparameter zur Multiplikation mit dem quadratischen Eingangsterm, d.h.

$$h(u(t)) = a_6 \cdot u^2(t).$$

Die Autoren beschreiben, dass das Modell für die Steuerung von Fahrradergometern verwendet werden kann.

Eine andere DGL wurde von Stirling et al. (2008) präsentiert. Hier werden Änderungen der HF als Funktion in Abhängigkeit von Geschwindigkeit (oder anderen Intensitätsmessungen) und Zeit modelliert. Ihr Modell basiert auf zwei grundlegenden Komponenten: Veränderungen der HF und Veränderungen des Trainingsbedarfs. Beide Komponenten hängen von Geschwindigkeit und Zeit ab und werden durch die minimalen und maximalen HF-Werte eines Probanden begrenzt. Es werden drei Differenzterme modelliert, die mit unterschiedlichen Parametern skaliert und anschließend multipliziert werden: die Differenz zwischen aktueller und minimaler HF, zwischen maximaler und aktueller HF sowie zwischen einer Funktion, die den aktuellen »Bedarf« an Herzfrequenzleistung in Abhängigkeit der Trainingsintensität und der Zeit berechnet, und der aktuellen HF. Die Parameter dienen der Skalierung und der Steuerung, wie schnell sich die HF der maximalen/minimalen HF annähert oder davon abweicht. Eine Veränderung der Parameter nach einer spezifischen Trainingszeit soll Informationen über den kardiovaskulären Zustand des Probanden liefern. Die Approximationsgenauigkeit wird nur visuell analysiert.

Verbesserte Versionen dieses Modells mit weniger Parametern werden von Zakyntinaki (2015) und Mazzoleni et al. (2016) vorgestellt: Zakyntinaki (2015) beschreibt, dass die HF-Dynamik als Reaktion auf Bewegung von einem einzigen Parameter abhängig sein sollte, der den kardiovaskulären Zustand beschreibt. Das Modell basiert auf dem DGL-Modell von Stirling et al. (2008), wird jedoch beispielsweise um verschiedene HF-Phasen und Zeitverzögerungen ergänzt; gleichzeitig werden die Parameter auf nur einen globalen Parameter reduziert, der den kardiovaskulären Zustand darstellen soll. Die Grundstruktur ihres Modells ist nach wie vor ein DGL-Modell, wobei auch hier verschiedene Differenzen zwischen aktueller HF, minimaler HF, maximaler HF und tatsächlichem Trainingsbedarf gebildet werden. So ist beispielsweise die Differenz zwischen tatsächlicher und maximaler HF Teil einer Sigmoidalfunktion ähnlich zu dem Modell von Cheng et al. (2008), anstatt wie zuvor diese Differenz mit einem Exponenten zu skalieren (d.h. $(HF - HF_{max})^A$ mit Parameter A). Die Anzahl der Parameter wurde durch verschiedene Versuche so reduziert, dass alle bis auf einen Parameter fixiert werden konnten. Zakyntinaki (2015) beschreibt, dass das Modell in der Lage sei, komplette Trainingseinheiten vorherzusagen. Die veröffentlichte Analyse basiert auf einem einzigen Protokoll für zwei männliche Probanden und die Ergebnisdarstellung erfolgt visuell. Das Modell wird von Zakyntinaki (2016) erneut verwendet, um verschiedene Belastungsverläufe für synthetische Daten vorherzusagen. Die Übertragbarkeit auf reale Trainingsdaten scheint noch nicht nachgewiesen und publiziert zu sein.

Mazzoleni et al. (2016) lassen ihr Modell ebenfalls auf dem DGL-Modell von Stirling et al. (2008) zur HF-Modellierung beim Radfahren basieren. Zusätzlich fügten sie einen Term hinzu, der den Drehmoment und die Trittfrequenz berücksichtigt; diese Aspekte betrachten sie als beim Radfahren entscheidende Parameter. Das Modell enthielt damit zunächst vierzehn Parameter, allerdings konnten Mazzoleni et al. (2016) die Anzahl der freien Parameter mit Hilfe einer auf Eigenwerten basierenden Stabilitätsanalyse auf elf Parameter reduzieren. Der

Wertebereich konnte bei zwei weiteren Parametern zusätzlich beschränkt werden. Die Parameter wurden auf Grundlage synthetischer Daten berechnet. In ihren Analysen erzielten Mazzoleni et al. (2016) ein Bestimmtheitsmaß von $r^2 = 0,90$, wenn sowohl die Trittfrequenz als auch die Ausgangsleistung als Modelleingangswerte verwendet werden.

Regressionsmodelle

Regressionsmodellen liegt die Idee zugrunde, dass eine abhängige Variable durch eine oder mehrere unabhängige Variablen bestimmt ist. Sofern es einen zwischen der abhängigen und einer oder mehrerer unabhängiger Variablen gibt, wird der Wert der abhängigen Variable vom Wert der unabhängigen Variable(n) beeinflusst.

Ein Ziel von probabilistisch-statistischen Ansätzen wie der multiplen Regression auf HF-Daten ist es, bestimmte Korrelationen zwischen der HF und anderen Parametern wie Messwerten oder weiteren Kennzahlen zu testen. Hoffmann und Wiemeyer (2017) verwenden mehrere Regressionsmethoden, um Faktoren zu finden, die neben der Trainingsbelastung einen signifikanten Einfluss auf Veränderungen der HF haben könnten. Sie analysieren 19 Variablen (wie die Erholbarkeit des Schlafes, Ernährung, aktuelle Stimmung und andere) in Bezug auf ihren Einfluss auf drei verschiedene Parameter der Bunc-Gleichung (Bunc et al., 1988): die HF zu Beginn des Trainings, die *steady state* HF, und einen Faktor, der in einem grundlegenden HF-Modell für die Steigung der HF-Kurve verwendet wird. In ihrer Untersuchung finden sie heraus, dass Einflüsse auf die HF-Reaktion sehr individuell sind, aber dass körperliche Gesundheit, negative Stimmung, die Anzahl der Trainingsintervalle sowie die Tageszeit Einfluss auf Veränderungen der HF zu haben scheinen.

Die Analyse von Jang et al. (2016) zielt darauf ab, einen Zusammenhang zwischen der Laufgeschwindigkeit und der HF mit Hilfe statistischer Regressionsmethoden zu finden. Anhand der Daten von 217 Proband*innen, die einen inkrementellen Stufentest durchführen, führen Jang et al. (2016) eine Regression für lineare und nichtlineare Komponenten der HF durch; letztere sind für sie aufgrund von Stoffwechselanforderungen und kardialen Drift-Effekten von besonderem Interesse. Sowohl in der Inter- als auch in der Intra-Subjektanalyse finden sie eine starke Korrelation zwischen der HF und der Laufgeschwindigkeit. Die kleinsten Fehler werden mit höheren Regressionsordnungen erreicht. Das Regressionsmodell vierter Ordnung ergibt eine Korrelation von 0,997 und einen mittleren Fehler in der HF-Differenz in Höhe von 2,04 bpm.

Fairbairn et al. (1994) haben in ihren Untersuchungen einen linearen Zusammenhang zwischen der HF und dem $\dot{V}O_2$ für verschiedene Altersgruppen von Männern und Frauen ermittelt. Dazu untersuchen sie die Daten von 231 Personen, die inkrementelle Fahrrad-Ergometertests durchführten.

Richards (1980) bietet einen guten Überblick, der unter anderem die HF-Analyse mit statistischen Maßen, multivariaten statistischen Methoden und Zeitreihenanalysen der HF mittels Autoregression umfasst. Zudem präsentiert Richards (1980) einen übersichtlichen Prozessablauf zur Auswahl geeigneter statistischer Methode für die Analyse der Rohdaten.

In der Analyse von Bennett et al. (1993) werden vier verschiedene autoregressive Methoden diskutiert, um HF-Zeitreihen basierend auf vergangenen HF-Werten und Rauschen anzupassen und vorherzusagen. Die besten Ergebnisse zur Beschreibung der HF-Dynamik erzielen sie mit einem bilinearen autoregressiven Modell. Dieses Modell wird mit einer Autoregression mit und ohne gleitenden Durchschnitt sowie polynomieller Autoregression verglichen. Doch auch das bilineare autoregressive Modell kann in der durchgeführten Vorhersage keine guten Ergebnisse erzielen. Eine ähnliche Analyse von Christini et al. (1995) bestätigt diese Ergebnisse. Bennett et al. (1993) und Christini et al. (1995) kommen zu dem Schluss, dass die Regelung von HF-Dynamiken nichtlinear modelliert werden sollte.

Wang et al. (2008) und Wang et al. (2009) verwenden die Methoden der linearen Regression und der Support Vektor Regression (SVR), um die Beziehung zwischen der der Sauerstoffaufnahme und anderen kardiovaskulären Variablen wie beispielsweise der HF zu untersuchen. Der Fokus der Regression liegt dabei konkret auf Zusammenhängen zwischen der Sauerstoffaufnahme und anderen kardiovaskulären Faktoren – es werden keine Schlussfolgerungen für Korrelationen zwischen der HF und anderen kardiovaskulären Faktoren gezogen. Ludwig et al. (2015) zeigen, dass die SVR grundsätzlich auch zur Simulation und Vorhersage der HF-Dynamik auf Grundlage früherer HF-Messungen verwendet werden kann.

Esmaeili und Ibeas (2016) wenden eine Partikelschwarmoptimierungsmethode auf das von Wang et al. (2008) vorgeschlagene SVR-Modell an und postulieren, durch diesen Ansatz bessere Modellparameter im Vergleich zu anderen Studien zu erreichen. Girard et al. (2016) verwenden dieses Modell, um die HF-Reaktion während des Trainings auf dem Laufband mit einem PID-Regler bei Laufbandgeschwindigkeiten unter 8 km/h erfolgreich zu regeln.

Hammerstein- und Wiener-Modelle

In Studien von Su et al. (2007a), Su et al. (2007b) und Su et al. (2010b) wird ein Hammerstein-Modell zur Modellierung der HF identifiziert. Die Modellidentifikation erfolgt für den linearen und nichtlinearen Teil des Modells getrennt voneinander. Dazu werden beide Teile mittels pseudozufälliger Binärsequenzen entkoppelt. Die Identifizierung beider Modellteile erfolgt mit maschinellen Lernalgorithmen (z.B. SVR), die auf gesammelten experimentellen Laufbandergometer-Daten basieren. Das Modell wird für die PID-Regelung des Laufbandes verwendet, die im Fokus der jeweiligen Arbeit steht.

Basierend auf diesen Hammerstein-Modellansätzen haben Mohammad et al. (2011b) ein modifiziertes Hammerstein-Modell vorgestellt und getestet. Ziel dieses Modells ist ebenfalls

eine geeignete PID-Regelung. Hierbei wird der nichtlineare Teil durch eine Polynomfunktion approximiert.

Gonzalez et al. (2016) konzentrieren sich mit ihrem Modell auf die Simulation und Vorhersage der $\dot{V}O_2$. Ergänzend zeigen sie, dass ihr identifiziertes Modell auch in der HF-Modellierung und Vorhersage anwendbar ist. In ihrer Arbeit analysieren sie verschiedene Arten von Modellen, darunter autoregressive Modelle mit und ohne gleitenden Durchschnitt, State-Space-Modelle und Hammerstein-Wiener-Modelle. In ihren Analysen zeigt ein Hammerstein-Wiener-Modell die besten Ergebnisse. Durch weitere Optimierung gelangen sie schließlich zu einem reinen Wiener-Modell. In einer Untersuchung mit fünf männlichen Probanden, die jeweils Trainings mit vier verschiedenen Fahrradergometerprotokollen durchführen, beträgt die durchschnittliche Approximationsgenauigkeit (Trainingsset) der HF 4,55 bpm und die durchschnittliche Vorhersagegenauigkeit (Testset) 7,46 bpm. Die Ergebnisse von Gonzalez et al. (2016) werden in Abschnitt 5.2 ausführlich mit einem eigenen Modell (»Faltungsmodell«, Abschnitt 5.1) auf dem in der Studie verwendeten Datensatz verglichen, der von den Autoren dankenswerter Weise zur Re-Analyse zur Verfügung gestellt wurde.

Weitere Arten von Modellen

Auch einige weitere Modellarten, die sich nicht direkt zu einer der oben genannten Kategorien zuordnen lassen, werden gelegentlich für die HF-Modellierung verwendet; um einen kurzen Eindruck von der Vielfalt zu vermitteln, werden in diesem Abschnitt einige der Modelle kurz erwähnt.

Dur-e Zehra Baig et al. (2010) vergleichen ein lineares zeitinvariantes (LTI) Modell mit einem linearen zeitvariablen (LTV) Modell zur HF-Approximation beim Gehen, Radfahren und Rudern. Für jede dieser Aktivitäten werden drei verschiedene Intensitäten gewählt, sodass sich ein Datensatz mit neun verschiedenen Aktivitäten pro Person ergibt. Die Analyse erfolgt mit zwei männlichen Probanden. Das LTV-Modell schneidet in allen analysierten Fällen besser ab: Der durchschnittliche mittlere quadratische Fehler beträgt $0,071 \text{ bpm}^2$, während mit dem LTI-Modell ein durchschnittlicher mittlerer quadratischer Fehler von $0,158 \text{ bpm}^2$ erzielt wird.

Le et al. (2009), Sinclair et al. (2009), und Yang et al. (2012) definieren die HF als Summe aus einem anfänglichen HF-Wert vor Beginn des Trainings und den belastungsinduzierten Änderungen zu jedem Zeitpunkt. Die Veränderungen in der HF werden in zwei Phasen unterteilt: eine Phase, in der die HF zunimmt, und eine Phase, in der der kardiale Drift auftritt. Während Le et al. (2009) zwischen moderaten und erschöpfenden Intensitäten für die Phase des Anstiegs unterscheiden, definiert Sinclair et al. (2009) eine stationäre HF-Phase einschließlich des kardialen Drifts und verwendet die akkumulierte Arbeit anstelle von reinen Belastungswerten. Le et al. (2009) und Yang et al. (2012) modellieren zusätzlich eine Erho-

lungsphase, definiert durch eine Exponentialfunktion (Yang et al., 2012), bzw. der Summe aus der HF am anaeroben Schwellenwert abzüglich berechneter HF-Werte bis zur Erschöpfung (Le et al., 2009), was letztlich das Gegenstück zu ihrer Umsetzung der HF-Erschöpfung darstellt. Die Phase der Erhöhung der jeweiligen HF bei moderater Intensität wird bei Sinclair et al. (2009) als ein einziger Parameter modelliert. Le et al. (2009) skalieren die Arbeitsbelastung und die Änderung der HF bezüglich des vorhergehenden Zeitpunktes durch Parameter und summieren diese auf. Yang et al. (2012) ergänzen zusätzliches Rauschen additiv. Die Drift wird von Sinclair et al. (2009) als einzelner Parameter modelliert, während Le et al. (2009) und Yang et al. (2012) eine skalierte Exponentialfunktion verwenden, die von der aktuellen bzw. vorhergehenden Arbeitsbelastung abhängt.

Endler (2013) passt ein Modell von Perl (2004) speziell für das Lauf-Training an, welches ursprünglich für die Modellierung von Trainingsprozessen allgemein entwickelt wurde: *PerPot-Run* verwendet die Lauf-Geschwindigkeit als Input und unterteilt sie antagonistisch in ein positives und negatives Potential. Das Modell bestimmt die HF als Output durch Strömungsgleichungen, bei denen positive und negative Potenziale die HF mit unterschiedlichen Verzögerungen beeinflussen. Für die Vorhersage der Nutzung des Modells muss es auf eine einzelne Person mit Hilfe eines inkrementell abgestuften Belastungstests kalibriert werden. *PerPot-Run* kann zur Berechnung der individuellen anaeroben Schwelle verwendet werden (Endler et al., 2017). Darüber hinaus kann es zur Optimierung von Langstreckenrennen und allgemeinem Lauf-Training eingesetzt werden. Endler und Friedrich (2016) präsentieren eine Erweiterung von *PerPot-Run*, mit der auch Steigungen und das Gefälle einer Strecke berücksichtigt werden.

Die Studie von Koenig et al. (2009) hat zum Ziel, die HF während des Gehens auf einem Lokomaten vorherzusagen. Dazu wird der durchschnittliche HF-Anstieg für verschiedene Geh-Geschwindigkeiten abzüglich eines HF-Wertes von vor Beginn des Trainings bestimmt. Das entwickelte Modell wird das als Relaisblockdiagramm mit 11 Parametern dargestellt, um die Auswirkungen der Eingangswerte zu skalieren. Dabei werden auch Faktoren wie die Ermüdung des Probanden berücksichtigt. Durch weitere Experimente können Koenig et al. (2009) die Anzahl der freien Parameter schließlich auf vier Parameter reduzieren.

Zusammenfassung der HF-Modellierung

Ein Überblick über die zuvor beschriebenen HF-Modelle ist in Tabelle 2.3 dargestellt. Die jeweiligen Modell-Eigenschaften sind dabei kurz zusammengefasst. Jedes Modell ist durch seine Eigenschaft spezifiziert, ein Modell mit maschinellem Lernverfahren, ein Regressionsanalysemodell oder ein (analytisches) parametrisches Modell zu sein. Die Anzahl der zu optimierenden Parameter ist nach Möglichkeit und sofern in der zugehörigen Literatur spezifiziert angegeben; bei künstlichen neuronalen Netzen ergibt sich die Anzahl der Parameter durch Multiplikation der Anzahl der versteckten Knoten mit der Anzahl der Ein- und

2.2. Modellierung der Herzfrequenz als akute Anpassung

Tabelle 2.3.: Übersicht über verschiedene Spezifikationen von HF-Modellen. Verwendete Abkürzungen: KNN = künstliches neuronales Netzwerk, DGL = Differentialgleichung, Reg. = Regression, und H/W = Hammerstein-/Wiener-Modell; Modelltypen: 'M' = Maschinelles Lernen, 'R' = Regressionsmodell, 'P' = Parametrisches Modell; Vorhersagetyp: 'K' = kurzfristige Vorhersage bis zu max. 30 Sek., 'P' = partielle Vorhersage einer Trainingseinheit, 'V' = vollständige Vorhersage einer Trainingseinheit; Reglungmodell: Implementierung eines Reglers; Probanden: Anzahl der Probanden; Freie Parameter: Anzahl der Parameter (ggf. nach Reduktion); HF-Verzögerung: verzögerte exponentielle exponentiell abgeschwächte Reaktion; S-förmig: S-förmige HF-Reaktion; K. Drift: Kardiovaskulärer Drift; HF_{init}: HF vor einem Training; HF_{max}: maximale HF; Belastung: modellierte Art der HF-Beanspruchung auf eine Belastung mit 'polyn.' = polynomielle, '(non-)lin' = (nicht-)lineare, und 'quad.' = quadratische Verwendung des Eingangssignals; 'x': zutreffend, '(x)': indirekt zutreffend, 'ø': in der Publikation nicht spezifiziert

Ansatz	Quelle	Spezifikation					Phänomene					Weitere Parameter		
		Modelltyp	Vorhersagetyp	Reglungmodell	Probanden	freie Parameter	HF-Verzögerung	S-förmig	C. Drift	Erschöpfung	Erholung	HF _{init}	HF _{max}	Belastung
ANN	Yuchi und Jo (2008)	M	K	ø	1	150	ø	ø	ø	ø	ø	ø	ø	non-lin.
	Xiao et al. (2011)	M	P	ø	ø	ø	ø	ø	ø	ø	ø	ø	ø	non-lin.
	Mutijarsa et al. (2016)	M	K	ø	9	1333	ø	ø	ø	ø	ø	ø	ø	non-lin.
DGL	Cheng et al. (2007b)	P	K	x	5	5	x	x	ø	ø	tanh	ø	ø	quad.
	Cheng et al. (2008)	P	K	x	6	5	x	x	ø	ø	ø	ø	ø	quad.
	Paradiso et al. (2013)	P	K	x	ø	6	x	x	(x)	ø	ø	ø	ø	quad.
	Stirling et al. (2008)	P	ø	ø	1	4	ø	x	ø	ø	ø	x	x	polyn.
	Zakynthinaki (2015)	P	ø	ø	2	1	x	x	ø	iAT	iAT	x	x	polyn.
	Mazzoleni et al. (2016)	P	ø	ø	4	14 (11)	ø	x	ø	ø	ø	x	x	polyn.
Reg.	Hoffmann und Wiemeyer (2017)	R	ø	ø	4	ø	ø	ø	ø	ø	ø	ø	ø	ø
	Jang et al. (2016)	R	ø	ø	217	ø	ø	ø	ø	ø	ø	ø	ø	ø
	Fairbairn et al. (1994)	R	ø	ø	462	ø	ø	ø	ø	ø	ø	ø	ø	ø
	Bennett et al. (1993)	R	K	ø	9	ø	ø	ø	ø	ø	ø	ø	ø	ø
	Christini et al. (1995)	R	ø	ø	9	ø	ø	ø	ø	ø	ø	ø	ø	ø
	Wang et al. (2008)	R	ø	ø	10	ø	ø	ø	ø	ø	ø	ø	ø	ø
H/W	Su et al. (2007b)	M	K	x	6	ø	ø	ø	ø	ø	ø	ø	ø	ø
	Mohammad et al. (2011b)	M	K	x	ø	12	ø	ø	ø	ø	ø	ø	ø	ø
	Gonzalez et al. (2016)	M	V	ø	5	ø	ø	ø	ø	ø	ø	ø	ø	ø
	Ludwig et al. (2019)	P	V	ø	5	4 (1)	x	x	ø	ø	ø	x	ø	polyn.
Weitere	Endler (2013)	P	V	ø	14	4	x	x	x	x	ø	(x)	(x)	lin.
	Dur-e Zehra Baig et al. (2010)	P	ø	ø	2	4 · time	ø	ø	ø	ø	ø	x	ø	lin.
	Koenig et al. (2009)	P	P	x	10	11 (4)	ø	ø	ø	ø	ø	x	ø	lin.
	Le et al. (2009)	P	P	x	ø	4	x	x	x	ø	konst.	ø	ø	lin.
	Sinclair et al. (2009)	P	V	ø	9	5	x	x	x	ø	x	iAT	ø	lin.
	Yang et al. (2012)	R	ø	ø	10	11	ø	x	x	ø	exp.	ø	ø	lin.

Ausgangsknoten (und dem Bias), da die hier betrachteten Netze mit genau einer versteckten Schicht (engl. *hidden layer*) aufgebaut sind. Zusätzlich ist angegeben, welche der vier eingangs erwähnten zentralen Effekte durch das jeweilige Modell modelliert werden können, d.h. ob die verzögerte exponentiell abgeschwächte Reaktion, die S-förmige Reaktion, der kardiovaskuläre Drift oder die vollständige Erschöpfung berücksichtigt werden. Die Einbeziehung möglicher Vorbelastungen der HF oder der maximalen HF einer Person in das Modell und die Art und Weise, wie die Belastung als Input einbezogen wird, ist ebenfalls dargestellt. Zusätzlich enthalten einige Modelle eine Komponente zur Wiederherstellung, die sich von der HR-Reaktion auf zunehmende Belastung unterscheidet. In diesem Fall ist die zur Wiederherstellung verwendete Funktion in der Tabelle angegeben. Die Anzahl der Parameter wird nicht angegeben, wenn ein Modell nicht explizit angegeben ist und die Anzahl der notwendigen Parameter für die Optimierung nicht in der entsprechenden Publikation angegeben ist. Alle Effekte oder Parameter, die nicht für ein Modell spezifiziert sind, werden zur Verdeutlichung mit »∅« markiert.

Anhand von Tabelle 2.3 wird insbesondere deutlich, dass die meisten hier betrachteten Modelle als Kontrollmodelle oder zur Analyse einer Korrelation zwischen der HF und einer Belastung genutzt werden. Eine Auswertung von Vorhersagen insbesondere vollständiger Trainingseinheiten ('V') wird in der aktuellen Literatur kaum berücksichtigt.

Bezüglich der abgebildeten Effekte fällt auf, dass die meisten Modelle sowohl eine exponentielle Reaktion auf die Belastung, als auch die S-förmige HF-Reaktion beinhalten. Viele Modelle verwenden eine initiale bzw. vor der Trainingseinheit gemessene HF, während die verbleibenden Effekte nur spärlicher betrachtet werden. Außerdem scheint die Belastung in den meisten Modellen polynomiell berücksichtigt zu werden.

Obwohl Modelle aus dem Maschinellen Lernen (z.B. künstliche Neuronale Netze) oder über von Toolboxen identifizierte, nicht näher spezifizierte Modelle in der Regel nicht physiologisch erklärt werden können, funktioniert die Simulation einer bestehenden HF-Kurve oder die Vorhersage der nächsten (wenigen) Sekunden häufig sehr gut und mit nur geringen Fehlern. Geht es hingegen um die Planung eines Trainings oder Wettkampfes, reicht die HF-Simulation von bereits existierenden Trainingseinheiten und die Vorhersage von nur wenigen Sekunden nicht mehr aus. Um eine ganze Trainingseinheit zu planen oder die Fähigkeiten einer Person in einem Wettbewerb zu simulieren, muss die Herzfrequenz über eine komplette Trainingseinheit hinweg vorhergesagt werden können. Solche Modelle neigen jedoch dazu, sich besonders stark an die zur Modellkalibrierung genutzten Daten anzupassen. Dies wird durch die hohe Anzahl von Parametern verursacht, wie sie allerdings auch häufig in analytisch-phänomenologischen HF-Modellen vorkommt (Ludwig et al. (2016)). Insbesondere die Interpretierbarkeit der Parameter eines Modells ist bei der HF-Vorhersage vorteilhaft: Die Modellierung tatsächlicher Effekte, die die HF beeinflussen, verbessert die Genauigkeit der Vorhersage im Vergleich zur Modellierung beliebiger, unerklärter Artefakte erheblich. Ein Vergleich verschiedener Arten phänomenologischer Modelle wurde von Lud-

wig et al. (2015) präsentiert; dabei wird auch die Genauigkeit von Simulationen und Vorhersagen der HF unter Berücksichtigung verschiedener Zeithorizonte dargestellt. Die dortigen Ergebnisse zeigen, dass eine gute Genauigkeit in der Simulation oder Vorhersage von wenigen Sekunden keinen Hinweis auf die Vorhersagegenauigkeit in einem kompletten Training liefert.

2.3. Zusammenhang zwischen Herzfrequenz und Leistung in der Modellierung: Eine Einordnung

Innerhalb einzelner Trainingseinheiten lässt sich die Beanspruchungsreaktion des Körpers auf die vorliegende Belastung über unterschiedliche physiologische Messwerte modellieren. Ein geeigneter Messwert ist die Herzfrequenz – trotz spezifischer Einschränkungen und Nachteile wie in »Messgrößen der Beanspruchung in der Diagnostik« (Unterabschnitt 1.2.3) und »Die Herzfrequenz als Indikator: Kritische Aspekte« (Unterabschnitt 1.3.1) diskutiert. Herzfrequenzmodelle werden üblicherweise genutzt, um ein Training anhand der Herzfrequenz als Maß für die individuelle Beanspruchung zu steuern. Für diese Anwendung genügt es, wenn die Herzfrequenzreaktion mit Hilfe des Modells simuliert und für wenige Sekunden vorhergesagt werden kann. Dazu sind wie in »Herzfrequenzmodellierung: Stand der Forschung« (Unterabschnitt 2.2.2) dargestellt zahlreiche Modelle publiziert.

Außerhalb einzelner Trainingseinheiten und im Kontext mehrwöchiger Zeiträume sind Modelle der Trainingswirkungsanalyse von Interesse. Ziel der Modellierung der Trainingswirksamkeit ist es, die Veränderung Leistungsfähigkeit als Folge der Trainingswirkung über einen längeren – in der Regel mehrwöchigen – Zeitraum zu untersuchen. Pfeiffer (2020) beschreibt Modelle der Trainingswirkungsanalyse als »Modelle zur prozessualen Abbildung der Relation von Training und Leistung«. Dabei ist das Verhältnis von externer Trainingsbelastung bzw. interner Trainingsbeanspruchung und der sportlichen Leistungsfähigkeit nicht linear; es wird durch unterschiedliche Anpassungsprozesse beeinflusst. In der Trainingswirkungsanalyse dienen die über mehrere Wochen hinweg erfassten Trainings- und Belastungsdaten als Eingabewerte für ein parametrisiertes Modell. Nach Individualisierung der Parameter über ein Kalibrierungsverfahren kann so die Leistungsfähigkeit approximiert werden. Ist das Modell an den Probanden angepasst, kann es optimalerweise schließlich auch zur Trainingsplanung genutzt werden, indem der gewünschte Leistungsverlauf vorgegeben und umgekehrt die dazu notwendige Belastung geschätzt wird. Insbesondere lässt sich auf diese Weise ein Trainingsverlauf planen, der eine Überlastung des Probanden vermeidet und die benötigte Belastung minimiert. Für die Modellierung der Leistungsfähigkeit innerhalb der Trainingswirkungsanalyse sind in der Regel spezifische Leistungstests in regelmäßigen Abständen erforderlich. Die Leistungstests können jedoch insbesondere als Ausbelastungstests das Training und den Ermüdungs- bzw. Erholungszustand einer Sportlerin oder eines

Sportlers beeinflussen. Gerade in Wettkampfphasen muss daher teilweise auf die Durchführung verzichtet werden. Zudem sind die Testergebnisse oft ungenau und abhängig von der Tagesform und der tagesaktuellen Motivation, sodass die Ergebnisse selbst für versierte Trainer*innen nicht immer leicht zu interpretieren sind.

Die Herzfrequenz eignet sich wie in Unterabschnitt 1.3.2 (»Gemeinsamkeiten akuter und chronischer Reaktionen auf Ausdauerbelastung«) dargelegt als Parameter zur Beschreibung der Trainingsintensität und auch der Leistung einer Athletin oder eines Athleten. In der Modellierung bleibt dieser Zusammenhang jedoch häufig unberücksichtigt. Obwohl das in Unterabschnitt 2.2.2 erwähnte DGL von Stirling et al. (2008) die HF vorhersagen kann und die Autoren einen Zusammenhang zwischen den Parametern und der generellen Ausdauerleistungsfähigkeit postulieren (»(...) the parameters give fundamental information about the condition of the individual subject across the spectrum of exercise intensities, not just a particular exercise intensity«), wird dieser Zusammenhang bislang nicht durch eine Studie belegt. Auch de Smet et al. (2016) haben ein Modell zur Modellierung und Vorhersage der HF auf Basis von Belastungsdaten entwickelt, dessen Parameter sie in zukünftigen Untersuchungen auf mögliche Korrelationen zu Leistungsmarkern aus Laboruntersuchungen analysieren wollen. Wenngleich also der Zusammenhang zwischen Herzfrequenz und Leistung physiologisch nahe liegt, scheint es bislang nicht gelungen zu sein, diesen über Modelle gezielt abzubilden.

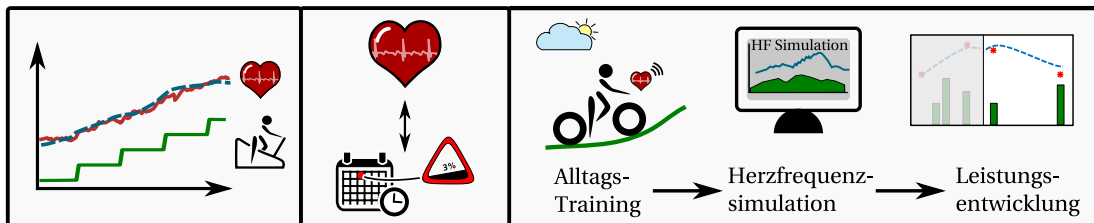
Basierend auf den Forschungsergebnissen dieser Arbeit konnte in den Jahren 2020-2021 das Forschungsprojekt KisS-BiS: »Künstliche Intelligenz für den Spitzensport im Spannungsfeld zwischen Big und Small Data« durchgeführt werden, in dem unter anderem die Kombination von Herzfrequenz und Leistung mit Hilfe von Methoden aus der künstlichen Intelligenz und des maschinellen Lernens untersucht wurde. Ein zentraler Aspekt dabei war die Idee, basierend auf den Modellierungen von kurzfristigen und langfristigen Beanspruchungen auf gegebene Belastungsreize eine »virtuelle Leistungsdiagnostik« durchzuführen. Grundlegend sind für die Entwicklung von geeigneten Trainingsplänen auf Basis der virtuellen Leistungsdiagnostik mehrere Schritte notwendig:

1. Die Modellierung der kurzfristigen Beanspruchungsreaktion – diese kann durch die Herzfrequenzsimulation mit geeigneten Modellen durchgeführt werden.
2. Die Modellierung der langfristigen Beanspruchungsreaktionen – dafür sind Modelle aus dem Bereich der Trainingswirkungsanalyse notwendig, die im Kontext dieser Arbeit nicht weiter betrachtet werden.
3. Die Verwendung der Modellierung der Beanspruchungsreaktion innerhalb einer Trainingseinheit für die Simulation der längerfristigen Leistungsfähigkeit – hierbei werden beide Modell-Typen nacheinander verwendet und die Ergebnisse miteinander verknüpft.

Aufbauend auf einer geeigneten Modellkalibrierung zur Herzfrequenzvorhersage lassen sich beliebige Beanspruchungsprotokolle vorgeben, für die die Herzfrequenz realistisch simuliert werden kann. Dabei kann die Belastung beispielsweise in Form eines Stufenprotokolls gewählt werden, wie es in der Leistungsdiagnostik verwendet wird. Basierend auf der Simulation der Herzfrequenz kann ein korrespondierender TRIMP-Wert berechnet werden, der eine (simulierte) Leistungsfähigkeit darstellt. Wird das Herzfrequenz-Modell iterativ nach jedem Trainingstag mit neuen Daten über einen gewissen Zeitraum kalibriert, lassen sich so für jeden Trainingstag simulierte Leistungswerte berechnen. Die simulierten Leistungswerte können anschließend in das Modell zur Trainingswirkungsanalyse eingespeist werden, um auch dieses Modell individuell zu kalibrieren. Die Kombination beider Modelltypen ermöglicht es schließlich, basierend auf alltäglichen Trainingsdaten und der Herzfrequenzmodellierung Rückschlüsse über die langfristige Leistungsentwicklung zu ziehen. (Hagg et al., 2021)

Das Konzept der virtuellen Leistungsdiagnostik wird in »Analyse zur Abschätzung der Leistungsveränderung auf Basis von Herzfrequenz-Simulationen: Profi-Sportler« (Abschnitt 7.1) näher erläutert und innerhalb einer der dort vorgestellten Studien in abgewandelter Form ohne die Verwendung von Modellen aus dem Bereich der Trainingswirkungsanalyse exemplarisch durchgeführt. Die Ergebnisse der nachfolgenden Studien stellen eine der Grundlagen dar, auf Basis derer das Projekt letztlich aufgebaut wurde.

Zusammenfassung des Forschungsstands und Fragestellungen



Abstract.

In diesem Kapitel werden kurz die zur Beantwortung der zentralen Fragestellung notwendigen Grundlagen rekapituliert. Aus den Lücken der Forschung ergeben sich darauf aufbauend drei Fragestellungen, die schließlich in den nachfolgenden Kapiteln untersucht und beantwortet werden. Das Zusammenspiel dieser drei abgeleiteten Fragestellungen liefert schließlich eine Antwort auf die zuvor formulierte Zielsetzung (vgl. S. 3) der vorliegenden Arbeit.

Das Ziel der eingangs formulierten zentralen Fragestellung dieser Arbeit ist die Untersuchung, ob es über Modelle möglich ist, eine Prognose zur Leistungsentwicklung beim Radfahren auch ohne vorhergehende ergometriebasierte Leistungsdiagnostik basierend auf alltäglichem Training zu stellen. In den vorangegangenen Kapiteln wurden ausführlich die physiologischen Grundlagen (Kapitel 1, »Leistungsphysiologische Grundlagen«) beim Ausdauertraining (Abschnitt 1.2, »Allgemeine Grundlagen der Trainingssteuerung im Ausdauersport«) sowie allgemein zur Leistungsfähigkeit und die Eignung der Herzfrequenz als möglicher Indikator (Abschnitt 1.3, »Herzfrequenzverhalten bei körperlicher Belastung«) dargestellt. Trotz der in »Die Herzfrequenz als Indikator: Kritische Aspekte« (Unterabschnitt 1.3.1) betrachteten Einschränkungen überwiegen die Vorteile der Nutzung der Herzfrequenz, so dass der Herzfrequenz als physiologischem Parameter im Rahmen dieser Arbeit der Vorzug vor möglichen Alternativen gegeben wird. Die Grundlagen der Modellierung (Kapitel 2, »Grundlagen zur Modellierung des Herzfrequenzverhaltens im Ausdauersport«) haben aufgezeigt, welchen Aspekten bei einer geeigneten Modellierung von Herzfrequenz oder Leistung eine besonders zentrale Bedeutung zukommt. Außerdem wurde in »Modellierung der Herzfrequenz als akute Anpassung« (Abschnitt 2.2) dargelegt, dass es zwar zahlreiche Modelle gerade im Bereich der Herzfrequenz-Modellierung gibt, diese sich jedoch kaum zur Vorhersage längerer Trainingseinheiten eignen. Insbesondere erscheint es ohne die Simulation und Vorhersage der Herzfrequenz vollständiger Trainingseinheiten nicht möglich, darauf basierend spezifische Aussagen über die Leistungsfähigkeit zu treffen.

Damit eine Leistungsentwicklung auf Basis alltäglicher herzfrequenzbasierter Trainingsdaten anhand eines Modellparameters modelliert werden kann, müssen zwei zentrale Voraussetzungen erfüllt sein: Erstens muss das Herzfrequenz-Modell die Herzfrequenz im Training zuverlässig modellieren und möglichst auch zuverlässig prognostizieren. Zweitens ist es notwendig, dass die Modellparameter sich nicht untereinander beeinflussen und damit eine mögliche Korrelation zur Leistungsentwicklung verschleiern. Dazu ist es hilfreich, wenn das Modell so wenig freie Parameter wie möglich enthält. Um die zentrale Fragestellung abschließend analysieren zu können, müssen daher im Vorfeld diese beiden Aspekte – die Simulation und Vorhersage der Herzfrequenz mit einem geeigneten Modell und die Reduktion der Parameter des Modells – untersucht werden.

Mit der Möglichkeit, allein durch das Herzfrequenzverhalten in regulären Trainingseinheiten Rückschlüsse auf die Leistungsentwicklung ziehen zu können, ließe sich die Trainingswirkungsanalyse vereinfachen: Insbesondere an Tagen, an denen Leistungstests beispielsweise die erwartete Wettkampfleistung einer Profi-Athletin oder eines Profi-Athleten negativ beeinflussen würden, könnten dennoch gewünschte Prognosen über gezielte Simulationen unmittelbar vor Wettkämpfen gestellt werden. Auch im Hobby- und Gesundheitssport könnte ein herzfrequenzbasiertes Verfahren eine leichtere Kontrolle der persönlichen Leistung ermöglichen. Dafür ist es jedoch wichtig, dass die zugrunde liegenden Modelle 1. keine

zu hohe Komplexität aufweisen und sich 2. mit möglichst geringem Aufwand sowie 3. mit einer geringen Anzahl an Datensätzen kalibrieren lassen.

Im Folgenden werden daher ergänzend zur zentralen Fragestellung zwei weitere Problemstellungen auf Grundlage der vorhergehenden Kapitel hergeleitet und es werden nacheinander alle drei Fragestellungen analysiert und diskutiert. Die Analysen finden jeweils in Form eigener Studien statt, bei denen insbesondere die Genauigkeit und Validität der betrachteten Modelle im Fokus stehen. Bei den verwendeten Daten handelt es sich mit einer Ausnahme stets um Re-Analysen bereits existierender Daten, sodass keine Einflussnahme auf die Trainingsmethodiken möglich war. Die sich daraus ergebenden Einschränkungen werden bei der Durchführung der empirischen Studien geeignet berücksichtigt.

Modellierung des Herzfrequenzverhaltens

Die Herzfrequenz wird häufig als guter Indikator für Beanspruchungsintensität einer Belastung betrachtet (Lucía et al., 2000), da insbesondere das Herz-Kreislauf-System eine leistungslimitierende Komponente darstellt (Hohmann et al., 2007). Wie in »Herzfrequenzverhalten bei körperlicher Belastung« (Abschnitt 1.3) dargestellt, reagiert die Herzfrequenz auf sportliche Belastungen und verändert sich sowohl kurz- als auch langfristig durch Ausdauersport. Trotz der in »Die Herzfrequenz als Indikator: Kritische Aspekte« (Unterabschnitt 1.3.1) betrachteten Limitationen stellt sie eine gut messbare und physiologisch sinnvolle Größe dar, mit der die Reaktion des Körpers auf sportliche Betätigung zumindest unter Berücksichtigung einiger Einschränkungen und je nach Fragestellung beschrieben werden kann. Dabei sind stets auch externe Einflussfaktoren (vgl. »Die Herzfrequenz als Indikator: Kritische Aspekte« Unterabschnitt 1.3.1) zu berücksichtigen.

In der Simulation beschränken sich die meisten Herzfrequenz-Modelle auf die reine Simulation der Modell-Kalibrierung oder die kurzfristige Vorhersagen weniger Sekunden, wie sie für die Regelung von Ergometern benötigt werden. In »Modellierung der Herzfrequenz als akute Anpassung« (Abschnitt 2.2) wurden verschiedene Modelle kurz beschrieben und auf ihre publizierte Nutzung längerer Vorhersagen hin analysiert. Die Modelle von Zakyntinaki (2015) und Mazzoleni et al. (2016) scheinen auch für die Vorhersage längerer Zeiträume vielversprechend, allerdings konnte keine Publikation einer Validierung auf empirischen Datensätzen gefunden werden. Zudem werden in dem Modell von Mazzoleni et al. (2016) elf Parameter genutzt, was insbesondere bei geringen Datenmengen zu einer Überanpassung an die Kalibrierungsdaten führen kann.

In Hinblick auf die Zielsetzung, die Herzfrequenz und ihre Modellierung zu nutzen, um letztlich Rückschlüsse auf die Leistungsfähigkeit einer Person ziehen zu können, scheinen die betrachteten Modelle jedoch auf Grund ihrer Vielzahl an Parameter und der meist fehlenden Möglichkeit einer physiologischen oder phänomenologischen Interpretation ungeeignet zu sein. In »Das FaltungsmodeLL: Ein neuer Ansatz zur Simulation und Vorhersage der

Herzfrequenz im Training« (Kapitel 5) wird daher ein selbst entwickeltes Modell (das »Faltungsmodell«) vorgestellt und auf seine Simulations- und Vorhersagequalität hin analysiert. Damit wird die folgende Problemstellung genauer untersucht:

Fragestellung 1: Vorhersage der Herzfrequenz im Training

Lassen sich mit dem Faltungsmodell bessere Vorhersagen der Herzfrequenz während körperlicher Belastung im Ausdauersport für ganze Trainingssequenzen ohne Rückkopplung gemessener Herzfrequenzwerte treffen, als mit bisherigen Modellen der Herzfrequenzanalyse?

Nach einer ausführlichen Modellbeschreibung und -validierung wird das Faltungsmodell schließlich in »Nutzen des parameterreduzierten Faltungsmodells bei der Leistungsbeurteilung« (Kapitel 7) erneut betrachtet, um einen modelltheoretischen Zusammenhang zwischen der Herzfrequenz und der Leistung näher untersuchen zu können.

Parameterreduktion im Herzfrequenzmodell

Vorbereitend auf die zentrale Zielsetzung, aus alltäglichen Trainingseinheiten basierend auf einem Modell zur Herzfrequenzmodellierung Rückschlüsse auf die Leistungsentwicklung zu ziehen, kann das Faltungsmodell weiter untersucht werden. Während das Faltungsmodell als neues Modell zur Modellierung des Herzfrequenzverhaltens über vollständige Trainingseinheiten innerhalb der vorherigen Fragestellung evaluiert wurde, kann zusätzlich die Frage untersucht werden, ob eine adäquate Modellierung und Vorhersage der Herzfrequenz über das Faltungsmodell auch mit weniger Modellparametern realisiert werden kann.

Aus einer möglichst geringen Anzahl freier Parameter resultieren unterschiedliche Vorteile. Ganz allgemein betrachtet sind weniger freie Parameter von Vorteil, da dadurch das Modellverhalten in der Regel stabiler wird und es weniger zu einer Überanpassung an die Kalibrierungsdaten kommt – vorausgesetzt, die Simulation und Vorhersage gehen im Vergleich zu einer größeren Parameteranzahl nicht mit einem starken Genauigkeitsverlust einher. Hinsichtlich der zentralen Zielsetzung ist es von zentraler Bedeutung, dass die Parameter des Herzfrequenzmodells sich nicht gegenseitig beeinflussen. Gleichzeitig ermöglichen weniger Parameter eine direktere Analyse, inwieweit basierend auf der Herzfrequenzmodellierung Rückschlüsse auf die Leistungsentwicklung möglich sind.

Daraus resultiert die folgende Problemstellung:

Fragestellung 2: Parameterreduktion im Faltungsmodell

Kann die Anzahl der Parameter des Faltungsmodells ohne signifikanten Verlust der Vorhersagegenauigkeit reduziert werden?

»Verbesserung des Faltungsmodells: Parameterreduktion« (Kapitel 6) widmet sich der Betrachtung und Analyse dieser Fragestellung. Dazu wird zunächst algorithmisch ein Verfahren beschrieben, mit dem sich Zusammenhänge zwischen den Parametern detektieren lassen. Anschließend wird die Parameterreduktion anhand von zwei Studien durchgeführt und hinsichtlich ihrer Genauigkeit im Vergleich zum vierparametrischen Modell analysiert.

Rückschlüsse auf die Leistungsentwicklung anhand der Herzfrequenzmodellierung

Wie in »Gemeinsamkeiten akuter und chronischer Reaktionen auf Ausdauerbelastung« (Unterabschnitt 1.3.2) dargestellt, besteht zwischen der Herzfrequenz und der allgemeine Leistungsfähigkeit einer Person physiologisch betrachtet durchaus ein Zusammenhang. Insbesondere die über 5000 Studien, die von Bellenger et al. (2016) in der bereits referenzierten Meta-Analyse in Hinblick auf eine solche Korrelation untersucht wurden, legt nahe, dass sich Eigenschaften der Herzfrequenz durchaus eignen können, um die Leistungsfähigkeit zu kontrollieren, zu beschreiben und entsprechend auch vorherzusagen.

In der Modellierung scheint dieser Zusammenhang zwar in manchen Forschungsgruppen durchaus von Interesse, allerdings fehlen bislang Publikationen über erfolgte Untersuchungen, die die Korrelation etwaiger Modellparameter mit der Leistung gezielt untersuchen (vgl. »Zusammenhang zwischen Herzfrequenz und Leistung in der Modellierung: Eine Einordnung«, Abschnitt 2.3). Ein Grund dafür könnte sein, dass ebendieser Zusammenhang, der in der Physiologie betrachtet und auch begründet werden kann, in bisherigen Modellen nicht auffindbar ist. Durch die Komplexität des menschlichen Organismus und die Vielfalt der Zusammenhänge und Wechselwirkungen im Körper ist es kaum verwunderlich, dass solche speziellen Aspekte nicht zwangsläufig auch in der Modellierung abgebildet werden. Es soll daher schließlich untersucht werden, ob es mit relativ einfachen Modellierungstechniken und einer gezielten Auswahl eines Modells zur Herzfrequenzmodellierung möglich ist, die (empirische oder wiederum modellierte) Leistung nachzubilden. Kapitel 7 zur »Nutzen des parameterreduzierten Faltungsmodells bei der Leistungsbeurteilung« ist daher aufbauend auf den vorherigen Ergebnissen der Analyse der zentralen Fragestellung gewidmet:

Fragestellung 3: Nutzen der Herzfrequenzmodellierung für die Bewertung der Leistung

Trägt der aus der Herzfrequenzmodellierung stammenden Parameter des parameterreduzierten Faltungsmodells basierend auf alltäglichem Training zur Gewinnung von Rückschlüssen auf die Leistungsentwicklung einer Athletin oder eines Athleten im Ausdauersport bei?

Dabei werden die Resultate und Erkenntnisse der vorherigen Untersuchungen aus den Kapiteln 5 und 6 zu »Das Faltungsmodell: Ein neuer Ansatz zur Simulation und Vorhersage der Herzfrequenz im Training« und »Verbesserung des Faltungsmodells: Parameterreduktion«

genutzt, um die Möglichkeiten von möglichst unmittelbaren Zusammenhängen zwischen dem Modellparameter des Faltungsmodells und der Leistungsentwicklung zu analysieren.

Teil II.

Empirische Studien

Die zentrale Zielsetzung der vorliegenden Arbeit ist die Analyse, in wie weit es mit Hilfe eines möglichst einfachen mathematisch-phänomenologischen Modells möglich ist, basierend auf regulären Trainingseinheiten und der Herzfrequenzmessung Rückschlüsse auf die Leistungsentwicklung einer Sportlerin oder eines Sportlers zu ziehen.

Aufbauend auf den in Teil I dargestellten physiologischen, modelltheoretischen und statistischen Grundlagen konnte die zentrale Fragestellung in drei aufeinander aufbauende Fragestellungen untergliedert werden. Nachfolgend werden wie in Abbildung 3 illustriert unterschiedliche empirische Studien präsentiert, die die einzelnen Fragestellungen behandeln.

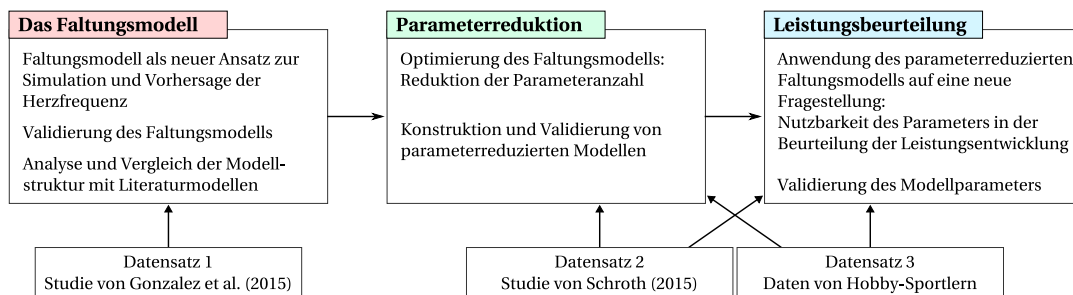


Abbildung 3.: Übersicht über die einzelnen Fragestellungen in den empirischen Studien

Nach einer Vorstellung der verwendeten Datensätze (Kapitel 4) wird in Kapitel 5 in einem ersten Schritt ein neu entwickeltes Modell (»Faltungsmodell«) zur Simulation und Vorhersage der Herzfrequenz während des Radtrainings präsentiert. Zahlreiche Modelle aus der Literatur werden lediglich zur Simulation der Herzfrequenz oder der Vorhersage weniger Sekunden (oft im Rahmen einer PID-Regelung) verwendet und basieren in der Regel auf einer Vielzahl freier Parameter. Das spezifisch zur Simulation und Vorhersage vollständiger Trainingseinheiten entwickelte Faltungsmodell kommt mit lediglich vier freien Parametern aus, die zudem phänomenologisch interpretiert werden können. Im Rahmen einer ersten empirischen Studie wird das Faltungsmodell validiert und die Ergebnisse mit denen eines Wiener-Modells aus der Literatur verglichen. Eine mathematisch-strukturelle Analyse, die das Faltungsmodell mit einigen ausgewählten Modellen aus der Literatur vergleicht, schließt das Kapitel ab.

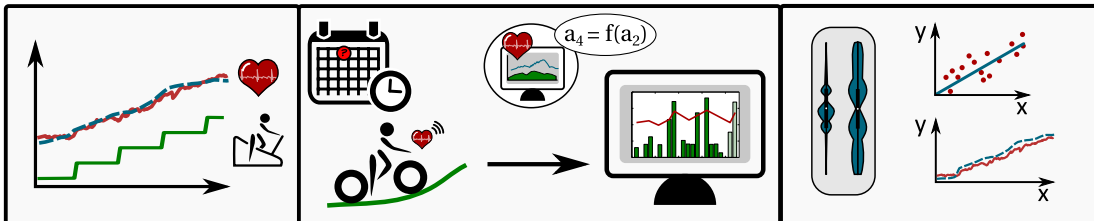
In einem zweiten Schritt wird das Faltungsmodell innerhalb von Kapitel 6 weiter optimiert. Dabei soll eine mögliche Überanpassung an die Kalibrierungsdaten verringert und die wechselseitige Beeinflussung verschiedener Parameter reduziert werden. Ziel dabei ist es, die Anzahl der Parameter im Faltungsmodell auf einen einzigen freien Parameter zu reduzieren. Dieser freie Parameter soll im weiteren Verlauf der Arbeit hinsichtlich eines Zusammenhangs zur Leistungsentwicklung analysiert werden. Dazu wird die Anzahl der Parameter des Faltungsmodells zunächst innerhalb von zwei empirischen Studien mit unterschiedlichen Probandendaten reduziert. Die Ergebnisse der daraus entstandenen parameterreduzierten Faltungsmodelle werden dabei sowohl für die Simulation als auch für die Vorhersage der

Herzfrequenz mit dem vierparametrischen Faltungsmodell qualitativ und quantitativ verglichen.

Mit Hilfe der identifizierten parameterreduzierten Faltungsmodelle mit je einem freien Parameter wird schließlich in Kapitel 7 ein möglicher Zusammenhang zwischen diesem freien Parameter und der Leistungsentwicklung über die Zeit untersucht. Analog zur Analyse der zweiten Fragestellung werden auch hier zwei empirische Studien mit unterschiedlichen Probandendaten durchgeführt.

Kapitel 8 stellt abschließend eine Zusammenfassung der Arbeit samt der Ergebnisse aus den empirischen Studien sowie einen Ausblick über die weitere Nutzungsmöglichkeit des einparametrischen Faltungsmodells dar.

Datenmaterial



Abstract.

Innerhalb der empirischen Arbeiten zur Analyse der formulierten Fragestellungen werden insgesamt drei verschiedene Datensätze verwendet. Die drei Datensätze werden in diesem Kapitel präsentiert und ihr Einsatzzweck wird vorgestellt.

Innerhalb der nachfolgenden empirischen Arbeiten zur Analyse der formulierten Fragestellungen werden insgesamt drei unterschiedliche Datensätze verwendet:

1. Der erste Datensatz beinhaltet Daten über die Wattbelastung, sowie Herzfrequenz- und Sauerstoffaufnahmen während der Trainingsdiagnostik auf Basis von vier unterschiedlichen Ergometer-Protokollen für drei männliche Hobby-Sportler. Dieser Datensatz wird zur Evaluation des Faltungsmodells verwendet.
2. Der zweite Datensatz beinhaltet reguläre Outdoor-Trainingseinheiten, Ergometer-tests zur Leistungsdiagnostik, Daten für die Herzfrequenzmodellierung sowie die Ergebnisse der Leistungsdiagnostik bei vier männlichen Profi-Athleten. Dieser Datensatz wird zur Evaluation des parameterreduzierten Faltungsmodells und der darauf folgenden Anwendung auf die Leistungsentwicklung verwendet.
3. Der dritte Datensatz beinhaltet Ergometer-tests zur Leistungsdiagnostik, Daten für die Herzfrequenzmodellierung sowie die Ergebnisse einer Laktatdiagnostik zur Leistungsüberprüfung bei drei männlichen Hobby-Sportlern. Dieser Datensatz wird ebenfalls zur Evaluation des parameterreduzierten Faltungsmodells und der darauf folgenden Anwendung auf die Leistungsentwicklung verwendet.

Tabelle 4.1 fasst die Einsatzzwecke der unterschiedlichen Datensätze für die verschiedenen empirischen Arbeiten nochmal zusammen.

Tabelle 4.1.: Übersicht der für die einzelnen empirischen Arbeiten verwendeten Datensätze

#	Empirische Arbeit	Datensatz	Einsatzzweck
1	Abschnitt 5.2	Datensatz 1	Validierung des Faltungsmodells
2	Abschnitt 6.2	Datensatz 2	Anwendung und Vereinfachung des Faltungsmodells auf den Daten von Profi-Athleten
3	Abschnitt 6.3	Datensatz 3	Anwendung und Vereinfachung des Faltungsmodells auf den Daten von Hobby-Sportlern
4	Abschnitt 7.1	Datensatz 2	Anwendung des parameterreduzierten Faltungsmodells auf die Leistungsfähigkeit auf den Daten von Profi-Athleten
5	Abschnitt 7.2	Datensatz 3	Anwendung des parameterreduzierten Faltungsmodells auf die Leistungsfähigkeit auf den Daten von Hobby-Sportlern

Die drei unterschiedlichen Datensätze werden nachfolgend präsentiert.

4.1. Datensatz 1: Daten aus einer Studie von Gonzalez et al.

In einer 2015 erstellten Studie von Gonzalez et al. (2015) wurden fünf männliche gesunde Probanden untersucht. Im Mittel waren die Probanden zu Studienbeginn $37,8 \pm 14,8$ Jahre alt mit einer Körpergröße von $180,4 \pm 10,1$ cm und einem Gewicht von $75,2 \pm 7,6$ kg. Eine detaillierte Übersicht über die zentralen Charakteristika der Probanden ist in Tabelle 4.2 dargestellt. Da es sich um die Re-Analyse einer Fremdstudie handelt, liegen weitere Informationen über die Probanden nicht vor.

Tabelle 4.2.: Charakteristiken der Teilnehmer der Studie, Datensatz 1

	P1	P2	P3	P4	P5
Alter (zu Studienbeginn)	30	61	28	26	44
Größe [cm]	193	182,5	183,5	176	166
Gewicht [kg]	81,9	72,6	83,4	64,7	73,4

Jeder Proband absolvierte vier verschiedene Belastungstests auf einem Cyclus2-Fahrradergometer (RBM Elektronik-Automation GmbH, Leipzig, Deutschland). Alle Probanden waren zum Zeitpunkt der Untersuchung gut trainierte Hobby-Athleten. Während der Tests wurden die Herzfrequenz, die erbrachte Watt-Leistung, sowie über ein Spirometer (Ergostik, Geratherm Respiratory GmbH, Bad Kissingen, Deutschland) die kontinuierliche Gaskonzentration jedes Atemzuges in Form von $\dot{V}O_2$ und $\dot{V}CO_2$ aufgezeichnet.

Die vier Testprotokolle wurden dabei wie folgt definiert:

1. Test 01: ein inkrementeller Stufentest. Das Protokoll startet mit einer Belastung von 80 W und wird alle 3 Minuten um 20 W erhöht.
2. Test 02: eine Kombination aus vier Sprints und einem Rampentest. Jeder Sprint dauert 6 s, wobei je zwei Sprints vor bzw. nach dem Rampentest durchgeführt werden. Damit sollen die maximale Wattleistung des Probanden sowie die $\dot{V}O_2$ -Profile in einem erholtten und in einem ermüdeten Zustand gemessen werden.
3. Test 03: ein variabler Stufentest. Die einzelnen Stufen variieren in ihrer Belastung und Dauer, wobei zwischen niedriger, moderater und hoher Intensität gewechselt wird. Auch die Dauer des linearen Anstiegs bzw. Abfalls der Belastung variiert zeitlich.
4. Test 04: ein synthetischer *hill climb*-Test, der über einen Simulator kontrolliert wird. Die Belastung wird hauptsächlich von der Steigung des simulierten Pfades sowie dem Gewicht des Probanden determiniert. Während eine festgelegte Kadenz beibehalten wurde, durften die Probanden die Intensität durch Umschalten des Gangs variieren.

Die vier Profile sind beispielhaft für einen Probanden in Abbildung 4.1 dargestellt. Die vorliegenden Daten wurden von Gonzalez et al. (2015) geglättet und bezüglich ihrer Abtastraten

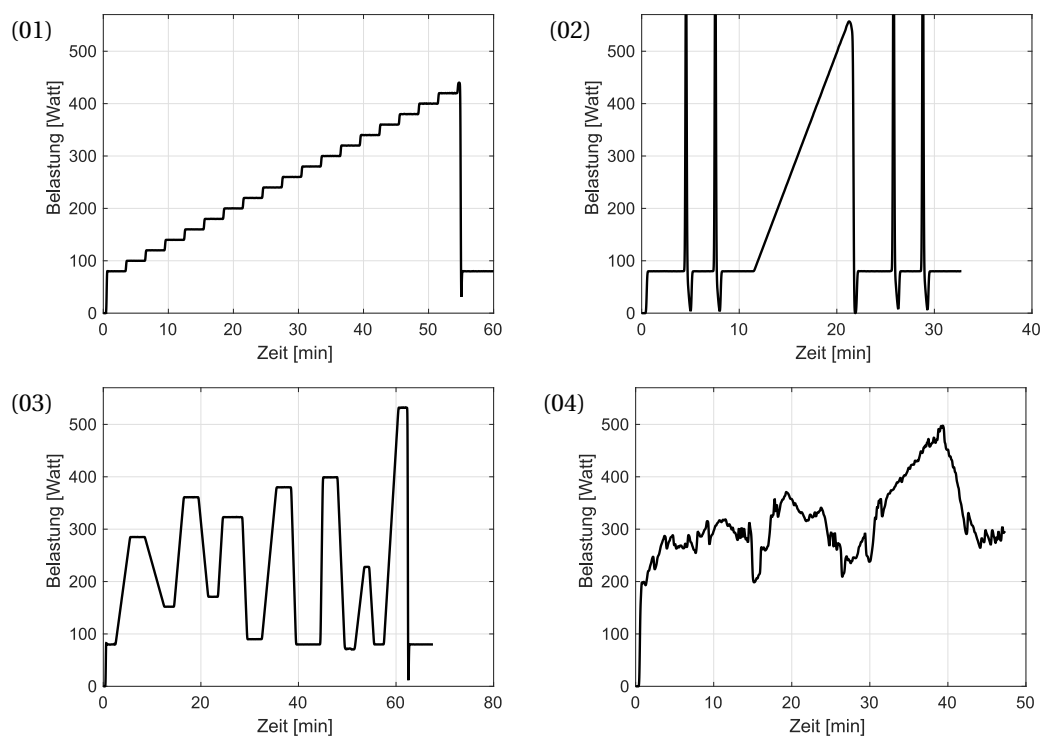


Abbildung 4.1.: Exemplarische Darstellung der vier Testprotokolle 01-04: (01) inkrementeller Stufentest, (02) Sprint- und Rampentest, (03) variabler Stufentest, (04) synthetischer *hill climb*-Test

umgerechnet, um das Rauschen innerhalb der Daten zu reduzieren und um unterschiedliche Messzeiten miteinander zu harmonisieren.

Bei der hier durchgeführten Forschungsarbeit handelt es sich um eine Re-Analyse dieser bereits existierenden Daten zu einer ähnlichen Fragestellung. Aus diesem Grund war eine Einflussnahme auf die Methodik der Studie nicht möglich. Es wurden die bereits vorverarbeiteten und Rausch-reduzierten Daten genutzt.

4.2. Datensatz 2: Daten aus einer Studie von Schrot mit Profi-Athleten

In der in den Jahren 2011-2012 erstellten Studie von Dr. Christian Schrot wurden vier Landeskader-StraßenradSPORTler untersucht. Eine detaillierte Charakteristik der Sportler kann der ursprünglichen Studie von Schrot (2015) (Studie 2) entnommen werden. Die im Rahmen dieser Forschungsarbeit relevanten Charakteristiken und Daten werden im Folgenden kurz zusammengefasst. Da es sich bei dieser Forschungsarbeit um eine Re-Analyse bereits existierender Daten aus einer Studie von 2011/2012 zu einer anderen Fragestellung handelt, war eine Einflussnahme auf die Methodik der Studie nicht möglich.

Alle Experimente basieren auf den Leistungs- und Trainingslastdaten von vier männlichen Radsportlern (P1-P4) der deutschen Straßenradnationalmannschaft (Tabelle 4.3), die über Zeiträume von 31 (P1) bzw. 35 Wochen (P2, P3, P4) gesammelt wurden. Die Anzahl der zur Analyse nutzbaren Trainingseinheiten (ergometerbasiert und freies Training) ist dabei zusätzlich aufgeführt¹.

Die Saison ist dabei in vier verschiedene Trainingsphasen unterteilt: Eine Vorbereitungszeit von sieben (P1) bzw. neun Wochen für grundlegendes Ausdauertraining einschließlich Kraftausdauertraining, eine zweite Vorbereitungszeit von sieben Wochen für grundlegendes Ausdauertraining einschließlich Maximalkrafttraining, acht Wochen zur Wettbewerbsvorbereitung einschließlich Intensitätstraining und sieben (P1) bzw. neun Wochen mit mehreren Wettkämpfen. Tage ohne Radfahrtraining am Ende der letzten Periode werden nicht betrachtet.

Tabelle 4.3.: Charakteristiken der Teilnehmer der Studie, Datensatz 2

	P1	P2	P3	P4
Alter (zu Studienbeginn)	17	17	18	18
Größe [cm]	174	180	182	179
Gewicht [kg]	61.2 ± 0.6	68.3 ± 0.9	66.1 ± 0.6	60.8 ± 0.6
Trainingspensum pro Jahr [km]	ca. 18'000	ca. 17'000	ca. 18'000	ca. 18'000
Ergometer-Trainings	21	18	23	19
Freie Trainingseinheiten	117	120	130	136

In unregelmäßigen Abständen fanden Ergometer-Tests zur Leistungsdiagnostik statt. Innerhalb der Leistungsdiagnostik wurde auf die Erhebung der Blutlaktatkonzentration und ähnlicher Leistungsparameter verzichtet. Stattdessen fanden die Tests bis zur maximalen Ausbelastung statt. Die vorliegenden empirischen Leistungswerte entsprechen der innerhalb der Ergometrie relativen Arbeit in kJ/kg, also der spezifischen Verbrennungsenergie in Relation zur Masse. Die Fahrradleistung als relative akkumulierte Arbeit (kJ/kg) wurde dabei bei den Probanden 1 und 2 während eines meist wöchentlichen inkrementellen Stufentests mit Ausbelastung auf einem Fahrradergometer quantifiziert. Der Widerstand betrug zu Beginn des Tests 150 Watt, gefolgt von einer Erhöhung von 10 W alle 10 Sekunden. Die Probanden 3 und 4 absolvierten zur Leistungsüberprüfung in regelmäßigen Abständen einen 5-Minuten-Time-Trial-Tests (All-Out-Test) auf einem Rollentrainer. Dabei sollte in jedem Test die maximal mögliche durchschnittliche Wattleistung über fünf Minuten erzielt werden. In der damaligen Studie wurden diese beiden Verfahren zur Leistungsfeststellung gewählt, da sie das reguläre Training verhältnismäßig wenig beeinflussen und nur einem möglichst geringen Einfluss der Motivation unterliegen. Ergometer-Trainingseinheiten liegen von allen vier Probanden vor.

¹Einige wenige Datensätze mussten auf Grund fehlender Herzfrequenzaufzeichnungen exkludiert werden.

4.3. Datensatz 3: Daten von Hobby-Sportlern

In einer 2014-2015 erstellten Studie haben drei männliche Probanden teilgenommen. Alle drei Probanden fuhren während der Studie regelmäßig Fahrrad. Über einen Zeitraum von sieben Monaten wurden alle zwei bis vier Wochen standardisierte Ergometer-Test auf einem Cyclus2-Fahrradergometer (RBM Elektronik-Automation GmbH, Leipzig, Deutschland) durchgeführt. Messungen, bei denen die Wattleistung, die Herzfrequenz oder Laktatmesswerte gänzlich oder zu mehr als 30% fehlen, werden von den Analysen ausgeschlossen.

Tabelle 4.4.: Charakteristiken der Teilnehmer der Studie, Datensatz 3

	P1	P2	P3
Alter (zu Studienbeginn)	46	23	33
Größe [cm]	160	180	178
Gewicht [kg]	68	85	65
Anzahl nutzbarer Testungen	4	4	6

Tabelle 4.4 gibt eine Übersicht über die Charakteristiken der einzelnen Teilnehmer. Während der Tests wurde am Ende einer jeden Stufe Laktat über einen Blutstropfen aus dem Ohrfläppchen mit dem »Lactate SCOUT +« (SensLab GmbH, Deutschland) gemessen. Die Herzfrequenz wurde kontinuierlich über Herzfrequenzsensoren von »Polar« erfasst.

Als Protokoll für die Fahrradergometer-Tests wurde ein Stufentest gewählt: Die Startbelastung betrug 25 Watt. Die Stufe wurde alle drei Minuten um 25 Watt bis zum individuellen Abbruch erhöht.

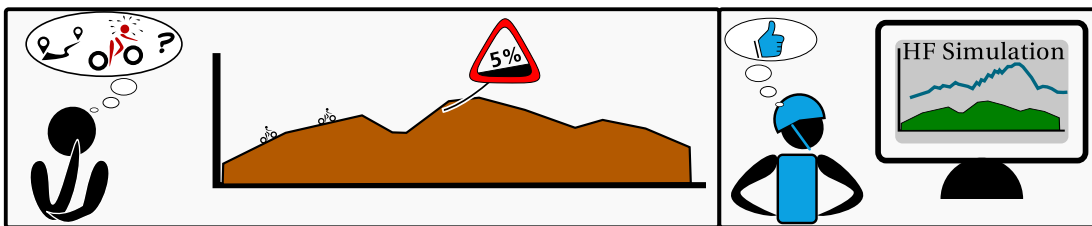
Da es sich bei allen drei Probanden um Hobby-Athleten handelt, kann nicht davon ausgegangen werden, dass der individuelle Abbruch des Tests einer Ausbelastung entspricht. Es handelt sich sehr wahrscheinlich um ausschließlich submaximale Belastungen, sodass die Laktatdiagnostik als Methode zur Bestimmung der Ausdauerleistungsfähigkeit gewählt wurde. Röcker (2013) zufolge eignet sich auch und gerade die alleinige Nutzung der Laktatdiagnostik zur Bestimmung der Anaeroben Schwelle:

»In der sportbezogenen Leistungsdiagnostik und Trainingssteuerung dürfte die alleinige Durchführung der Laktatdiagnostik [im Vergleich zur simultanen Anwendung der Spiroergometrie] überlegen sein, da bei Gesunden und insbesondere bei Sportlern Referenzpunkte wie die »Anaerobe Schwelle« mit Laktat zuverlässiger und auch in direkter Relation zur Leistung bestimmt werden können.«

Das gewählte Protokoll entspricht den Empfehlungen zur Laktatabnahme bei Freizeitsportlern von Röcker (2013).

Basierend auf den einzelnen Laktatmessungen wurde die Belastung an der individuellen anaeroben Schwelle (IAT, vgl. Unterabschnitt 1.2.3, »Messgrößen der Beanspruchung in der Diagnostik«) über die Software »Ergonizer« (Ergonizer, Freiburg, Deutschland) bestimmt. Der Schwellenbestimmung im *Ergonizer* liegt die individuellen anaeroben Schwelle nach Wasserman und McIlroy (1964) und Dickhuth et al. (1991) zugrunde – die IAT entspricht einem Nettoanstieg der Blutlaktatkonzentration von $1,5 \text{ mmol/l}$ basierend auf der Konzentration an der LT. Die LT zeigt den Beginn des Anstiegs der Blutlaktatkonzentration an (vgl. »Schwellenkonzepte«, Unterabschnitt 1.2.4). Die Messdaten werden mit Splines interpoliert – das Verfahren wird von Roecker et al. (1998) näher beschrieben.

Das Faltungsmodell: Ein neuer Ansatz zur Simulation und Vorhersage der Herzfrequenz im Training



Abstract.

Die Simulation und Vorhersage des Herzfrequenzverhaltens über vollständige Trainingseinheiten erfordert eine geeignete Modellierung. Durch das Fehlen interpretierbarer Modelle, die zu diesem Zweck geeignet sind, wurde ein Faltungsmodell konzipiert. Das Faltungsmodell wird in diesem Kapitel vorgestellt und phänomenologisch motiviert.

Die Auswertung anhand einer Studie zeigt, dass das Faltungsmodell zur Simulation und Vorhersage unterschiedlicher Beanspruchungsparameter (Herzfrequenz, $\dot{V}O_2$ und $\dot{V}CO_2$) im Rahmen vollständiger Trainingseinheiten auf dem Ergometer gut geeignet ist. Fragestellung 1 lässt sich schließlich mit den in diesem Kapitel erzielten Resultaten bestätigen.

Eine geeignete und möglichst zuverlässige Modellierung des Herzfrequenzverhaltens bei körperlicher Belastung findet insbesondere im Ausdauersport zahlreiche Anwendungsmöglichkeiten. Dabei ist je nach spezifischer Fragestellung auch die Vorhersage des Herzfrequenzverhaltens über einen definierten Zeithorizont relevant. Die genaue Ausgestaltung des zeitlichen Horizonts obliegt dem konkreten Anwendungsfall. Für eine unmittelbare Belastungssteuerung auf einem Ergometer oder Laufband beispielsweise reicht es oftmals, einige wenige Sekunden vorherzusagen. Beispiele dieser Nutzungsvariante finden sich bei Koenig et al. (2009), Dur-e Zehra Baig et al. (2010), Cheng et al. (2008), Mazzoleni et al. (2016) und anderen. Bei diesem Anwendungsszenario kann die Belastung kurzfristig und in Abhängigkeit der für die nächsten Sekunden vorhergesagten Herzfrequenz reduziert oder gesteigert werden, um eine Trainerin oder einen Trainer dabei zu unterstützen, das Training bestmöglich zu gestalten. Besonders im Reha-Bereich stehen dabei Sicherheitsaspekte und die Vermeidung einer Überlastung im Fokus. Speziell im Kontext der Vorhersage weniger Sekunden gibt es zahlreiche Herzfrequenzmodelle, die dafür sehr gut geeignet sind. Dazu gehören die Modelle aus den soeben genannten Publikationen und viele weitere Modelle wie beispielsweise von Mohammad et al. (2011b), Stork et al. (2017), Le et al. (2008), Argha et al. (2016), Hunt et al. (2015) und anderen publiziert. Einige dieser Modelle wurden exemplarisch bereits in »Modellierung der Herzfrequenz als akute Anpassung« (Abschnitt 2.2) vorgestellt und diskutiert. Wie in ebenjenem Kapitel dargelegt, eignen sich diese Modelle jedoch kaum dafür, das Herzfrequenzverhalten über längere Belastungszeiträume hinweg mit hinreichender Genauigkeit zu prognostizieren. Die Vorhersage längerer Zeiträume ohne Echtzeit-Rückkopplung ist jedoch relevant, sobald eine Planung vorangehend zu einem Training gestaltet werden soll oder wenn eine gute Modellgüte ohne Überanpassung (*overfitting*) erzielt werden soll. Dies kann beispielsweise der Fall sein, wenn

- die Beanspruchungsreaktion für eine bestimmten Strecke simuliert werden soll;
- bei der vorab-Planung einer Belastung im Reha-Bereich abgeschätzt werden muss, ob eine bestimmte Strecke für eine Person in der aktuellen Kondition realistisch ist;
- allgemein eine »virtuelle Leistungsdiagnostik« (d.h. die Vorhersage der Beanspruchung eines diagnostischen Leistungstests auf Basis regulärer Trainingseinheiten) durchgeführt werden soll.

Insbesondere kann die Prognose der Herzfrequenz dabei unterstützen, eine mögliche Verknüpfung zwischen Herzfrequenz und Leistung zu identifizieren, zu analysieren und zu nutzen oder mit Hilfe virtueller Leistungsdiagnostiken Rückschlüsse auf die Leistung zu ziehen. Das *Faltungsmo­dell* wurde entwickelt, solchen umfangreicheren Vorhersagen gerecht werden zu können. In diesem Kapitel wird die erste Fragestellung untersucht:

Fragestellung 1:

Lassen sich mit dem Faltungsmodell bessere Vorhersagen der Herzfrequenz während körperlicher Belastung im Ausdauersport für ganze Trainingssequenzen ohne Rückkopplung gemessener Herzfrequenzwerte treffen, als mit bisherigen Modellen der Herzfrequenzanalyse?

Zur Analyse dieser Fragestellung wird das Faltungsmodell im Folgenden vorgestellt und phänomenologisch motiviert (»Faltungsmodell«, Abschnitt 5.1). Anschließend werden die Resultate von Simulations- und Prädiktionsergebnissen des Faltungsmodells untersucht und mit den entsprechenden Ergebnissen eines Hammerstein-Wiener-Modells verglichen (»Analyse des Faltungsmodells: Modellvergleich«, Abschnitt 5.2). Da in der Literatur häufig Differentialgleichungsmodelle in allgemeiner oder spezieller Form (LTI-Modelle, Hammerstein-Wiener-Modelle) zur Herzfrequenzvorhersage genutzt werden, wird das Faltungsmodell in »Mathematische Betrachtung und Vergleich zu ähnlichen Modelltypen« (Abschnitt 5.3) exemplarisch mit der Struktur einiger in diesem Bereich typischer DGL-Modelle verglichen. Dazu wird das Faltungsmodell selbst in die Form eines DGL-Modells und eines Hammerstein-Wiener-Modells transformiert. Die Ergebnisse werden in einem »Zusammenfassung und Fazit« (Abschnitt 5.4) abschließend zusammengefasst.

Teile dieses Kapitels beruhen auf den Publikationen von Ludwig et al. (2016) und Ludwig et al. (2019).

5.1. Faltungsmodell

Wie in Abschnitt 2.2 zur »Modellierung der Herzfrequenz als akute Anpassung« dargelegt, gibt es unterschiedliche phänomenologische Aspekte, die bei der Modellierung eines Modells zur Simulation des Herzfrequenzverhaltens bei Belastung berücksichtigt werden können. Um zunächst mit einem möglichst einfachen und phänomenologisch interpretierbaren Modell zu starten, wurde die Konzeption des Faltungsmodells am Fitness-Fatigue Modell von Calvert et al. (1976) angelehnt, da in diesem Modell aus mathematischer Sicht einige grundlegende Eigenheiten sowohl für die kurz- als auch für die langfristige Modellierung relevant sind. Dazu gehören:

- ein Startwert als Entsprechung des Zustands (bezüglich Trainingsleistung oder Herzfrequenz) vor dem Training;
- eine zeitliche Abschwächung der wirkenden Belastung (beispielsweise durch einen exponentiellen Filter), die abbildet, dass eine kürzliche Belastung einen größeren Einfluss auf die aktuelle Situation hat als eine Belastung, die etwas länger zurück liegt;
- sowie die Möglichkeit, die Belastung im Rahmen des funktionalen Zusammenhangs zu berücksichtigen und auf verschiedene Arten zu skalieren.

Die zeitlich verzögerte Reaktion des menschlichen Organismus auf einen Reiz (vgl. »Herzfrequenzverhalten bei körperlicher Belastung«, Abschnitt 1.3) lässt sich mathematisch beispielsweise über eine Faltung modellieren.

Definition 5.1: Faltungsmodell.

Seien u und v Zeitreihen mit Länge T und sei $t \in [0, T]$. Seien ferner $a_1, a_2, a_3, a_4 \in \mathbb{R}$ Parameter des Modells. Dann ist der Output y des Faltungsmodells schrittweise definiert durch:

$$y(t) = a_2 \cdot \left[\frac{1}{a_1} (u * e^{-\bullet/a_1})(t) \right]^{a_4} + a_3, \quad (5.1)$$

wobei der Term $(u * v)(t)$ die Faltung innerhalb des Definitionsbereichs der Funktionen u und v beschreibt, d.h. $(u * v)(t) = \int_0^t u(s) \cdot v(t-s) ds$ zum Zeitpunkt t . Außerdem gilt: $e^{-\bullet/a_1} : (\tau \mapsto e^{-\tau/a_1})$. Im diskreten Fall entspricht das der diskreten Faltung $\sum_s u(s) \cdot v(t-s)$ zum Zeitpunkt $t \in [0, T]$, wobei s den gleichen Definitionsbereich hat wie die Funktionen u und v .

Die Funktion u stellt dabei die Belastungssequenz (z.B. in Watt) dar, während t ein (beliebiger) Zeitpunkt innerhalb dieser Sequenz ist. Auf phänomenologischer Ebene lassen sich die verwendeten Parameter dabei wie folgt interpretieren:

- a_1 beschreibt die Auswirkung der früheren Belastung auf die aktuelle Herzfrequenz, d.h. wie viel Einfluss die frühere Belastung hat. Je größer der Wert für a_1 , desto mehr Einfluss haben Werte aus früheren Belastungen auf den aktuell simulierten Herzfrequenzwert zum Zeitpunkt t , sodass jeder einzelne Belastungswert abgeschwächt berücksichtigt wird. Es ergibt sich eine stärkere Glättung der Simulationskurve.
- a_2 erklärt den proportionalen Effekt der zunehmenden Belastung der Herzfrequenz, d.h. dieser Parameter zeigt, wie stark die Beanspruchungsreaktion auf die Belastung ist und wie stark die Herzfrequenz steigt.
- a_3 entspricht einem *Levelparameter*, der als additive Konstante verwendet wird, um die vorhergesagte Herzfrequenz auf ein geeignetes Niveau anzuheben. Phänomenologisch kann der Parameter als Höhe der Herzfrequenz vor der Belastung interpretiert werden. Ähnlich zu der spezifischen Ruheherzfrequenz einer Person kann hier die Herzfrequenz vor einer Belastung verwendet werden, auf deren Basis die Herzfrequenz unter Belastung ansteigt.
- a_4 ermöglicht eine nichtlineare Reaktion der Herzfrequenz auf eine zunehmende Belastung und nahe der persönlichen Leistungsgrenze. Dieser Parameter kann phänomenologisch eine Idee vermitteln, wie die körperliche Reaktion auf Belastung sich insbesondere zu Beginn und Ende eines Trainings sowie bei Belastungsspitzen verhält.

Im Folgenden wird das Faltungsmodell auf seine Eignung zur Herzfrequenzvorhersage im Rahmen einer Studie untersucht und mit Resultaten eines anderen publizierten Hammerstein-Wiener-Modells verglichen.

5.2. Analyse des Faltungsmodells: Modellvergleich

Die Fragestellung, ob sich mit dem Faltungsmodell bessere Vorhersagen der Herzfrequenz während körperlicher Belastung im Ausdauersport für ganze Trainingssequenzen und ohne Rückkopplung gemessener Herzfrequenzwerte treffen lassen, lässt sich geeignet durch den direkten Vergleich mit anderen Modellen überprüfen.

In der Vergangenheit wurden bereits einige unterschiedliche Modelle (DGL-Modell von Cheng et al. (2007b), DGL-Modell von Paradiso et al. (2013), LTI-Modell von Baig et al. (2010) und Takagi-Sugeno-Modell von Mohammad et al. (2011b)) ausführlich in der Studie von Ludwig et al. (2015) hinsichtlich ihrer Güte zur Herzfrequenzvorhersage beim Ergometertraining verglichen. Da die Vergleichsmodelle allesamt nicht für die Vorhersage größerer Zeithorizonte konzipiert wurden und ursprünglich auf die Vorhersage der Herzfrequenz von nur wenigen Sekunden (z.B. zur unmittelbaren Steuerung eines Ergometers) beschränkt wurden, wurde auch die Vorhersage für mehrere unterschiedliche Zeithorizonte (alle 10 s bis zu einer Minute, zusätzlich 90 s, 120 s und komplette Trainingseinheiten mit einer Länge von rund einer halben Stunde) betrachtet. Dabei schwankten die mittleren RMSE-Fehlerwerte je nach Modell zwischen 6,13 bpm (LTI-Modell) und 9,90 bpm (DGL-Modell von Cheng) für einen Vorhersagezeitraum von bis zu 60 s. Bei der Vorhersage vollständiger Trainingseinheiten ergaben sich Fehler im Bereich von 8,64 bpm (Takagi-Sugeno) bis 11,46 bpm (LTI-Modell). Wenngleich die Modelle damit grundsätzlich auch zur Simulation und Vorhersage vollständiger Trainingseinheiten verwendet werden können, ist eine physiologische Interpretation der Modelle kaum oder nur eingeschränkt möglich. Durch die Vielzahl der dort verwendeten freien Parameter neigen die Modelle zu einer Überanpassung an die Kalibrierungsdaten, was insbesondere bei einer Kalibrierung auf regulären Trainingseinheiten und einer anschließenden Herzfrequenzprognose problematisch werden kann.

In Ludwig et al. (2016) wurden schließlich zwei dieser Modelle (das LTI-Modell von Baig et al. (2010) und das Takagi-Sugeno-Modell von Mohammad et al. (2011b)) auf einem anderen Datensatz (Fahrradergometer) mit dem Faltungsmodell verglichen. Eine Übersicht über die dort erzielten Ergebnisse mit den verglichenen Modellen ist in Abbildung 5.1 basierend auf dem RMSE gegeben. Dabei wurden insgesamt 30 Vorhersagen mit unterschiedlich vielen Kalibrierungsdatensätzen durchgeführt. Die beste Prognosegüte mit der geringsten Varianz und ohne große Ausreißer zeigt sich mit dem Faltungsmodell. Das Takagi-Sugeno-Modell erreicht eine bessere Prognosegüte als das LTI-Modell, wobei einige Ausreißer mit Fehlerwerten von über 25 bpm zu unbrauchbaren Vorhersagen führen. Insbesondere eine

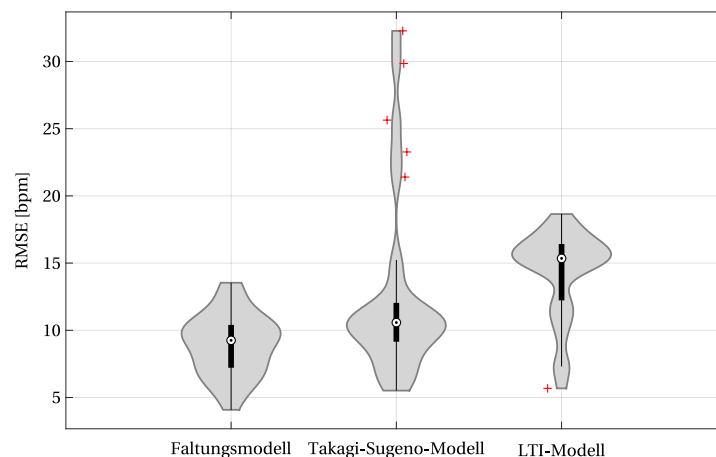


Abbildung 5.1.: Vergleich der Prognosegüte vollständiger Trainingseinheiten zwischen dem LTI-Modell, dem Takagi-Sugeno-Modell und dem Faltungsmodell aus der Studie von Ludwig et al. (2016) als Violinplot

Vorhersage mit dem Takagi-Sugeno-Modell ist damit unzuverlässig: Bei einer echten Vorhersage ohne Validierungsmöglichkeit sind Aussagen darüber, welche Ergebnisse valide sind und welche als Ausreißer gewertet werden müssen nur in Ausnahmefällen (wie beispielsweise im Falle unrealistisch hoher oder niedriger Herzfrequenzwerte) möglich. Die Ergebnisse legen die Schlussfolgerung nahe, dass das Faltungsmodell mit mittleren Fehlerwerten von unter 10 bpm Abweichung zur Vorhersage vollständiger Trainingssequenzen grundsätzlich geeignet ist und insbesondere im Vergleich zu den betrachteten Herzfrequenz-Prognosemodellen präzisere Vorhersagen trifft.

Es bleibt zu untersuchen, ob das Faltungsmodell auch zu solchen Vorhersagemodellen vergleichbare Ergebnisse erzielen kann, die spezifisch zum Zweck der Modellierung und Vorhersage vollständiger Trainingssequenzen konzipiert wurden. In der Zwischenzeit wurde von Gonzalez et al. (2016) ein Hammerstein-Wiener-Modell entwickelt, das wie das Faltungsmodell das Ziel verfolgt, unterschiedliche Beanspruchungsreaktionen (Herzfrequenz, $\dot{V}O_2$ und $\dot{V}CO_2$) vollständiger Trainingseinheiten zu simulieren und vorherzusagen. Die nachfolgende Studienarbeit stellt daher den Vergleich mit dem genannten Hammerstein-Wiener-Modell dar, um die Validität des Faltungsmodell nun auch mit einem Modell ähnlicher Zielsetzung zu vergleichen. Die verwendeten Trainingsdaten entsprechen den Daten aus Gonzalez et al. (2016) und wurden von den Autoren zur Re-Analyse zur Verfügung gestellt.

Innerhalb der Studie wird sich zeigen, dass das Faltungsmodell vollumfänglich vergleichbare und in Bezug auf die Vorhersage einer kompletten Trainingssequenz valide Resultate erzielt. Das Faltungsmodell ist damit zur Verwendung von Simulationen und Prognosen der Herzfrequenz wie auch für andere Beanspruchungsreaktionen ($\dot{V}O_2$ und $\dot{V}CO_2$) gut geeignet.

5.2.1. Experimenteller Aufbau und Methoden

Zur Analyse der Simulations- und Vorhersagegüte des Faltungsmodells wurden zwei verschiedene Arten von Vorhersageexperimenten durchgeführt. Die zugrundeliegenden Daten wurden in »Datensatz 1: Daten aus einer Studie von Gonzalez et al.« (Abschnitt 4.1) beschrieben. Die Vorhersageexperimenten können einerseits alleinstehend betrachtet und bewertet werden. Gleichzeitig können Teile der Experimente auch mit korrespondierenden Ergebnissen des Hammerstein-Wiener-Modells aus einer Folgestudie von Gonzalez et al. (2016) verglichen werden. Für die Betrachtung der Simulationsgüte wurde das Modell für jeden Test eines jeden Probanden einzeln kalibriert. Für die Vorhersage wurden zwei unterschiedliche Szenarien betrachtet:

1. Vorhersageexperiment *A*: Das Modell wird pro Person auf einem der vier spezifizierten Testprotokolle kalibriert. Die empirischen Belastungsdaten der anderen drei Protokolltypen werden anschließend vorhergesagt. Einzelne Experimente werden entsprechend desjenigen Tests, auf dem die Modellkalibrierung stattfindet, benannt: Bei einer Kalibrierung auf Testart 01 wird das Experiment mit *Vorhersageexperiment A.1* bezeichnet und analog für die anderen Kalibrierungen.
2. Vorhersageexperiment *B*: Es wird eine »Leave-One-Out«-Kreuzvalidierung durchgeführt: Das Modell wird auf drei Tests je einer Person kalibriert und der verbleibende Test wird vorhergesagt. Dieses Vorgehen wird für alle möglichen Kombinationen der Kalibrierung wiederholt.

Ein Vergleich mit den Ergebnissen von Gonzalez et al. (2016) ist für die Kalibrierung der einzelnen Datensätze und im Rahmen der Vorhersage für Experiment *A.3* möglich. Für die anderen der durchgeführten Experimente liegen aus der Studie von Gonzalez et al. (2016) keine Ergebnisse zur Simulations- oder Vorhersagegenauigkeit vor, sodass diese hier nicht verglichen werden können. In ihrer Studie sind Gonzalez et al. (2016) zwar ursprünglich von einem Hammerstein-Wiener-Modell ausgegangen, konnten dieses im Verlauf ihrer Experimente jedoch zu einem reinen Wiener-Modell reduzieren. Dieses haben sie mit einer Anpassung ihres zuvor in Gonzalez et al. (2015) publizierten »dynamischen Modells« verglichen. Im Rahmen dieser Studie werden die Ergebnisse dieser beiden Modelle zum Vergleich aufgeführt.

Für die Modellkalibrierung wird die Methode der kleinsten Quadrate von Levenberg-Marquardt genutzt, wie von Busso et al. (1997) empfohlen und wie in »Zeitreihenanalyse und statistische Parameter« (Unterabschnitt 2.1.1) beschrieben. Die Evaluation der Ergebnisse findet unter Verwendung des RMSE und des MAPE statt.

5.2.2. Ergebnisse

Im Folgenden werden zunächst die Simulationsergebnisse dargestellt. Anschließend werden die Ergebnisse der unterschiedlichen Vorhersageexperimente präsentiert, die bei allen drei physiologischen Beanspruchungsparametern eine zu Gonzalez et al. (2016) vergleichbare Genauigkeit liefern. Dabei ist zu beachten, dass sehr geringe Fehler in der Kalibrierung auf eine Überanpassung hindeuten können und die Ergebnisse der Vorhersage von größerer Relevanz sind.

Simulation.

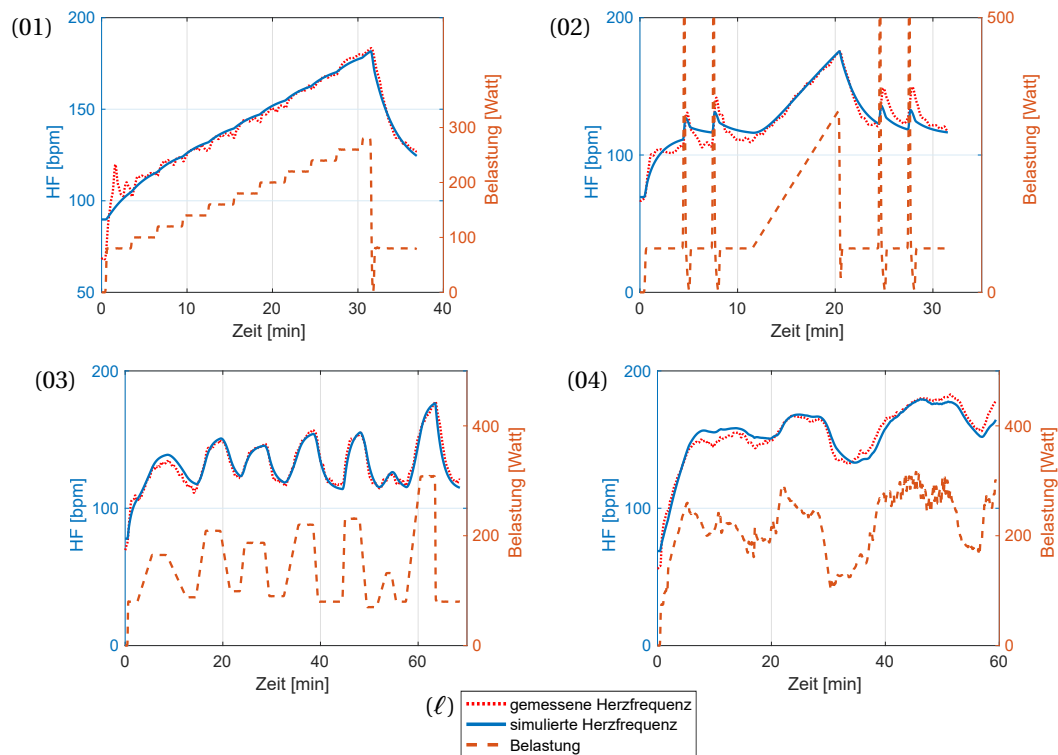


Abbildung 5.2.: Eine Simulation der Testprotokolle 01-04 mit dem Faltungsmodell ist beispielhaft für die Herzfrequenz-Beanspruchung (empirisch und simuliert) von Proband #05 zur entsprechenden Belastung dargestellt. Die Fehlerwerte sind dabei: (01) 3.69%, (02) 2.29%, (03) 2.11%, (04) 2.95%

Zunächst wird das Modell für jede Person auf allen vier Testprotokollen separat kalibriert. Die Simulation entspricht dabei den Ergebnissen der Modellkalibrierung, d.h. die Parameter werden auf dem jeweils betrachteten Datensatz kalibriert, auf dem die Modellgüte analysiert wird. Es werden Durchschnittswerte über alle Probanden und alle Tests berechnet. Die gesamten Simulationsergebnisse sind in Tabelle 5.1 dargestellt. Die Ergebnisse von Gonzalez et al. (2016) werden zum Vergleich zitiert. Für alle betrachteten physiologischen Reaktionen (Herzfrequenz und Sauerstoffkinetik) einer Person zeigte das von Gonzalez et al.

Tabelle 5.1.: Durchschnittliche RMSE- und MAPE-Fehlerwerte für die Simulation aller Tests über alle Probanden und für die drei betrachteten Beanspruchungsmaße. Zellen mit den geringsten Fehlern sind für jedes Beanspruchungsmaß hervorgehoben. Referenzwerte des Dynamischen Modells und des Wiener Modells stammen aus Gonzalez et al. (2016)

	HF		$\dot{V}O_2$		$\dot{V}CO_2$	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
Dynamisches Modell	4.55	2.5	0.09	3.1	0.12	4.7
Wiener Modell	n/a	0.9	0.06	1.8	0.08	2.9
Faltungsmodell	4.95	2.92	0.13	4.0	0.18	6.60

(2016) vorgeschlagene Wiener Modell die besten Ergebnisse mit den geringsten Fehlerwerten. Beim direkten Vergleich zwischen dem Dynamischen Modell und dem Faltungsmodell sind die Fehler des dynamischen Modells etwas geringer.

Die Herzfrequenzsimulation mit dem Faltungsmodell ist in Abbildung 5.2 für einen Probanden beispielhaft dargestellt. In diesem Beispiel hat das Testprotokoll 01 einen etwas höheren Kalibrierungsfehler in Höhe von 3,69%. Es fällt auf, dass die empirische Herzfrequenz zu Beginn der Belastung einige Minuten zur Stabilisierung benötigt. Innerhalb der ersten drei Minuten weist die HF einen kurzen Spitzenwert auf, der vom Faltungsmodell nicht simuliert wird. Die Spitzen bei den Sprints in Test 02 werden vom Faltungsmodell ebenfalls nicht exakt abgebildet. In der Simulation von Test 03 werden die insgesamt kleinsten Fehlerwerte erzielt.

Prädiktion.

Im ersten Vorhersageexperiment (Experiment A) wird das Modell auf je einem Test für jeden Probanden einzeln kalibriert. Die physiologischen Beanspruchungsreaktionen des jeweiligen Probanden werden für die drei verbleibenden Tests vorhergesagt. Für das Faltungsmodell ist die resultierende Vorhersage von Test 04 in Abbildung 5.3 für einen Probanden exemplarisch dargestellt.

Die Vorhersageergebnisse bezüglich der Fehlerwerte (RMSE und MAPE im Median) sind in Tabelle 5.2 aufgeführt. Die Ergebnisse von Gonzalez et al. (2016) werden zum Vergleich zitiert und entsprechen Vorhersageexperiment A.3 (für die anderen Vorhersageexperimente aus A liegen keine von ihnen publizierten Ergebnisse vor). Im Vergleich des Wiener Modells mit dem Faltungsmodell sind die Fehler des Faltungsmodells für alle drei physiologischen Beanspruchungsmaße innerhalb von Experiment A.3 geringer. Im Vergleich des dynamischen Modells mit dem Faltungsmodell sind die Fehler des Faltungsmodells für die Herzfrequenz etwas höher und für die Sauerstoffkinetik etwas geringer. Insgesamt sind die Fehlerwerte bei allen Modellen in einem ähnlichen Bereich.

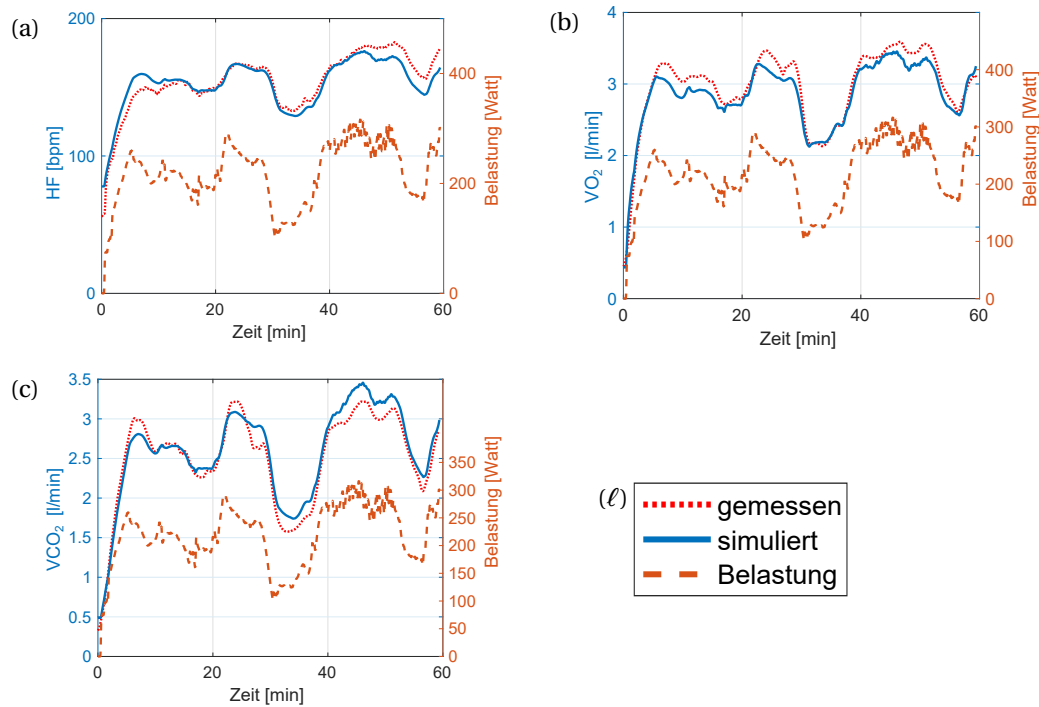


Abbildung 5.3.: Vorhersage von Testprotokoll 04 in Vorhersageexperiment A für (a) HF, (b) $\dot{V}O_2$, (c) $\dot{V}CO_2$ von Proband #05. Dabei sind die empirischen Werte, die simulierten Werte und die zugehörigen Belastungswerte über die Zeit dargestellt

Generell zeigen sich in Vorhersageexperiment A Unterschiede in der Vorhersagegenauigkeit zwischen verschiedenen Tests, die für die Modellkalibrierung verwendet werden. Die Kalibrierung basierend auf Test 02 liefert beispielsweise für jeden Probanden die besten Ergebnisse in der Herzfrequenzvorhersage. Gleichzeitig werden in Vorhersageexperiment A.02

Tabelle 5.2.: Prädiktion verschiedener physiologischer Beanspruchungsmaße. Die durchschnittlichen RMSE- und MAPE-Fehler sind für die unterschiedlichen Vorhersageexperimente in Abhängigkeit des zur Kalibrierung genutzten Tests dargestellt. Zellen mit den geringsten Fehlern sind für jedes Beanspruchungsmaß hervorgehoben. Referenzwerte des Dynamischen Modells und des Wiener Modells stammen aus Gonzalez et al. (2016)

Experiment	Modell	HF		$\dot{V}O_2$		$\dot{V}CO_2$	
		RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
A.3	Dynamisches Modell	7.46	4.4	0.30	8.8	0.41	12.6
	Wiener Modell	n/a	4.8	0.27	7.5	0.44	14.2
A.1	Faltungsmodell	8.61	5.28	0.30	9.03	0.39	14.32
A.2	Faltungsmodell	7.53	4.50	0.24	7.66	0.49	19.96
A.3	Faltungsmodell	8.05	4.65	0.24	6.79	0.32	11.36
A.4	Faltungsmodell	8.08	5.06	0.25	7.53	0.36	12.68
B	Faltungsmodell	7.34	4.41	0.24	7.01	0.35	12.71

die schlechtesten Ergebnisse für die Vorhersage von $\dot{V}CO_2$ erzielt. Beim Vergleich aller verschiedenen Vorhersagekonfigurationen hat die Testwahl für die Kalibrierung im Falle der HF keine auffällige Auswirkung: Abweichungen der Fehler liegen bei bis zu 0,87%. Für die Sauerstoffkinetik variieren die Fehler der Vorhersageexperimente um 2,24% für $\dot{V}O_2$ und um 8,6% für $\dot{V}CO_2$. Der Fehler von fast 20% bei der Betrachtung von $\dot{V}CO_2$, der erreicht wird, wenn das Modell auf Test 02 kalibriert wird, unterscheidet sich deutlich von den Werten aus den Ergebnissen von Vorhersageexperiment A.3: Beim Vergleich der MAPE-Werte von Vorhersageexperiment A.2 mit denen aus A.3 lässt sich mit dem Welch-Test die Nullhypothese aufstellen, dass die Daten beider Gruppen die gleichen Mittelwerte haben. Mit »Gruppen« werden in diesem Kontext die Ergebnisse der unterschiedlichen Experimente bezeichnet, deren Ähnlichkeit statistisch miteinander verglichen wird. Die Testentscheidung basiert auf der Empfehlung von Rasch et al. (2011), wie in Unterabschnitt 2.1.3 über die »Kritische Betrachtung statistischer Parameter« dargelegt. Auf eine Überprüfung der Varianzhomogenität und der Normalverteilung wird dabei verzichtet. Zum Signifikanzniveau $\alpha = 0,05$ kann die Nullhypothese mit $p < 0,01$ verworfen werden. Mit 95%-iger Wahrscheinlichkeit unterscheiden sich die Mittelwerte der betrachteten Gruppen folglich signifikant voneinander.

Die Vorhersage der Herzfrequenz liefert für verschiedene Testprotokolle stabile Ergebnisse, unabhängig von den für die Modellkalibrierung verwendeten Daten. Abbildung 5.4 a, b und c zeigt exemplarisch die Vorhersage der Tests 01 (a), 02 (b) aus Vorhersageexperiment A.3 sowie die Vorhersage von Test 04 (c) aus Vorhersageexperiment A.1.

In Bezug auf die Kreuzvalidierung (Vorhersageexperiment B in Tabelle 5.2) sind die Vorhersagefehler für die HF kleiner als die Fehler in Experiment A, bei denen das Modell auf nur jeweils einem Test kalibriert wurde. Dies wird exemplarisch durch Abbildung 5.4 d veranschaulicht. Bei der Sauerstoffkinetik sind die Fehler in der Kreuzvalidierung im Median verglichen mit dem besten Ergebnis aus Experiment A um etwa 0,22% höher für $\dot{V}O_2$ und um etwa 1,35% höher für $\dot{V}CO_2$.

5.2.3. Interpretation und Diskussion

Um die Reaktion einer Person auf eine Belastung lediglich zu simulieren, d.h. zu reproduzieren, erreicht das Wiener Modell mit Fehlerwerten von maximal 3% für die Sauerstoffkinetik und mit unter 1% für die Herzfrequenzsimulation sehr präzise Ergebnisse. Da bei der Betrachtung der Vorhersagen hingegen die Ergebnisse des Wiener Modells vergleichsweise höhere Fehler aufweisen als die analytischen Modelle, muss davon ausgegangen werden, dass die hohe Simulationsgenauigkeit auf eine Überpassung zurückgeführt werden kann. Häufig sind maschinelle Lernansätze für die alleinige Simulation besser geeignet als analytische Modelle, da sie auch kleine Veränderungen – ebenso wie Ungenauigkeiten durch Rauschen und Artefakte – oft gut abbilden können. Ähnliche Beobachtungen wurden bereits in früheren Studien gemacht Füller et al. (2015); Gonzalez et al. (2016); Ludwig et al. (2015). Vor-

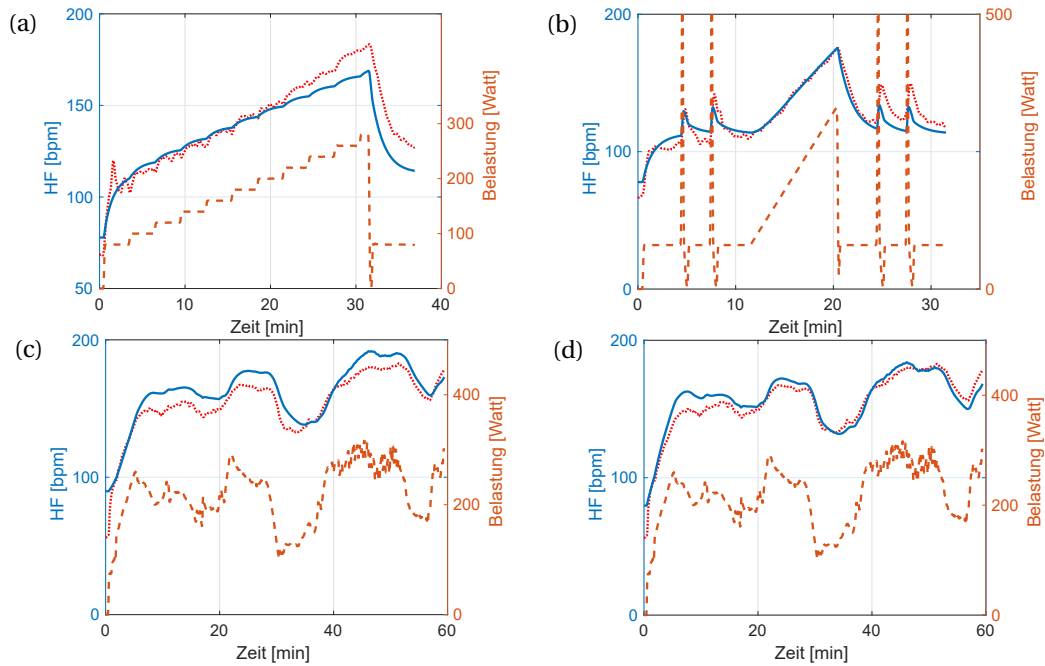


Abbildung 5.4.: Vorhersage der HF von Proband #05: (a) Test 01 in Vorhersageexperiment A.3, (b) Test 02 in Vorhersageexperiment A.3, (c) Test 04 in Vorhersageexperiment A.1, und (d) Test 04 in Vorhersageexperiment B

hersageexperiment A.3 zeigt, dass die beiden analytischen Modelle, d.h. das dynamische Modell von Gonzalez et al. (2016) und das Faltungsmodell, zueinander vergleichbare Ergebnisse liefern. Während das dynamische Modell in der Simulation eine geringfügig höhere Genauigkeit aufweist (eine höhere Genauigkeit von etwa 0,42% für die HF, 0,9% für $\dot{V}O_2$ und 1,9% für $\dot{V}CO_2$), sind die Prognoseergebnisse der beiden Modelle sehr ähnlich. Es ist denkbar, dass die etwas genauere Anpassung des dynamischen Modells in der Simulation auf der Verwendung von 12 Parametern basiert. Das Faltungsmodell hat mit lediglich vier Parametern weniger Freiheitsgrade.

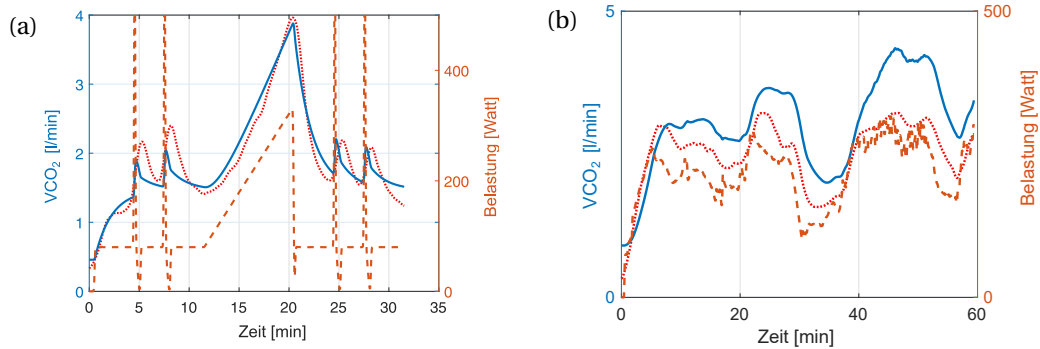


Abbildung 5.5.: Simulation von Test 02 und Vorhersage von Test 04 in Vorhersageexperiment A.1 für $\dot{V}CO_2$, Proband #05

Insgesamt zeigt Vorhersageexperiment A, dass der Art des Ergometertests keine besondere Bedeutung zukommt, auf dessen Grundlage das Modell kalibriert wird: Die Vorhersagequalität der HF zeigt sich weitgehend unabhängig und führt für alle durchgeführten Tests zu ähnlichen, unter Verwendung des Welch-Tests zum 5%-Niveau nicht signifikant voneinander abweichenden Fehlerwerten. Abweichungen in der $\dot{V}O_2$ -Vorhersage waren bis auf einen Vergleich ebenfalls nicht signifikant: Die Vorhersage von $\dot{V}CO_2$ in Experiment A.2 weist signifikant höhere Fehlerwerte auf als die Ergebnisse von Vorhersageexperiment A.3. Ein Beispiel für eine Vorhersage der $\dot{V}CO_2$ -Kinetik im vierten Testprotokoll ist in Abbildung 5.5 dargestellt. Während die Simulation auf Testprotokoll 02 nah an den empirischen $\dot{V}CO_2$ -Werten liegt und diese insbesondere bei den kurzen Sprints sogar unterschätzt, kann das Verhalten der $\dot{V}CO_2$ -Kurve in der Vorhersage des synthetischen *hill climb*-Tests kaum geeignet nachgebildet werden. Die Dynamik, die durch den Kalibrierungsdatensatz für die Parameter ermittelt wurde, ist in diesen Testarten eine völlig andere: Während die Kombination aus Rampentest und Sprints zu sehr spontanen, aber kurzfristigen Änderungen der Dynamik führt, provoziert der synthetische *hill climb*-Test unterschiedliche Belastungen im vergleichsweise langsamen Wechsel. Dies macht sich in der Beanspruchungsreaktion bemerkbar. Der Anstieg der $\dot{V}CO_2$ -Reaktion beim Rampentest wird über die Dauer steiler. Die abflachende Beanspruchung innerhalb der eingeschobenen Sprints wird grundlegend falsch modelliert: Bei den ersten Sprints kann die Höhe der $\dot{V}CO_2$ -Reaktion nicht abgebildet werden, während die abflachende Reaktion langsamer modelliert wird, als sie empirisch messbar ist. Bei den beiden letzten Sprints hingegen wird zwar die Höhe der maximalen Belastung gut modelliert, die Abflachung der Reaktion findet jedoch auch hierbei in der Modellierung zu langsam statt. Dieser Umstand einer langsam sinkenden modellierten Beanspruchungsreaktion nach einer höheren Belastung zeigt sich deutlich beim synthetischen *hill climb*-Test. Wird die erste Steigung noch passend modelliert, erhöht sich die Modellkurve systematisch nach jedem synthetischen Tal – die modellierte Beanspruchung sinkt zu langsam ab, während umgekehrt die Reaktion auf eine ansteigende Belastung im Verlauf passend zur Reaktion aus dem Rampentest immer stärker ansteigt. Infolge dessen steigt das simulierte $\dot{V}CO_2$ -Niveau im Verlauf des *hill climb*-Tests sukzessive an und die Diskrepanz zur empirischen Beanspruchung wird immer größer. Physiologisch lässt sich die Reaktion damit erklären, dass durch die zwischendurch sinkende Belastung die Atmung normalisiert werden kann und einer möglichen Hyperventilation entgegengewirkt wird. Die Parameter des Faltungsmodells hingegen sind durch die Kalibrierung auf dem Rampentest mit Sprints an die verhältnismäßig steilere und im Verlauf der Zeit steiler ansteigende Beanspruchungsreaktion mit zu flachem Absinken bei nachlassender Belastung optimiert.

Auch bei den Experimenten von A.1 gibt es bei der Vorhersage von Test 04 trotz fehlender signifikanter Unterschiede zu den anderen Experimenten eine Auffälligkeit: Wie bereits zuvor in Abbildung 5.4 (c) dargestellt, wird die empirische Beanspruchung des synthetischen *hill climb*-Protokolls (Test 04) in Vorhersageexperiment A.1 leicht überschätzt. Dies ist nicht

nur im dargestellten Beispiel der Fall, sondern trifft auf alle Probanden und alle Beanspruchungsmaße zu. Ähnlich wie im soeben diskutierten Beispiel lässt sich dieser Kalibrierungsfehler dadurch erklären, dass wechselnde Belastungen mit hoher Dynamik im Stufentest nicht vorhanden sind, während die Beanspruchung mit der Zeit steiler ansteigt. Insbesondere ein regelmäßiges Absinken der Belastung findet in dieser Testart nicht statt und wird folglich bei der Kalibrierung der Parameter nicht berücksichtigt. Dieser Effekt zeigt sich in der Vorhersage deutlich: Auch hier steigt die als Vorhersage simulierte Beanspruchung über die Zeit steiler an als die zugrunde liegende Belastung.

Ein genauerer Blick auf Vorhersageexperiment *B* deutet darauf hin, dass die Prognosegüte der Beanspruchungsreaktion von einer größeren Datenmenge zur Modellkalibrierung profitieren kann: Das Faltungsmodell hat in diesem Experiment eine nochmal höhere Genauigkeit erzielen können, wie in den einzelnen Vorhersageexperimenten aus Teil *A*. Insbesondere die Vorhersage der $\dot{V}CO_2$ scheint von der größeren Datenmenge bei der Kalibrierung zu profitieren. Ausreißer und unzureichende Prognosesimulationen wie bei Experiment *A.3* treten in Experiment *B* nicht auf. Die »leave-one-out«-Kreuzvalidierung aus Experiment *B* hat im Vergleich zu den Experimenten aus *A* den Vorteil, dass durch die unterschiedlichen Ergometerprofile in jedem Fall eine sehr unterschiedliche Dynamik der Beanspruchungsreaktion in der Kalibrierung berücksichtigt wird. Dieses Vorgehen macht eine Überanpassung an einzelne Daten sehr unwahrscheinlich, sodass die Prognosegüte verbessert wird.

5.2.4. Fazit zur Studie

Die Vorhersage von physiologischen Messungen wie HF, $\dot{V}O_2$, oder $\dot{V}CO_2$ basierend auf einer begrenzten Datenmenge kann von einer Reduktion der Anzahl der freien Modellparameter profitieren. Komplexe Modelle mit einer großen Anzahl an Parametern neigen eher dazu, besonders bei kleinen Datenmengen auf die gegebenen Daten zu überoptimieren (overfitting). Typischerweise führt die Überanpassung zwar zu sehr präzisen Ergebnissen bei der Kalibrierung, umgekehrt jedoch häufig zu einem erheblichen Genauigkeitsverlust bei der Prädiktion neuer Daten. Das vorgestellte Faltungsmodell hat nur vier freie Parameter. Die geringe Anzahl der freien Parameter führt zu einer Verbesserung der Vorhersagegenauigkeit, während sich maschinelle Lernansätze insbesondere zur reinen Simulation etwas besser zu eignen scheinen.

Die Trainingszonen für aerobes und anaerobes Training liegen (je nach Alter) rund 15-20 bpm bzw. 10% der maximalen Herzfrequenz voneinander entfernt: Die aerobe Trainingszone wird in der Regel mit 70 – 80% der maximalen Herzfrequenz angegeben, die anaerobe Trainingszone mit 80 – 90% der maximalen Herzfrequenz (Marx et al., 2018). Eine gute Vorhersage sollte entsprechend so geringe Fehler haben, dass eine individuelle Einschätzung der Trainingszone prinzipiell möglich ist. Vorhersagen mit Fehlern über 10% sind folglich unbrauchbar und Fehlerwerte von rund 5% oder weniger im ambitionierten Hobbysport-

Bereich wünschenswert. Mit dem Faltungsmodell konnten bei drei von fünf der durchgeführten Experimente mittlere Fehlerwerte von unter 5% erreicht werden. Die Fehlerwerte der anderen beiden Experimente (A.1 und A.4) sind mit 5,28% bzw. 5,06% ebenfalls gering.

Insgesamt zeigt sich das Faltungsmodell zur Vorhersage der Beanspruchung über vollständige, ergometerbasierte Trainingseinheiten als gut geeignet. Dabei sind die Ergebnisse im Vergleich mit dem Wiener Modell und einem dynamischen Modell vergleichbar, wobei das Faltungsmodell deutlich weniger freie Parameter beinhaltet.

5.3. Mathematische Betrachtung und Vergleich zu ähnlichen Modelltypen

Der Vergleich des Faltungsmodells mit einem Wiener-Modell im vorangehenden Kapitel hat gezeigt, dass das Faltungsmodell vergleichbar gute Ergebnisse besonders bei der Vorhersage unterschiedlicher Beanspruchungsparameter erzielt. Da Hammerstein-Wiener-Modelle wie in »Modellierung der Herzfrequenz als akute Anpassung« (Abschnitt 2.2) beschrieben und auch das Faltungsmodell Spezialfälle von Differentialgleichungsmodellen sind, wird das Faltungsmodell im Folgenden strukturell mit den DGL-Modellen und dem Wiener-Modell von Gonzalez et al. (2016) verglichen. Aus einem derartigen Vergleich lassen sich möglicherweise Erkenntnisse gewinnen, warum sich das Faltungsmodell und das untersuchte Wiener-Modell besser zur Simulation und Vorhersage vollständiger Trainingseinheiten eignen, als die referenzierten DGL-Modelle aus der Literatur.

Es wird sich zeigen, dass sich die Modellgleichung des Faltungsmodells aus Gleichung 5.1 auch als Differentialgleichung und im Speziellen als Wiener-Modell darstellen lässt und dass die betrachteten DGL-Modelle aus der Literatur strukturell Hammerstein-Modellen entsprechen.

5.3.1. Das Faltungsmodell als Differentialgleichungsmodell

Um das Faltungsmodell strukturell besser mit den in Unterabschnitt 2.2.2 (»Herzfrequenzmodellierung: Stand der Forschung «) betrachteten DGL-Modellen von Cheng et al. (2008) bzw. Scalzi et al. (2012) und Paradiso et al. (2013) vergleichen zu können, bietet es sich an, das Faltungsmodell selbst zunächst in die Form von einem DGL-Modell zu bringen.

Eine intuitive und in ihrer Struktur möglichst einfach gehaltene Differentialgleichung für eine Modellierung der Veränderung \dot{x} einer körperlichen Beanspruchungsreaktion x in Abhängigkeit einer Belastung u über die Zeit t lässt sich wie in Gleichung 5.2 darstellen:

$$\dot{x} = b(u - x) \tag{5.2}$$

Dabei ist $b \in \mathbb{R}$ eine beliebige Konstante, die zur Skalierung der Differenz genutzt wird. Das Modell lässt sich intuitiv so interpretieren, dass die Veränderung \dot{x} einer körperlichen Beanspruchung x beschrieben wird als die mit b skalierte Abweichung zwischen der Belastung u und der aktuellen Beanspruchung x . Die in Gleichung 5.3 dargestellte Funktion x löst die DGL.

$$x(t) = b \cdot \left(u * e^{-b \cdot \bullet} \right) (t) \quad (5.3)$$

Der Beweis dieser Lösung lässt sich wie nachfolgend gezeigt über die Ableitung \dot{x} darstellen: Es wird sich zeigen, dass die Ableitung der postulierten Lösung x aus Gleichung 5.3 exakt die Lösung der ursprünglichen Differentialgleichung aus Gleichung 5.2 mit Randbedingung $\forall t : u(t) \geq 0$ ist. Der Zusammenhang zwischen dem DGL bzw. seiner Lösung und dem Faltnungsmodell wird im Anschluss betrachtet.

Bei der Anwendung als Herzfrequenz- oder auch Leistungsentwicklungsmodell bleiben die Werte einer beliebigen Belastung u für alle Zeitpunkte t stets positiv oder sind gleich Null, sodass $\forall t : u(t) \geq 0$ angenommen werden kann.

Außerdem lässt sich die Funktion $x(t)$ wie folgt umformen:

$$\begin{aligned} x(t) &= b(u * e^{-b \cdot \bullet})(t) \\ \Leftrightarrow x(t) &= b \cdot \int_0^t u(s) \cdot e^{-b \cdot (t-s)} ds \\ \Leftrightarrow x(t) &= b \cdot \int_0^t u(s) \cdot e^{-bt+bs} ds \\ \Leftrightarrow x(t) &= b \cdot \int_0^t u(s) \cdot e^{-bt} \cdot e^{bs} ds \\ \Leftrightarrow x(t) &= b \cdot e^{-bt} \cdot \int_0^t u(s) \cdot e^{bs} ds \end{aligned} \quad (5.4)$$

Für die Ableitung ergibt sich damit:

$$\begin{aligned}
 \dot{x}(t) &= b \cdot \left(-b \cdot e^{-bt} \cdot \int_0^t u(s) \cdot e^{bs} ds + e^{-bt} \cdot \left[u(s) \cdot e^{bs} \right]_0^t \right) \\
 \Leftrightarrow \dot{x}(t) &= b \cdot \left(-b \cdot e^{-bt} \cdot \int_0^t u(s) \cdot e^{bs} ds + e^{-bt} \cdot \left[u(t) \cdot e^{bt} - u(0) \cdot e^{b \cdot 0} \right] \right) \quad | u(0) = 0 \\
 \Leftrightarrow \dot{x}(t) &= b \cdot \left(-b \cdot e^{-bt} \cdot \int_0^t u(s) \cdot e^{bs} ds + e^{-bt} \cdot \left[u(t) \cdot e^{bt} - 0 \right] \right) \\
 \Leftrightarrow \dot{x}(t) &= b \cdot \left(-b \cdot \int_0^t u(s) \cdot e^{-bt} \cdot e^{bs} ds + e^{-bt} \cdot u(t) \cdot e^{bt} \right) \quad | e^{-bt} \cdot e^{bt} = e^0 = 1 \tag{5.5} \\
 \Leftrightarrow \dot{x}(t) &= b \cdot \left(-b \cdot \int_0^t u(s) \cdot e^{-b \cdot (t-s)} ds + u(t) \right) \\
 \Leftrightarrow \dot{x}(t) &= -b \cdot \left(b \cdot (u * e^{-b \cdot \bullet})(t) \right) + b \cdot u(t) \quad | x(t) = b \cdot (u * e^{-b \cdot \bullet})(t) \\
 \Leftrightarrow \dot{x}(t) &= -b \cdot x(t) + b \cdot u(t) \\
 \Leftrightarrow \dot{x}(t) &= b \cdot (u(t) - x(t))
 \end{aligned}$$

□

Die Ableitung der postulierten Lösung entspricht also wie angekündigt genau der DGL aus Gleichung 5.2. Die Lösung des DGLs enthält demnach eine Faltung und eine Exponentialfunktion und hat bereits ohne weitere Anpassungen gewisse Ähnlichkeiten zum Faltungsmodell. Tatsächlich lässt sich das Faltungsmodell unmittelbar über die Funktionsgleichung

$$f(x(t)) = a_2 x(t)^{a_4} + a_3 \tag{5.6}$$

und mit Wahl des Parameters $b = \frac{1}{a_1}$ erhalten.

Als DGL lässt sich das Faltungsmodell damit formulieren über das in Gleichung 5.7 dargestellte Modell.

$$\begin{aligned}
 \dot{x} &= \frac{1}{a_1} (u - x) \\
 y &= a_2 x(t)^{a_4} + a_3
 \end{aligned} \tag{5.7}$$

Die Funktion $f(x(t)) = a_2 x(t)^{a_4} + a_3$ (bzw. das DGL-Modell in Gleichung 5.7) entspricht demnach unter Verwendung der Lösung x aus Gleichung 5.3 genau dem Faltungsmodell $y(t)$:

$$\begin{aligned}
 f(x(t)) &= a_2 x(t)^{a_4} + a_3 \\
 &= a_2 \left(b(u * e^{-b \cdot \bullet})(t) \right)^{a_4} + a_3 \quad | b = \frac{1}{a_1} \\
 &= a_2 \cdot \left(\frac{1}{a_1} (u * e^{-\bullet/a_1})(t) \right)^{a_4} + a_3 \\
 &= y(t)
 \end{aligned}
 \tag{5.8}$$

Ausgehend von der intuitiven Modellierung der Veränderung \dot{x} der körperlichen Beanspruchungsreaktion x in Abhängigkeit einer Belastung u lässt sich die Herzfrequenz $f(x(t))$ entsprechend als eine Skalierung dieser Beanspruchung deuten. Die phänomenologische Interpretation der Skalierungsparameter ändert sich durch die Darstellung als DGL nicht.

Basierend auf dieser Form des Faltungsmodells kann das Modell nun im Folgenden mit den Differentialgleichungsmodellen aus der Literatur verglichen werden. Wenngleich die zugrunde liegende Struktur des Faltungsmodells und der DGL-Modelle aus der Literatur weiterhin unterschiedlich ist, lassen sich einige Gemeinsamkeiten und Unterschiede finden.

Für einen besseren Überblick wird die Struktur der betrachteten Differentialgleichungen aus der Literatur nachfolgend zusammenfassend dargestellt. Die drei unterschiedlichen in Unterabschnitt 2.2.2 betrachteten Differentialgleichungen genügen alle der Form:

$$\begin{aligned}
 \dot{x}_1(t) &= -a_1 \cdot x_1(t) + a_2 \cdot x_2(t) + h(u(t)) \\
 \dot{x}_2(t) &= -a_4 \cdot x_2(t) + g(x_1(t), x_2(t)) \\
 y(t) &= x_1(t)
 \end{aligned}
 \tag{5.9}$$

Tabelle 5.3 fasst für einen besseren Überblick nochmal die Spezifikationen der Unterfunktionen h und g für die jeweiligen Modelle zusammen.

Tabelle 5.3.: Überblick über die Spezifikationen der Unterfunktionen h und g für die betrachteten DGL-Modelle aus der Literatur

DGL-Modell	$h(u(t))$	$g(x_1(t), x_2(t))$
Cheng et al. (2007b)	$= \frac{u^2(t)}{1 + \exp(-u(t) + a_3)}$	$= a_4 \cdot \tanh(x_2(t)) + a_5 \cdot x_1(t)$
Cheng et al. (2008)	$= a_2 \cdot u^2(t)$	$= \frac{a_4 \cdot x_1(t)}{1 + \exp(-(x_1(t) - a_5))}$
Paradiso et al. (2013)	$= a_6 \cdot u^2(t)$	$= a_4 \cdot \tanh(x_2(t)) + a_5 \cdot x_1(t)$

Sowohl das Faltungsmodell als DGL-Modell als auch die Literaturmodelle haben gemeinsam, dass die Herzfrequenz $y(t)$ in einem zweistufigen Verfahren berechnet wird: In einem ersten Schritt wird eine Differentialgleichung bzw. ein Differentialgleichungssystem erster Ordnung bestimmt. Im zweiten Schritt wird anschließend $y(t)$ basierend auf einer der Variablen aus dem DGL berechnet.

Die beiden Modelltypen unterscheiden sich in drei wesentlichen Punkten:

1. Die Modelle aus der Literatur sind Differentialgleichungssysteme mit zwei Variablen x_1 und x_2 , während das DGL-Faltungsmodell mit einer einzigen Variablen auskommt. Ein solches DGL-System mit zwei Variablen entspricht letztlich einer DGL zweiter Ordnung, sodass in den Literaturmodellen durch die höhere Komplexität der Modelle auch Veränderungen höherer Ordnung modelliert werden.
2. In den Differentialgleichungssystemen, wie sie in Gleichung 5.9 abstrahiert dargestellt sind, wird die Herzfrequenz $y(t)$ direkt mit einer der Variablen gleichgesetzt ($y(t) = x_1(t)$). Im DGL-Faltungsmodell wird die Variable $x(t)$ hingegen als allgemeine Beanspruchungsreaktion des Körpers interpretiert und unter Verwendung der Parameter in die Modellstruktur des ursprünglichen Faltungsmodells eingesetzt, um die Herzfrequenz als ein konkretes Maß für die Beanspruchung zu berechnen: $y(t) = f(x(t)) = a_2 x(t)^{a_4} + a_3$.
3. Die Belastung $u(t)$ wird in den DGL-Systemen aus der Literatur innerhalb einer Funktion auf nicht-lineare Weise transformiert (in der vereinfachten Variante von Cheng et al. (2008) geschieht dies über die Funktion $h(u(t)) = a_2 \cdot u^2(t)$, vgl. Gleichung 2.5 auf S. 97). Im Faltungsmodell fließt sie unverändert ein.

Da eine explizite Lösung der DGL-Modell aus der Literatur dort nicht angegeben wird und sich nicht trivial ermitteln lässt, wird auf einen entsprechenden Vergleich verzichtet.

Während Systeme, die mit linearen DGLen mit konstanten Koeffizienten beschrieben werden können, linear und zeitinvariant sind, verhalten sich diese Eigenschaften bei den betrachteten DGL-Modell aus der Literatur etwas anders. Tabelle 5.4 stellt in einer Übersicht dar, welche der Modellvariationen welche der Eigenschaften Linearität und Zeitinvarianz erfüllen, und ob sie sich einem Hammerstein- oder einem Wienermodell zuordnen lassen.

Tabelle 5.4.: Übersicht erfüllter Eigenschaften der betrachteten DGL-Modelle im Vergleich zum Faltungsmodell

Modell/Eigenschaft	Hammerstein	Wiener	linear	zeitinvariant
DGL: Cheng et al. (2007b)	✓	–	–	✓
DGL: Cheng et al. (2008)	✓	–	–	✓
DGL: Paradiso et al. (2013)	✓	–	–	✓
LTI: Dur-e Zehra Baig et al. (2010)	–	–	✓	✓
Faltungsmodell	–	✓	–	✓

Strukturell beinhalten sowohl Hammerstein- als auch Wiener-Modelle eine Nichtlinearität. In den betrachteten Modellen aus der Literatur tritt die Nichtlinearität (mit Ausnahme des LTI-Modells) bei der Transformierung des Eingangssignals u innerhalb von $h(u)$ auf, sodass nur das LTI-Modell die Eigenschaft der Linearität erfüllt. Die innere DGL des Faltungsmodells verhält sich linear – allerdings tritt hier die Nichtlinearität durch die anschließende Transformation mit dem Exponentialfilter und dem additiven Parameter auf. Eine Zeitab-

hängigkeit des Eingangssignals existiert in keinem der betrachteten Modelle – sie erfüllen alle die Eigenschaft der Zeitinvarianz. Während das Faltungsmodell durch seine Struktur die Eigenschaften eines Wiener-Modells erfüllt (vgl. nachfolgend Unterabschnitt 5.3.2: »Das Faltungsmodell als Wiener-Modell«), sind die drei DGL-Modelle von Cheng et al. (2007b), Cheng et al. (2008) und Paradiso et al. (2013) Hammerstein-Modelle: Die Nichtlinearität tritt innerhalb von \dot{x}_1 und \dot{x}_2 auf, wohingegen das Ausgangssignal mit $y = x_1$ linear im Eingangssignal u ist. Das LTI-Modell von Dur-e Zehra Baig et al. (2010) ist weder ein Hammerstein- noch ein Wienermodell, da es lediglich der linearen Systemkomponente ohne nichtlineare Transformation besteht.

5.3.2. Das Faltungsmodell als Wiener-Modell

Im vorangegangenen Abschnitt wurde das Faltungsmodell bereits in der allgemeinen Struktur eines DGL-Modells betrachtet. Insbesondere hat sich gezeigt, dass das Faltungsmodell äquivalent ist zu einer nichtlinearen Transformation der Lösung einer simplen DGL; die Herzfrequenz kann also wie in Gleichung 5.10 beschrieben werden als:

$$\begin{aligned} y(t) &= f(x(t)) \\ &= a_2 x(t)^{a_4} + a_3 \end{aligned} \tag{5.10}$$

mit $x(t) = \frac{1}{a_1} (u * e^{-\bullet/a_1})(t)$.

Dabei wird ein eingehendes Signal wie die Belastung u zunächst über die Differentialgleichung linear transformiert und anschließend (innerhalb von $x(t)$) einer nichtlinearen Skalierung unterzogen. Es handelt sich folglich strukturell bereits um ein Wiener-Modell.

Im Hammerstein-Wiener-Modell von Gonzalez et al. (2016) wird das Eingangssignal nichtlinear verarbeitet, woraufhin ein linearer Block und eine weitere nichtlineare Komponente für das Ausgangssignal folgen. Die Autoren haben ihr Modell mit Hilfe der »System Identification Toolbox« aus MATLAB automatisiert ermitteln lassen – eine explizite Formel wird nicht angegeben. Insofern lässt sich für einen Modellvergleich mit dem Faltungsmodell lediglich feststellen, dass das Modell von Gonzalez et al. (2016) im Vergleich zum Faltungsmodell eine nichtlineare Komponente am Eingangssignal aufweist.

5.4. Zusammenfassung und Fazit

Ziel dieses Kapitels war es, die erste Fragestellung zu untersuchen:

Fragestellung 1:

Lassen sich mit dem Faltungsmodell bessere Vorhersagen der Herzfrequenz während körperlicher Belastung im Ausdauersport für ganze Trainingssequenzen ohne Rückkopplung gemessener Herzfrequenzwerte treffen, als mit bisherigen Modellen der Herzfrequenzanalyse?

Das Faltungsmodell wurde selbständig dazu entwickelt, basierend auf einer gegebenen Belastung die Herzfrequenz (oder andere Beanspruchungsparameter) für komplette Trainingseinheiten geeignet zu simulieren und vorherzusagen.

Zur Analyse der Fragestellung wurde das Faltungsmodell detailliert präsentiert, phänomenologisch motiviert und seine Güte anhand unterschiedlicher Experimente untersucht.

Wie in »Analyse des Faltungsmodells: Modellvergleich« (Abschnitt 5.2) basierend auf den Studien aus Ludwig et al. (2015) und Ludwig et al. (2016) dargestellt, erzielt das Faltungsmodell im Vergleich zu Modellen, die inhärent zur Vorhersage sehr kurzer Zeithorizonte von wenigen Sekunden konzipiert wurden, eine tendenziell bessere Vorhersagegenauigkeit. Im Vergleich mit einem Modellen zur Vorhersage vollständiger Trainingssequenzen sind die Ergebnisse des Faltungsmodells vollumfänglich vergleichbar. Dabei können zur Simulation und Vorhersage der Beanspruchung sowohl die Herzfrequenz als auch Parameter der Sauerstoffkinetik verwendet werden.

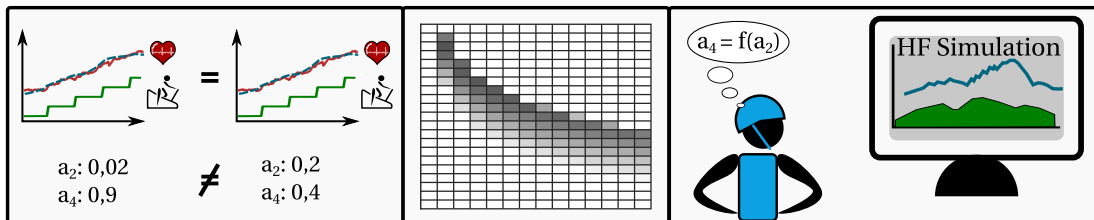
In »Mathematische Betrachtung und Vergleich zu ähnlichen Modelltypen« (Abschnitt 5.3) konnte gezeigt werden, dass der innere Teil des Faltungsmodells die Lösung einer einfachen Differentialgleichung darstellt, die auf möglichst intuitive Weise die Veränderung einer körperlichen Beanspruchungsreaktion in Abhängigkeit einer Belastung im zeitlichen Verlauf beschreibt. Das Faltungsmodell ist dabei äquivalent zu einer nichtlinearen Transformation der Lösung dieses Differentialgleichungsmodells. Insbesondere entspricht das Faltungsmodell durch die sich ergebende Struktur einer Nichtlinearität angewandt auf eine lineare Transformation einem Wiener-Modell.

Auffällig ist, dass sich sowohl das Wiener-Modell von Gonzalez et al. (2016) als auch das Faltungsmodell – was ebenfalls ein Wiener-Modell ist – gut zur Simulation und Vorhersage vollständiger Trainingseinheiten eignen. Im Gegensatz dazu sind die betrachteten DGL-Modelle aus der Literatur allesamt Hammerstein-Modelle, die besser für den Einsatz einer Vorhersage weniger Sekunden geeignet sind.

Die durchgeführten Experimente haben damit gezeigt, dass Fragestellung 1 positiv beantwortet werden kann: Mit dem Faltungsmodell lassen sich bessere oder vergleichbar gute Vorhersagen der Herzfrequenz während körperlicher Belastung im Ausdauersport für ganze Trainingssequenzen ohne Rückkopplung gemessener Herzfrequenzwerte treffen als mit bisherigen Modellen der Herzfrequenzanalyse. Dabei hat das Faltungsmodell eine phänomenologisch interpretierbare Struktur und kommt mit lediglich vier freien Parametern aus.

Im nachfolgenden Kapitel wird das Faltungsmodell gezielt dahingehend untersucht, ob die Anzahl der freien Parameter weiter reduziert werden kann, ohne einen signifikanten Genauigkeitsverlust vor allem in der Vorhersage zu erleiden.

Verbesserung des Faltungmodells: Parameterreduktion



Abstract.

Das Faltungsmodell eignet sich wie bereits gezeigt gut zur Simulation und Vorhersage von unterschiedlichen Beanspruchungsparametern. Mit vier freien Parametern ist jedoch die Komplexität für eine spätere Evaluation eines Zusammenhangs zur Leistungsentwicklung noch hoch. Insbesondere ist bei vier oder mehr freien Parametern die Wahrscheinlichkeit hoch, dass einzelne Parameter sich gegenseitig beeinflussen und bereits intern voneinander abhängen. Aus diesem Grund wird eine mehrschrittige Vorgehensweise zur Reduktion der Parameteranzahl (kurz: Parameterreduktion) vorgestellt.

Die Auswertungen anhand zweier unterschiedlicher Studien zeigen, dass die parameterreduzierten Varianten mit lediglich einem freien Parameter in der Vorhersage der Herzfrequenz in ergometriebasierten Trainingseinheiten im Vergleich zum vierparametrischen Modell keinen signifikanten Genauigkeitsverlust aufweisen. Auch auf Basis von regulären Trainingseinheiten (»Outdoor-Trainings«) kann eine Vorhersage mit dem parameterreduzierten Faltungsmodell mit plausiblen Resultaten durchgeführt werden. Fragestellung 2 lässt sich mit den in diesem Kapitel erzielten Resultaten bestätigen.

Der Idee von *Ockhams Rasiermesser* (bzw. dem »Sparsamkeitsprinzip« oder auch »Prinzip der Parsimonie«) in der Modellierung folgend (vgl. z.B. MacKay (2003), Kapitel 28), sollte mit dem Faltungsmodell ein möglichst einfaches Modell entwickelt werden, um die individuellen physiologischen Beanspruchungsreaktionen einer Person auf Belastung geeignet abzubilden. Insbesondere bei wenigen Datensätzen führt eine hohe Anzahl von Parametern schnell zu einer Überoptimierung an die zur Kalibrierung genutzten Daten und damit zu ungenauen Vorhersage-Ergebnissen. Daher sollte die Anzahl der Parameter so gering wie möglich sein, ohne die Genauigkeit signifikant zu verschlechtern.

Gleichzeitig kann ein Modell mit nur einem freien Parameter genutzt werden, um die Existenz eines direkten Zusammenhangs zwischen diesem Parameters und der Leistungsentwicklung einer Athletin oder eines Athleten zu untersuchen. In diesem Kapitel wird daher die zweite Fragestellung untersucht:

Fragestellung 2:

Kann die Anzahl der Parameter des Faltungsmodells ohne signifikanten Verlust der Vorhersagegenauigkeit reduziert werden?

Grundsätzlich stehen mehrere methodische Ansätze zur Verfügung, um die Anzahl der freien Parameter zu reduzieren. Ein »klassisches« Verfahren aus dem Bereich der multivariaten Statistik ist die Durchführung einer Hauptkomponentenanalyse. Dabei werden die Variablen der Daten (z.B. die Modellparameter) durch Linearkombinationen von weniger Variablen angenähert. Eine andere Möglichkeit stellt die Reduktion durch eine manuelle Parameteranalyse dar. Diese lässt sich durch geeignete Visualisierungen unterstützen und vereinfachen. Das Vorgehen zur Reduktion der Parameter über eine manuelle Vorgehensweise sowie mithilfe einer geeigneten Visualisierung werden nachfolgend beschrieben, die Hauptkomponentenanalyse wird ergänzend erwähnt. Für jede dieser Methoden – d.h. die manuelle Reduktion, die visuelle Unterstützung bei der Parameterreduktion und auch die Hauptkomponentenanalyse – ist zuvor eine Parameteroptimierung des vierparametrischen Faltungsmodells notwendig. Auf diese Weise kann ein geeigneter Wertebereich der Parameter bereits im Vorfeld eingegrenzt werden.

Für die Parametervisualisierung in Form von Heatmaps wird dabei auf Grund der geringen Datenmengen im Vorfeld eine Rastersuche (engl. »*grid search*«) mit den vier Parametern durchgeführt: Für jeden der vier Parameter wird für die betrachtete Probandengruppe ein geeigneter Bereich basierend auf den kalibrierten Parametern definiert. Jeder Parameter wird anschließend manuell auf das Intervall $[0,1]$ normiert. Durch die Integration von Vorwissen über die Parameter (anhand der durchgeführten Kalibrierung des vierparametrischen Modells und Kenntnissen über die Datenstruktur) lassen sich die Abstände zwischen den Rasterpunkten unterschiedlich fein wählen. Für jede Kombination aus Rasterpunkten für

jeden der vier Parameter wird anschließend eine Modellsimulation mit Fehlerberechnung durchgeführt.

Exkurs: Beispiel einer Rastersuche

Angenommen, bei einer Kalibrierung des Faltungsmodells über mehrere Datensätze unterschiedlicher Proband*innen haben sich die in Tabelle 6.1 dargestellten Bereiche für die einzelnen Parameter ergeben:

Tabelle 6.1.: Beispielparameter über mehrere Datensätze unterschiedlicher Proband*innen. Für jeden Parameter sind Minimum, Maximum und der Median aufgeführt

Parameter	Minimum	Maximum	Median
a_1	0,5849	4,1683	2,2382
a_2	0,0000084	0,0056	0,00034
a_3	60,6755	107,5408	85,3274
a_4	0,8361	1,2064	1,0558

Für diese Parameter können beispielsweise die in Tabelle 6.2 dargestellten Bereiche über die Rastersuche näher betrachtet und zur Normierung linear skaliert werden.

Tabelle 6.2.: Bereiche und Normalisierung der Parameter, wobei die Werte der Rastersuche in \hat{a}_i eingesetzt werden und a_i wie gewohnt die Modellparameter sind

Parameter	Bereich	Normalisierung	Umkehrfunktion
a_1	0,5 – 4,5	$\hat{a}_1 = (a_1 - 0,5) \cdot 0,25$	$a_1 = (\hat{a}_1 \cdot 4) + 0,5$
a_2	0 – 0,04	$\hat{a}_2 = (a_2 - 0) \cdot 25$	$a_2 = (\hat{a}_2 \cdot 0,04) + 0$
a_3	60 – 110	$\hat{a}_3 = (a_3 - 60) \cdot 0,02$	$a_3 = (\hat{a}_3 \cdot 50) + 60$
a_4	0,75 – 1,25	$\hat{a}_4 = (a_4 - 0,75) \cdot 2$	$a_4 = (\hat{a}_4 \cdot 0,5) + 0,75$

Je nach gewünschter Auflösung des Rasters lässt sich anschließend für jeden Parameter die Feingranularität für die Suche festlegen. Durch die Integration von Vorwissen können die einzelnen Dimensionen der Parameter unterschiedlich groß gewählt werden. Wird für Parameter a_1 beispielsweise der Bereich von Null bis Eins in 0,05-er Schritten betrachtet, entspricht das – bezogen auf den ursprünglichen Wertebereich ohne Normalisierung – den 21 Werten: 0,5; 0,7; 0,9; 1,1; ...; 4,3; 4,5, bei einem Suchbereich in 0,025-er Schritten ergeben sich entsprechend 42 Werte: 0,5; 0,6; ... 4,4; 4,5.

Sind die Normalisierung und die Auflösung für jeden Parameter festgelegt, wird die Modellsimulation mit Berechnung der entsprechenden Fehlerwerte für jede mögliche Parameterkombination durchgeführt.

Eine Reduktion der Modellparameter bietet mehrere Vorteile, von denen zwei zur Beurteilung der aufgestellten Fragestellungen von besonderem Interesse sind: Wie bereits erläu-

tert können zwischen den Parametern Modell-interne Abhängigkeiten und Korrelationen bestehen, die die Stabilität der Parameter möglicherweise negativ beeinflussen. Die Reduktion der Parameter kann somit zur Modellstabilität beitragen. Außerdem lässt sich der verbleibende Parameter eines einparametrischen Modells hinsichtlich einer Korrelation zur empirischen Leistungsfähigkeit einer Person analysieren. Auch die Frage, ob basierend auf Herzfrequenz-Simulationen aus dem Training Rückschlüsse auf die Leistungsfähigkeit einer Person möglich sind, lässt sich mit einem stabilen Modell untersuchen.

6.1. Parameterreduziertes Faltungsmodell

Während das Faltungsmodell mit vier Parametern (vgl. Kapitel 5, »Das Faltungsmodell: Ein neuer Ansatz zur Simulation und Vorhersage der Herzfrequenz im Training«) bereits gute Ergebnisse erzielt, stellt sich die Frage, ob es ohne signifikanten Genauigkeitsverlust möglich ist, die Anzahl der Parameter für die Herzfrequenzmodellierung weiter einzuschränken und interne Abhängigkeiten zwischen den Parametern zu reduzieren. Interne Abhängigkeiten haben stets das Problem, dass die Kalibrierung der Parameter schon bei kleinsten Änderungen der Daten (d.h. beispielsweise auch durch Rauschen in den Daten) zu unterschiedlichen Parametern führen kann. Insbesondere kann es sogar dazu kommen, dass exakt die gleichen Daten bei mehreren Kalibrierungen zu unterschiedlichen Modellparametern führen. Die Parameter im Modell sind also nicht stabil. Eine Modellierung mit nur einem einzigen freien Parameter bietet den Vorteil, dass keine internen Abhängigkeiten zwischen mehreren Parametern bestehen können.

Nachfolgend werden zwei unterschiedliche Vorgehensweisen vorgestellt: eine manuelle Reduktion der Parameter und eine visuell unterstützte Parameteranalyse. Ergänzend wird die Möglichkeit einer Hauptkomponentenanalyse zur Parameterreduktion erwähnt – diese findet jedoch im Rahmen dieser Arbeit keine Anwendung. Auf Basis der unterschiedlichen Methoden wird schließlich ein iteratives Vorgehen aus einer Kombination der visuell unterstützten Parameteranalyse und einer manuellen Reduktion vorgestellt. Diese Vorgehensweise wird im Anschluss in zwei empirischen Studien zur Parameterreduktion verwendet.

Manuelle Reduktion der Parameter

Eine ausschließlich manuelle Reduktion der Parameter kann insbesondere als »proof-of-concept« dienen, um zu zeigen, dass eine Parameterreduktion für das Faltungsmodell »geeignet« möglich ist. Geeignet bedeutet in dem Fall, dass über den Welch-Test keine signifikanten Unterschiede zwischen den betrachteten Modellvarianten zum Niveau $\alpha = 5\%$ bei der Vorhersage festgestellt werden können, oder dass die Fehlerwerte weder bei der Modellkalibrierung noch bei einer späteren Vorhersage einen spezifischen prozentualen Fehler deutlich überschreiten. Als Grenzwert dieses Fehlers wurde ein Fehlerwert von 15% gewählt;

dieser entspricht dem zweifachen mittleren Fehler aus typischen Simulationsergebnissen einer Vorhersage mit dem Faltungsmodell.

Im Folgenden wird die phänomenologische Interpretation der Parameter in Hinblick auf eine mögliche (manuelle) Fixierung einiger Parameter betrachtet. Insbesondere wird anhand eines Teilerperiments und der Probandendaten aus Abschnitt 5.2 (d.h. »Datensatz 1: Daten aus einer Studie von Gonzalez et al«, wie er in Abschnitt 4.1 beschrieben wird, eingeschränkt auf die Daten zur Herzfrequenz) untersucht, wie sich die Parameter des Faltungsmodells grundsätzlich und abgestimmt auf eine spezifische Probandengruppe (mit vergleichbarem Leistungsniveau) reduzieren lassen. Eine allgemeine Reduktion unabhängig von der Probandengruppe und quer durch Leistungs- und Hobbysportler hinweg erscheint nicht zielführend, wie die phänomenologische Betrachtung verdeutlichen wird.

Wie in Abschnitt 5.1 beschrieben, entspricht der additive Levelparameter a_3 phänomenologisch der Höhe der Herzfrequenz vor der Belastung. Es liegt daher nahe, diesen Parameter individuell pro Proband zu fixieren. Dafür kann je nach konkreter Simulation der erste gemessene Herzfrequenzwert vor Beginn der Belastung des aktuellen oder vorhergehenden Trainingseinheiten genutzt werden.

Der exponentielle Parameter a_4 , der insbesondere Belastungsspitzen modelliert, schien in zahlreichen Experimenten nur einen geringen Einfluss zu haben und hat in festen Probandengruppen mit ähnlichem Leistungsniveau häufig ähnliche Werte (unabhängig von den restlichen Parametern) angenommen. Auch hier liegt ein Festsetzen dieses Parameters nahe. Dabei ist jedoch zu beachten, dass das Ausmaß von Belastungsspitzen mit zunehmender Fitness abnehmen kann, da die Leistungsreserven während des Ausdauertrainings zunehmen.

Der multiplikative Parameter a_2 hängt besonders stark von der betrachteten physiologischen Reaktion ab: Im Modell sorgt er für die Höhe der Ausschläge in der Simulationskurve. In den Analysen der Studie aus Abschnitt 5.2 hat sich beispielsweise gezeigt, dass dieser Parameter bei der Simulation der Sauerstoffkinetik in völlig anderen Bereichen liegt, als bei der Simulation der Herzfrequenz. Auch hier ist zu berücksichtigen, dass sich der Organismus bei Verbesserung der Fitness einer Person besser an auftretende Belastungen adaptieren und beispielsweise Sauerstoff effektiver nutzen kann. Dadurch nimmt das Ausmaß der körperlichen Reaktion auf die Belastung in der Regel ab, sodass auch der multiplikative Parameter möglicherweise geringere Werte annimmt.

Der Parameter des Exponentialfilters, a_1 , ist phänomenologisch schwieriger zu interpretieren. Mathematisch beeinflusst dieser Parameter, wie stark weiter zurückliegende Werte aus der Faltung berücksichtigt werden. Mit dem Parameter des Exponentialfilters a_1 lässt sich steuern, wie stark der exponentiell abfallende Einfluss vergangener (berechneter) Beanspruchungswerte zur Bestimmung der aktuellen Beanspruchung noch ist. Je mehr Werte aus der

Faltung einen hohen Einfluss auf das Ergebnis haben, desto mehr Daten aus der Vergangenheit werden für die Berechnung insgesamt verwendet.

Als Datenbasis wurden beispielhaft die Daten aus Vorhersage-Experiment A.3 aus den Experimenten aus der »Analyse des Faltungsmodells: Modellvergleich« (Abschnitt 5.2) verwendet. Zur Vermeidung von Abhängigkeiten der einzelnen Modellparameter zur generellen »Höhe« der simulierten Herzfrequenz wurde zunächst der Levelparameter a_3 fixiert – auf den minimalen Wert, der vor Beginn des Trainings gemessen wurde. Frühere Messwerte oder eine tatsächliche Ruheherzfrequenz liegen für den Datensatz nicht vor. In einem zweiten Schritt wurde die Simulation für alle Probanden mit dem jeweiligen fixierten Wert für a_3 durchgeführt und der Exponentialparameter a_4 über die Probanden hinweg betrachtet. Innerhalb dieser Studiendaten variierte a_4 geringfügig um den Wert 0,6 für die Simulation der Herzfrequenz, unabhängig von den restlichen Parametern. Entsprechend wurde der Exponentialparameter a_4 auf diesen Wert fixiert. Das parameterreduzierte Faltungsmodell hat schließlich noch zwei freie Parameter, den Exponentialfilter a_1 und den multiplikativen Faktor a_2 .

Entsprechend der Vorgehensweise, wie sie auch in Ludwig et al. (2016) präsentiert wurde, lässt sich zusätzlich auch ein – für die Probandengruppe individualisiertes – Einparameter-Modell finden. Hierzu werden a_3 und a_4 wie soeben beschrieben fixiert. Für die verbleibenden Parameter a_1 und a_2 werden darauf aufbauend unterschiedliche – möglichst einfache – Variationen untersucht:

1. Einer der beiden Parameter wird ebenfalls auf einen festen Wert fixiert;
2. Ein möglicher linearer Zusammenhang der beiden Parameter wird analysiert;
3. Ein möglicher polynomieller Zusammenhang der beiden Parameter wird analysiert.

Die Analyse eines möglichen linearen Zusammenhangs kann prinzipiell durch eine Hauptkomponentenanalyse [engl. *Principal Component Analysis*] (PCA) unterstützt werden; Bei der Analyse eines möglichen polynomiellen Zusammenhangs kann eine visuelle Parameteranalyse ergänzend von Nutzen sein. Beide Varianten werden in den nachfolgenden Unterabschnitten erläutert, wurden innerhalb der manuellen Parameterreduktion jedoch nicht verwendet. Die rein manuell durchgeführte Parameterreduktion dient hierbei als »proof-of-concept«, um die grundsätzliche Möglichkeit der Parameterreduktion für das Faltungsmodell zu zeigen.

Bei den betrachteten Daten werden die stabilsten Resultate mit der Fixierung des Exponentialfilter-Parameters a_1 erzielt. Konkret wird $a_1 = 100$ für die Simulation der Herzfrequenz und $a_1 = 60$ für die Simulation von $\dot{V}O_2$ und $\dot{V}CO_2$ gesetzt. Der multiplikative Parameter a_2 verbleibt als freier Parameter.

Mit diesen Parametereinstellungen ergibt sich im Fitting über alle Tests und Probanden im Mittel ein RMSE von 6,20 bpm und im Mittel ein MAPE von 3,82% für die Herzfrequenz.

Tabelle 6.3 zeigt für die Vorhersage der drei anderen Testprotokolle die mittleren Ergebnisse des RMSE und des MAPE. Dabei werden die einzelnen Schritte der Parameterreduktion gesondert aufgeführt.

Tabelle 6.3.: Durchschnittliche Vorhersageergebnisse für drei physiologische Beanspruchungsmaße. Das Modell wurde exemplarisch auf Test 03 gefittet (entspricht Vorhersageexperiment A.3) und die verbleibenden Testprotokolle wurden für jeden Probanden vorhersagt. Vor dem Fitting fand eine schrittweise Reduktion der Parameter statt. Die Ergebnisse aus A.3 des vierparametrischen Faltungsmodells sind zur besseren Vergleichbarkeit in der ersten Zeile angegeben

Fixierte Parameter	RMSE	MAPE
–	8,05	4,65
a_3	8,24	4,68
a_3, a_4	8,04	4,43
a_1, a_3, a_4	7,73	4,26

In diesem experimentellen Setting sind die Ergebnisse für das Faltungsmodell mit einem und zwei Parametern vergleichbar. Insbesondere weist das einparametrische Faltungsmodell geringere Fehlerwerte auf als das originale vierparametrische Faltungsmodell im zugehörigen Vorhersageexperiment A.3. Auch dann, wenn das einparametrische Faltungsmodell innerhalb einer Leave-One-Out-Kreuzvalidierung verwendet wird, bleiben die Ergebnisse zum vierparametrischen Modell ohne signifikante Unterschiede (auf Basis des Welch-Tests zum Signifikanzniveau $\alpha = 0,05$) und damit vergleichbar.

Für die manuelle Erstellung eines parameterreduzierten Faltungsmodells ergeben sich also die folgenden aufeinander aufbauenden Schritte, sofern keine weiteren Analysen ergänzend durchgeführt wurden die die Reduktion allein auf Basis manueller Entscheidungen getroffen werden soll:

1. Fixierung des Levelparameters a_3 auf den kleinsten gemessenen Wert, möglichst vor Trainingsbeginn;
2. Unter Berücksichtigung des fixierten Parameters a_3 : Analyse des Exponentialparameters a_4 über die Probanden hinweg und Fixierung auf einen geeigneten Wert, der zu keinem deutlichen Genauigkeitsverlust führt;
3. Unter Berücksichtigung der so fixierten Parameter a_3 und a_4 : Analyse, ob sich analog zu Schritt 2 ein weiterer Parameter (a_1 oder a_2) festsetzen lässt; andernfalls Analyse, ob ein linearer oder polynomieller Zusammenhang zwischen den beiden Parametern (möglichst über die gesamte Probandengruppe) gefunden werden kann¹.

Es empfiehlt sich allerdings, die manuelle Parameterreduktion mindestens durch eine geeignete visuelle Darstellung der Parameter zu unterstützen, wie sie im nachfolgenden Abschnitt erläutert wird.

¹Eine solche Analyse wurde im Rahmen des vorangegangenen Beispiels nicht durchgeführt

Visuell unterstützte Parameteranalyse

Eine geeignete Möglichkeit, die Parameterreduktion visuell zu unterstützen, ist die Verwendung von Heatmaps (vgl. »Visualisierung« in Unterabschnitt 2.1.2). Zur Veranschaulichung werden als Datenbasis die gleichen Daten wie für die manuelle Reduktion verwendet (d.h. die Herzfrequenz-Daten aus den Experimenten aus der »Analyse des Faltungsmodells: Modellvergleich« (Abschnitt 5.2)).

Entgegen der obigen Analyse werden im Folgenden alle Trainingseinheiten der Probanden genutzt anstatt die Ergebnisse nur auf Basis des Fittings auf Datensatz 03 zu analysieren. Auf diese Weise können exemplarisch geeignete Mittelungen der Daten erläutert werden. Durch die Betrachtung aller Datensätze ergeben sich allerdings Veränderungen in der Wahl der Parameterbereiche, sodass die Analysen nicht direkt zusammenhängen.

Vor der Analyse der Parameter werden die Definitionsbereiche jedes einzelnen Parameters zunächst linear auf das Intervall $[0,1]$ normiert. Anschließend wird eine Rastersuche angelehnt an die zuvor durch die Modellkalibrierung (mit der Methode der kleinsten Quadrate) ermittelte Wertebereiche durchgeführt. Für die hier beispielhaft betrachteten Daten ergibt sich die in Tabelle 6.4 dargestellte Normierung der Parameter.

Tabelle 6.4.: Bereiche und Normalisierung der Parameter, wobei die Werte der Rastersuche in \hat{a}_i eingesetzt werden und a_i wie gewohnt die Modellparameter sind

Parameter	Bereich	Normalisierung	Umkehrfunktion
a_1	10 – 25	$\hat{a}_1 = (a_1 - 10) \cdot \frac{1}{15}$	$a_1 = (\hat{a}_1 \cdot 15) + 10$
a_2	0,001 – 0,02	$\hat{a}_2 = (a_2 - 0,001) \cdot \frac{1}{0,019}$	$a_2 = (\hat{a}_2 \cdot 0,019) + 0,001$
a_3	65 – 95	$\hat{a}_3 = (a_3 - 65) \cdot \frac{1}{30}$	$a_3 = (\hat{a}_3 \cdot 30) + 65$
a_4	0,7 – 1,0	$\hat{a}_4 = (a_4 - 0,7) \cdot \frac{1}{0,3}$	$a_4 = (\hat{a}_4 \cdot 0,3) + 0,7$

Basierend auf den Ergebnissen der Fittingsimulationen mit den Parametern aus der Rastersuche wird für einen ersten Überblick eine Heatmap für jede Kombination von zwei Parametern erstellt. In Abbildung 6.1 sind exemplarisch zwei Beispiele einer solchen Heatmap dargestellt, wobei die Parameter in der Beschriftung zur besseren Interpretation zurücktransformiert werden. Für die Kombination aus je zwei Parametern werden die minimalen MAPE-Fehlerwerte aus der Rastersuche für jeden Datensatz eines jeden Probanden über die verbleibenden beiden Parameter betrachtet. Dabei wird der Median über die einzelnen Datensätze gebildet und als »median(min_MAPE)« für die Einfärbung der Rechtecke innerhalb der Heatmap gewählt.

Anhand der exemplarisch dargestellten Heatmaps in Abbildung 6.1 lassen sich mehrere Aspekte erkennen:

1. In Abbildungsteil (a) sind leichte Grauschattierungen erkennbar: Die MAPE-Fehlerwerte sind für $a_4 \in \{0,88; 0,91; 0,97; 1\}$ höher als für andere Werte von a_4 . Für

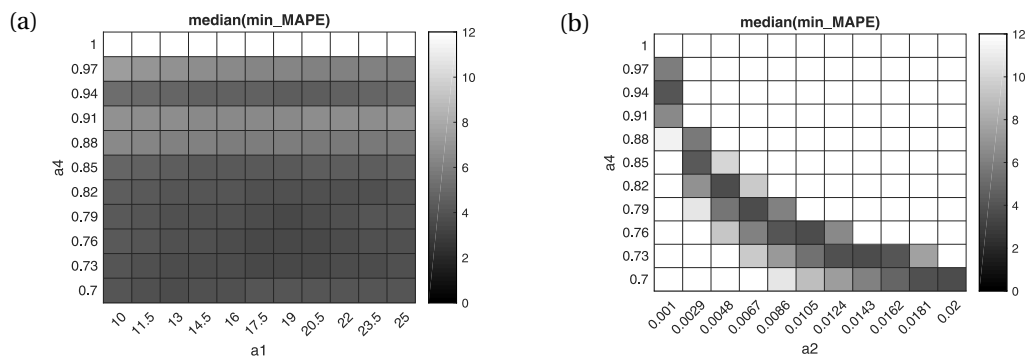


Abbildung 6.1.: Heatmap-Darstellung für exemplarisch zwei Kombinationen aus je zwei Parametern. Für jedes Wertepaar aus der betrachteten Parameterkombination wird der Median der minimalen MAPE-Werte für alle Trainingseinheiten aller Probanden über die Einfärbung betrachtet

alle anderen betrachteten Werte von a_4 existiert (mindestens) eine Parameterkombination von a_2, a_3 , mit der geringe Fehlerwerte ermittelt werden können.

2. Abbildungsteil (b) hat in diesem Beispiel die meiste Struktur. Die Parameter a_2 und a_4 scheinen funktionell (z.B. logarithmisch) zusammenzuhängen. Dieser funktionale Zusammenhang sollte für die Parameterreduktion näher untersucht werden.
3. Es scheint für keine der hier paarweisen Parameterkombinationen einen eindeutigen linearen Zusammenhang zu geben.

Statt den Median über alle Probanden zu betrachten, können für Detailanalysen analog Heatmaps für jeden Probanden einzeln mit Median über die Trainingseinheiten oder auch für jeden Probanden und jede Trainingseinheit einzeln (gänzlich ohne Medianbildung) dargestellt und analysiert werden. Innerhalb der hier betrachteten Daten sind die Heatmaps für die einzelnen Trainingseinheiten eines jeden Probanden in der Grundstruktur tendenziell ähnlich wie im oben betrachteten Median, sodass sich durch eine Detailbetrachtung in diesem Fall keine fundamentalen Unterschiede in der Interpretation ergeben.

Insbesondere verdeutlicht die visuelle Parameteranalyse auf Basis des vierparametrischen Faltungmodells, dass eine frühzeitige visuell unterstützte Analyse sinnvoll ist: Statt den Exponentialparameter a_4 wie in dem Beispiel zur manuellen Reduktion konstant zu wählen, bietet es sich an, bereits vor der Wahl konstanter Werte den über die Heatmap gefundenen funktionalen Zusammenhang detailliert zu untersuchen. Im Anschluss daran kann für ein geringfügig parameterreduziertes Modell (beispielsweise mit noch 3 freien Parametern) erneut eine Rastersuche durchgeführt und die visuelle Analyse wiederholt werden.

Heatmaps können die Parameteranalyse und darauf aufbauend die Parameterreduktion also in mehreren Schritten unterstützen:

1. Für einen ersten Überblick können jeweils zwei der Parameter auf die Achsen der Heatmap geplottet werden; die Rechtecke werden dabei anhand der über die verblei-

benden Parameter minimal erzielten Fehlerwerte (z.B. MAPE oder RMSE) eingefärbt. Je nach Beschaffenheit der Daten können die Heatmaps dabei beispielsweise für jeden Datensatz einzeln, gemittelt über die Datensätze der einzelnen Proband*innen oder auch gemittelt über alle Datensätze und alle Proband*innen betrachtet werden.

2. Je nach Struktur der Heatmaps und resultierender Interpretation können ggf. zwei Parameter funktional miteinander verknüpft oder einzelne Parameter auf einen konstanten Wert gesetzt werden.
3. Kann das Modell um mindestens einen freien Parameter reduziert werden, ohne signifikant an Genauigkeit zu verlieren, kann erneut eine Rastersuche durchgeführt und jeweils zwei der verbleibenden Parameter analog zu den Schritten 1 und 2 analysiert werden.
4. Die Schritte 2 und 3 können je nach Zielsetzung der Parameterreduktion und Struktur der Daten ggf. iterativ durchgeführt werden.

Ein Nachteil der hier betrachteten visuellen Unterstützung ist die Laufzeit der vorherigen Datenverarbeitung: Durch die Rastersuche ergeben sich schnell sehr große Mengen an Daten zur Fittingsimulation, deren Durchführung entsprechend viel Zeit in Anspruch nimmt. Wird beispielsweise für jeden der Parameter wie in obigem Beispiel ein eher grobes Raster aus lediglich 21 Werten betrachtet, ergeben sich für jede Trainingseinheit jedes Probanden $21^4 = 194\,481$ Fittingsimulationen. Bei fünf Probanden mit jeweils vier Trainingseinheiten resultiert die gesamte Rastersuche entsprechend in fast vier Millionen ($21^4 \cdot 5 \cdot 4 = 3\,889\,620$) Kombinationen aus Parametern und Fehlerwerten. Dieser Aspekt ist auch zu berücksichtigen, wenn die nachfolgend beschriebene Hauptkomponentenanalyse basierend auf den Ergebnissen einer Rastersuche durchgeführt werden soll.

Hauptkomponentenanalyse

Die PCA stellt ein statistisches Verfahren dar, mit dem Datensätze vereinfacht werden sollen. Mit Hilfe der PCA sollen Linearkombinationen gefunden werden, die die maximale Varianz aus den Daten beschreiben. Dabei werden über die PCA lineare Zusammenhänge der statistischen Variablen extrahiert. Durch die Betrachtung derjenigen Linearkombinationen, die die maximale Varianz beschreiben, kann die Menge der korrelierten beobachteten Variablen auf weniger unabhängige und zusammengesetzte Variablen reduziert werden. Mathematisch wird dabei eine *Hauptachsentransformation* durchgeführt. Für eine detaillierte Beschreibung der PCA wird auf Wolff und Bacher (2010) verwiesen.

Weder im oben betrachteten Beispiel noch in den nachfolgenden Studien hat es sich ergeben, dass eine PCA für die weitere Analyse zielführend gewesen wäre, sodass auf eine ausführliche Darstellung der Vorgehensweise verzichtet wird.

Zusammenfassung: Vorgehen zur Parameterreduktion im Faltungsmodell

Für das Faltungsmodell wird basierend auf den dargestellten Methoden folgendes iteratives Vorgehen zur Parameterreduktion gewählt:

1. Kalibrierung der Parameter in Form einer Fitting-Simulation der zu analysierenden Trainingsdaten über die Methode der kleinsten Quadrate – jeder Datensatz wird dabei einzeln betrachtet und kalibriert;
2. Extraktion geeigneter Parameter-Bereiche aus der zuvor durchgeführten Kalibrierung
3. Fitting-Simulation über die Rastersuche, wobei die zuvor identifizierten Parameter-Bereiche auf $[0, 1]$ normiert werden und einen Anhaltspunkt geben, wie fein das Raster gewählt wird;
4. Visuelle Parameteranalyse der Ergebnisse aus der Rastersuche über Heatmaps; dabei:
 - (a) Sofern die visuelle Parameteranalyse Hinweise auf einen linearen Zusammenhang ergibt: Durchführung einer Hauptkomponentenanalyse und ggf. Ersetzen der Parameter durch die reduzierte Anzahl der resultierenden Hauptkomponenten;
 - (b) Identifikation geeigneter funktionaler Zusammenhänge oder konstant wählbarer Parameter;
 - (c) Sukzessive Reduktion der Parameter (durch geeignete Kombination der Informationen aus manuell integrierbaren Annahmen über die Daten, visueller Parameteranalyse und ggf. PCA)
5. Überprüfung der Reduktion der Anzahl freier Parameter: Wie weit weichen Ergebnisse einer Fitting-Simulation mit dem parameterreduzierten Faltungsmodell von den Ergebnissen aus dem vierparametrischen Faltungsmodell ab?

Nachdem die Anzahl der Parameter erfolgreich reduziert werden konnte, kann die Möglichkeit einer weiteren Parameterreduktion iterativ komplett oder partiell nach diesem Verfahren wiederholt werden.

Ein Beispiel für eine »integrierbare Annahme über die Daten« ist eine Wahl von a_3 in Relation zum kleinsten oder ersten gemessenen Herzfrequenzwert.

6.2. Konstruktion eines einparametrischen Faltungsmodells: Studiendaten von Profi-Sportlern

Basierend auf Leistungstest-Daten von vier männlichen professionellen Radsportlern wird im Folgenden analysiert, ob es möglich ist, die Anzahl der Parameter im Faltungsmodell der-

art zu reduzieren, dass im Vergleich zum vierparametrischen Faltungsmodell kein signifikanter Genauigkeitsverlust entsteht. Die Daten wurden in »Datensatz 2: Daten aus einer Studie von Schrot mit Profi-Athleten« (Abschnitt 4.2) beschrieben.

Es wird sich zeigen, dass auf Basis der vorliegenden Daten eine plausible Parameterreduktion für das Faltungsmodell durchgeführt und ein einparametrisches Modell ohne signifikanten Genauigkeitsverlust konstruiert werden kann.

6.2.1. Durchführung der Parameterreduktion und Ergebnisse

Die Parameterreduktion für das Faltungsmodell wird nachfolgend schrittweise anhand der zuvor identifizierten und am Ende von Abschnitt 6.1 (»Parameterreduziertes Faltungsmodell«) zusammengefassten Vorgehensweise durchgeführt. Dabei werden zunächst ausschließlich Trainingsdaten aus Ergometerfahrten verwendet. Das parameterreduzierte Modell wird anschließend zusätzlich auf den vorhandenen Outdoor-Trainingseinheiten evaluiert. Auf diese Weise wird die Wahrscheinlichkeit einer Überanpassung reduziert und die Menge der Daten für die Rastersuche zu Gunsten der Laufzeit eingeschränkt.

Zur leichteren Lesbarkeit wird nachfolgend ein n -parametrisches Faltungsmodell mit n P-Faltungsmodell abgekürzt.

Schritte 1-2: Kalibrierung und Identifikation der Parameterbereiche

Auf Basis der Kalibrierung der einzelnen Datensätze haben sich die in Tabelle 6.5 dargestellten Bereiche für die Parameter und die entsprechend gewählte Normierung ergeben. Als Grenzen werden gerundete Werte der minimalen und maximalen Werte aus der Parameterkalibrierung betrachtet.

Tabelle 6.5.: Bereiche und Normalisierung der Parameter auf Basis einer vorangehenden Kalibrierung, wobei die Werte der Rastersuche in \hat{a}_i eingesetzt werden und a_i wie gewohnt die Modellparameter sind

Parameter	Bereich	Normalisierung	Umkehrfunktion
a_1	0,05 – 4,55	$\hat{a}_1 = (a_1 - 0,05) \cdot \frac{1}{4,5}$	$a_1 = (\hat{a}_1 \cdot 4,5) + 0,05$
a_2	0 – 1,1	$\hat{a}_2 = (a_2 - 0) \cdot \frac{1}{1,1}$	$a_2 = (\hat{a}_2 \cdot 1,1) + 0$
a_3	50 – 155	$\hat{a}_3 = (a_3 - 50) \cdot \frac{1}{105}$	$a_3 = (\hat{a}_3 \cdot 105) + 50$
a_4	0,35 – 1,6	$\hat{a}_4 = (a_4 - 0,35) \cdot \frac{4}{5}$	$a_4 = (\hat{a}_4 \cdot 1,25) + 0,35$

Die gewählte Normierung bildet die Basis für die Fitting-Simulationen über die Rastersuche. Dabei wird basierend auf den Ergebnissen einer vorherigen Kalibrierung mit dem 4P-Modell zunächst für Parameter a_1 eine Schrittweite von 0,5, für a_2, a_4 eine Schrittweite von 0,1 und für Parameter a_3 eine Schrittweite von etwa 5 gewählt – pro Proband und Trainingseinheit entstehen damit $10 \cdot 12 \cdot 22 \cdot 13 = 34\,320$ Fittingsimulationen. Für alle Da-

tensätze aller Probanden ergeben sich entsprechend rund 2,78 Millionen Kombinationen ($34\,320 \cdot (21 + 18 + 23 + 19) = 2\,779\,920$) aus Parametern und Fehlerwerten.

Schritte 3-5: Fitting-Simulation über die Rastersuche, visuelle Parameteranalyse (Heatmaps) und sukzessive Überprüfung

Basierend auf den Ergebnissen der Fittingsimulationen mit den Parametern aus der Rastersuche wird für einen ersten Überblick eine Heatmap mit Medianwerten über die einzelnen Datensätze erstellt. Der »median(min_MAPE)« ist für die Einfärbung der Rechtecke innerhalb der Heatmap gewählt. Während die Rastersuche wie zuvor beschrieben auf $[0,1]$ -normierten Parametern durchgeführt wurde, sind die Parameter in der Beschriftung zur besseren Interpretation bereits zurücktransformiert.

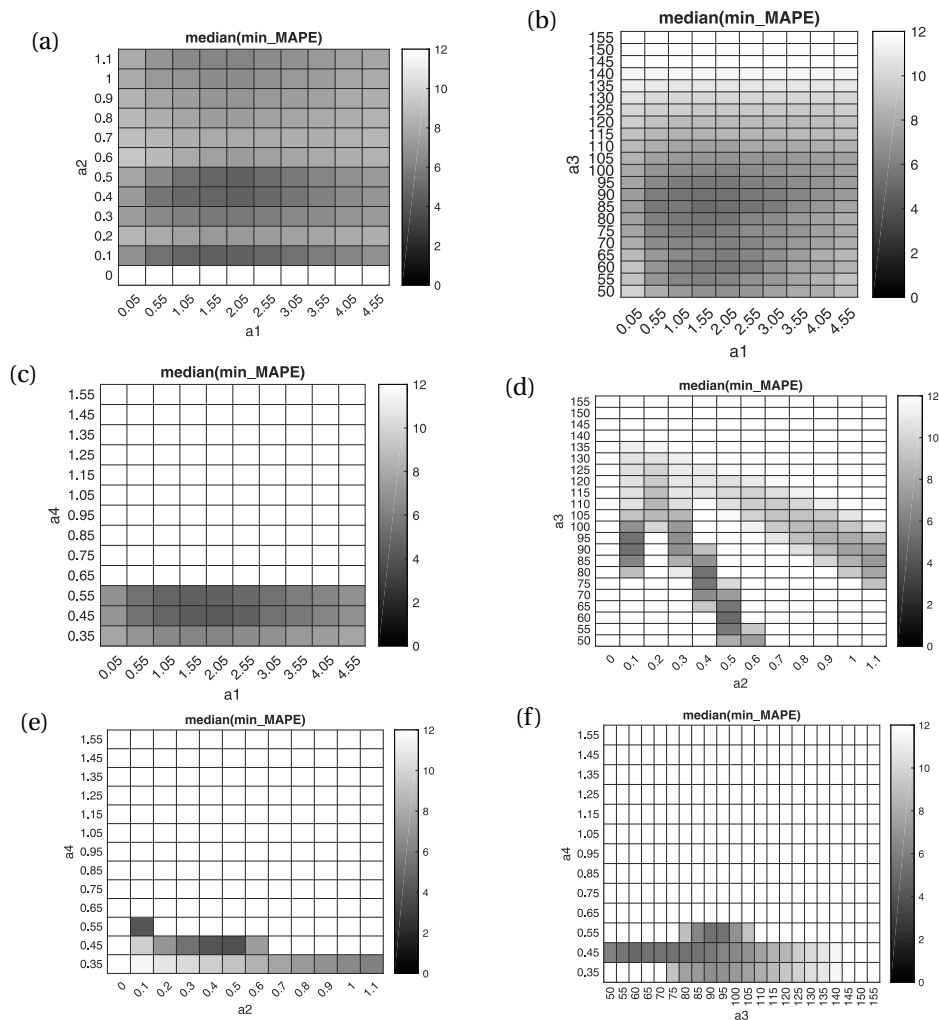


Abbildung 6.2.: Heatmap-Darstellung für die Kombination aus je zwei Parametern. Für jedes Wertepaar aus der betrachteten Parameterkombination wird der Median der minimalen MAPE-Werte für alle Trainingseinheiten aller Probanden über die Einfärbung betrachtet

Anhand der Heatmaps in Abbildung 6.2 lassen sich mehrere Aspekte erkennen:

1. Für den betrachteten Parameterbereich gibt es hinsichtlich des Parameters a_1 keine deutlichen Unterschiede in den Abbildungen (a), (b) und (c). Für jede Kombination aus Parametern a_1, a_2 sowie a_1, a_3 und a_1, a_4 (im betrachteten Wertebereich) gibt es eine Kombination aus den verbleibenden Parameter (a_3, a_4 bzw. a_2, a_4 bzw. a_2, a_3), mit der ein ähnlich geringer Fehlerwert erzielt werden kann. Es gibt in allen drei Abbildungsteilen horizontale Grauschattierungen gibt, die leichte Unterschiede für die Wahl von a_1 zeigen – der Bereich zwischen 1,05 und 2,55 scheint etwas geringere Fehlerwerte aufzuweisen als die anderen Bereiche von a_1 .
2. In Abbildungsteil (b) zeigen sich vereinzelt dunklere Strukturen (d.h. geringere MAPE-Werte im Median) für einige Kombinationen von a_1 und a_3 . Eine eindeutige Struktur ist nicht erkennbar.
3. Abbildungsteil (c) zeigt deutlich, dass für Parameter a_4 eine Einschränkung von $a_4 \leq 0,55$ detaillierter untersucht werden sollte. Für größere Werte von a_4 ist keine Struktur erkennbar, der gesamte Bereich von $a_4 > 0,55$ hat einheitlich für alle Werte von a_1 Fehlerwerte von mindestens 12%.
4. Abbildungsteil (d) zeigt diagonal einige strukturelle Unterschiede für die einzelnen Parameterkombinationen. Die Struktur entspricht einer Bifurkation. Ein möglicher Zusammenhang von a_2 und a_3 wird zu einem späteren Zeitpunkt erneut untersucht, nachdem ein erster Parameter festgesetzt werden kann.
5. Abbildungsteil (e) zeigt erneut deutlich, dass eine Einschränkung des Parameters $a_4 \leq 0,55$ sinnvoll ist. Basierend auf diesem Ergebnis wird nachfolgend eine erneute Rastersuche mit veränderten Parameterbereichen und einer genaueren Auflösung für a_2 und a_4 durchgeführt.
6. Abbildungsteil (f) bestätigt erneut den Eindruck aus Teil (c) und (e) hinsichtlich eines geeigneten Wertebereichs für a_4 . Für höhere Werte von a_4 werden unabhängig von a_3 keine MAPE-Werten $< 12\%$ gefunden, ebenso verhält es sich für $a_3 \geq 145$ selbst bei kleinen Werten für a_4 . Bezüglich a_3 bietet sich eine Einschränkung des Wertebereichs mit $a_3 < 135$ an – in Kombination mit den Resultaten aus (d) wird nachfolgend $60 \leq a_3 \leq 110$ gewählt.

Um einen besseren Eindruck der Struktur zu erhalten, wird eine neue Rastersuche mit feinerer Auflösung für a_2 und a_4 durchgeführt. Dabei wird $a_1 \in [1;3]$ mit einer Schrittweite von 0,5 gewählt, $a_2 \in [0;2,6]$ und einer Schrittweite von 0,2, $a_3 \in [60;110]$ mit einer Schrittweite von 10 und $a_4 \in [0,15;0,55]$ mit einer Schrittweite von 0,02. Es ergibt sich eine Rastersuche mit $5 \cdot 11 \cdot 6 \cdot 21 = 6\,930$ Fehlerwerten pro Datensatz, d.h. insgesamt rund 560 000 Kombinationen aus Parametern und Fehlerwerten ($6\,930 \cdot (21 + 18 + 23 + 19) = 561\,330$). Abbildung 6.3 (a) zeigt die resultierende Heatmap im Median über alle Datensätze für die Parameterkombina-

tion a_2, a_4 . Die weiteren Heatmaps aus dieser Rastersuche sind der Vollständigkeit halber im Anhang, Seite 251, in Abbildung A.1 aufgeführt.

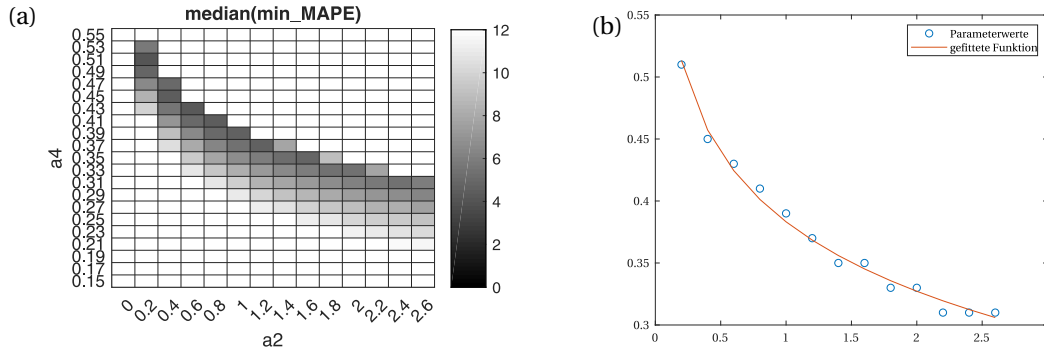


Abbildung 6.3.: Heatmaps im Median über alle Datensätze für die Parameterkombination a_2, a_4 und Curvefitting

Zur Bestimmung des zugrunde liegenden funktionalen Zusammenhangs zwischen den Parametern a_2 und a_4 wird ein automatisiertes Curvefitting in MATLAB über die Funktion *fit* durchgeführt. Da die Struktur logarithmisch (skaliert und verschoben) aussieht, wird dafür nach manueller Funktionsidentifikation die Struktur $p_1 \cdot (1 - \log(p_2 \cdot a_2)) + p_3$ vorgegeben (mit $a_2 \neq 0$ und $p_2 \neq 0$). Falls in den zugrunde liegenden Parameterkombinationen mehrere Fehlerwerte für einen Wert von a_2 (mit unterschiedlichen Werten von a_4) vorliegen, wird diejenige Parameterkombination mit dem kleinsten Fehlerwert gewählt. Es ergibt sich die Funktion $a_4 = 0,0808 \cdot (1 - \log(0,4622 \cdot a_2)) + 0,2400$ (für $a_2 \neq 0$), deren Parameter für nachfolgende Versuche auf $a_4 = 0,08 \cdot (1 - \log(0,46 \cdot a_2)) + 0,24$ (für $a_2 \neq 0$) gerundet werden.

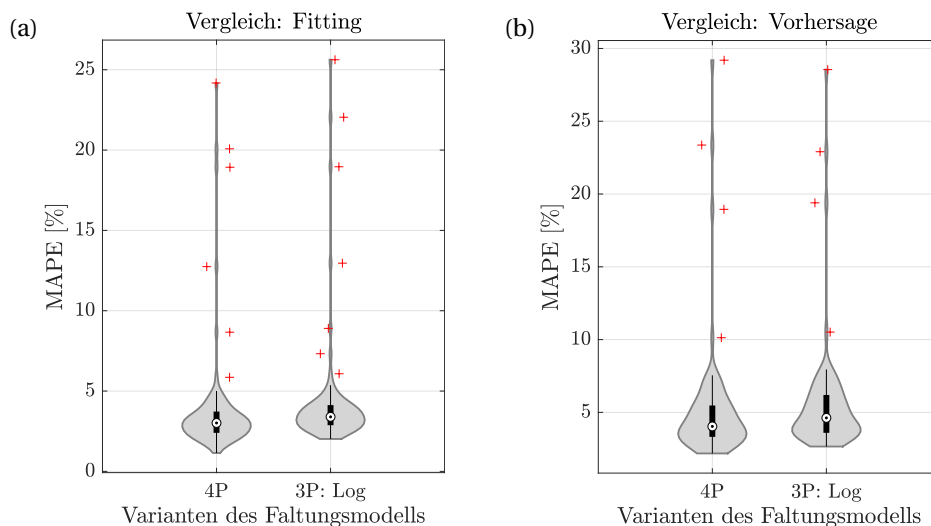


Abbildung 6.4.: Violinenplot: Vergleich Faltungsmoell 4 Parameter und 3 Parameter für (a) Fitting und (b) Vorhersage

Bevor die Anzahl der Parameter weiter reduziert wird, wird überprüft, ob das Faltungsmoell mit drei Parametern und der logarithmischen Funktion für a_4 im Fitting und in der

Vorhersage vergleichbare Ergebnisse zum 4P-Faltungsmodell liefert. Die Fittingsimulation findet auf jedem Datensatz einzeln statt. Für die Simulation der Vorhersage werden die Parameter iterativ in einem »sliding window«-Ansatz auf drei Trainingsdatensätzen kalibriert und der zeitlich darauffolgende Datensatz wird vorhergesagt. Für die n Trainingseinheiten eines Probanden bedeutet das, dass das Modell zunächst auf den Datensätzen 1, 2, 3 kalibriert wird, um Datensatz 4 vorherzusagen. Anschließend wird das Modell auf den Datensätzen 2, 3, 4 kalibriert und Datensatz 5 wird vorhergesagt. Auf diese Weise wird das »Fenster« für die Kalibrierung und die Vorhersage über die gesamten Trainingsdaten eines jeden Probanden geschoben, bis schließlich auf den Datensätzen $n - 3$, $n - 2$, $n - 1$ kalibriert und Datensatz n vorhergesagt wird. Für die einzelnen Probanden mit n Trainingseinheiten entstehen damit $n - 3$ Vorhersagesimulationen (d.h. $21 - 3 = 18$ Vorhersagesimulationen für Proband 1, $18 - 3 = 15$ für Proband 2, $23 - 3 = 20$ für Proband 3 und $19 - 3 = 16$ für Proband 4). Die Violinenplots der MAPE-Fehlerwerte für die beiden Modellvarianten sind in Abbildung 6.4 dargestellt, wobei (a) die Ergebnisse der Fittingsimulation und (b) die Ergebnisse der vergangenheitsbasierten Prädiktion darstellt. Die prozentualen Fehlerwerte des 3P-Modells sind in der Fittingsimulation und in der Vorhersagesimulation nur geringfügig höher sind als beim originalen Faltungsmodell. Die Violinenplots haben eine ähnliche Form und weichen insgesamt nur wenig voneinander ab. Es ergibt sich keine signifikante Verschlechterung im Fitting oder in der Vorhersage durch die Reduktion der Parameteranzahl von vier auf drei: Die Nullhypothese, dass der Mittelwert der MAPE-Fehlerwerte des 4P-Modells und des 3P-Modells gleich sind, kann zum Niveau $\alpha = 5\%$ mit dem Welch-Test sowohl im Fitting ($p = 0,43$) als auch in der Vorhersage ($p = 0,57$) nicht verworfen werden.

Da keine signifikante Verschlechterung durch den funktionalen Zusammenhang der Parameter a_2 und a_4 feststellbar ist, wird der logarithmische Zusammenhang beibehalten. Eine weitere offensichtliche Struktur zwischen anderen Parameterkombinationen konnte wie bereits beschrieben im visuellen Parametervergleich nicht detektiert werden, sodass erneut eine Rastersuche mit höherer Auflösung in den Schrittweiten für die Parameter a_1 , a_2 und a_3 unter Verwendung des logarithmischen Zusammenhangs von a_2 und a_4 durchgeführt wird.

Dabei wird $a_1 \in [0,5; 3]$ mit einer Schrittweite von 0,1 gewählt, $a_2 \in [0,00025; 2,5025]$ und einer Schrittweite von 0,05 und $a_3 \in [50; 120]$ mit einer Schrittweite von 5. Es ergibt sich eine Rastersuche mit $26 \cdot 51 \cdot 15 = 19\,890$ Fehlerwerten pro Datensatz, d.h. insgesamt rund 1,6 Millionen Kombinationen ($19\,890 \cdot (21 + 18 + 23 + 19) = 1\,611\,090$) aus Parametern und Fehlerwerten.

Anhand der Heatmaps in Abbildung 6.5 und Abbildung 6.6 lassen sich einige weitere Aspekte erkennen:

1. Für den betrachteten Parameterbereich gibt es hinsichtlich des Parameters a_1 weiterhin keine deutlichen Unterschiede in Abbildung (a). In Abbildungsteil (b) lassen sich horizontal minimale Schattierungen erkennen. Der Bereich zwischen etwa 1,7 und 2,2

scheint etwas dunkler gefärbt, sodass es für diesen Parameterbereich von a_1 bei Werten von a_3 zwischen 70 und 85 einen Parameter a_2 gibt, für den der MAPE-Fehlerwert etwas geringer ausfällt, als bei anderen Wertebereichen von a_1 und a_3 . Auch für größere Werte von a_3 (100 – 110) zeichnen sich etwas dunklere Schattierungen in einen Bereich von $a_1 \approx 2$ ab. Die Wahl von $a_1 = 2$ (und einigen weiteren Werten im Bereich 1,7 – 2,2) wird daher nachfolgend für eine weitere Reduktion der Parameteranzahl untersucht.

2. Abbildungsteil (c) zeigt eine deutliche Struktur. Für den minimal betrachteten Wert von $a_2 = 0,0025$ muss der additive Parameter $a_3 > 90$ gewählt werden, damit Fehlerwerte $< 12\%$ (im Median über alle Datensätze) erreicht werden können. Für größere Werte von $a_2 > 0,5$ entsteht ein breites »Band« für a_3 im Bereich 55 bzw. 60 bis etwa 75 bzw. 80. Die Grauschattierung im Bereich $a_2 \in \{0,00025; 0,05025; 0,10025\}$ erscheint dunkler als für andere Werte von a_2 , sodass für diesen Bereich eine weitere Raster-suche mit feinerem Raster erfolgt. Das Ergebnis ist in Abbildung 6.6 dargestellt. Die Bifurkation aus der ersten Rastersuche mit vier freien Parametern ist nicht mehr erkennbar.
3. Abbildung 6.6 zeigt das Ergebnis einer weiteren Rastersuche für Werte von a_2 im Bereich 0,0025 – 0,2 mit einer Schrittweite von 0,0025. Auch hier zeigt sich ähnlich wie in Teil (c) ein breites Band für die Werte von a_3 , in dem es keine erkennbaren Schattierungen gibt. Werte für a_3 im Bereich 75 – 85 scheinen in diesem Wertebereich von a_2 für viele Werte von a_2 geringe Fehlerwerte zu ermöglichen.

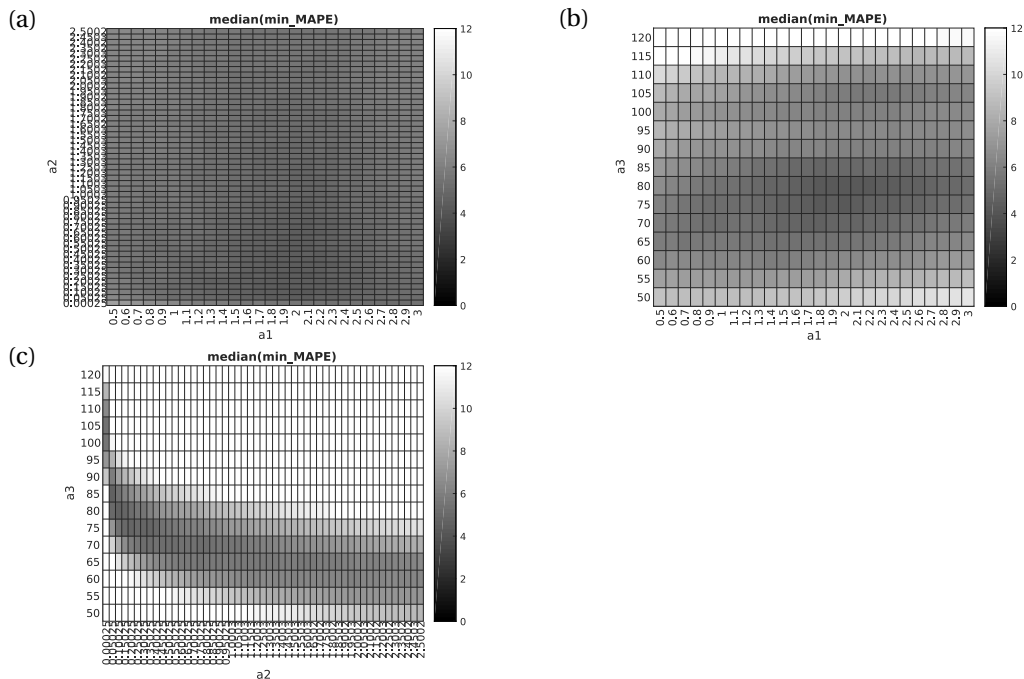


Abbildung 6.5.: Heatmap für die Parameterkombination a_1, a_2, a_3 , Studie 3

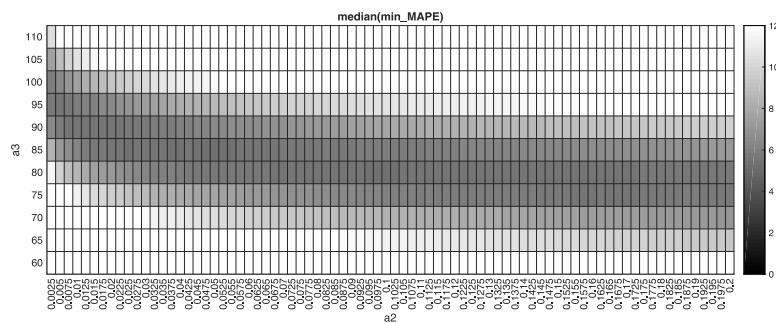


Abbildung 6.6.: Heatmap einer Detail-Rastersuche für a_2, a_3

Im nächsten Schritt wird daher in unterschiedlichen Experimenten jeweils einer der beiden Parameter versuchsweise manuell festgesetzt. Innerhalb der visuellen Parameteranalyse hat sich deutlich gezeigt, dass Parameter a_1 im betrachteten Bereich keinen großen Einfluss zu haben scheint. Für jeden Wert von a_1 wurde eine geeignete Kombination der restlichen Parameter gefunden, mit dem geringe MAPE-Werte erzielt werden. Wie bereits beschrieben werden daher innerhalb des geeigneten Bereichs rund um $a_1 = 2$ mehrere Werte für a_1 testweise festgesetzt und die Ergebnisse für Fitting und vergangenheitsbasierte Vorhersage miteinander und mit dem bereits identifizierten 3P-Faltungsmodell verglichen. Exemplarisch werden die Werte $a_1 \in \{1,5; 1,7; 2; 2,2; 2,5\}$ betrachtet. Die Ergebnisse sind als Violinenplots der MAPE-Fehlerwerte in Abbildung 6.7 dargestellt.

Die Violinenplots bestätigen dein Eindruck aus den Heatmaps, dass die spezifische Wahl von a_1 im Faltungsmodell mit noch zwei weiteren freien Parametern faktisch keinen Einfluss auf die Güte der Ergebnisse hat – weder im Fitting, noch in der Vorhersage. Die MAPE-Werte für $a_1 = 2$ und $a_1 = 2,2$ sind minimal niedriger als für die anderen betrachteten Werte von a_1 , sodass nachfolgend $a_1 = 2$ gewählt wird – eine andere Wahl ist jedoch ebenso möglich, was in den nachfolgenden Analysen bei der Interpretation berücksichtigt werden muss.

Die Nullhypothese, dass der Mittelwert der MAPE-Fehlerwerte des 3P-Modells und dieses 2P-Modells gleich sind, kann zum Niveau $\alpha = 5\%$ mit dem Welch-Test in der Vorhersage ($p = 0,35$) nicht verworfen werden. Im Fitting ergibt sich eine signifikante Abweichung mit $p = 0,02$. Da jedoch vor allem die Vorhersagegenauigkeit eine zentrale Rolle spielt, während es im Fitting insbesondere mit einer höheren Anzahl an freien Parametern leicht zu einer Überanpassung an die Kalibrierungsdaten kommen kann und innerhalb der Vorhersage keine signifikante Abweichung feststellbar ist, wird die Reduktion beibehalten. Auch mit einer anderen Wahl von a_1 ergibt sich für das Fitting ein signifikantes Testergebnis mit $p \leq 0,02$, während in der Vorhersage keine Signifikanz feststellbar ist ($p \geq 0,22$).

Innerhalb des neu gebildeten 2P-Modells wird nun Parameter a_3 versuchsweise fixiert. Dieser Parameter entspricht phänomenologisch der Start-Herzfrequenz vor Beginn der Trainingseinheit. In der bereits untersuchten visuellen Parameteranalyse erschienen Werte im Bereich $a_3 \in \{65, 70, 75, 80, 85\}$ grundsätzlich eher geeignet zu sein. Daher werden nachfol-

6. Verbesserung des Faltungsmodells: Parameterreduktion

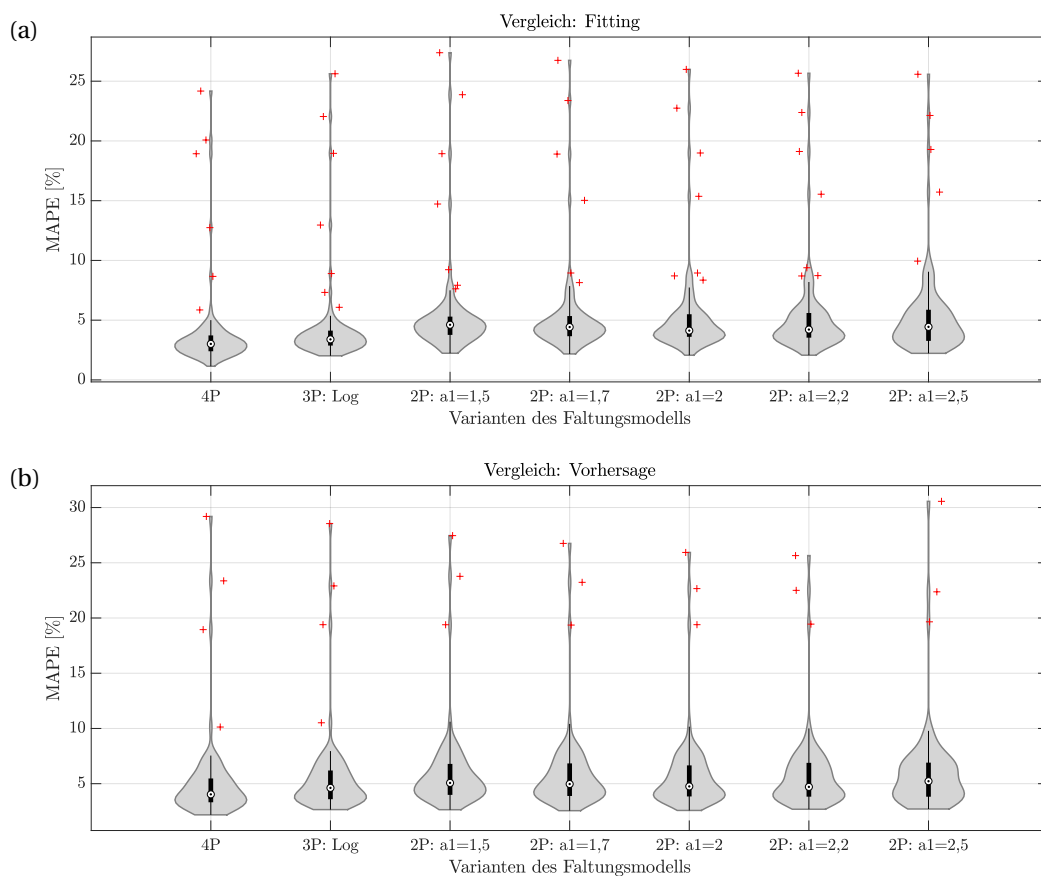


Abbildung 6.7.: Violinenplot: Vergleich mehrerer Varianten des Faltungsmodells mit unterschiedlicher Parameteranzahl und Parametereinstellung für (a) Fitting und (b) Vorhersage

gend die Ergebnisse von sieben weiteren Varianten eines 2P-Faltungsmodells ergänzt, bei denen a_3 nacheinander auf einem dieser fünf Werte, zusätzlich exemplarisch auf 100 und im siebten Fall auf den ersten gemessenen Herzfrequenzwert festgesetzt wird. Die Violinenplots der MAPE-Fehlerwerte für die sieben Modellvarianten sind in Abbildung 6.8 dargestellt. Bei der Wahl von a_3 als erster gemessener Herzfrequenzwert stellt sich für das 2P-Faltungsmodell nur ein geringer Genauigkeitsverlust im Fitting im Vergleich zum 2P-Faltungsmodell ein – die Vorhersage hingegen weist im Vergleich leicht höhere Fehlerwerte auf. Andere Einstellungen für a_3 erreichen weder im Fitting noch in der Vorhersage zuverlässige Ergebnisse und weisen MAPE-Fehlerwerte von über 15% ($a_3 = 100$) bis hin zu fast 40% ($a_3 = 65$) im Median auf.

Die Nullhypothese, dass der Mittelwert der MAPE-Fehlerwerte des 2P-Modells und eines dieser 1P-Modells gleich sind, kann zum Niveau $\alpha = 5\%$ mit dem Welch-Test sowohl im Fitting ($p \ll 0,0001$) als auch in der Vorhersage ($p \ll 0,0001$) für alle Einstellungen von $a_3 \in \{65, 70, 75, 80, 85\}$ verworfen werden. Wird der erste gemessene HF-Wert für a_3 gewählt, tritt lediglich im Fitting im Vergleich zum 2P-Modell mit $p = 0,054$ keine Signifikanz zum 5%-Niveau ein.

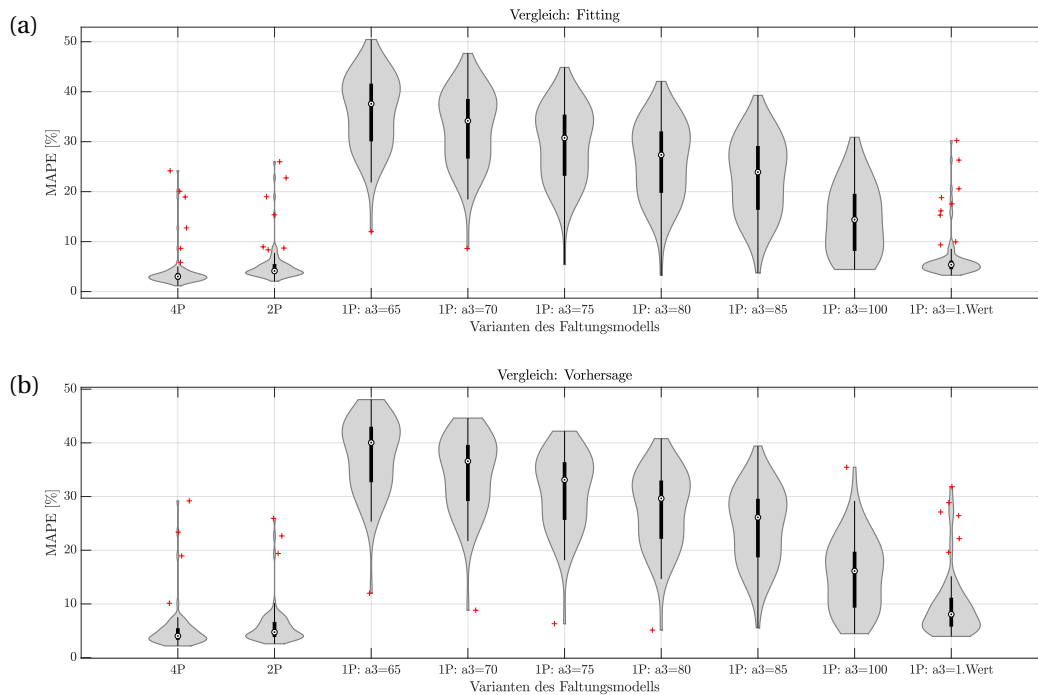


Abbildung 6.8.: Violinplot: Vergleich mehrerer Varianten des Faltungsmodells mit unterschiedlicher Parameteranzahl und Parametereinstellung für (a) Fitting und (b) Vorhersage

Die Fixierung von a_3 auf einen festen Wert für alle Probanden und alle Datensätze erweist sich damit als ungeeignet. Die Verwendung des ersten Messwertes kommt prinzipiell in Betracht, geht allerdings mit einer signifikanten Verschlechterung der Fehlerwerte in der Vorhersage einher.

Anstelle von a_3 wird daher versuchsweise Parameter a_2 zur Bildung eines Einparametermodells fixiert. Auf Basis der betrachteten Heatmap-Darstellungen wird hierzu $a_2 \in \{0,0025; 0,025; 0,1; 0,25; 2\}$ betrachtet. Abbildung 6.9 zeigen die Violinplots der unterschiedlichen Parametereinstellungen für Fitting und Vorhersage im Vergleich. Ähnlich wie bereits bei der Fixierung von a_1 zeigen sich auch hier kaum Unterschiede in den einzelnen Einstellungen des Parameters. Die Fehlerwerte für Fitting und Vorhersage liegen für alle betrachteten Werte in einem zum 2P-Faltungsmodell vergleichbaren Bereich, das seinerseits wie bereits diskutiert eine zum vier- und 3P-Faltungsmodell vergleichbare Genauigkeit modelliert. Die Nullhypothese, dass der Mittelwert der MAPE-Fehlerwerte des 2P-Modells und eines der 1P-Modelle gleich sind, kann zum Niveau $\alpha = 5\%$ mit dem Welch-Test lediglich für $a_2 = 4$ im Fitting ($p = 0,04$) schwach signifikant verworfen werden, wobei der Violinplot etwas höhere Fehlerwerte für diese Wahl von a_2 zeigt. Für alle anderen betrachteten Werte von a_2 kann die Nullhypothese sowohl im Fitting ($p \geq 0,25$) als auch in der Vorhersage ($p \geq 0,37$) nicht verworfen werden. Die MAPE-Werte in Fitting und Vorhersage für $a_2 = 0,25$ sind minimal niedriger als für die anderen betrachteten Werte von a_2 , sodass nachfolgend

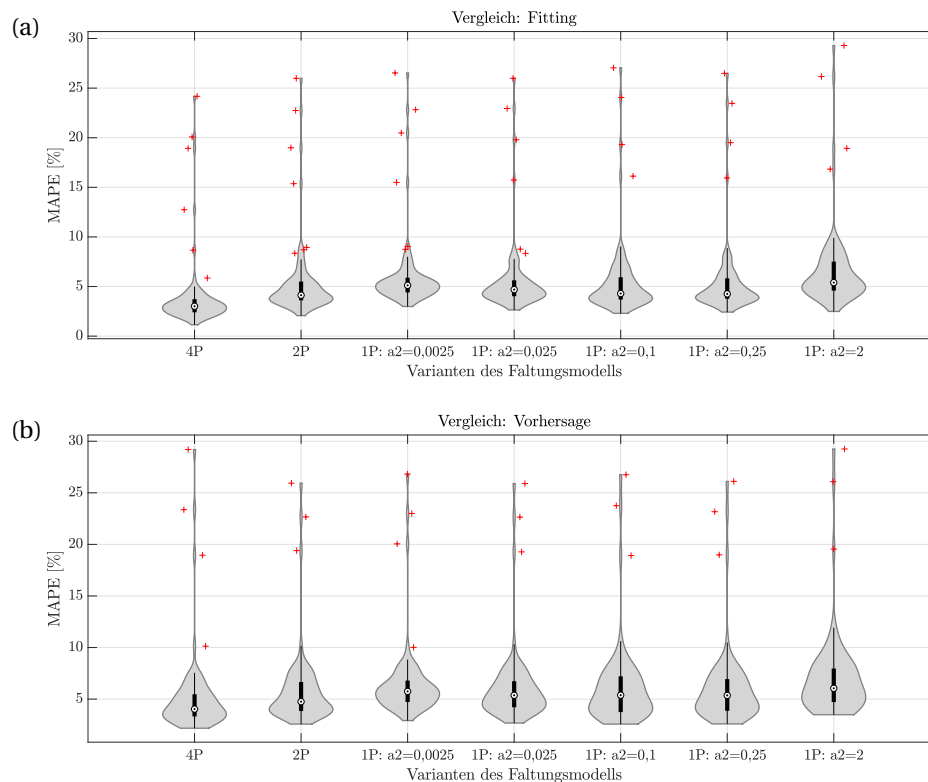


Abbildung 6.9.: Violinplot: Vergleich mehrerer Varianten des Faltungsmodells mit unterschiedlicher Parameteranzahl und Parametereinstellung für (a) Fitting und (b) Vorhersage

$a_2 = 0,25$ gewählt wird – eine andere Wahl (außer $a_2 = 4$) ist jedoch ebenso möglich, was in den nachfolgenden Analysen bei der Interpretation berücksichtigt werden muss.

Im Vergleich dieses 1P-Modells (d.h. mit $a_2 = 0,25$) mit dem 4P-Modell kann die Nullhypothese auf gleiche Mittelwerte im Fitting mit $p = 0,0085$ verworfen werden. In der Vorhersage sind mit $p = 0,25$ keine signifikanten Unterschiede mit dem Welch-Test identifizierbar. Wie bereits zu Beginn und im Kontext der zuvor identifizierten 1P-Faltungsmodellvariante erläutert, spielt vor allem die Vorhersagegenauigkeit eine zentrale Rolle, während es im Fitting insbesondere mit einer höheren Anzahl an freien Parametern leicht zu einer Überanpassung an die Kalibrierungsdaten kommen kann. Das 1P-Faltungsmodell weist in der Vorhersage keine signifikante Verschlechterung zum 4P-Faltungsmodell auf und kann demnach ebenfalls als parameterreduzierte Variante verwendet werden. Die QQ-Plots beider Varianten sind im Anhang in Abbildung A.2 (S. 252) dargestellt.

Evaluation auf Outdoor-Trainingseinheiten

Die Parameterreduktion wurde auf den insgesamt 81 Ergometer-Datensätzen der vier Probanden durchgeführt. Ergänzend wird das identifizierte Einparameter-Faltungsmodell nun anhand der restlichen Trainingsdaten (Outdoor-Trainingseinheiten und Trainingsfahrten

auf der Rolle) evaluiert. Dazu werden das originale Faltungsmodell mit vier Parametern und das 1P-Faltungsmodell verwendet und die Ergebnisse für Fitting und Vorhersage miteinander verglichen. Die Ergebnisse des Fittings beziehen sich auf die Kalibrierung der Modellparameter bzw. des Modellparameters auf jeweils einer Trainingseinheit. Für die Vorhersage wird das jeweilige Modell auf je drei Trainingseinheiten kalibriert, und die Herzfrequenz des darauf folgenden Trainings wird prognostiziert – diese Form der Prädiktion wurde bereits auf den Ergometer-Datensätzen für die Identifikation einer geeigneten Parameterreduktion verwendet. Die Ergebnisse über alle Probanden sind in Form von Violinenplots in Abbildung 6.10 dargestellt.

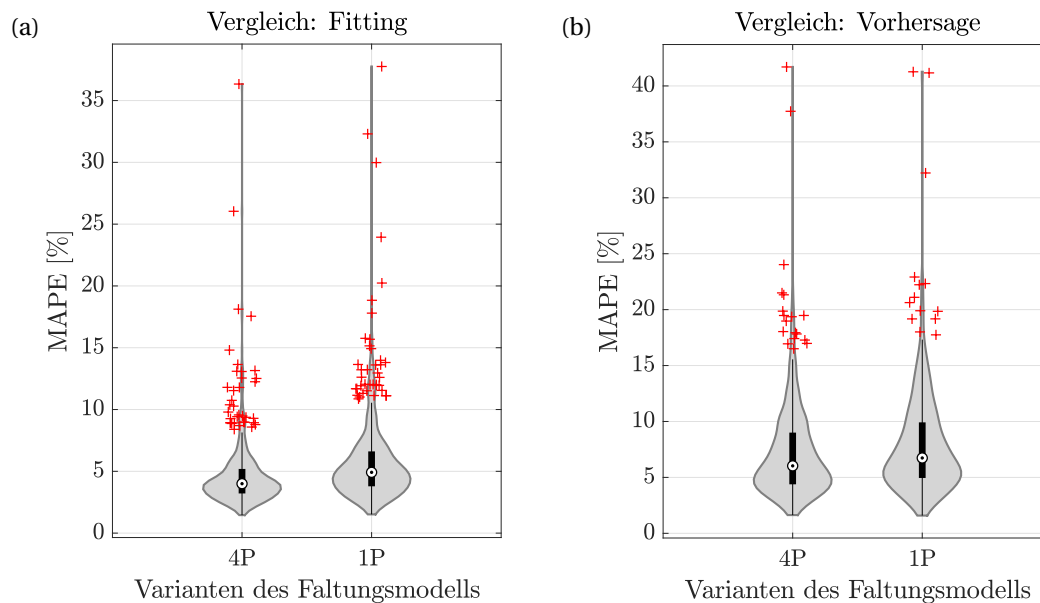


Abbildung 6.10.: Violinenplots zur Evaluation der Parameterreduktion auf Outdoor-Trainingsdaten für (a) Fitting und (b) Vorhersage

Sowohl im Fitting als auch in der Vorhersage erzielt das 1P-Faltungsmodell im Vergleich zum 4P-Faltungsmodell nur minimal höhere Fehlerwerte (Fitting: 4,63% im Vergleich zu 5,37% im Median, Vorhersage: 6,62% im Vergleich zu 7,22% im Median). Auch die Anzahl und Fehlerhöhe der Ausreißer ist für beide Modelle sehr ähnlich. Es kann folglich davon ausgegangen werden, dass das parameterreduzierte Modell nicht an die zur Parameterreduktion verwendeten Trainingsdaten überangepasst wurde. Allerdings zeigt sich im Welch-Test ein signifikanter Unterschied für die Mittelwerte: Sowohl im Fitting ($p < 0,0001$) als auch für die Vorhersage ($p = 0,02$) wird die Nullhypothese auf gleiche Mittelwerte mit deutlicher Signifikanz verworfen. Über den QQ-Plot in Abbildung 6.11 zeigt sich, dass sich fast alle Quantil-Werte auf der Diagonalen befinden und es sowohl im Fitting als auch in der Vorhersage nur drei Ausreißer gibt. Trotz der signifikanten Abweichung vom 4P-Modell in den Fehlerwerten scheint das 1P-Modell auch bei der Verwendung für Outdoordaten insgesamt niedrige Fehler und valide Ergebnisse zu liefern.

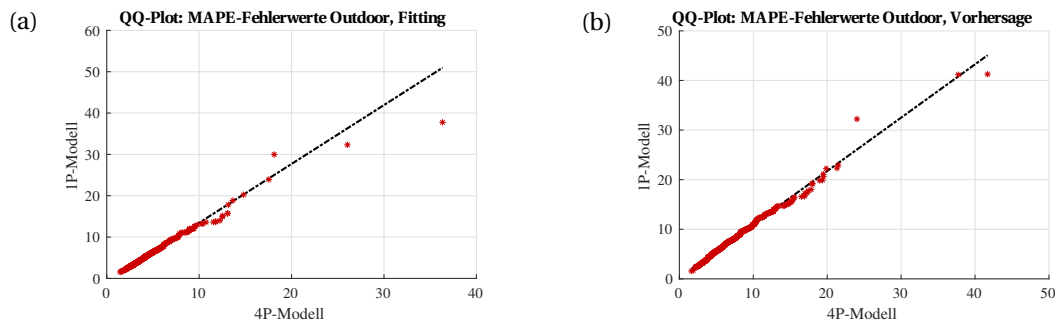


Abbildung 6.11.: QQ-Plots zur Evaluation der Parameterreduktion auf Outdoor-Trainingsdaten

Es treten vereinzelt Fälle in der Vorhersage auf, die mit beiden Modellvarianten nicht adäquat gefittet und vorhergesagt werden können – es werden dabei maximale Fehlerwerte von 41,70% (4P) bzw. 41,26% (1P) erzielt.

Eine qualitative Analyse der Vorhersagesimulationen zeigt, dass die Vorhersage selbst durchweg geeignet erscheint, die empirischen Daten jedoch in den Fällen mit hohen Fehlern starke Messfehler mit unrealistische Schwankungen aufweisen. Abbildung 6.12 zeigt exemplarisch vier Vorhersagesimulationen: Abbildung 6.12 (a) und (b) zeigen das Ergebnis einer typischen Vorhersagesimulation am Beispiel einer Ergometrie-Einheit von Proband 1. Abbildungsteil (a) zeigt dabei die Vorhersage mit dem 4P-Faltungsmodell, Abbildung (b) die gleiche Vorhersage mit dem 1P-Faltungsmodell. Die Fehler sind bei beiden Modellen sehr niedrig mit (a) 2,84% bzw. (b) 3,12%. Beide Modellvarianten sind in der Lage, die Herzfrequenzdynamik auch über den Zeitraum von rund zwei Stunden adäquat auf Basis einer Kalibrierung anhand vorheriger Trainingseinheiten zu prognostizieren. Im Verlauf des Herzfrequenzverhaltens lässt sich visuell kein Unterschied zwischen den Vorhersagen mit den beiden Modellen erkennen – lediglich der erste Wert der Herzfrequenz ist im 1P-Modell deutlich niedriger als im 4P-Modell.

Abbildung 6.12 (c) und (d) zeigen das Ergebnis eines typischen Beispiels einer »schlechten« Vorhersage von Proband 2. Der Fehler des 4P-Modells in Abbildungsteil (c) liegt bei rund 24,02%, der des 1P-Modells in Abbildungsteil (d) bei rund 41,17%. Es zeigt sich deutlich, dass die gemessenen Daten bzw. die Vielzahl an fehlenden Messwerten für den größten Teil des Fehlers verantwortlich ist. Es liegen lediglich für einen kurzen Abschnitt Messungen der Wattleistung vor (in den erste ca. 10 Minuten sowie etwa von Minute 140 bis 170). Zu den Zeitpunkten, an denen sowohl Messungen der empirischen Herzfrequenz wie auch der Wattleistung vorliegen, sind beide Modelle in der Lage, das Herzfrequenzverhalten annähernd adäquat zu prognostizieren. Dabei fällt auf, dass das 1P-Modell die Herzfrequenz chronisch unterschätzt.

Es ist innerhalb der Analyse der Outdoor-Trainingseinheiten kein Fall innerhalb der Vorhersage aufgetreten, in dem nur eine der beiden Modellvarianten besonders hohe oder besonders niedrige Fehlerwerte erzielt hat. Bei allen Fällen mit hohen Fehlerwerten liegen starke

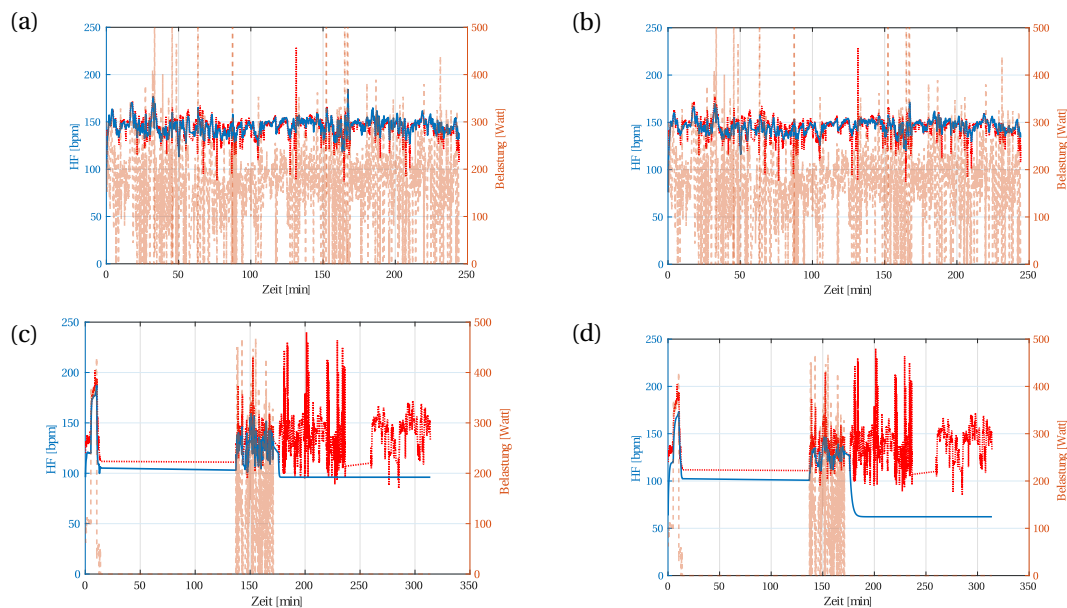


Abbildung 6.12.: Exemplarische Vorhersagesimulationen auf den Outdoor-Trainingsdaten mit der zugrunde liegenden Belastung in Watt (orange, halb transparent), der empirischen Herzfrequenz (rot) und der simulierten Herzfrequenz (blau). (a): Beispiel des 4P-Faltungsmodells mit geringem Fehler (MAPE 2,84%); (b): Beispiel der gleichen Simulation wie in (a) des 1P-Faltungsmodells mit geringem Fehler (MAPE 3,12%); (c): Beispiel des 4P-Faltungsmodells mit hohem Fehler (MAPE 24,02%); (d): Beispiel der gleichen Simulation wie in (c) des 1P-Faltungsmodells mit sehr hohem Fehler (MAPE 41,17%)

Schwankungen und Messfehler der empirischen Herzfrequenz oder in der Wattmessung vor. Die insgesamt gering höheren Vorhersagefehler beim 1P-Faltungsmodell lassen sich durch eine leichte Unterschätzung der empirischen Herzfrequenz und einem tendenziell zu niedrig kalibrierten Parameter a_3 erklären.

Evaluation einer virtuellen Leistungsdiagnostik

Die »Virtuelle Leistungsdiagnostik« meint eine Vorhersage von ergometriebasierten Testeinheiten auf Basis einer Modellkalibrierung auf vorhergehenden »klassischen« Trainingseinheiten (als »Outdoor«-Feld-Training oder auch als Trainingseinheit auf der Rolle). Ziel der virtuellen Leistungsdiagnostik ist es, basierend auf klassischen Trainingseinheiten zu prognostizieren, wie sich die Leistung einer Probandin oder eines Probanden entwickelt. Ein erster Schritt dazu ist es, basierend auf den regulären Trainingseinheiten die Herzfrequenz einer Probandin oder eines Probanden in einem Ergometertest vorherzusagen. Anschließend können bei der Verwendung standardisierter Testprotokolle beispielsweise TRIMP-Werte als Leistungswerte berechnet werden. Darüber soll es schließlich möglich sein, mit Hilfe einer versierten Trainerin oder eines versierten Trainers Rückschlüsse auf die Leistungsfähigkeit ziehen zu können, ohne die Ergometrie tatsächlich durchzuführen. Im Rah-

men der in dieser Arbeit durchgeführten Analysen wird nur der erste Schritt betrachtet (Vorhersagesimulation), in dem anhand von klassischen Trainingseinheiten die Herzfrequenz auf dem Ergometer vorhergesagt wird.

Dazu werden für jedes vorliegende Ergometer-Training die vorangegangenen Trainingseinheiten betrachtet, die maximal 14 Tage zurückliegen. Der Parameter des 1P-Faltungsmodells und die Parameter des 4P-Faltungsmodells werden auf diesen maximal 14 vergangenen Trainingseinheiten kalibriert und anschließend dazu verwendet, das Ergometer-Training zu prognostizieren. Ein Vergleich mit den empirischen Herzfrequenzwerten des Ergometer-Trainings zeigt, wie valide die virtuelle Leistungsdiagnostik mit dem Faltungsmodell durchgeführt werden kann.

Als weitere Fragestellung wird in diesem Kontext untersucht, inwieweit es anhand der regulären Trainingsdaten möglich ist, einen Ergometertest vorherzusagen. Abbildung 6.13 zeigt die quantitativen Ergebnisse des Fittings und der Vorhersage für das 4P- und das 1P-Faltungsmodell im Violinenplot. Die Auswertung zeigt, dass beide Modellvarianten gute Ergebnisse im Fitting auf den regulären (meist »Outdoor« durchgeführten) Trainingseinheiten erzielen.

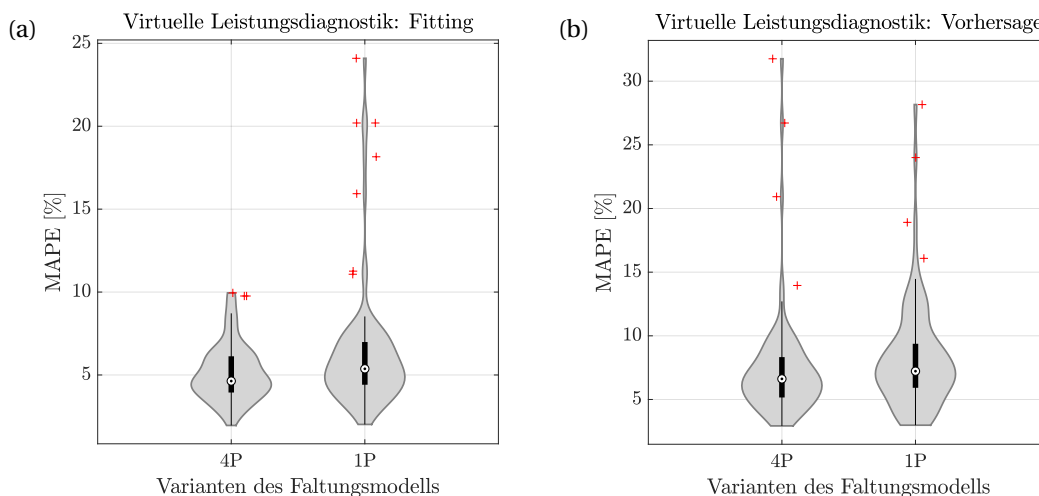


Abbildung 6.13.: Violinenplots zur Evaluation der virtuellen Leistungsdiagnostik für (a) Fitting und (b) Vorhersage

Die prozentualen MAPE-Fehlerwerte liegen im Median bei 4,63% (4P) bzw. 5,37% (1P). Der maximale Fehler für das 4P-Modell liegt bei 9,95%. Mit dem 1P-Modell konnten einige einzelne Trainingseinheiten nicht akkurat gefittet werden, sodass vereinzelt Fehler bis zu 24,10% auftreten. Diese Effekt, dass nur mit dem 1P-Modell derartige Ausreißer auftreten, nivelliert sich in der Vorhersage: Der prozentuale Fehler im Median ist auch in der Vorhersage für beide Modelle vergleichbar und nur geringfügig höher als beim Fitting mit Fehlerwerten von 6,62% (4P) bzw. 7,22% (1P). Vereinzelt treten in der Vorhersage für beide Modellvarianten Fälle von Trainingseinheiten auf, die nicht vorhergesagt werden können – es werden maximale Fehlerwerte von 31,76% (4P) bzw. 28,16% (1P) erzielt. Während die Nullhypothe-

se auf gleiche Mittelwerte mit dem Welch-Test zum Signifikanzniveau $\alpha = 5\%$ im Fitting mit $p = 0,01$ abgelehnt wird, zeigt sich in der Vorhersage mit $p = 0,22$ kein signifikanter Unterschied. Die QQ-Plots sind im Anhang in Abbildung A.2 (S. 252) dargestellt.

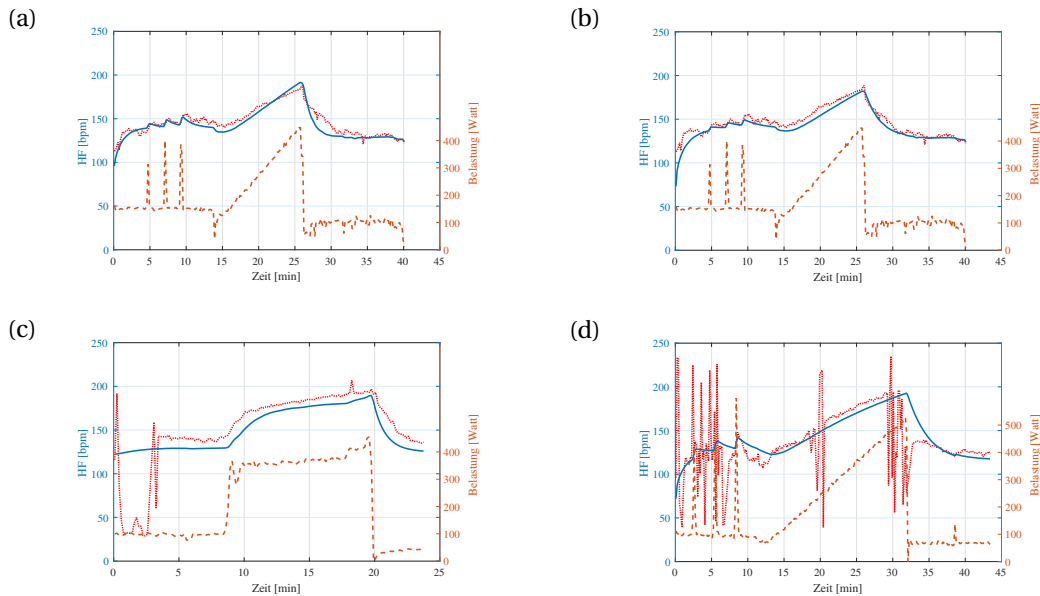


Abbildung 6.14.: Exemplarische Vorhersagesimulationen der virtuellen Leistungsdiagnostik mit der zugrunde liegenden Belastung in Watt (orange), der empirischen Herzfrequenz (rot) und der simulierten Herzfrequenz (blau). (a): Beispiel des 4P-Faltungsmodells mit geringem Fehler (MAPE 3,41%); (b): Beispiel der gleichen Simulation wie in (a) des 1P-Faltungsmodells mit geringem Fehler (MAPE 3,46%); (c): Beispiel des 4P-Faltungsmodells mit hohem Fehler (MAPE 31,76%); (d): Beispiel des 1P-Faltungsmodells mit hohem Fehler (MAPE 18,91%)

Eine qualitative Analyse der Vorhersagesimulationen zeigt, dass die Vorhersage selbst durchweg geeignet erscheint, die empirischen Daten jedoch in den Fällen mit hohen Fehlern starke Messfehler mit unrealistische Schwankungen aufweisen. Abbildung 6.14 zeigt exemplarisch vier Vorhersagesimulationen: Abbildung 6.14 (a) und (b) zeigen das Ergebnis einer typischen Vorhersagesimulation am Beispiel einer Ergometrie-Einheit von Proband 1. Abbildungsteil (a) zeigt dabei die Vorhersage mit dem 4P-Faltungsmodell, Abbildung (b) die gleiche Vorhersage mit dem 1P-Faltungsmodell. Die Fehler liegen für beide Modelle bei (a) 3,41% bzw. (b) 3,46%. Während die Herzfrequenz-Peaks der kurzen Sprint-Einheiten bis Minute 10 mit dem Vierparameter-Modell etwas exakter abgebildet werden können als mit dem Einparameter-Modell, wird der Belastungsabfall bei Minute 26 vom Einparameter-Modell realistischer nachgebildet. Es fällt auf, dass insbesondere der erste Wert der modellierten Herzfrequenz beim Einparameter-Modell deutlich zu niedrig liegt. Diese Beobachtung wird im Rahmen der »Interpretation und Diskussion« näher beleuchtet.

Abbildungsteil (c) zeigt ein typisches Beispiel einer »schlechten« Vorhersage mit dem 4P-Faltungsmodell anhand von Proband 2. Der Fehler liegt bei rund 31,76%. Während das Mo-

dell die empirische Herzfrequenz über die gesamte Zeit leicht unterschätzt, entsteht der größte Teil des Fehlers durch die offensichtlichen Messfehler der empirischen Herzfrequenz in den ersten ca. vier Minuten der Trainingseinheit. Derartige Fehler könnten durch eine geeignete Vorverarbeitung der Rohdaten vermieden werden. Das 1P-Modell (nicht abgebildet) erzielt in diesem Beispiel einen Vorhersagefehler von rund 28,16%.

Abbildungsteil (d) zeigt ein typisches Beispiel einer »schlechten« Vorhersage mit dem 1P-Faltungsmodell anhand von Proband 4. Der Fehler liegt bei rund 18,91%. Bei der empirischen Messung zeigen sich über den gesamten Zeitraum der Trainingseinheit immer wieder zahlreiche Messfehler mit Schwankungen zwischen unter 50 bpm und fast 250 bpm. Der Verlauf der simulierten Vorhersage mit dem 1P-Faltungsmodell hingegen erscheint hinsichtlich der zugrunde liegenden Wattleistung realistisch. Das 4P-Modell (nicht abgebildet) erzielt in diesem Beispiel einen Vorhersagefehler von rund 20,93%.

Es ist innerhalb der Analyse der virtuellen Leistungsdiagnostik kein Fall innerhalb der Vorhersage aufgetreten, in dem nur eine der beiden Modellvarianten besonders hohe oder besonders niedrige Fehlerwerte erzielt hat. Bei allen Fällen mit hohen Fehlerwerten liegen starke Schwankungen und Messfehler der empirischen Herzfrequenz vor.

6.2.2. Interpretation und Diskussion

Durch das iterative Vorgehen der Heatmap-Analyse und der manuellen Fixierung von Parametern konnte die Parameterreduktion erfolgreich durchgeführt und ein Faltungsmodell mit lediglich einem freien Parameter konstruiert werden. Dabei haben sich einige Auffälligkeiten ergeben, die im folgenden interpretiert und diskutiert werden.

Effekt des Parameters a_1

Sowohl in den ersten Heatmap-Abbildungen vor der funktionalen Kopplung der Parameter a_2 und a_4 hat sich gezeigt, dass der spezifische Wert von a_1 innerhalb des betrachteten Wertebereichs keinen starken Einfluss auf das Finden einer Parameterkombination mit niedrigem Fehlerwert hat. Dieser Eindruck hat sich auch in den nachfolgenden Analysen bestätigt. Beim Vergleich des 2P-Modells mit unterschiedlichen konstanten Werten für a_2 mit dem 3P- und 4P-Faltungsmodell zeigten sich weder im Fitting noch in der Vorhersage signifikante Unterschiede zwischen den einzelnen Werten für a_1 . Der Parameter dient im Modell dazu, die Faltung zu gewichten und damit die simulierte Herzfrequenzkurve zu glätten. Ein hoher Wert von a_1 sorgt dafür, dass innerhalb der Berechnung der Faltung mehr Werten aus der Vergangenheit ein höheres Gewicht beigemessen wird. Umgekehrt sorgt ein niedriger Wert von a_1 dafür, dass vergleichsweise wenige Werte aus der Vergangenheit einen hohen Einfluss auf die Berechnung der aktuellen Herzfrequenz haben. Auf diese Weise wird die simulierte Kurve durch hohe Werte von a_1 stärker geglättet, während die Schwankun-

gen bei niedrigen Werten von a_1 den zugrunde liegenden Eingabewerten aus der Belastung ähnlicher sind. Das spezifische Ausmaß der Glättung scheint innerhalb des betrachteten Wertebereichs keinen großen Einfluss auf das Ergebnis zu haben. Die spezifische Wahl von a_1 nimmt demzufolge innerhalb größerer Wertebereiche keinen deutlichen Einfluss auf das Ergebnis – global betrachtet ist der Wert von a_1 natürlich relevant, da sehr kleine Werte die Struktur der Belastung kopieren und sehr hohe Werte jegliche Struktur durch die Glättung verschwinden lassen.

Wahl des Parameters a_3

Bei der Wahl von a_3 zur Konstruktion eines 2P-Faltungsmodells hat sich im Vergleich mit dem 4P- und 3P-Faltungsmodell gezeigt, dass für a_3 ausschließlich die Wahl des ersten Wertes aus der jeweiligen Trainingseinheit in Frage kommt. Wird a_3 auf einem konstanten Wert fixiert, der von der spezifischen Trainingseinheit und dem Probanden unabhängig ist, sind die Fehlerwerte sowohl bei der Modellkalibrierung als auch in der Vorhersage mit Fehlerwerten von $> 15\%$ bis hin zu rund 40% im Median signifikant höher. Wird a_3 hingegen auf den ersten Wert der Trainingseinheit als eine Art »Herzfrequenz unmittelbar vor Trainingsbeginn« gesetzt, ist die Fehlerverteilung des 2P-Faltungsmodells im Fitting mit der aus dem 3P- und 4P-Faltungsmodell vollumfänglich vergleichbar. In der Vorhersage zeigen sich jedoch auch hier auffällig höhere Fehlerwerte als bei den Modellen mit mehr Parametern.

Dies legt den Schluss nahe, dass der additive Parameter a_3 , der die Basis-Höhe der Simulation vorgibt, eine zentrale Rolle für die individuelle Modellierung der Herzfrequenzreaktion spielt. Während eine leichte Individualisierung bereits für die Kalibrierung unabdingbar ist, deutet der höhere Fehler in der Vorhersage darauf hin, dass eine Fixierung dieser Art für a_3 ungeeignet ist. Dies hat zusätzlich den Vorteil, dass der freie Parameter des identifizierten parameterreduzierten Faltungsmodells gut interpretierbar ist: Wie bereits in den Grundlagen beschrieben lassen sich Leistungsveränderungen an der Höhe der Herzfrequenz erkennen. Während eine Verbesserung der Leistung zu einer Optimierung des Herzschlags und damit in der Regel zu einem Absinken der Herzfrequenz führt, hat eine Verschlechterung der Leistung häufig eine leichte Erhöhung des Herzfrequenz-Niveaus zur Folge. Parameter a_3 als freier Parameter im 1P-Faltungsmodell bietet damit am ehesten die Möglichkeit, einen geeigneten Zusammenhang zur Leistungsentwicklung herzustellen. Ob sich diese Vermutung bestätigen lässt, wird in Abschnitt 7.1: »Analyse zur Abschätzung der Leistungsveränderung auf Basis von Herzfrequenz-Simulationen: Profi-Sportler« ausführlich untersucht.

Wahl des Parameters a_2 im Zusammenspiel mit a_4

Während sich zunächst recht schnell ein funktionaler Zusammenhang zwischen a_2 und a_4 über die Heatmap-Darstellung gezeigt hat, konnte im weiteren Verlauf der Analysen gezeigt

werden, dass Parameter a_2 – und damit automatisch auch Parameter a_4 – sogar konstant gewählt werden kann. Dabei kann a_2 ähnlich wie bereits Parameter a_1 innerhalb eines spezifischen Wertebereichs beliebig gewählt werden – die Effekte werden durch den Zusammenhang zu a_4 ausreichend kompensiert, sodass sich keine signifikanten Unterschiede in der Fehlerverteilung feststellen lassen.

Grundsätzlich erscheint ein logarithmischer Zusammenhang von a_2 und a_4 auf den ersten Blick wenig überraschend: Im 4P-Faltungsmodell ist a_2 ein multiplikativer Faktor vor einem Term, bei dem sich a_4 im Exponenten befindet. Durch die Logarithmierung von a_4 entsteht ein einfacher linearer Zusammenhang zu a_2 . Überraschend dabei ist jedoch, dass ein ebensolcher linearer Zusammenhang zwischen diesen beiden Parametern eine Parameterkombination darstellt, die zu geringen Fehlern führt und mit der sich die Ergebnisse des 4P-Faltungsmodells nicht signifikant von denen des 3P-Faltungsmodells unterscheiden. Aus mathematisch-analytischer Sicht wäre es möglicherweise interessant, das 3P-Faltungsmodell mit dem identifizierten parametrischen Zusammenhang zwischen a_2 und a_4 strukturell mit dem ursprünglichen 4P-Faltungsmodell zu vergleichen. Dadurch, dass Parameter a_2 im weiteren Verlauf der Parameterreduktion jedoch konstant gewählt werden konnte und sich somit auch für a_4 lediglich ein konstanter Wert ergibt, wird auf eine derartige Untersuchung der Struktur verzichtet.

Bifurkation in der Heatmap von a_2 und a_3

Innerhalb der ersten Rastersuche für das Faltungsmodell mit vier freien Parametern hat sich in der Heatmap der Parameter a_2 und a_3 deutlich eine bifurkale Struktur gezeigt. Diese ist in der anschließenden Rastersuche für das 3P-Faltungsmodell nicht mehr aufgetreten. Grund dafür sind die Wechselwirkungen zwischen den einzelnen Parametern: Parameter a_3 verschiebt als additiver Wert die simulierte Herzfrequenzkurve in der Höhe, während Parameter a_2 die Steigung der Herzfrequenz als Reaktion auf die Belastung modelliert. Wird a_3 – ausgehend von einem der empirischen Herzfrequenz angepassten Wert – leicht verringert, kann die dadurch »zu niedrige« Simulation durch eine leichte Verstärkung der Steigung (d.h. einen etwas größeren Wert von a_2) selbst ohne Anpassung der übrigen Parameter weitgehend ausgeglichen werden. Dies erklärt auch die logarithmisch anmutende Struktur aus der Rastersuche des 3P-Modells. Insbesondere zu niedrige Werte von a_3 können durch etwas höhere Werte von a_2 hinsichtlich der Fehlerwerte damit annähernd ausgeglichen werden. Solange die übrigen Parameter a_1, a_4 ebenfalls frei variiert werden können, sind unterschiedliche Wertebereiche für die Kombination aus a_2 und a_3 möglich. Auf diese Weise ergeben sich die drei »Äste« der Bifurkation. Durch den funktionalen Zusammenhang von a_4 und a_2 im 3P-Modell reduzieren sich die Möglichkeiten für Wertepaare a_2 und a_3 , mit denen niedrige Fehlerwerte erzielt werden können deutlich, sodass die Bifurkation in der Struktur verschwindet. Wird a_3 ausgehend von einem der empirischen Herzfrequenz angepassten

Wert erhöht, ist eine Anpassung durch a_2 zu einem gewissen Grad ebenfalls möglich: Wird die simulierte Herzfrequenz von vorneherein zu hoch angenommen, kann die Modellkurve über einen niedrigen Wert von Parameter a_2 abgeflacht werden. Die Dynamik der empirischen Herzfrequenz lässt sich auf diese Weise jedoch nur eingeschränkt nachbilden. Wird das 3P-Faltungsmodell mit logarithmischem Zusammenhang zwischen a_2 und a_4 verwendet, muss a_2 sehr klein gewählt werden, um auch den exponentiellen Anstieg durch Parameter a_4 auszugleichen und so die simulierte Herzfrequenz in einem passenden Bereich zu halten. Abbildung 6.15 zeigt exemplarisch den Vergleich zwischen (a) einer gefitteten Modellkurve, (b) einer Simulation mit zu hoch angenommenem Parameter a_3 ohne Anpassung weiterer Parameter, (c) einer Simulation mit zu hohem Wert für a_3 mit exemplarisch angepasstem Wert von a_2 und (d) einer Simulation mit zu hohem Wert für a_3 mit exemplarisch angepasstem Wert von a_2 unter Berücksichtigung des logarithmischen Zusammenhangs zu a_4 aus dem 3P-Modell.

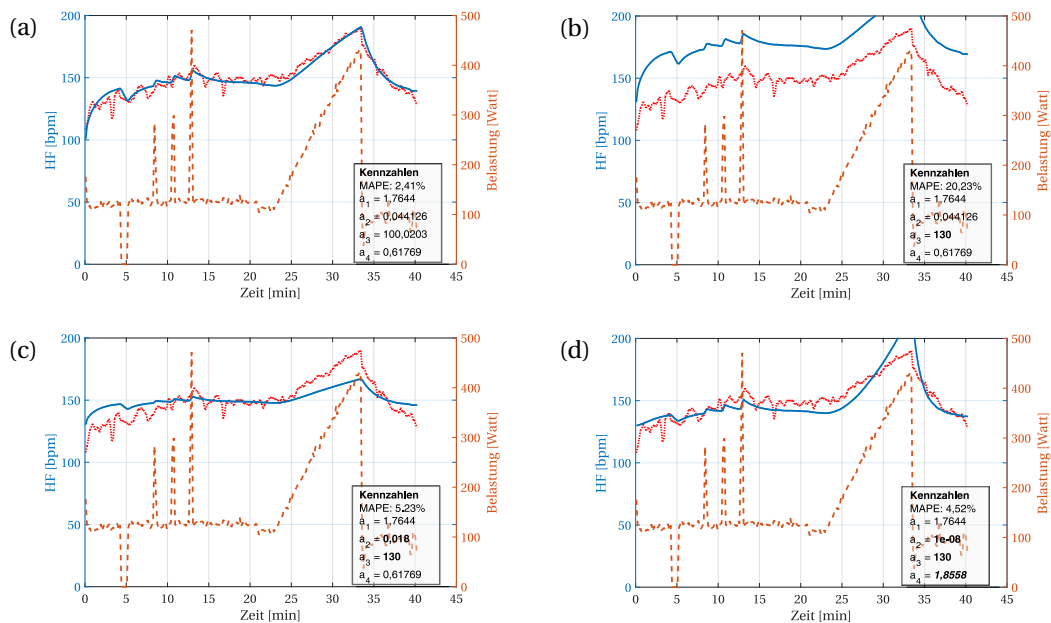


Abbildung 6.15.: Ergebnisse der Modellierung mit dem Faltungsmodell. Die jeweiligen Parameter- und Fehlerwerte sind in der unteren rechten Ecke angegeben, veränderte Parameter sind fett gedruckt. (a) Optimale, kalibrierte Parameter; (b) Variation von a_3 ohne Anpassung von a_2 ; (c) Variation von a_3 mit Anpassung von a_2 ; (d) Variation von a_3 mit Anpassung von a_2 unter Berücksichtigung des logarithmischen Zusammenhangs zu a_4 im 3P-Faltungsmodell

Verwendung bei Outdoor-Trainingseinheiten

Durch die Evaluation auf Outdoor-Trainingseinheiten hat sich gezeigt, dass sowohl das 4P-Faltungsmodell als auch das 1P-Faltungsmodell dazu geeignet ist, auch verrauschte Outdoor-Trainingseinheiten mit im Vergleich zu Ergometer-Trainingseinheiten deutlich län-

gerer Dauer adäquat zu modellieren und vorherzusagen. Dabei führt die Verwendung des 1P-Faltungsmodells lediglich zu geringfügig und nicht signifikant höheren Fehlerwerten als das 4P-Faltungsmodell. Bei beiden Modellvarianten treten sowohl in der Kalibrierungsphase als auch in der Vorhersage einige hohe Fehlerwerte mit MAPE-Werten von deutlich über 30% auf. Derart hohe Fehler konnten in der qualitativen Analyse als Messfehler der empirischen Herzfrequenzwerte oder als fehlende Wattleistungswerte identifiziert werden.

Beim 1P-Faltungsmodell wird der freie Parameter a_3 innerhalb der Kalibrierung tendenziell niedrig gewählt. Dadurch kommt es in manchen Vorhersagen zu einer Unterschätzung der empirischen Herzfrequenzleistung im Verlauf der Trainingseinheit. Im Vergleich zum 4P-Faltungsmodell ist es dem 1P-Faltungsmodell nicht möglich, die Dynamik der Herzfrequenz über die anderen Parameter individuell anzupassen. Da es sich bei Parameter a_3 um einen additiven Parameter handelt, stellt dieser Wert eine untere Schranke für die Herzfrequenz dar, auf den die Simulation bei ausbleibender Belastung zurückfällt. Die Messungen in den vorliegenden Daten starten in der Regel bereits mit einer Wattleistung größer Null, während der erste Herzfrequenzwert der Simulation keine vorherigen Wattleistungsmesswerte zur Berechnung zur Verfügung hat. Zur Berechnung des ersten Herzfrequenzwertes muss daher eine zugrundeliegende Leistung von 0 Watt angenommen werden, sodass das Modell den Parameter a_3 entsprechend niedrig wählt.

Virtuelle Leistungsdiagnostik: Vorhersagesimulation

Innerhalb der Vorhersagesimulationen mit den beiden Modellvarianten des Faltungsmodells hat sich gezeigt, dass der erste Wert der modellierten Herzfrequenz beim Einparameter-Modell deutlich niedriger liegt, als (a) beim Vierparameter-Modell und (b) der empirische Herzfrequenzwert. Eine mögliche Erklärung liegt darin begründet, dass der additive Parameter a_3 als untere Schranke für die Simulation der Herzfrequenz fungiert. Ebendieser Parameter verbleibt in dem hier identifizierten parameterreduzierten Faltungsmodell jedoch zur Individualisierung frei. Entsprechend muss a_3 niedrig genug gewählt werden, sodass die Simulation der Herzfrequenz insgesamt zum Verlauf der zur Kalibrierung genutzten empirischen Daten »passt«. Beim 4P-Faltungsmodell fungiert a_3 zwar ebenfalls als untere Schranke. Die Dynamik der Herzfrequenz kann jedoch durch die übrigen drei Parameter besser nachgebildet werden, ohne a_3 derart niedrig anzusetzen. Dies scheint bei der gewählten Parameterkombination für das 1P-Modell nicht der Fall zu sein.

Insgesamt zeigt sich die Vorhersage der Herzfrequenz von Ergometer-Tests basierend auf einer Kalibrierung der Parameter auf regulären Trainingseinheiten für beide Modellvarianten sehr zuverlässig. Sporadisch aufgetretene hohe Fehlerwerte konnten in der qualitativen Analyse als Messfehler der empirischen Herzfrequenzwerte identifiziert werden.

6.2.3. Fazit zur Studie

In einem mehrschrittigen Verfahren konnte die Anzahl der Parameter des Faltungsmodells sukzessive für das betrachtete Probandenkollektiv reduziert werden. Es wurde eine parameterreduzierte Version des Faltungsmodells mit einem verbleibenden freien Parametern konstruiert, das im Vergleich zum originalen Faltungsmodell mit vier Parametern weder im Fitting noch in der Vorhersage zu deutlich sichtbar größeren Fehlerwerten führt. Die Unterschiede sind dabei für die Vorhersage statistisch nicht signifikant. Dieser Eindruck hat sich auch bei der anschließenden Evaluation auf einer Vielzahl von Trainingsdaten bestätigt, die nicht zur Parameterreduktion verwendet wurden (Evaluation auf Outdoor-Trainingsdaten und Evaluation des ersten Schritts einer virtuellen Leistungsdiagnostik). Für die weitere Analyse im Zusammenhang mit der Leistungsdiagnostik wird das parameterreduzierte Faltungsmodell mit $a_1 = 2$, $a_2 = 0,25$ und $a_4 = 0,493 = 0,08 \cdot (1 - \log(0,46 \cdot a_2)) + 0,24$ verwendet, wobei Parameter a_3 zur Individualisierung und zur weiteren Analyse im Kontext der Leistungsdiagnostik genutzt wird.

6.3. Konstruktion eines einparametrischen Faltungsmodells: Studiendaten von Hobby-Sportlern

Basierend auf Leistungstest-Daten von drei männlichen Hobby-Sportlern wird im Folgenden analysiert, ob es auch auf einer geringen Datenmenge und bei einer Probandengruppe aus Hobbysportlern möglich ist, ein einparametrisches Faltungsmodell zu konstruieren. Die Daten wurden in »Datensatz 3: Daten von Hobby-Sportlern« (Abschnitt 4.3) beschrieben.

Es wird sich zeigen, dass auf Basis der vorliegenden Daten ebenfalls eine plausible Parameterreduktion für das Faltungsmodell durchgeführt werden kann.

6.3.1. Durchführung der Parameterreduktion und Ergebnisse

Die Parameterreduktion für das Faltungsmodell wird nachfolgend schrittweise anhand der zuvor identifizierten und am Ende von Abschnitt 6.1 (»Parameterreduziertes Faltungsmodell«) dargestellten Vorgehensweise durchgeführt.

Zur leichteren Lesbarkeit wird nachfolgend ein n -parametrisches Faltungsmodell mit n -Faltungsmodell abgekürzt.

Schritte 1-2: Kalibrierung und Identifikation der Parameterbereiche

Auf Basis der Kalibrierung der einzelnen Datensätze haben sich die in Tabelle 6.6 dargestellten Bereiche für die Parameter und die entsprechend gewählte Normierung ergeben.

Tabelle 6.6.: Bereiche und Normalisierung der Parameter auf Basis einer vorangehenden Kalibrierung, wobei die Werte der Rastersuche in \hat{a}_i eingesetzt werden und a_i wie gewohnt die Modellparameter sind

Parameter	Bereich	Normalisierung	Umkehrfunktion
a_1	0,5 – 4,5	$\hat{a}_1 = (a_1 - 0,5) \cdot \frac{1}{4}$	$a_1 = (\hat{a}_1 \cdot 4) + 0,5$
a_2	0 – 0,0025	$\hat{a}_2 = (a_2 - 0) \cdot 400$	$a_2 = (\hat{a}_2 \cdot \frac{1}{400}) + 0$
a_3	60 – 110	$\hat{a}_3 = (a_3 - 60) \cdot \frac{1}{50}$	$a_3 = (\hat{a}_3 \cdot 50) + 60$
a_4	0,75 – 1,25	$\hat{a}_4 = (a_4 - 0,75) \cdot 2$	$a_4 = (\hat{a}_4 \cdot 0,5) + 0,75$

Die gewählte Normierung bildet die Basis für die Fitting-Simulationen über die Rastersuche. Dabei werden für Parameter a_1, a_2 Schrittweiten 0,05 und für Parameter a_3, a_4 Schrittweiten von 0,1 gewählt – pro Proband und Trainingseinheit entstehen damit $21 \cdot 21 \cdot 11 \cdot 11 = 53\,361$ Fittingsimulationen. Für alle Datensätze aller Probanden ergeben sich entsprechend rund 750 Tausend Kombinationen $((21 \cdot 21 \cdot 11 \cdot 11) \cdot (4 + 4 + 6) = 747\,054)$ aus Parametern und Fehlerwerten.

Schritte 3-5: Fitting-Simulation über die Rastersuche, visuelle Parameteranalyse (Heatmaps) und sukzessive Überprüfung

Basierend auf den Ergebnissen der Fittingsimulationen mit den Parametern aus der Rastersuche wird für einen ersten Überblick eine Heatmap mit Medianwerten über die einzelnen Datensätze erstellt. Der »median(min_MAPE)« ist für die Einfärbung der Rechtecke innerhalb der Heatmap gewählt. Während die Rastersuche wie zuvor beschrieben auf $[0,1]$ -normierten Parametern durchgeführt wurde, sind die Parameter in der Beschriftung zur besseren Interpretation bereits zurücktransformiert.

Anhand der Heatmaps in Abbildung 6.16 lassen sich mehrere Aspekte erkennen:

1. Für den betrachteten Parameterbereich gibt es hinsichtlich Parameter a_1 keine deutlichen Unterschiede in den Abbildungen (a), (b) und (c). Für jede Kombination aus Parametern a_1, a_2 sowie a_1, a_3 und a_1, a_4 (im betrachteten Wertebereich) gibt es eine Kombination aus den verbleibenden Parameter (a_3, a_4 bzw. a_2, a_4 bzw. a_2, a_3), mit der ein geringer Fehlerwert erzielt werden kann. Vor allem in den Teilgraphiken (a) und (c) gibt es keine deutlichen horizontalen Strukturen, sodass der Wert für a_1 keine große Bedeutung zu haben scheint.
2. In Abbildungsteil (b) zeigen sich leicht dunklere Bereiche (d.h. geringere MAPE-Werte im Median) für $a_3 \in \{70; 75; 80\}$.
3. Abbildungsteil (c) zeigt deutlich, dass für Parameter a_4 eine Einschränkung auf den Bereich zwischen 0,9 und 1,15 sinnvoll ist. Für kleinere oder größere Werte von a_4 kann in dem betrachteten Parameterbereich von a_1 keine Kombination der Parameter a_2, a_3 mit MAPE-Werten unter 12% gefunden werden. Innerhalb dieses Bereiches

nimmt $a_4 = 1,1$ die höchsten MAPE-Werte an, was sich durch die hellere Grauschattierung zeigt.

4. Abbildungsteil (d) zeigt sich eher undeutlich. Hier wird nachfolgend eine detaillierte Betrachtung ohne Median vorgenommen.
5. Abbildungsteil (e) zeigt deutlich einen funktionalen Zusammenhang der Parameter a_2 und a_4 . Dieser Zusammenhang wird für die Parameterreduktion näher untersucht.
6. Abbildungsteil (f) bestätigt den Eindruck aus Teil (c) hinsichtlich eines geeigneten Wertebereichs für a_4 . Im Bereich $a_4 \in \{0,95; 1; 1,05\}$ werden dabei etwas niedrigere MAPE-Werte erzielt als für $a_4 \in \{0,9; 1,1; 1,15\}$. Bezüglich a_3 bestätigt sich ebenfalls der bisherige Eindruck aus Teil (b): Für $a_3 \in \{70; 75; 80\}$ lassen sich Parameterkombinationen a_1, a_2 finden, die im Median über alle Datensätze einen niedrigeren MAPE-Wert annehmen als für andere der betrachteten Werte von a_3 .
7. Es scheint für keine der paarweisen Parameterkombinationen einen eindeutigen linearen Zusammenhang zu geben. Auf die Durchführung einer PCA wird daher verzichtet.

Eine nähere Betrachtung von Abbildung 6.16 (d) für jeden einzelnen Probanden (mit Median-Bildung über die verschiedenen Trainingseinheiten der einzelnen Probanden) ist in Abbildung 6.17 dargestellt. Hier ist die Struktur deutlicher zu erkennen: Die Struktur derjenigen Parameterkombinationen mit niedrigen Fehlerwerten fächern sich für alle Probanden in einer Art Bifurkation auf. Es wird deutlich, dass die beiden Parameter a_2 und a_3 voneinander abhängen, wobei kein direkter linearer Zusammenhang zu bestehen scheint. Durch die Struktur wird gleichzeitig ein großer Bereich der beiden Parameter abgedeckt – unabhängig davon, ob a_2 oder a_3 auf einen spezifischen Wert fixiert wird, lässt sich in den meisten Fällen ein korrespondierender Parameter finden, sodass der Fehlerwert niedrig ist.

Alle einzelnen Heatmaps für jede Trainingseinheit eines jeden Probanden ohne Median-Bildung finden sich im Anhang (Abbildung A.3, Abbildung A.4 und Abbildung A.5)

Der in Abbildung 6.16 (e) beobachtete funktionale Zusammenhang zwischen den Parametern a_2 und a_4 ist logarithmischer Struktur. Um einen besseren Eindruck der Struktur zu erhalten, wird eine neue Rastersuche mit feinerer Auflösung für a_4 durchgeführt. Dabei wird $a_1 \in [1; 3,5]$ mit einer Schrittweite von 0,5 gewählt, a_2 wie zuvor nur ohne $a_2 = 0$, d.h. mit $a_2 \in [0,000125; 0,0025]$ und einer Schrittweite von 0,000125, $a_3 \in [70; 85]$ mit einer Schrittweite von 5 und $a_4 \in [0,85; 1,2]$ mit einer Schrittweite von 0,01. Es ergibt sich eine Rastersuche mit $6 \cdot 21 \cdot 4 \cdot 36 = 18\,144$ Fehlerwerten pro Datensatz, d.h. insgesamt $18\,144 \cdot (4 + 4 + 6) = 308\,448$ Kombinationen aus Parametern und Fehlerwerten. Abbildung 6.18 zeigt die resultierende Heatmap im Median über alle Datensätze für die Parameterkombination a_2, a_4 .

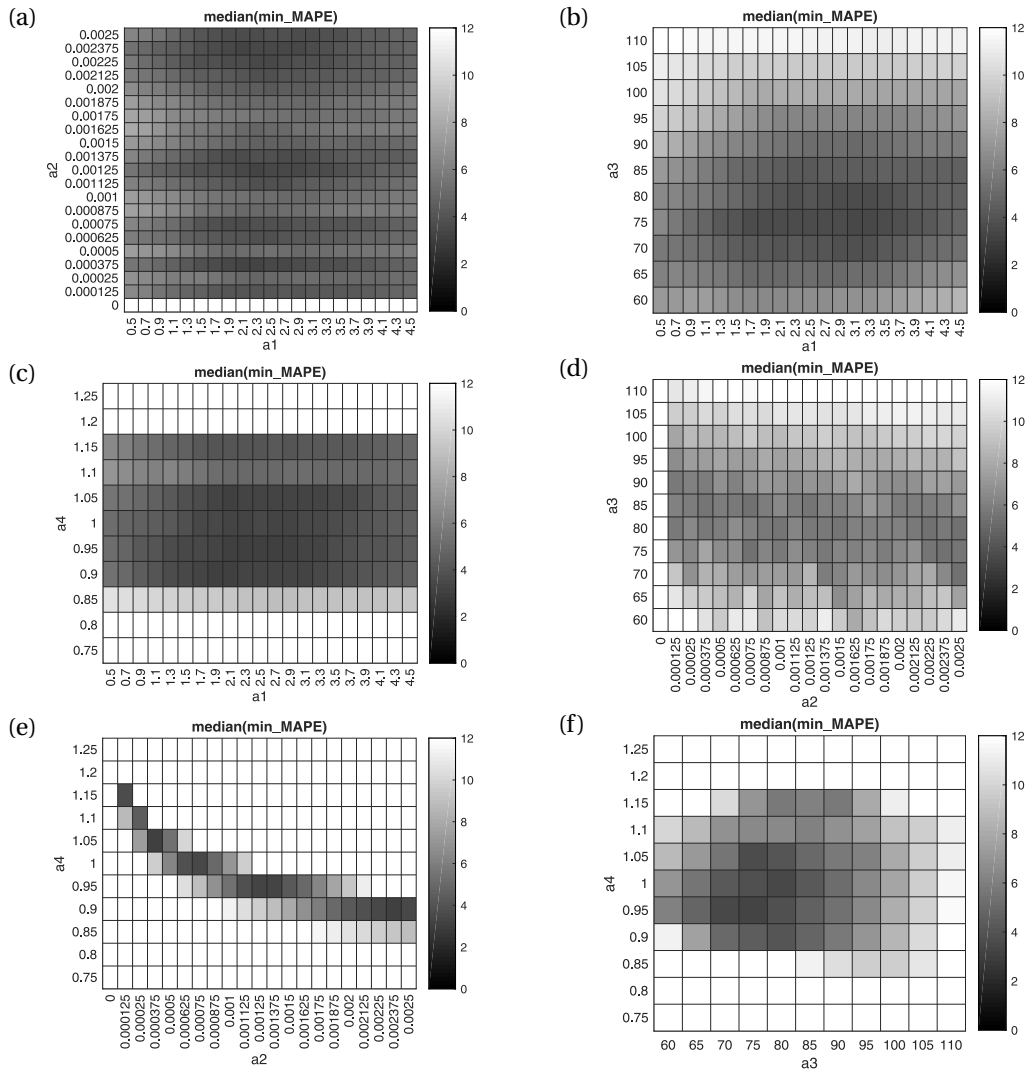


Abbildung 6.16.: Heatmap-Darstellung für die Kombination aus je zwei Parametern. Für jedes Wertepaar aus der betrachteten Parameterkombination wird der Median der minimalen MAPE-Werte für alle Trainingseinheiten aller Probanden über die Einfärbung betrachtet

Zur Bestimmung eines geeigneten funktionalen Zusammenhangs der Parameter a_2 und a_4 wird ein automatisiertes Curvefitting in MATLAB über die Funktion *fit* durchgeführt. Da die Struktur logarithmisch (skaliert und verschoben) und ähnlich zu der Struktur aus Abschnitt 6.2: »Konstruktion eines einparametrischen Faltungsmodells: Studiendaten von Profisportlern« aussieht, wird dafür erneut die Struktur $p_1 \cdot (1 - \log(p_2 \cdot a_2)) + p_3$ vorgegeben (mit $a_2 \neq 0$ und $p_2 \neq 0$). Falls in den zugrunde liegenden Parameterkombinationen mehrere Fehlerwerte für einen Wert von a_2 (mit unterschiedlichen Werten von a_4) vorliegen, wird diejenige Parameterkombination mit dem kleinsten Fehlerwert gewählt. Es ergibt sich die Funktion $a_4 = 0,0843 \cdot (1 - \log(14,7236 \cdot a_2)) + 0,5353$ (für $a_2 \neq 0$).

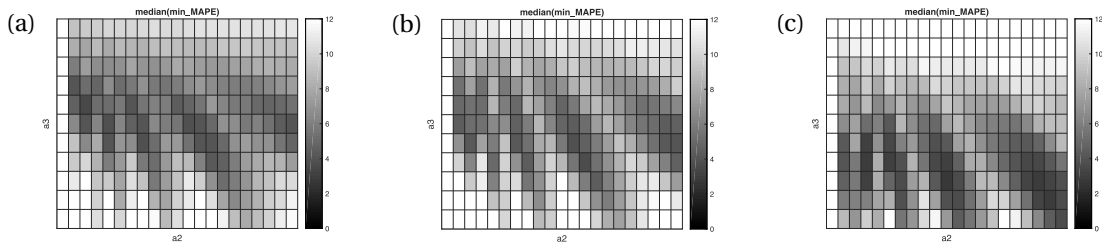


Abbildung 6.17.: Heatmaps der drei Probanden für die Parameterkombination a_2, a_3

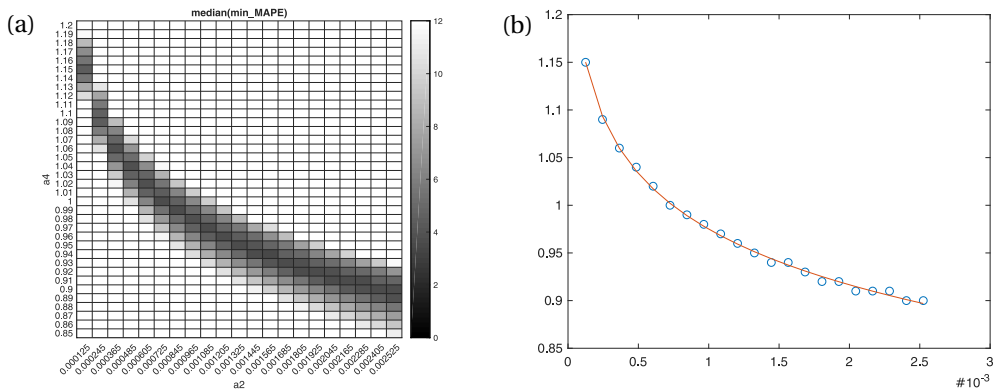


Abbildung 6.18.: Heatmaps im Median über alle Datensätze für die Parameterkombination a_2, a_4 und Curvefitting

Bevor die Anzahl der Parameter weiter reduziert wird, wird überprüft, ob das Faltungsmodell mit drei Parametern und der logarithmischen Funktion für a_4 im Fitting und in der Vorhersage vergleichbare Ergebnisse zum 4P-Faltungsmodell liefert. Die Fittingsimulation findet auf jeden Datensatz einzeln statt. Die Simulation der Vorhersage findet vergangenheitsbasiert basierend auf stets mindestens zwei Kalibrierungsdatensätzen statt: Für die vier Trainingseinheiten von Proband 1 bedeutet das, dass drei Vorhersageexperimente durchgeführt werden können. Zunächst wird das jeweilige Modell auf den ersten beiden Trainingseinheiten kalibriert und die Herzfrequenz der (zeitlich) nachfolgenden zwei Trainings wird vorhergesagt. Anschließend werden die ersten drei Trainingseinheiten zur Kalibrierung verwendet und das verbleibende letzte Training wird vorhergesagt. Die Durchführung für die anderen beiden Probanden verläuft analog – für Proband 3 mit sechs Trainingseinheiten resultieren entsprechend $4 + 3 + 2 + 1 = 10$ Vorhersagesimulationen. Die Violinenplots der MAPE-Fehlerwerte für die beiden Modellvarianten sind in Abbildung 6.19 dargestellt, wobei (a) die Ergebnisse der Fittingsimulation und (b) die Ergebnisse der vergangenheitsbasierten Prädiktion darstellt. Während die prozentualen Fehlerwerte des 3P-Modells in der Fittingsimulation minimal höher sind als beim originalen Faltungsmodell, sind sie in den Vorhersageexperimenten sogar im Median leicht besser. Auch der maximal erzielte Fehler ist mit dem 3P-Faltungsmodell etwas geringer als beim 4P-Faltungsmodell. Dies ist ein Hinweis für eine leichte Überanpassung durch das 4P-Modell. Die Nullhypothese, dass der Mittelwert der MAPE-Fehlerwerte des 4P-Modells und des 3P-Modells gleich sind, kann zum

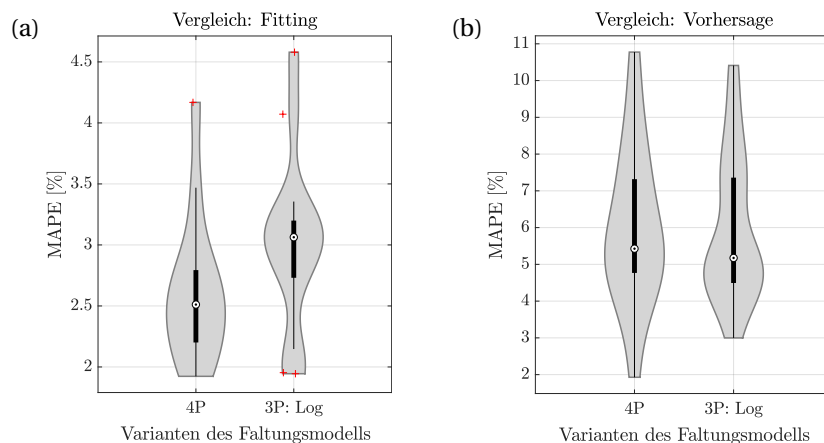


Abbildung 6.19.: Violinenplot: Vergleich Faltungsmodell 4 Parameter und 3 Parameter für (a) Fitting und (b) Vorhersage

Niveau $\alpha = 5\%$ mit dem Welch-Test sowohl im Fitting ($p = 0,14$) als auch in der Vorhersage ($p = 0,93$) nicht verworfen werden.

Da bei der Vorhersage keine Verschlechterung durch den funktionalen Zusammenhang der Parameter a_2 und a_4 aufgetreten ist, wird der logarithmische Zusammenhang beibehalten. Eine weitere offensichtliche Struktur zwischen anderen Parameterkombinationen konnte wie bereits beschrieben im visuellen Parametervergleich nicht detektiert werden; es ergab sich lediglich der Eindruck, dass die konkrete Wahl von a_1 im betrachteten Parameterbereich keinen großen Einfluss hat. Dieser Parameter erscheint daher als einziger freier Parameter ungeeignet.

Im nächsten Schritt wird daher sowohl eine manuelle Fixierung der Kombination der Parameter a_1 und a_2 wie auch von a_1 und a_3 betrachtet. Für einen der Parameter lassen sich Kenntnisse über die Struktur der Daten und physiologische Grundlagen integrieren, sodass dieser zuerst betrachtet wird: Parameter a_3 entspricht phänomenologisch der Start-Herzfrequenz vor Beginn der Trainingseinheit. In der bereits untersuchten visuellen Parameteranalyse erschienen Werte im Bereich $a_3 \in \{70, 75, 80\}$ grundsätzlich eher geeignet zu sein. Daher werden nachfolgend die Ergebnisse von vier weiteren Varianten eines 2P-Faltungsmodells ergänzt, bei denen einmal nacheinander $a_3 = 70, a_3 = 75, a_3 = 80$ und im vierten Fall für a_3 der erste gemessene Herzfrequenzwert gewählt wird. Die Violinenplots der MAPE-Fehlerwerte für die vier Modellvarianten sind in Abbildung 6.20 dargestellt. Bei der Wahl von a_3 als erster gemessener Herzfrequenzwert stellt sich für das 2P-Faltungsmodell nur ein geringer Genauigkeitsverlust im Fitting und keinerlei Genauigkeitsverlust in der Vorhersage im Vergleich zum 4P- oder 3P-Faltungsmodell ein.

Die Nullhypothese, dass der Mittelwert der MAPE-Fehlerwerte des 3P-Modells und dieses 2P-Modells gleich sind, kann zum Niveau $\alpha = 5\%$ mit dem Welch-Test sowohl im Fitting ($p = 0,06$) als auch in der Vorhersage ($p = 0,18$) nicht verworfen werden.

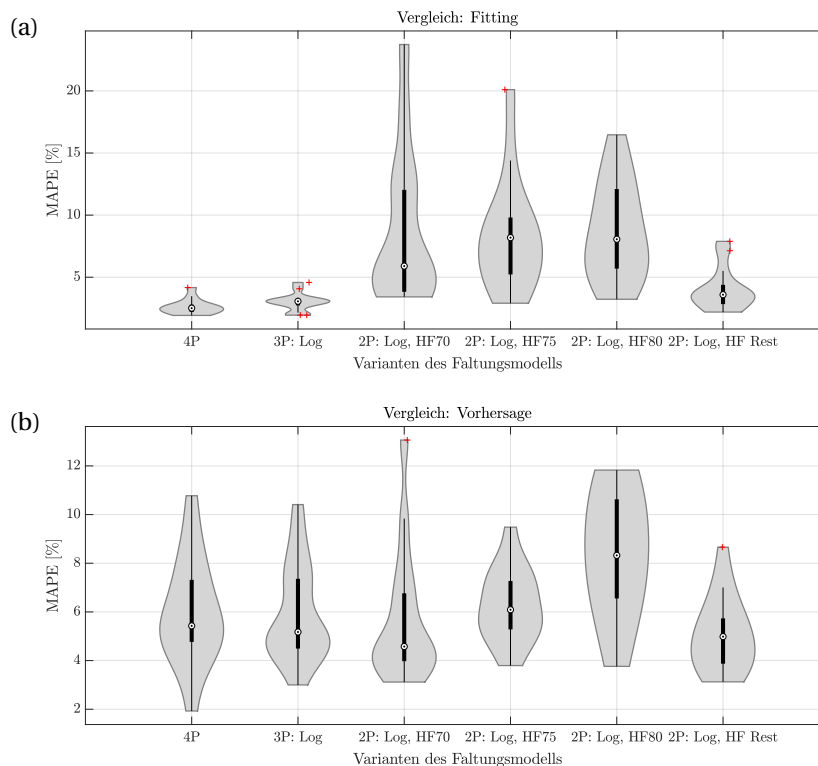


Abbildung 6.20.: Violinenplot: Vergleich mehrerer Varianten des Faltungsmodells mit unterschiedlicher Parameteranzahl und Parametereinstellung für (a) Fitting und (b) Vorhersage

Innerhalb der visuellen Parameteranalyse hat sich wie bereits dargelegt gezeigt, dass Parameter a_1 im betrachteten Bereich keinen großen Einfluss zu haben scheint. Für jeden Wert von a_1 wurde eine geeignete Kombination der restlichen Parameter gefunden, mit dem geringe MAPE-Werte erzielt werden. Ähnlich wie zur Fixierung von a_3 werden innerhalb des geeigneten Bereichs mehrere Werte für a_1 testweise festgesetzt und die Ergebnisse für Fitting und vergangenheitsbasierte Vorhersage miteinander und mit dem bereits identifizierten 2P-Faltungsmodell verglichen. Exemplarisch werden die Werte $a_1 \in \{1; 2; 2,5; 3; 3,5; 4\}$ betrachtet. Die Ergebnisse sind als Violinenplots der MAPE-Fehlerwerte in Abbildung 6.21 dargestellt. Auch nach Fixierung von a_3 bestätigt sich durch die einzelnen Fitting- und Vorhersageexperimente der Eindruck aus der visuellen Parameteranalyse: a_1 hat innerhalb des gewählten Parameterbereiches für die betrachteten Datensätze keinen großen Einfluss. Die Nullhypothese für den paarweisen Vergleich der einzelnen konstanten Werte für $a_1 \in \{2; 2,5; 3; 3,5; 4\}$ zeigt, dass die Nullhypothese (Gleichheit der jeweiligen Mittelwerte der MAPE-Fehlerwerte) mit dem Welch-Test zum 5%-Niveau weder im Fitting ($p \geq 0,06$) noch in der Vorhersage ($p \geq 0,10$) verworfen werden kann. Lediglich für $a_1 = 1$ zeigen sich außer im Vergleich mit $a_1 = 4$ signifikante Unterschiede in der Vorhersage ($p \leq 0,027$) – der Violinenplot zeigt, dass die Ergebnisse für $a_1 = 1$ damit in der Vorhersage signifikant schlechter sind, als für Werte von $a_1 \in \{2; 2,5; 3; 3,5\}$.

6. Verbesserung des Faltungmodells: Parameterreduktion

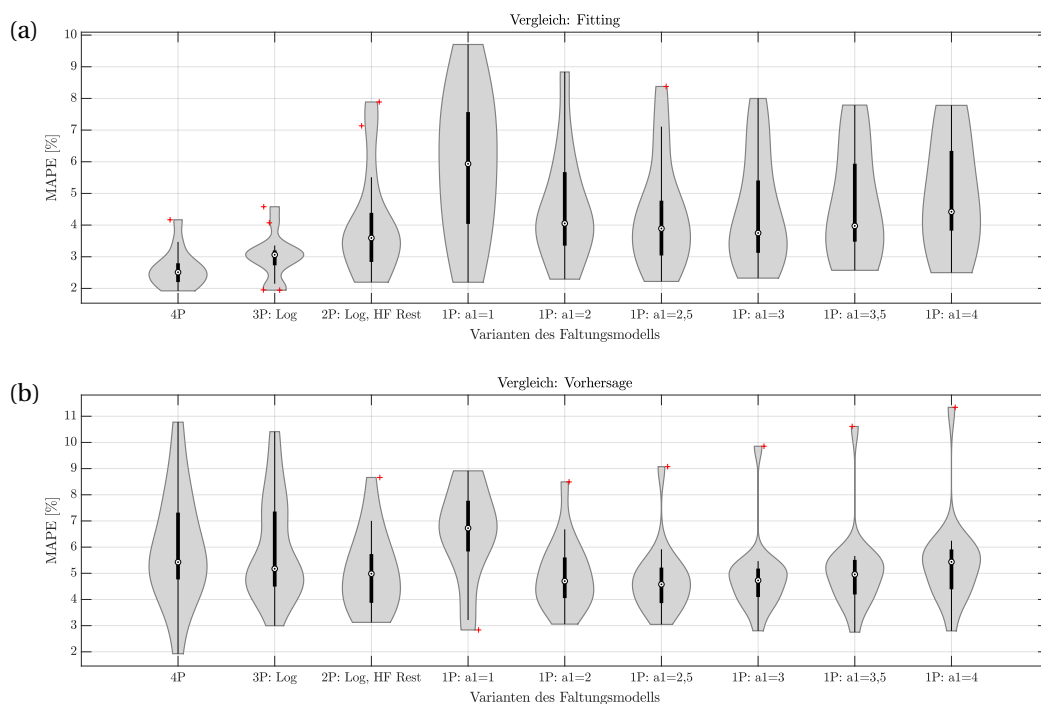


Abbildung 6.21.: Violinenplot: Vergleich mehrerer Varianten des Faltungmodells mit unterschiedlicher Parameteranzahl und Parametereinstellung

Da $a_1 = 2$ minimal bessere Ergebnisse innerhalb der Vorhersage liefert, wird der Parameter nachfolgend auf diesen Wert festgesetzt. Eine andere Wahl ist jedoch ebenso möglich, was in den nachfolgenden Analysen insbesondere in Abschnitt 7.2 (»Analyse zur Abschätzung der Leistungsveränderung auf Basis von Herzfrequenz-Simulationen: Hobby-Sportler«) und bei der Interpretation berücksichtigt werden muss. Parameter a_2 verbleibt zur Individualisierung frei.

Die Nullhypothese, dass der Mittelwert der MAPE-Fehlerwerte des 2P-Modells und dieses 1P-Modells mit $a_1 = 2$ gleich sind, kann zum Niveau $\alpha = 5\%$ mit dem Welch-Test sowohl im Fitting ($p = 0,46$) als auch in der Vorhersage ($p = 0,81$) nicht verworfen werden.

Im Vergleich des 1P-Modells mit dem 4P-Modell kann die Nullhypothese auf gleiche Mittelwerte im Fitting mit $p = 0,0019$ verworfen werden. In der Vorhersage sind mit $p = 0,108$ keine signifikanten Unterschiede mit dem Welch-Test identifizierbar. Wie bereits zu Beginn erläutert, spielt die Vorhersagegenauigkeit eine wichtige Rolle, wohingegen es im Fitting insbesondere mit einer höheren Anzahl an freien Parametern leicht zu einer Überanpassung an die Kalibrierungsdaten kommen kann. Das 1P-Faltungmodell weist in der Vorhersage keine signifikante Verschlechterung zum 4P-Faltungmodell auf und kann als parameterreduzierte Variante verwendet werden. Die zugehörigen QQ-Plots für den Vergleich des 4P-Modells mit dem 1P-Modell sind im Anhang in Abbildung A.6 (S. 255) dargestellt.

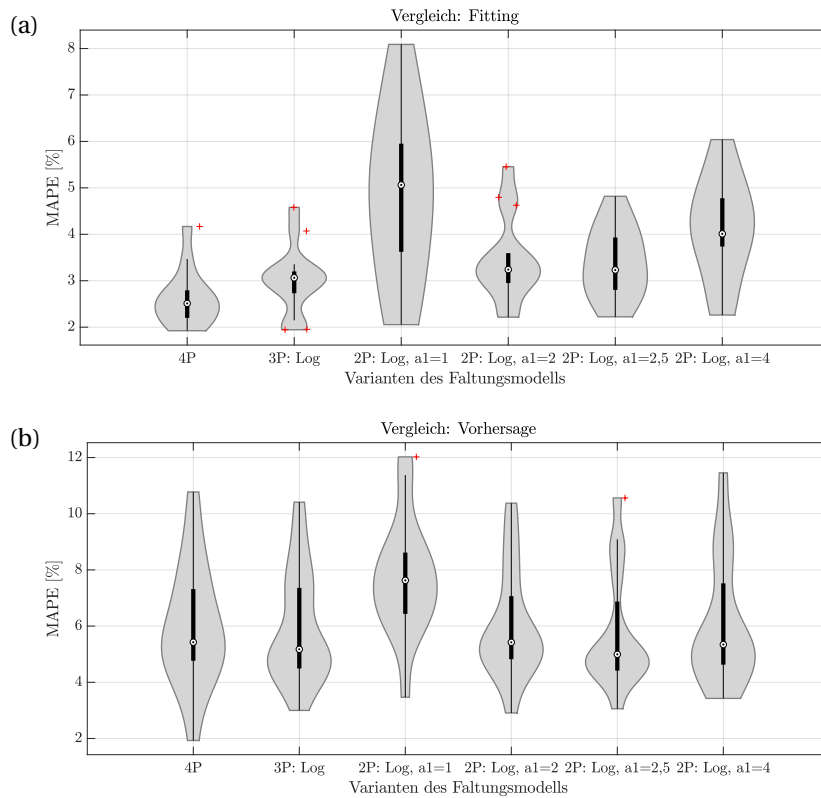


Abbildung 6.22.: Violinenplot: Vergleich mehrerer Varianten des Faltungsmodells mit unterschiedlicher Parameteranzahl und Parametereinstellung für (a) Fitting und (b) Vorhersage

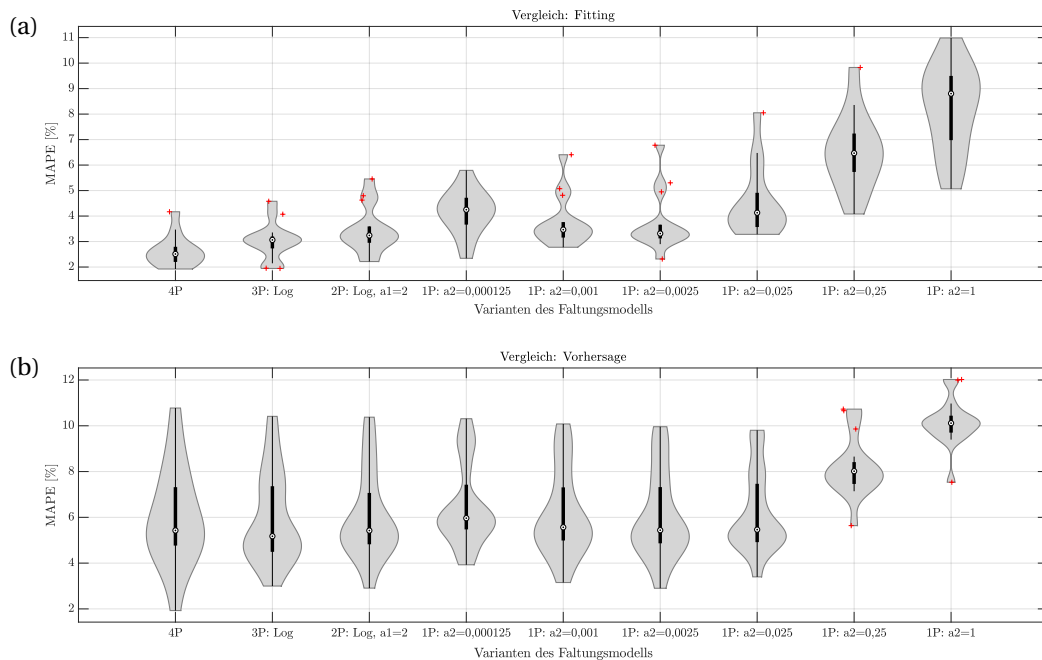


Abbildung 6.23.: Violinenplot: Vergleich mehrerer Varianten des Faltungsmodells mit unterschiedlicher Parameteranzahl und Parametereinstellung für (a) Fitting und (b) Vorhersage

Wenngleich damit bereits eine 1P-Variante des Faltungsmodells identifiziert wurde, soll auch die Möglichkeit eines 1P-Faltungsmodells mit a_3 als freiem Parameter in Betracht gezogen werden: Da es weder für Parameter a_1 noch für Parameter a_2 eine phänomenologische Interpretation gibt, die die Wahl spezifischer Werte für einen der Parameter begründen würde, werden zunächst für a_1 ähnlich wie zuvor exemplarisch die Werte $a_1 \in \{1; 2; 2,5; 3; 3,5; 4\}$ betrachtet. Die Ergebnisse sind als Violinenplots der MAPE-Fehlerwerte in Abbildung 6.22 dargestellt. Die Ergebnisse für alle betrachteten Werte von $a_1 \neq 1$ sind sowohl im Fitting als auch in der Vorhersage sehr ähnlich und unterscheiden sich nicht signifikant voneinander. Im paarweisen Vergleich der einzelnen konstanten Werte für a_1 liegt lediglich für $a_1 = 1$ eine signifikante Abweichung zu den meisten anderen konstanten Werten vor (p-Wert im Fitting und in der Vorhersage bei maximal $p \leq 0,04$; keine signifikante Abweichung zu $a_1 = 4$ im Fitting mit $p = 0,25$). Aus dem Violinenplot wird dabei deutlich, dass die Fehlerwerte für $a_1 = 1$ im Median höher sind als für die anderen konstanten Werte von a_1 . Für die restlichen paarweisen Vergleiche liegt der p-Wert im Fitting und in der Vorhersage bei mindestens $p \geq 0,08$, sodass die Nullhypothese nicht verworfen werden kann. Nachfolgend wird zur besseren Vergleichbarkeit mit dem zuvor identifizierten 1P-Modell erneut $a_1 = 2$ gewählt. Dabei ist zu beachten, dass eine andere Wahl von a_1 ebenso plausibel möglich ist.

In der bereits untersuchten visuellen Parameteranalyse erschienen alle Werte im Bereich $0,000125 \leq a_2 \leq 0,0025$ ähnlich gut geeignet. Höhere Werte wurden dabei nicht betrachtet. Es werden daher exemplarisch drei Werte aus diesem Wertebereich, sowie drei deutlich größere Werte für a_2 betrachtet: $a_2 \in \{0,000125; 0,001; 0,0025; 0,025; 0,25; 1\}$. Die Violinenplots der MAPE-Fehlerwerte für die sechs Modellvarianten sind – zum Vergleich gemeinsam mit den zuvor identifizierten Reduktionen und dem 4P-Modell – in Abbildung 6.23 dargestellt. Die Ergebnisse der Parameter $a_2 \in \{0,000125; 0,001; 0,0025; 0,025\}$ unterscheiden sich nicht signifikant voneinander ($p \geq 0,1$ im Fitting und $p \geq 0,63$ in der Vorhersage); die betrachteten höheren Werte für a_2 hingegen führen zu signifikant höheren Fehlerwerten sowohl im Fitting ($p \leq 0,0095$) als auch in der Vorhersage ($p \leq 0,00006$). Die niedrigsten Werte im Fitting und in der Vorhersage werden für $a_2 = 0,0025$ erzielt, sodass der Parameter auf diesen Wert fixiert wird. Erneut ist hierbei zu beachten, dass eine andere Wahl für a_2 ebenso möglich und plausibel ist, was in den nachfolgenden Analysen insbesondere in Abschnitt 7.2 (»Analyse zur Abschätzung der Leistungsveränderung auf Basis von Herzfrequenz-Simulationen: Hobby-Sportler«) und bei der Interpretation berücksichtigt werden muss. Parameter a_3 verbleibt zur Individualisierung frei.

Im Vergleich des 1P-Modells mit dem 4P-Modell kann die Nullhypothese auf gleiche Mittelwerte im Fitting mit $p = 0,0051$ verworfen werden. In der Vorhersage sind mit $p = 0,998$ keine signifikanten Unterschiede mit dem Welch-Test identifizierbar. Wie bereits zu Beginn und im Kontext der zuvor identifizierten 1P-Faltungsmodellvariante erläutert, spielt vor allem die Vorhersagegenauigkeit eine zentrale Rolle, während es im Fitting insbesondere mit einer höheren Anzahl an freien Parametern leicht zu einer Überanpassung an die Kalibrie-

rungsdaten kommen kann. Das 1P-Faltungsmodell weist in der Vorhersage keine signifikante Verschlechterung zum 4P-Faltungsmodell auf und kann demnach ebenfalls als parameterreduzierte Variante verwendet werden. Die zugehörigen QQ-Plots für den Vergleich des 4P-Modells mit dem 1P-Modell sind im Anhang in Abbildung A.6 (S. 255) dargestellt.

Es können demnach unterschiedliche Varianten eines parameterreduzierten Faltungsmodells mit einem freien Parameter identifiziert werden, das im Vergleich zum originalen Faltungsmodell mit vier Parametern weder im Fitting noch in der Vorhersage zu signifikant größeren Fehlerwerten führt. Für die weitere Analyse im Zusammenhang mit der Leistungsdiagnostik werden zwei Versionen des parameterreduzierten Faltungsmodells betrachtet:

1. Parameter a_2 verbleibt zur Individualisierung und zur weiteren Analyse im Kontext der Leistungsdiagnostik frei. Es werden die folgenden Parameter verwendet: $a_1 = 2$, $a_3 =$ »erster gemessener Herzfrequenzwert des Trainings« und $a_4 = 0,0843 \cdot (1 - \log(14,7236 \cdot a_2)) + 0,5353$ (für $a_2 \neq 0$);
2. Parameter a_3 verbleibt zur Individualisierung und zur weiteren Analyse im Kontext der Leistungsdiagnostik frei. Es werden die folgenden Parameter verwendet: $a_1 = 2$, $a_2 = 0,0025$ und $a_4 = 0,0843 \cdot (1 - \log(14,7236 \cdot a_2)) + 0,5353 = 0,8980$.

6.3.2. Interpretation und Diskussion

Bereits beim Vergleich des 4P-Faltungsmodells mit dem 3P-Faltungsmodell hat sich gezeigt, dass innerhalb der Vorhersage durch die Parameterreduktion um einen freien Parameter ein geringerer Fehler im Median wie auch maximal erzielt werden konnte, während die Fehler im Fitting für das 4P-Modell geringer sind. Dies deutet darauf hin, dass es durch die Verwendung des 4P-Faltungsmodells zu einer geringfügigen Überanpassung der Modellparameter an die Kalibrierungsdaten kommt. Es kann nicht ausgeschlossen werden, dass eine solche Überanpassung auch beim 3P-Modell stattfindet. Da jedoch die Fehler in der Vorhersage kleiner sind als beim 4P-Modell, ist davon auszugehen, dass der Einfluss einer möglichen Überanpassung zumindest im Vergleich geringer ausfällt. Durch die weitere Parameterreduktion zeigt sich, dass die Qualität der Vorhersage – je nach spezifischer Wahl der Werte für die fixierten Parameter – noch weiter verbessert werden kann. Dabei zeigt sich stets, dass bei der Verwendung eines Modells mit weniger freien Parametern die Fehler im Fitting leicht ansteigen, in der Vorhersage jedoch geringer werden. Dies legt den Schluss nahe, dass das Faltungsmodell durch die Parameterreduktion an Stabilität gewinnt und eine bestehende Überanpassung an die Kalibrierungsdaten reduziert werden kann.

Ähnlich wie in der vorangegangenen empirischen Studie in Abschnitt 6.2: »Konstruktion eines einparametrischen Faltungsmodells: Studiendaten von Profi-Sportlern« scheint Parameter a_1 in dem betrachteten Wertebereich keine große Bedeutung einzunehmen. Für die meisten untersuchten Werte von a_1 konnten sowohl im Fitting als auch in der Vorhersage

mit Ausnahme von $a_1 = 1$ und bis auf leicht variable Extremwerte vergleichbare Fehlerwerte berechnet werden.

Bei der Analyse eines möglichen Zusammenhangs zwischen den Parametern a_2 und a_3 hat sich bei der Betrachtung derjenigen Parameterkombinationen mit niedrigen MAPE-Werten andeutungsweise die Struktur einer Bifurkation ergeben. Diese kommt durch die Funktionsweise dieser beiden Parameter zustande: Parameter a_3 verschiebt als additiver Wert die simulierte Herzfrequenzkurve in der Höhe, während Parameter a_2 die Steigung der Herzfrequenz als Reaktion auf die Belastung modelliert. Wird a_3 – ausgehend von einem der empirischen Herzfrequenz angepassten Wert – leicht verringert, kann die dadurch »zu niedrige« Simulation durch eine leichte Verstärkung der Steigung (d.h. einen etwas größeren Wert von a_2) selbst ohne Anpassung der übrigen Parameter weitgehend ausgeglichen werden. Sinkt der Wert von a_3 hingegen zu deutlich ab, führt eine stärkere Steigung durch einen größeren Wert von a_2 nur für einzelne Zeitfenster des Trainings dazu, dass die Simulation auf ein der empirischen Herzfrequenz entsprechendes »Level« angehoben werden kann. Über den gesamten Zeitverlauf betrachtet führt ein zu steiler Anstieg jedoch insgesamt zu einer großen Abweichung zwischen der simulierten und der empirischen Herzfrequenz, sodass der Fehlerwert steigt. Dieser Zusammenhang zwischen a_2 und a_3 ist in Abbildung 6.24 exemplarisch für vier unterschiedliche Parameterpaare von a_2 und a_3 illustriert. Die Möglichkeit, insbesondere zu niedrige Werte von a_3 durch etwas höhere Werte von a_2 hinsichtlich der Fehlerwerte annähernd auszugleichen, führt in der Heatmap zu den sichtbaren, diagonal nach unten verlaufenden »dunkleren Steifen« mit vergleichsweise niedrigeren Fehlerwerten. Durch Modulationen der anderen Parameter a_1, a_4 entstehen schließlich mehrere dieser »Streifen«, da mit anderen Werten von a_1 und a_4 mehr unterschiedliche Werte von a_2 möglich sind.

Insgesamt konnten zwei unterschiedliche Varianten eines 1P-Faltungsmodells identifiziert werden. Ähnlich wie bereits in der vorangegangenen Studie (Abschnitt 7.2) hat sich dabei gezeigt, dass für Parameter a_3 ausschließlich die Wahl des ersten Wertes aus der jeweiligen Trainingseinheit in Frage kommt. Die Fixierung von a_3 auf einem konstanten Wert führt in der Kalibrierung zu signifikant höheren Fehlerwerten von über 20%. In der Vorhersage sind die Unterschiede hingegen weniger stark und insbesondere nicht signifikant ausgeprägt. Als allgemeinen Tendenz zeigt sich dabei, dass innerhalb der Prädiktion ein eher niedriger Wert oder der erste Wert der Trainingseinheit besser geeignet ist als die betrachteten höheren Werte für a_3 . Dies legt erneut den Schluss nahe, dass der additive Parameter a_3 als Basis-Höhe der Simulation eine zentrale Rolle für die individuelle Modellierung der Herzfrequenzreaktion spielt.

Dieser Eindruck lässt sich durch die zweite 1P-Modellvariante mit a_3 als freiem Parameter verstärken – auch in diesem Fall werden für hinreichend kleine Werte von $a_2 \leq 0,025$ vergleichbare Fehlerwerte ohne signifikante Unterschiede erzielt.

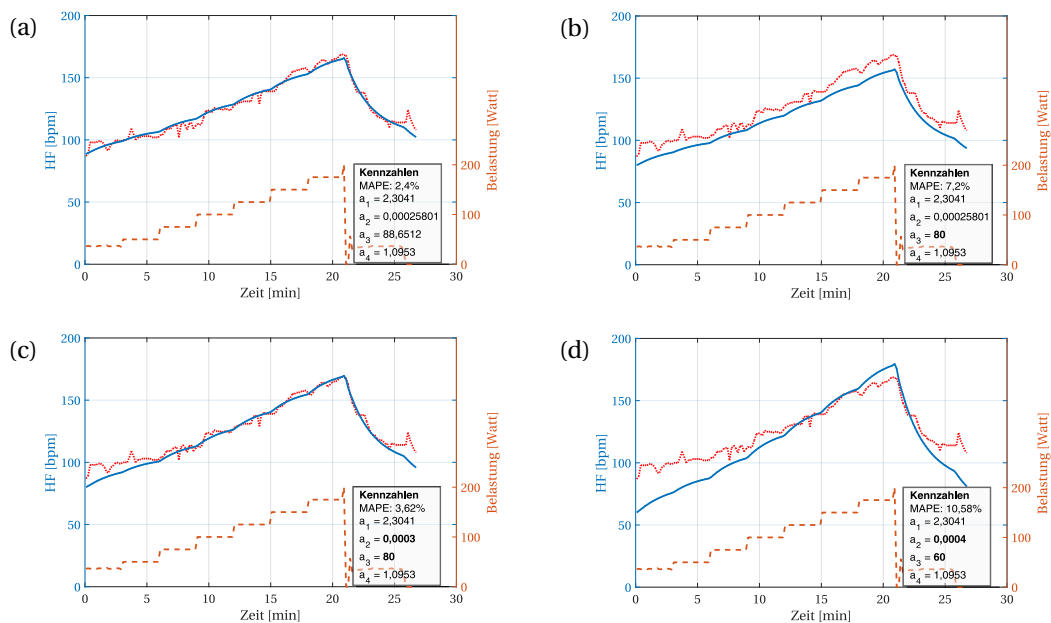


Abbildung 6.24.: Ergebnisse der Modellierung mit dem Faltungsmodell. Die jeweiligen Parameter- und Fehlerwerte sind in der unteren rechten Ecke angegeben, veränderte Parameter sind fett gedruckt. (a) Optimale, kalibrierte Parameter; (b) Variation von a_3 ohne Anpassung von a_2 ; (c) Variation von a_3 mit Anpassung von a_2 ; (d) Starke Variation von a_3 mit Anpassung von a_2 , wobei ein höherer Fehler bestehen bleibt

6.3.3. Fazit zur Studie

Für die Daten der drei Hobbysportler konnte trotz der geringen Datenmenge ein adäquates Faltungsmodell mit nur einem freien Parameter identifiziert werden. In einem mehrschrittigen Verfahren konnte die Anzahl der Parameter des Faltungsmodells sukzessive für das betrachtete Probandenkollektiv reduziert werden. Für das Faltungsmodell konnten zwei unterschiedliche 1P-Varianten identifiziert werden: Ein Modell, in dem der multiplikative Parameter a_2 zur Individualisierung frei bleibt, und eine Modellvariante mit freiem additiven Parameter a_3 . Beide Modellvarianten weisen im Vergleich zum originalen Faltungsmodell mit vier Parametern weder im Fitting noch in der Vorhersage deutlich sichtbar größere Fehlerwerte auf. Die Unterschiede sind dabei für die Vorhersage statistisch nicht signifikant. Für die weitere Analyse im Zusammenhang mit der Leistungsdiagnostik werden beide parameterreduzierten 1P-Faltungsmodell-Varianten verwendet, d.h. sowohl das Modell mit $a_1 = 2$, $a_3 =$ »erster gemessener Herzfrequenzwert des Trainings« und $a_4 = 0,0843 \cdot (1 - \log(14,7236 \cdot a_2)) + 0,5353$ (für $a_2 \neq 0$), als auch die Modellvariante mit $a_1 = 2$, $a_2 = 0,0025$ und $a_4 = 0,0843 \cdot (1 - \log(14,7236 \cdot a_2)) + 0,5353 = 0,8980$. Dabei wird in der ersten Variante wird Parameter a_2 zur Individualisierung und zur weiteren Analyse im Kontext der Leistungsdiagnostik verwendet und in der zweiten Variante Parameter a_3 .

Es konnte damit gezeigt werden, dass es auch für ergometerbasierte Herzfrequenzdaten von Hobbysportlern möglich ist, die Anzahl der Parameter des Faltungsmodells geeignet

zu reduzieren und stabile Ergebnisse für die Kalibrierung und die Vorhersage mit einem Einparameter-Faltungsmodell zu generieren.

6.4. Zusammenfassung und Fazit

Ziel dieses Kapitels war es, die zweite Fragestellung zu untersuchen:

Fragestellung 2:

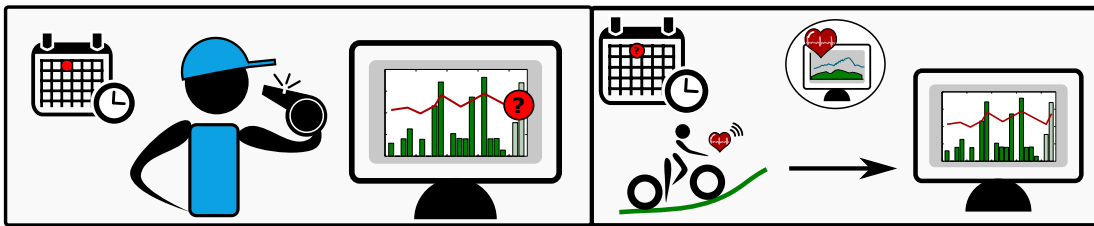
Kann die Anzahl der Parameter des Faltungsmodells ohne signifikanten Verlust der Vorhersagegenauigkeit reduziert werden?

Zur Analyse der Fragestellung wurde das zuvor in Kapitel 5: »Das Faltungsmodell: Ein neuer Ansatz zur Simulation und Vorhersage der Herzfrequenz im Training« entwickelte und evaluierte Faltungsmodell weiter untersucht. Dabei sollte eine Modellvariante des Faltungsmodells entwickelt werden, die letztendlich die Möglichkeit bietet, basierend auf den Modellparametern Rückschlüsse über die Leistungsfähigkeit einer Person ziehen zu können. Zu diesem Zweck wurde eine iterative Vorgehensweise präsentiert, mit der die Anzahl der Modellparameter sukzessive und individuell für ein spezifisches Probandenkollektiv reduziert werden kann. Eine derartige Parameterreduktion wurde anschließend in zwei unterschiedlichen empirischen Studien durchgeführt. Dabei wurden die einzelnen Schritte der Parameterreduktion stets im Vergleich zum ursprünglichen Faltungsmodell analysiert und evaluiert.

In beiden empirischen Studien konnte schließlich (mindestens) eine Variante für ein Faltungsmodell mit nur noch einem freien Parameter identifiziert werden, das im Vergleich zum 4P-Faltungsmodell mit keinem signifikanten Genauigkeitsverlust innerhalb der Vorhersage einhergeht. In der zweiten Studie konnte die Vorhersagequalität durch die Parameterreduktion sogar verbessert werden, wobei diese Verbesserung statistisch nicht signifikant war.

Die durchgeführten Experimente haben damit gezeigt, dass auch Fragestellung 2 positiv beantwortet werden kann: Die Anzahl der Parameter des Faltungsmodells lässt sich ohne signifikanten Genauigkeitsverlust reduzieren. Die identifizierten 1P-Faltungsmodellvarianten können daher in der nachfolgenden empirischen Arbeit (Kapitel 7) verwendet werden, um sie hinsichtlich eines Zusammenhangs zur Leistungsentwicklung als einen neuen Anwendungsfall zu analysieren.

Nutzen des parameterreduzierten Faltungsmodells bei der Leistungsbeurteilung



Abstract.

Die Simulation und Vorhersage der Leistungsentwicklung über mehrere Tage oder Wochen erfordert in der Regel Kalibrierungsdaten einer vorhergehenden spezifischen Leistungsdiagnostik. Um eine Alternative zu den üblicherweise verwendeten Ausbelastungstests und Laktatdiagnostiken zu eruieren, die keinerlei Wechselwirkungen mit den regulären Trainingseinheiten aufweisen, wird das Faltungsmodell auf seine Nutzbarkeit für diesen Anwendungsfall hin untersucht. Dazu wird analysiert, ob der spezifische Parameter des Faltungsmodells geeignet mit den Leistungsveränderungen korreliert werden kann und ob sich ein dem Leistungsverlauf entsprechender Trend im zeitlichen Verlauf des Parameters abzeichnet.

Die Auswertungen zweier Studien zeigen, dass sich Leistungsentwicklung mit dem Parameter des zuvor identifizierten einparametrischen Faltungsmodell unter gewissen Fällen nachbilden lässt. Fragestellung 3 lässt sich eingeschränkt bestätigen: Der aus der Herzfrequenzmodellierung stammende Parameter des parameterreduzierten Faltungsmodells lässt basierend auf alltäglichem Training grundsätzlich Rückschlüsse auf die Leistungsentwicklung zu.

Die Modellierung der Herzfrequenz wie auch der Leistungsfähigkeit einer Sportlerin oder eines Sportlers können bereits einzeln betrachtet zu einem Erkenntnisgewinn bezüglich des Trainingsfortschritts oder spezifischen trainingswissenschaftlichen Fragestellungen beitragen. Wie in Kapitel 3 zur »Zusammenfassung des Forschungsstands und Fragestellungen« erläutert, ließe sich die Modellierung der Trainingsleistungsentwicklung insbesondere auch im Hobby- und Breitensport durch den Verzicht auf aufwändige Leistungsdiagnostik vereinfachen.

Selbst bei reduzierter Genauigkeit der modellierten Leistungsfähigkeit einer Person könnten so Hobbysportlerinnen und Hobbysportler dabei unterstützt werden, ihre persönlichen Tendenzen im Training und damit ihre Trainingsplanung auch ohne tiefgehende trainingswissenschaftliche Kenntnisse messbarer zu gestalten. Im Leistungssport ließe sich die Trainingsentwicklung tagesaktuell ohne Leistungsdiagnostik und damit ohne Intervention in das reguläre Trainingsprogramm verfolgen.

Die Herzfrequenz bietet trotz ihrer in »Die Herzfrequenz als Indikator: Kritische Aspekte« (Unterabschnitt 1.3.1) diskutierten Schwächen ein leicht messbares und daher gut nutzbares Maß für derartige Zwecke. Wie in Kapitel 5 über die »Das Faltungsmodell: Ein neuer Ansatz zur Simulation und Vorhersage der Herzfrequenz im Training« gezeigt, hat das Faltungsmodell auf Grund der geringen Fehler bei Vorhersagen der Herzfrequenz das Potenzial einer vielfältigen Nutzung – insbesondere auch in Hinblick auf mögliche Rückschlüsse individueller Leistungsentwicklungen. Daher widmet sich dieses Kapitel der dritten und somit letzten Fragestellung:

Fragestellung 3:

Trägt der aus der Herzfrequenzmodellierung stammenden Parameter des parameterreduzierten Faltungsmodells basierend auf alltäglichem Training zur Gewinnung von Rückschlüssen auf die Leistungsentwicklung einer Athletin oder eines Athleten im Ausdauersport bei?

Zur Analyse von Fragestellung 3 wird im Folgenden basierend auf der parameterreduzierten Variante eine Korrelationsanalyse zwischen diesem Parameter im Zeitverlauf und der empirischen Leistung verschiedener Probanden durchgeführt. Dazu werden zunächst in einer »Analyse zur Abschätzung der Leistungsveränderung auf Basis von Herzfrequenz-Simulationen: Profi-Sportler« die Daten von vier männlichen Profi-Athleten betrachtet. Dabei werden unterschiedliche Methoden zur Verknüpfung der Modellparameter des Faltungsmodells mit den empirischen Leistungsdaten untersucht. Auch das Konzept einer »virtuellen Leistungsdiagnostik« wird präsentiert und deren Ergebnisse innerhalb der Studie diskutiert. Ergänzend wird eine vergleichbare Analyse auf den Herzfrequenz- und Leistungsdaten von drei männlichen Hobby-Athleten in Abschnitt 7.2 innerhalb der »Analyse zur Abschätzung der Leistungsveränderung auf Basis von Herzfrequenz-Simulationen: Hobby-

Sportler« durchgeführt, sodass sowohl eine grundlegend andere Probandengruppe als auch andere empirische Leistungswerte betrachtet werden können.

Insgesamt wird damit in den nachfolgenden Kapiteln analysiert,

1. ob der verbleibende Parameter des parameterreduzierten Faltungsmodells bei Kalibrierung auf Ergometer-Trainingseinheiten einen direkten Rückschluss auf die empirische Leistungsentwicklung zulässt;
2. ob der verbleibende Parameter des parameterreduzierten Faltungsmodells bei Kalibrierung auf regulären Outdoor-Trainingseinheiten im Rahmen einer virtuellen Leistungsdiagnostik direkten Rückschluss auf die empirische Leistungsentwicklung zulässt;
3. ob der Trendverlauf des verbleibenden Parameter des parameterreduzierten Faltungsmodells in einer der beiden Varianten Rückschlüsse auf den Trendverlauf der empirischen Leistungsentwicklung zulässt.

Anhand der Ergebnisse beider Studien wird abschließend Fragestellung 3 diskutiert und beurteilt (Abschnitt 7.3: »Zusammenfassung und Fazit«).

7.1. Analyse zur Abschätzung der Leistungsveränderung auf Basis von Herzfrequenz-Simulationen: Profi-Sportler

Basierend auf Leistungstest-Daten von vier männlichen professionellen Radsportlern wird im Folgenden analysiert, ob der verbleibende Parameter des einparametrischen Faltungsmodells direkt, im Rahmen einer virtuellen Leistungsdiagnostik oder in einer der beiden Varianten hinsichtlich des Trendverlaufs mit der empirisch ermittelten Leistungsfähigkeit der Probanden zusammenhängt. Dazu wird die zuvor identifizierte parameterreduzierte Variante des Faltungsmodells aus Abschnitt 6.2 (»Konstruktion eines einparametrischen Faltungsmodells: Studiendaten von Profi-Sportlern«) verwendet. Die Daten wurden in »Datensatz 2: Daten aus einer Studie von Schrot mit Profi-Athleten« (Abschnitt 4.2) beschrieben.

Es wird sich zeigen, dass schon beim direkten Vergleich die Tendenz eines Zusammenhangs zwischen dem verbleibenden Parameter und den Leistungswerten erkennbar ist, die basierend auf einer Ausbelastung ermittelt wurden. Beim Vergleich des Parameters aus einer virtuellen Leistungsdiagnostik zeigt sich ein grundsätzlich sehr ähnlicher Verlauf, dessen Korrelation zur empirischen Leistung insbesondere durch die Betrachtung des Trends deutlich wird.

7.1.1. Experimenteller Aufbau und Methoden

Zur Analyse der Problemstellung, über die modellgestützte Analyse der Herzfrequenz mögliche Rückschlüsse auf die Leistungsfähigkeit eine Person im Radsport zu ziehen, wird nachfolgend der zuvor in Abschnitt 6.2 identifizierte verbleibende freie Parameter des Faltungsmodells genauer analysiert.

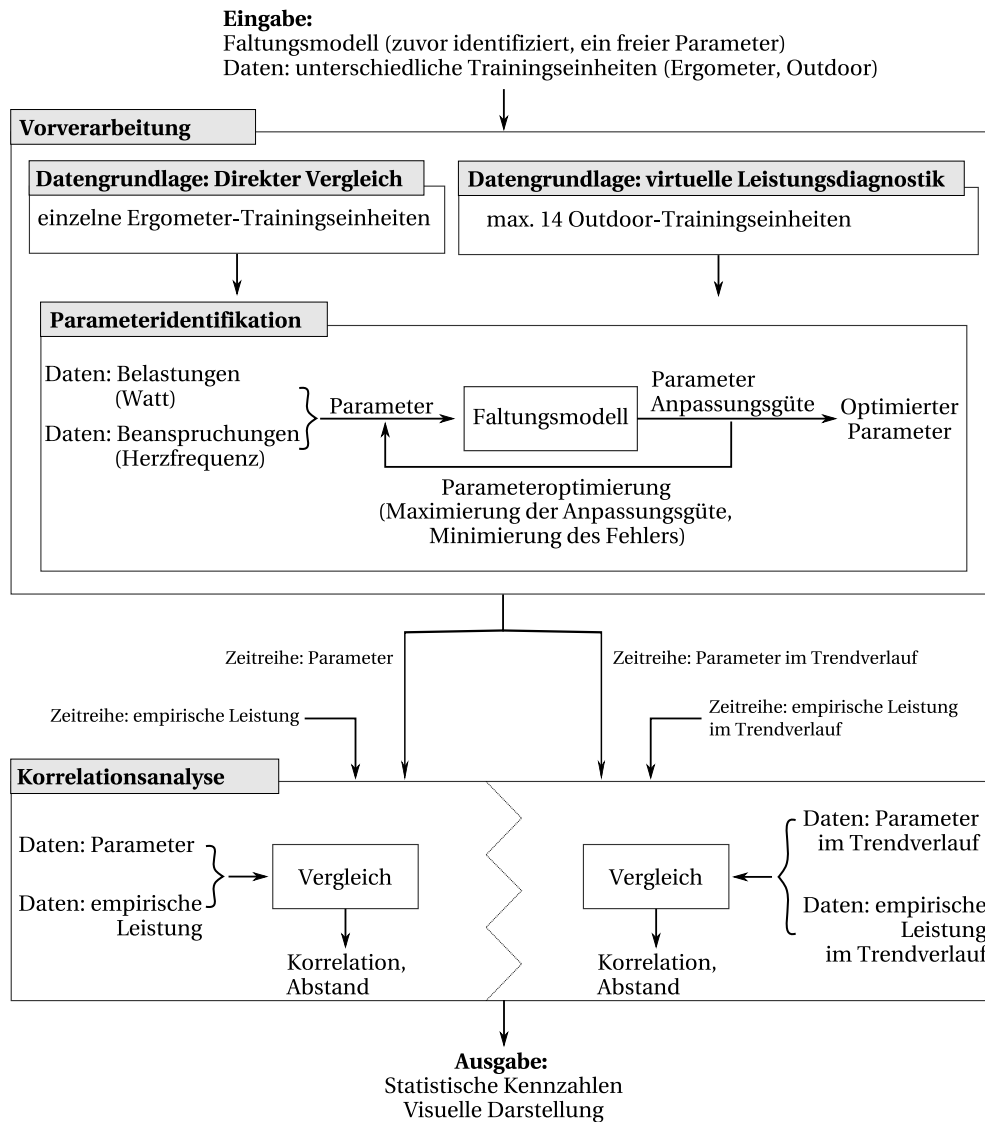


Abbildung 7.1.: Vorgehensweise zur Korrelationsanalyse zwischen der Leistung und dem Parametern der HF-Modellierung bei Profi-Athleten

Dazu werden zwei unterschiedliche Ansätze mit je zwei Variationen betrachtet: Ein direkter Vergleich zwischen der empirischen Leistung im zeitlichen Verlauf und dem Modellparameter aus einer Kalibrierung auf Ergometer-Trainingseinheiten sowie mit dem Modellparameter aus einer virtuellen Leistungsdiagnostik, d.h. basierend auf der Kalibrierung auf regulären Outdoor-Trainingseinheiten und einer Simulation der Ergometer-Trainingseinheit.

Für beide Verfahren werden die Daten zunächst direkt analysiert. In einem zweiten Schritt wird der Trend der betrachteten Zeitreihen extrahiert und der Trendverlauf der empirischen Daten mit dem Trendverlauf des Parameters verglichen. Diese vier Ansätze werden nachfolgend im Detail beschrieben und erläutert.

Für die Beschreibung und Interpretation der Ergebnisse wird dabei die folgende Notation verwendet:

1. Ansatz 1a: Direkter Vergleich des Parameters mit der empirischen Leistung über die Zeit
2. Ansatz 1b: Trendanalyse der beiden Zeitreihen (direkter Vergleich und empirische Leistung) durch Vergleich der jeweiligen Trends zueinander
3. Ansatz 2a: Vergleich des Parameters aus der virtuellen Leistungsdiagnostik mit der empirischen Leistung über die Zeit
4. Ansatz 2b: Trendanalyse der beiden Zeitreihen (virtuellen Leistungsdiagnostik und empirische Leistung) durch Vergleich der jeweiligen Trends zueinander

Abbildung 7.1 veranschaulicht den Ablauf des durchgeführten Prozesses: Die vorliegenden HF-Daten werden zur Kalibrierung des parameterreduzierten Faltungsmodells genutzt. Dabei wird für Ansatz 1 jeder Datensatz einer Ergometer-Trainingseinheit einzeln und unabhängig von anderen Trainingseinheiten betrachtet. Für Ansatz 2 werden die HF-Daten aus maximal 14 Outdoor-Trainingseinheiten zur Modellkalibrierung verwendet und der kalibrierte Modellparameter wird der zeitlich nachfolgenden Ergometer-Trainingseinheit zugeordnet. Auf diese Weise kann zu jeder Ergometer-Trainingseinheit ein Parameter identifiziert und im späteren Verlauf der Analyse mit der empirischen Leistung verglichen werden.

Parallel dazu werden Leistungswerte der Probanden aus der empirischen Leistungsdiagnostik betrachtet.

Der zuvor identifizierte freie Parameter des einparametrischen Faltungsmodells wird anschließend in Hinblick auf seine Korrelation zur Leistungsentwicklung hin analysiert. Ergänzend zur Analyse der beiden Zeitreihen werden die Trendverläufe über die Zeit ohne Tageschwankungen betrachtet (Variante b).

Vorverarbeitung der Daten für die Analyse

Die empirischen Leistungswerte aus der Ergometrie (kJ/kg) befinden sich in einem Wertebereich von 1,4–5,92 kJ/kg, wobei die Schwankungsbreite pro Proband bei maximal 0,86 kJ/kg liegt. Der freie Parameter des Modells entspricht phänomenologisch einem Wert nahe der Herzfrequenz zu Beginn der Trainingseinheit. Der Wertebereich des Parameters befindet sich daher bei den vorliegenden Daten zwischen 60 bpm und 150 bpm. Werden die empirischen Leistungswerte und die Werte des Modellparameters über die Zeit aufgetragen,

entstehen zwei Zeitreihen, die miteinander verglichen werden können. Um einen solchen Vergleich zu ermöglichen, werden die Daten zuvor auf das Intervall $[0,1; 1,1]$ ¹ linear normalisiert.

Wie in »Kritische Betrachtung statistischer Parameter« im Abschnitt über »Korrelationskoeffizienten« (Unterabschnitt 2.1.3) beschrieben, stellt die Analyse der Leistungsentwicklung für die Trendbereinigung eine Besonderheit dar. Da es hierbei vorrangig um den Trend der Entwicklung geht und Tagesschwankungen nicht von Interesse sind, wird für die beiden nachfolgenden Ansätze explizit *keine* Trendbereinigung durchgeführt.

Unabhängig vom betrachteten Ansatz ist es zur Berechnung der statistischen Kennzahlen notwendig, für beide miteinander verglichenen Zeitreihen die gleiche Anzahl an Datenpunkten zu betrachten. Nach der Vorverarbeitung beider Zeitreihen werden daher als Start- und Enddatum der jeweils erste und letzte Tag ausgewählt, für den sowohl ein empirischer Leistungswert als auch ein Modellparameter existiert. Für vereinzelte Tage der Leistungsdiagnostik liegen keine Protokolle mit Herzfrequenzaufzeichnung vor, sodass für diese keine Simulation mit dem Faltungsmodell durchgeführt werden kann. Umgekehrt enthalten die untersuchten Trainingsdaten Aufzeichnungen von Ergometrie-Trainingseinheiten, bei denen keine Leistungsdiagnostik stattgefunden hat, sodass für diese Tage kein empirischer Leistungswert vorliegt. Um eine größere Anzahl an Messwerten zur statistischen Analyse zur Verfügung zu haben, findet für die Tage dazwischen eine lineare Interpolation derjenigen Zeitreihe statt, für die weniger empirische Werte vorliegen. Für die Betrachtung eines langfristigen Trendverlaufs führt dieses Vorgehen zu keiner maßgeblichen Verzerrung der Daten – für eine Analyse tagesaktueller Schwankungen und der Betrachtung von trendbereinigten Daten wäre dieses Vorgehen hingegen nicht geeignet. Da für alle Probanden mehr empirische Leistungswerte als Aufzeichnungen von Ergometrie-Trainingseinheiten vorliegen, wird ausschließlich die Zeitreihe des freien Parameters aus dem Faltungsmodell interpoliert.

Die Zeitreihen – bestehend aus allen Tagen, für die ein empirischer Leistungswert im betrachteten Zeitintervall aus dem gewählten Start- und Enddatum vorhanden ist – werden anschließend hinsichtlich ihrer Korrelation zueinander (entsprechend der betrachteten Methodik) untersucht. Der Verlauf und eine mögliche Korrelation werden dabei sowohl anhand unterschiedlicher statistischer Kennzahlen wie auch über eine visuelle Darstellung der Zeitreihen analysiert.

Ansatz 1: Direkter Vergleich

Für den direkten Vergleich zwischen der anhand der durchgeführten Leistungsdiagnostik ermittelten Leistung und dem freien Parameter des Einparameter-Faltungsmodells werden die Daten nur geringfügig wie im vorherigen Abschnitt beschrieben vorverarbeitet.

¹Das Intervall $[0;1]$ ist in diesem Fall ungeeignet, da bei der Berechnung der statistischen Parameter der Wert 0 zu Problemen führen würde.

Dabei wird das zuvor identifizierte einparametrische Faltungsmodell auf den Ergometer-Trainingseinheiten kalibriert und der dort identifizierte freie Parameter über die Zeit direkt mit der empirischen Leistung verglichen.

Ansatz 2: Virtuelle Leistungsdiagnostik

Die »Virtuelle Leistungsdiagnostik« bezeichnet wie bereits in Abschnitt 6.2 (»Konstruktion eines einparametrischen Faltungsmodells: Studiendaten von Profi-Sportlern«) dargestellt eine Vorhersage von ergometriebasierten Testeinheiten auf Basis einer Modellkalibrierung auf vorhergehenden »klassischen« Trainingseinheiten (als »Outdoor«-Feld-Training oder auch als Trainingseinheit auf der Rolle).

Dazu werden für jedes vorliegende Ergometer-Training die vorangegangenen Trainingseinheiten betrachtet, die maximal 14 Tage zurückliegen. Der Parameter des einparametrischen Faltungsmodells wird auf diesen maximal 14 vergangenen Trainingseinheiten kalibriert und anschließend im zeitlichen Verlauf mit den empirischen Leistungswerten verglichen. Dass die Simulation der nachfolgenden Ergometer-Trainingseinheit valide Ergebnisse liefert, wurde bereits in Abschnitt 6.2 dargestellt, sodass der verbleibende Parameter plausibel erscheint und verwendet werden kann.

Wie im direkten Vergleich werden beide Zeitreihen normalisiert und nicht trendbereinigt.

Variante b: Trendanalyse

Als Variation für beide Ansätze wird eine reine Trendanalyse durchgeführt. Diese dient dazu, mögliche Tagesschwankungen in den Daten zu ignorieren und den Fokus auf die zugrundeliegende, längerfristige Leistungsentwicklung zu legen. Zu diesem Zweck wird eine Trendbereinigung der Daten berechnet und die Trendbereinigung wird aus den Daten entfernt:

Für eine Zeitreihe X sei X_{detrend} die linear trendbereinigte Zeitreihe (vgl. »Kritische Betrachtung statistischer Parameter«, Unterabschnitt 2.1.3, im Abschnitt »Korrelationskoeffizienten«). Dann lässt sich der lineare Trend beschreiben als $X_{\text{trend}} = X - X_{\text{detrend}}$.

Die Trendanalyse wird sowohl für den Parameter aus dem direkten Vergleich als auch für den Parameter aus der virtuellen Leistungsdiagnostik betrachtet.

7.1.2. Ergebnisse

Zur Analyse der Fragestellung, ob es möglich ist, über die modellgestützte Analyse der Herzfrequenz Rückschlüsse auf die Leistungsfähigkeit eine Person im Radsport zu ziehen, werden nachfolgend die Ergebnisse der vier unterschiedlichen Ansätze dargestellt. Die Beschreibung der Ergebnisse findet in den jeweiligen Abschnitten und gemeinsam mit der vi-

suellen Darstellung statt und werden anschließend nocheinmal zusammenfassend tabellarisch dargestellt.

Ansatz 1a: Direkter Vergleich

Für den direkten Vergleich wird das einparametrische Faltungsmodell verwendet und auf jedem Ergometrie-Datensatz einzeln kalibriert. Die Zeitreihen des kalibrierten Modellparameters und der empirischen Leistung werden wie in Abschnitt »Experimenteller Aufbau und Methoden« beschrieben normalisiert und an geeigneten Stellen zur Berechnung der statistischen Kennzahlen interpoliert.

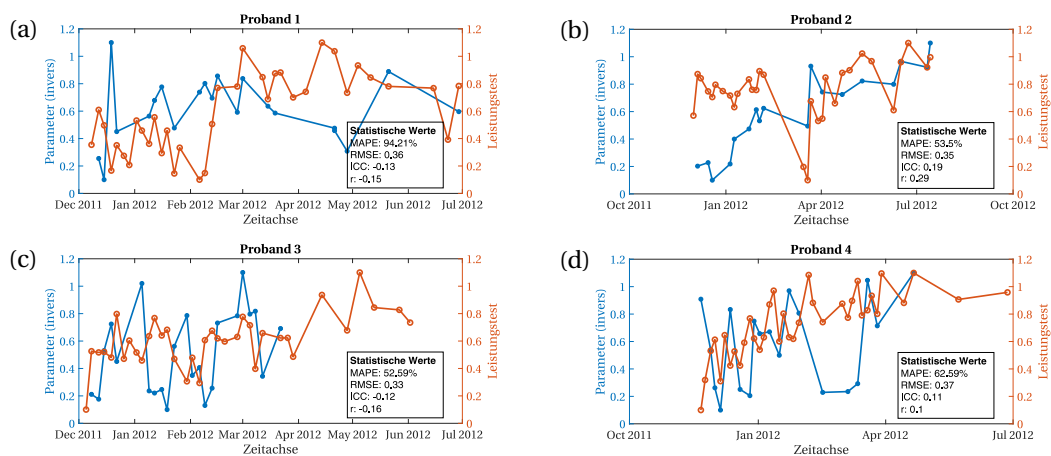


Abbildung 7.2.: Direkter Vergleich zwischen dem Parameter aus dem Faltungsmodell und der empirischen Leistung für die vier Probanden nach Normierung

Der Verlauf beider Zeitreihen ist für jeden der vier Probanden in Abbildung 7.2 dargestellt. Eine Korrelation ist für keinen der Probanden feststellbar. Sowohl der Korrelationskoeffizient r als auch der Intra-Korrelationskoeffizient ICC sind für alle vier Probanden sehr klein ($|r| < 0,3$ und $|ICC| < 0,2$). Für die Probanden 1 und 3 sind zudem beide Korrelationskoeffizienten negativ, während sie für die Probanden 2 und 4 positiv ausfallen. Der RMSE liegt für alle Probanden nahe 0,35 und der prozentuale mittlere Fehler (MAPE) schwankt zwischen Werten von über 50% bis hin zu knapp 95%. Visuell wird deutlich, dass die Werte beider Zeitreihen stark schwanken. Ein Trend ist ohne weitere Analyse weder für den Verlauf des Parameters noch für die Leistungsentwicklung erkennbar.

Ansatz 1b: Trendanalyse für den direkten Vergleich

Für die Trendanalyse der Werte aus dem direkten Vergleich wird das einparametrische Faltungsmodell wie zuvor auf jedem Ergometrie-Datensatz einzeln kalibriert. Die Zeitreihen des kalibrierten Modellparameters und der empirischen Leistung sind identisch mit denen aus Ansatz 1a. Zu Betrachtung der Trendverläufe beider Zeitreihen ist der Trend aus diesen

wie in Unterabschnitt 7.1.1 (»Experimenteller Aufbau und Methoden«) beschrieben durch Subtraktion der jeweiligen trendbereinigten Zeitreihe extrahiert.

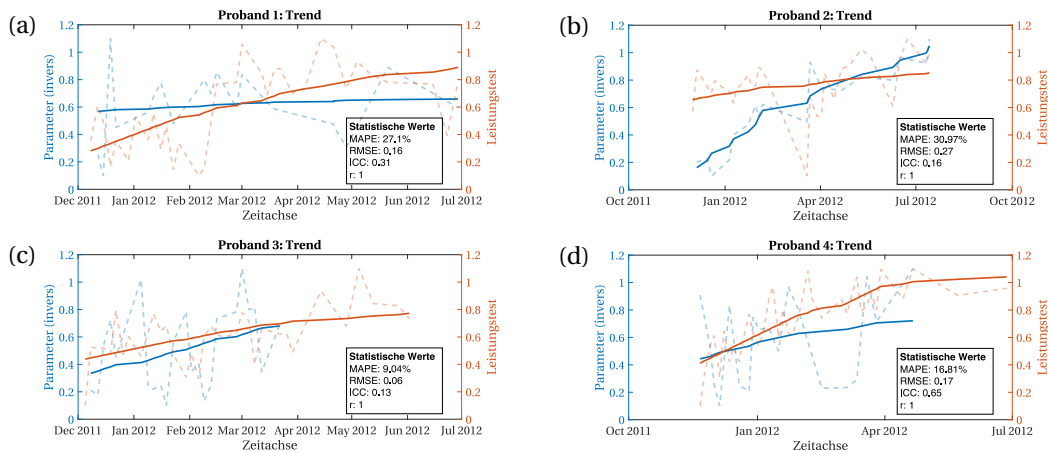


Abbildung 7.3.: Trendanalyse des Parameters aus dem direkten Vergleich zwischen dem Parameter aus dem Faltungsmodell und der empirischen Leistung für die vier Probanden nach Normierung auf $[0,1; 1,1]$

Der Trendverlauf beider Zeitreihen ist für jeden der vier Probanden in Abbildung 7.3 dargestellt. Die ursprünglichen Zeitreihen wie aus Ansatz 1a sind zum besseren Vergleich zusätzlich schwach im Hintergrund und mit gestrichelter Linie dargestellt. Die Abweichung der beiden Kurven zueinander fällt bei der Betrachtung der Trends gering aus: Während der MAPE einen maximalen prozentualen Fehler von knapp 31% (Proband 2) aufweist und für Proband 3 bei unter 10% liegt, weist auch der RMSE mit Werten von 0,09 bis maximal 0,27 auf vergleichsweise geringe Abweichungen hin. Der Intra-Korrelationskoeffizient ICC ist für alle Probanden positiv, schwankt jedoch je nach Proband zwischen 0,13 und 0,65. Damit liegen alle Werte des ICC unter 0,7, sodass nicht von einer guten Übereinstimmung der beiden Zeitreihen ausgegangen werden kann. Der Korrelationskoeffizient r beträgt für alle Probanden exakt 1, was für eine »perfekte Korrelation« spricht, deren Bedeutung in Abschnitt zur Interpretation näher erläutert wird. Visuell zeigt sich für die Probanden 3 und 4 am ehesten ein ähnlicher Verlauf der beiden aufsteigenden Trends. Dabei zeichnet sich ab, dass sich die Trendkurven mit Parameterwerten an weiteren Tagen stärker von der Trendkurve der empirischen Messwerte entfernen werden, sofern der aktuell berechnete Trend sich dadurch nicht verändert. Die Steigung der Trendverläufe des Parameters und der empirischen Leistung unterscheidet sich bei allen Probanden visuell deutlich voneinander.

Ansatz 2a: Virtuelle Leistungsdiagnostik

Für die virtuelle Leistungsdiagnostik (Ansatz 2) wird die Herzfrequenz aus jedem Ergometrie-Datensatz prognostiziert. Dabei findet die vorangehende Kalibrierung wie

in Abschnitt »Experimenteller Aufbau und Methoden« beschrieben auf den Outdoor-Trainingseinheiten der vergangenen 14 Tage und damit auf maximal 14 Datensätzen statt.

In der virtuellen Leistungsdiagnostik entstammt der Modellparameter dem Fitting auf den regulären Trainingseinheiten zur Vorhersage des Ergometertests am jeweils aufgetragenen Tag. Die Zeitreihen des kalibrierten Modellparameters und der empirischen Leistung werden wie in Abschnitt »Experimenteller Aufbau und Methoden« beschrieben normalisiert und an geeigneten Stellen zur Berechnung der statistischen Kennzahlen interpoliert.

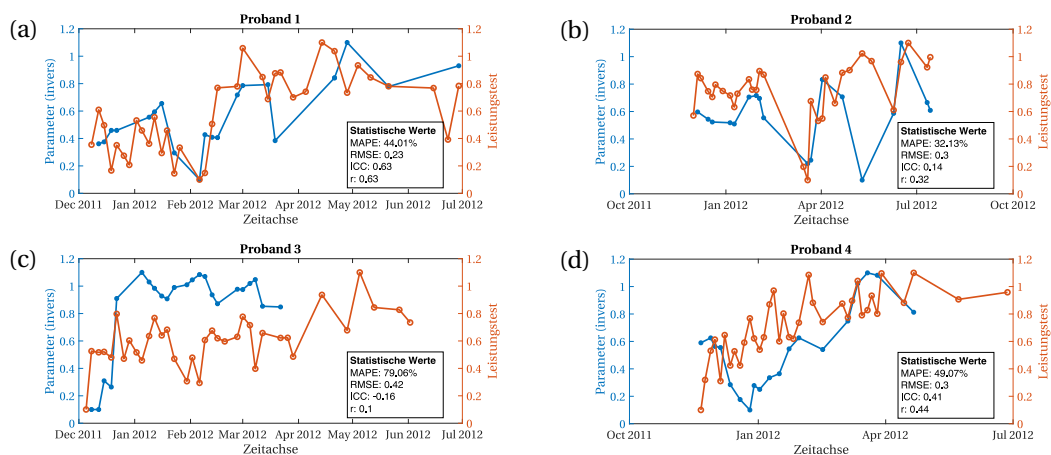


Abbildung 7.4.: Vergleich zwischen dem Parameter aus dem Faltungsmodell aus der virtuellen Leistungsdiagnostik und der empirischen Leistung für die vier Probanden nach Normierung

Der Verlauf der Zeitreihen für diesen Parameter und die empirische Leistung ist für jeden der vier Probanden in Abbildung 7.4 dargestellt. Eine eindeutige Korrelation ist für keinen der Probanden feststellbar. Pearson's Korrelationskoeffizient r schwankt zwischen Werten, die auf gar keine Korrelation schließen lassen (0,1 bei Proband 3) und Werten, die für einen deutlichen linearen Zusammenhang sprechen (0,63 bei Proband 1). Für den Intra-Korrelationskoeffizient ICC verhält es sich ähnlich – auch dieser schwankt je nach betrachtetem Proband zwischen 0,14 und 0,63. Für einen Probanden ist der ICC mit einem Wert von -0,16 negativ, was als Wert von 0 gewertet wird. Visuell wird deutlich, dass auch hier die Werte beider Zeitreihen stark schwanken. Ein Trend ist ohne weitere Analyse weder für den Verlauf des Parameters noch für die Leistungsentwicklung eindeutig erkennbar, wobei sich in allen Fällen insgesamt ein leichter Aufwärtstrend erahnen lässt.

Ansatz 2b: Trendanalyse aus der virtuellen Leistungsdiagnostik

Für die Trendanalyse der Werte aus der virtuellen Leistungsdiagnostik wird das einparametrische Faltungsmodell wie zuvor auf den Outdoor-Trainingseinheiten der vergangenen 14 Tage und damit auf maximal 14 Datensätzen kalibriert. Die Zeitreihen des kalibrierten Modellparameters und der empirischen Leistung sind identisch mit denen aus Ansatz 2a. Zur

Betrachtung der Trendverläufe beider Zeitreihen wird der Trend aus diesen wie in Unterabschnitt 7.1.1 (»Experimenteller Aufbau und Methoden«) beschrieben durch Subtraktion der jeweiligen trendbereinigten Zeitreihe extrahiert.

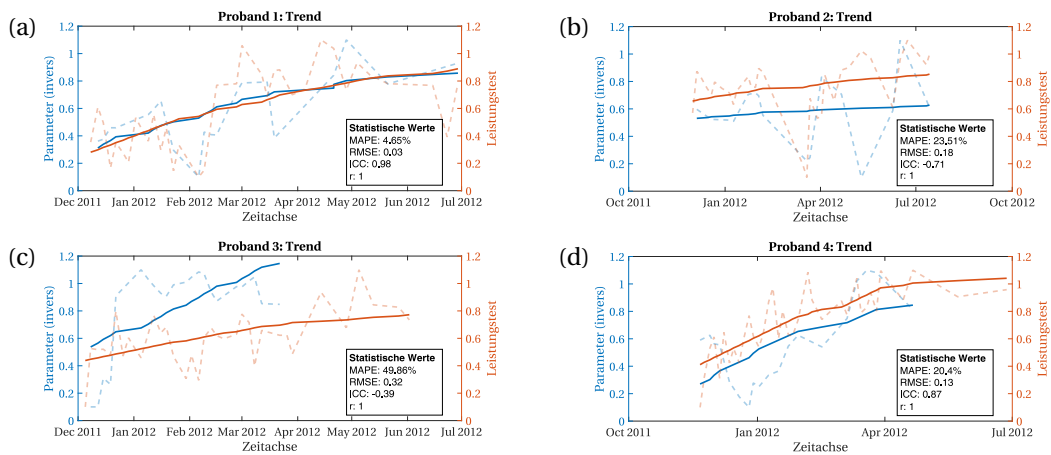


Abbildung 7.5.: Trendanalyse des Parameters aus der virtuellen Leistungsdiagnostik; Vergleich zwischen dem Parameter aus dem Faltungsmodell und der empirischen Leistung für die vier Probanden nach Normierung

Der Trendverlauf beider Zeitreihen ist für jeden der vier Probanden in Abbildung 7.5 dargestellt. Die ursprünglichen Zeitreihen wie aus Ansatz 2a sind zum besseren Vergleich zusätzlich schwach im Hintergrund und mit gestrichelter Linie dargestellt. Die Abweichung der beiden Kurven zueinander fällt bei der Betrachtung der Trends für die einzelnen Probanden sehr unterschiedlich aus: Der MAPE weist für Proband 3 eine hohe Abweichung mit einem prozentualen Fehler von knapp 50% auf, während sich bei Proband 1 mit einem MAPE-Wert von unter 5% kaum eine Abweichung zeigt. Für die Probanden 2 und 4 liegt der MAPE bei rund 21-25%. Für den RMSE zeigt sich ein ähnliches Bild mit Schwankungen zwischen 0,03 (Proband 1) bis zu Werten von 0,32 (Proband 3). Der Intra-Korrelationskoeffizient ICC fällt für die Probanden 1 und 4 mit Werten von 0,98 bzw. 0,87 sehr hoch aus. Dies weist für die beiden Probanden auf eine sehr hohe Übereinstimmung der beiden Zeitreihen hin. Für die Probanden 2 und 3 ist der Wert des ICC im negativen Bereich und wird entsprechend unabhängig von der Höhe seines Absolutwertes als Null gewertet. Pearson's Korrelationskoeffizient r beträgt wie in Ansatz 1b für alle Probanden exakt 1, was für eine »perfekte Korrelation« spricht, deren Bedeutung im Abschnitt zur Interpretation näher erläutert wird. Visuell zeigt sich für Proband 1 eine nahezu perfekte Übereinstimmung der beiden Trendverläufe. Auch die Verläufe bei Proband 4 sind sich sehr ähnlich – sie verlaufen mit einem kleinen Abstand annähernd parallel zueinander. Bei den Probanden 2 und 3 sind die Trendverläufe zwar allesamt ansteigend, die Steigung der einzelnen Trends selbst weicht jedoch sichtbar voneinander ab.

Zusammenfassender Überblick der Ergebnisse

Tabelle 7.1 und Tabelle 7.2 fassen die statistischen Kennzahlen für alle betrachteten Ansätze zur besseren Übersicht zusammen. Für den RMSE ist zu beachten, dass der RMSE bei maximalem Abstand beider Zeitverläufe zueinander auf Grund der durchgeführten Normalisierung exakt 1 beträgt und der angegebene Wert demzufolge auch als prozentualer Fehler interpretiert werden kann. Da es sich bei der Betrachtung des MAPE und RMSE um Fehlermaße handelt, weisen kleinere Werte auf einen größeren Zusammenhang und geringere Abstände der Zeitreihen hin. Bei den Korrelationsmaßen entsprechen größere Werte einem stärkeren Zusammenhang.

Tabelle 7.1.: MAPE und RMSE als statistische Maße der vier Probanden P1-P4 mit den vier unterschiedlichen Ansätzen für das parameterreduzierte Faltungsmodell mit a_3 als freiem Parameter

Methode / Kennzahl	MAPE				RMSE			
	P1	P2	P3	P4	P1	P2	P3	P4
Ansatz 1a	94,21%	53,5%	52,59%	62,59%	0,36	0,35	0,33	0,37
Ansatz 1b	27,1%	30,97%	9,04%	16,81%	0,16	0,27	0,06	0,17
Ansatz 2a	44,01%	32,13%	79,06%	49,07%	0,23	0,30	0,42	0,30
Ansatz 2b	4,65%	23,51%	49,86%	20,4%	0,03	0,18	0,32	0,13

Tabelle 7.2.: Zwei Korrelationsmaße (ICC und r) der vier Probanden P1-P4 mit den vier unterschiedlichen Ansätzen für das parameterreduzierte Faltungsmodell mit a_3 als freiem Parameter

Methode / Kennzahl	ICC				r			
	P1	P2	P3	P4	P1	P2	P3	P4
Ansatz 1a	-0,13	0,19	-0,12	0,11	-0,15	0,29	-0,16	0,1
Ansatz 1b	0,31	0,16	0,13	0,65	1	1	1	1
Ansatz 2a	0,63	0,14	-0,16	0,41	0,63	0,32	0,1	0,44
Ansatz 2b	0,98	-0,71	-0,39	0,87	1	1	1	1

7.1.3. Interpretation und Diskussion

Nachfolgend werden die gezeigten Ergebnisse der vier Ansätze interpretiert und diskutiert.

Ansatz 1: Direkter Vergleich

Eine Korrelationsanalyse kann nur für die gleiche Anzahl von Datenpunkten der beiden Zeitreihen durchgeführt werden. Durch den Umstand, dass nicht für jeden Tag mit empirischer Leistung ein Ergometer-Test vorliegt und nicht für jeden Tag mit Ergometer-Test eine empirische Leistung bekannt ist, wurden die Zeitreihen im Vorfeld zur Analyse auf diejenigen

Tage eingeschränkt, zu denen sowohl ein Parameter als auch ein empirischer Leistungswert betrachtet werden kann. Diese Einschränkung der betrachteten Tage kann die berechneten statistischen Werte jedoch verfälschen. Bei den Probanden 3 und 4 enden die zur Analyse nutzbaren Daten der Ergometrietest und damit auch die kalibrierten Parameterwerte mehrere Monate früher als die Daten der empirischen Leistungsmessungen. Bei Proband 3 hat sich ein deutlicher Aufwärtstrend der empirischen Messungen allerdings erst in dem Zeitabschnitt gezeigt, für den auf Grund der fehlenden Datenaufzeichnungen keine Parameterwerte des Faltungsmodells mehr kalibriert werden konnten. Hier wäre es durchaus interessant zu sehen, ob die Leistungssteigerung auch über den Parameter des Faltungsmodells detektiert werden kann.

Obwohl die in der ursprünglichen Studie gewählte Leistungsdiagnostik explizit mit Hinblick auf die Vermeidung starker Einflüsse von Motivation und Tagesform ausgewählt wurde, können diese Einflüsse nicht ausgeschlossen werden. Auch der insgesamt geringe Wertebereich der empirischen Leistung von (ohne Normalisierung) maximal 0,86 kJ/kg führt dazu, dass bereits kleine Effekte in der Tagesform einen möglicherweise erheblichen Einfluss auf die ermittelte empirische Leistung haben. Die Validität der Leistungswerte, mit denen der Parameter verglichen wird, ist dadurch nicht vollständig gesichert.

Insbesondere zeigt sich durch die hohen Schwankungen der empirischen Leistung, dass der diagnostisch ermittelte Wert stark von der jeweils aktuellen Tagesform beeinflusst wurde. Das wird insbesondere durch die Betrachtung der Trendverläufe (Ansatz 1b) deutlich: Sowohl die Intraklassen-Korrelation über den ICC als auch die Fehlerwerte in Form von MAPE- und RMSE-Werten weisen deutlich geringere Fehler und eine höhere Korrelation auf als in Ansatz 1 a. Für alle Probanden lässt sich sowohl über die empirische Leistung als auch über den Modellparameter ein Leistungszuwachs erkennen, der durch die starken Schwankungen der Daten ohne Trendanalyse nicht direkt erkennbar war. Dabei fällt jedoch auch auf, dass die Steigung der Trends von empirischer und modellierter Leistung sehr unterschiedlich ausfällt.

Die starken Schwankungen innerhalb der Daten werfen zusätzlich die Frage auf, wie »fair« der durchgeführte Vergleich ist: Bereits innerhalb der empirischen Leistungsmessung können von einem zum nächsten Test extrem hohe Schwankungen von rund 80% auftreten (z.B. fällt der normierte empirische Leistungswert bei Proband 2 Ende März von rund 1 auf rund 0,2 ab). Eine Vielzahl der Datenpunkte, die für die Modellparameter-Kalibrierung zur Verfügung standen, entstammen Ergometereinheiten von Tagen, an denen kein empirischer Leistungswert vorliegt, d.h. an denen lediglich ein Training auf dem Ergometer ohne Leistungsdiagnostik absolviert wurde. Fraglich ist damit, ob die verwendeten Daten inhärent bereits zu starke tagesformabhängige Unterschiede aufweisen. Zusätzlich ist dabei zu beachten, dass der freie Modellparameter sehr stark von der Herzfrequenz vor Trainingsbeginn abhängt. Wie bereits in Unterabschnitt 1.3.1: »Die Herzfrequenz als Indikator: Kritische Aspekte« dargestellt, unterliegt die Herzfrequenz zahlreichen Einflussfaktoren. Eine

Kalibrierung des Parameters auf einem einzelnen Datensatz führt damit unweigerlich dazu, dass der Einfluss von leistungsunabhängigen Faktoren auf die Herzfrequenz einen großen Einfluss auf die Kalibrierung nimmt. Es bleibt demnach fraglich, in wie weit das Fehlen eines linearen Zusammenhangs in Ansatz 1 a möglicherweise mit den starken Abweichungen der einzelnen Tage zusammenhängt.

Eine Antwort darauf ließe sich im Rahmen der Validierung des Modellparameters nur über die Verwendung von geeigneten Datensätzen finden, für die die empirische Leistungsmessung und die Kalibrierung der Modellparameter auf den gleichen diagnostischen Testdaten basiert. Ein derartiger Datensatz wird in Abschnitt 7.2: »Analyse zur Abschätzung der Leistungsveränderung auf Basis von Herzfrequenz-Simulationen: Hobby-Sportler« untersucht.

Ansatz 2: Virtuelle Leistungsdiagnostik

Wenngleich eine Anpassung der verwendeten Daten auch im Rahmen der virtuellen Leistungsdiagnostik zur Berechnung der Korrelations- und Fehlerwerte notwendig ist, damit die gleiche Anzahl von Datenpunkten der beiden Zeitreihen vorliegt, scheint der Einfluss der Tagesschwankungen innerhalb dieser Analysen schwächer als im direkten Vergleich zu sein. Dadurch, dass die Parameterkalibrierung nicht auf einem einzelnen Datensatz basiert, sondern auf unterschiedlichen Messungen von bis zu 14 Tagen durchgeführt wurde, scheint der Verlauf insgesamt weniger starken Schwankungen zu unterliegen.

Bei Proband 1 zeigt sich bereits ohne spezifische Trendanalyse ein sehr ähnlicher Verlauf beider Zeitreihen (ICC und r bei 0,63) – in der Trendanalyse ist kaum ein Unterschied zwischen dem Trendverlauf der empirischen Leistung und dem ermittelten Parameter des Faltungmodells zu erkennen. Dieser visuelle Eindruck bestätigt sich auch in den statistischen Kennzahlen, beispielsweise durch einen MAPE $< 5\%$ und eine ICC von 0,98.

Auch für Proband 2 verlaufen die beiden Zeitreihen tendenziell sehr ähnlich. Lediglich Mitte bis Ende Mai weicht ein einzelner Parameter des Faltungmodells deutlich vom restlichen Trend und der empirischen Leistung ab. Allerdings lagen für die Kalibrierung des Parameters für diesen Tag lediglich drei Outdoor-Trainingsdatensätze vor; alle anderen Parameter von Proband 2 konnten auf durchschnittlich 6,9 ($\pm 2,4$) Outdoor-Trainingsdatensätzen kalibriert werden. Aufgrund dessen, dass der sonstige Verlauf des Parameters über die Zeit dem Verlauf der empirischen Leistungsentwicklung visuell sehr ähnlich ist, könnte es sich bei diesem Wert um einen Ausreißer mit hohem Einfluss auf die statistischen Ergebnisse handeln.

Ein ähnlich interessantes Bild zeigt sich bei Proband 3: Sowohl in den empirischen Messungen als auch im Verlauf des Parameters zeigt sich ein starker Anstieg etwa Mitte Dezember. Zu diesem Zeitpunkt scheint das Training zu einem leichten Leistungszuwachs zu führen. Die vier vorherigen Werte des Parameters sind jedoch vergleichsweise sehr niedrig und wei-

chen stark von den anschließenden Werten ab. Ähnlich wie bei Proband 2 zeigt sich auch hier, dass die Kalibrierung des Parameters bei den vier ersten Werten auf deutlich weniger Outdoor-Trainingseinheiten stattgefunden hat: Für die ersten beiden Werte standen für die Parameterkalibrierung zwei Datensätze zur Verfügung, für den dritten und vierten Wert jeweils vier. Allerdings standen bei diesem Probanden generell weniger Kalibrierungsdaten für die einzelnen Parameterwerte zur Verfügung – der Mittelwert liegt hier bei 5,3 ($\pm 2,0$) Outdoor-Trainingseinheiten, die zur Kalibrierung verwendet werden konnten. Dabei fällt zusätzlich auf, dass in der Zeitreihe der Modellparameter deutlich weniger Daten für die Analyse vorliegen. Die Zeitreihe endet bereits Ende März, während für die empirischen Messwerte Daten bis Ende Juni berücksichtigt werden konnten. Dadurch, dass sich die Werte nach dem sprunghaften Anstieg des Parameters Mitte Dezember von unter 0,3 (vierter Parameterwert) auf 0,9 (fünfter Parameterwert) auf einem neuen Niveau oberhalb von 0,8 eingependelt haben, wird der Trendverlauf verzerrt. Bei den empirischen Messwerten liegen im Verhältnis zum Anstieg Mitte Dezember deutlich mehr Messdaten auf dem neuen Leistungsniveau vor als für den Modellparameter, sodass die anfänglich niedrigeren Werte weniger stark ins Gewicht fallen. Wie stark sich der Effekt der deutlich abweichenden Anzahl der Datenpunkte beider Zeitreihen auf die Analyse auswirkt und diese ggf. verzerrt ist dabei abhängig davon, wie sich der Parameter im weiteren zeitlichen Verlauf verhalten würde.

Für Proband 4 zeigt sich ähnlich wie bei Proband 1 ein hoher Zusammenhang bei der Analyse des Trendverlaufs. Obwohl der Parameter zu Beginn einen deutlichen Abfall beschreibt, bevor der Wert ab Ende Dezember nahezu kontinuierlich ansteigt, verlaufen die beiden Trends insgesamt annähernd parallel zueinander. Werden die Tagesschwankungen durch die Trendanalyse ignoriert, ergibt sich eine ICC von 0,87, die auf eine hohe Ähnlichkeit beider Messreihen hindeutet. Doch auch bereits auf Basis von Ansatz 2a weisen die ICC mit einem Wert von rund 0,41 und der r -Wert von 0,44 auf einen mäßigen linearen Zusammenhang hin. Auch hier bestätigt sich der eingangs diskutierte Eindruck, dass der Parameter möglicherweise durch die Kalibrierung auf mehreren Datensätzen weniger starken Tagesschwankungen unterliegt und sich dadurch insgesamt stabiler verhält.

Ergänzend zu Ansatz b: Pearson's r und ICC

Der Korrelationskoeffizient nach Pearson ist bei Zeitreihen mit Vorsicht zu betrachten. Üblicherweise sollten Zeitreihen einer Trendbereinigung unterzogen werden, bevor eine Korrelationsanalyse durchgeführt wird. Die zugrunde liegende Problematik wurde bereits ausführlich in Unterabschnitt 2.1.3: »Kritische Betrachtung statistischer Parameter« im Abschnitt über Korrelationskoeffizienten dargestellt. Da wie eingangs erläutert bei der Analyse der Leistungsentwicklung jedoch explizit der Trend von Interesse ist, während Variationen in der Tagesform im Verlauf kaum von Bedeutung sind, wurde in der vorangegangenen Analyse auf eine Trendbereinigung verzichtet. Es wurde im Gegenteil sogar explizit eine

Analyse des Trendverlaufs durchgeführt. Damit wird Pearson's r in diesem spezifischen Anwendungsfall jedoch zu einem unzuverlässigen Maß. Während es innerhalb von Ansatz a noch für eine Tendenz des linearen Zusammenhangs unter Berücksichtigung anderer statistischer Kennzahlen verwendet werden kann, ist dieser Korrelationskoeffizient innerhalb der Trendanalyse weitgehend unbrauchbar. In den durchgeführten Analysen ergibt die Berechnung des r -Werts bei der Trendanalyse stets einen Wert von 1 und weist damit theoretisch auf einen »perfekten linearen Zusammenhang« hin. Dieser lässt sich anhand der anderen statistischen Kennzahlen und der visuellen Analyse nicht bestätigen. Mit dem r -Wert lässt sich in diesem Fall lediglich feststellen, dass beide Zeitreihen – die Zeitreihe des Modellparameters und die der empirischen Leistungsdaten – im Zeitverlauf dem gleichen linearen Trend in Form eines streng monotonen Anstiegs folgen.

Im Gegensatz zu Pearson's r ist die ICC auch für die Trendanalyse geeignet. Die ICC bewertet Unterschiede bezüglich Streuung und Mittelwerten zwischen den betrachteten Datenreihen und innerhalb der Datenreihen. Für die ICC besteht keine Notwendigkeit einer Trendbereinigung. Während bei der Trendanalyse des direkten Vergleichs (Ansatz 1b) lediglich für Proband 4 ein hoher ICC-Wert $> 0,6$ berechnet werden konnte, zeigt sich in der Trendanalyse der virtuellen Leistungsdiagnostik zusätzlich für Proband 1 ein sehr starker Zusammenhang durch einen ICC von 0,98. Bei den Probanden 2 und 3 liegen im Vergleich dazu deutlich stärkere Schwankungen vom Mittelwert vor, was sich auch in den niedrigen, teils negativen Werten der ICC widerspiegelt.

Aussagekraft der Abstandsmaße (MAPE, RMSE)

Wenngleich der MAPE und der RMSE als Fehlermaße grundsätzlich auch für Zeitreihen geeignet sind, müssen bei der Interpretation der Ergebnisse zwei Aspekte besonders berücksichtigt werden:

1. Die betrachteten Daten wurden normiert;
2. Während die empirische Leistung (ohne Normierung) in kJ/kg angegeben werden kann, sind die Parameter abstrakter Natur und dimensionslos.

Die Art Normierung kann bereits Einfluss darauf haben, wie nah die empirische und die simulierte Kurve beieinander liegen. Aus diesem Grund wurden beide Zeitreihen in den durchgeführten Experimenten ohne nähere Betrachtung linear auf das Intervall $[0,1; 1,1]$ normiert. Der Wert 0 wurde dabei nicht verwendet, um eine adäquate Umrechnung der Parameter zu gewährleisten. Letztendlich spielt die genaue Positionierung der Zeitreihen zueinander jedoch auch nur eine untergeordnete Rolle: Ziel der Untersuchung war es, die Ähnlichkeit im Verlauf zwischen den beiden Zeitreihen zu analysieren. Dabei spielen weder die spezifische Skalierung, noch die Abstände untereinander eine wesentliche Rolle. Zentral bei der Betrachtung des Verlaufs sind die grundlegenden Trendentwicklungen und ob

es mit Hilfe des Parameters möglich ist, einen Leistungszuwachs und eine Leistungsabnahme tendenziell über die Zeit voneinander unterscheiden zu können. In diesem Sinne ist auch das Fehlen einer Dimensionsgröße bei dem freien Modellparameter kein Hindernis: Der Vergleich, ob und in wie weit sich der Parameter über die Zeit ähnlich zu empirischen Leistungswerten entwickelt, ist nicht von einer spezifischen Einheit des Parameters abhängig und kann ebenso gut abstrakt betrachtet werden.

Umgekehrt folgt daraus jedoch auch, dass die Aussagekraft der exakten Fehlerwerte durch die Normierung nicht als allgemeingültig angenommen werden kann. Die Fehlermaße können als Vergleichswerte lediglich in einem ebenso aufgebauten Setting verwendet werden. Für die hier durchgeführten Vergleiche unterschiedlicher Ansätze und zur besseren Einschätzung der Ähnlichkeit zwischen den beiden Zeitreihen erscheinen sie grundsätzlich geeignet.

7.1.4. Fazit zur Studie

Beim direkten Vergleich der Werte des Modellparameters mit der empirischen Leistung (Ansatz 1) ist sowohl visuell als auch basierend auf der statistischen Analyse keine Korrelation zwischen den beiden Zeitreihen erkennbar. Die Trendanalyse hat jedoch gezeigt, dass beide Zeitreihen dabei deutlich von Tagesschwankungen beeinflusst werden. Obwohl der Trend für alle Werte beider Zeitreihen und über alle Probanden hinweg insgesamt monoton ansteigt, variiert die Höhe der Steigung noch deutlich voneinander, sodass nicht von einem eindeutigen linearen Zusammenhang zwischen dem Parameter und den empirischen Messwerten ausgegangen werden kann. Insbesondere dadurch, dass die Werte des Modellparameters ausschließlich auf einzelnen Kalibrierungsdaten beruhen, übt die tagesaktuelle Herzfrequenz in diesem Fall einen sehr hohen Einfluss aus.

Dieser Einfluss kann durch die Analyse der virtuellen Leistungsdiagnostik (Ansatz 2) verringert werden: Durch die Kalibrierung auf mehreren Outdoor-Trainingseinheiten wirkt der Modellparameter im zeitlichen Verlauf insgesamt robuster. Innerhalb der virtuellen Leistungsdiagnostik zeigte sich außerdem für zwei der Probanden bereits ohne spezifische Trendanalyse ein mäßiger bis deutlicher linearer Zusammenhang zwischen dem Verlauf des Parameters und dem der empirisch gemessenen Leistung. Die Trendanalyse konnte diesen ersten Eindruck bestätigen und verstärken. Für die beiden Probanden, bei denen auch hier der Zusammenhang weniger deutlich erkennbar ist, liegen mehrere Schwächen innerhalb der zugrundeliegenden Daten vor, die das Ergebnis möglicherweise negativ beeinflussen.

In einer zweiten Studie wird nachfolgend in »Analyse zur Abschätzung der Leistungsveränderung auf Basis von Herzfrequenz-Simulationen: Hobby-Sportler« (Abschnitt 7.2) untersucht, ob auch bei Trainingsdaten von Hobby-Sportlern ein Zusammenhang gefunden werden kann und ob der Parameter des Faltungsmodells mit anderen Parametern der Leistungsdiagnostik möglicherweise sogar besser korreliert.

7.2. Analyse zur Abschätzung der Leistungsveränderung auf Basis von Herzfrequenz-Simulationen: Hobby-Sportler

Die Untersuchung des Modellparameters aus dem 1P-Faltungsmodell hinsichtlich eines Zusammenhangs zur Leistungsentwicklung wird wie bereits im vorangegangenen Abschnitt erneut mit anderen Daten durchgeführt: Basierend auf Leistungstest-Daten von drei männlichen Hobby-Sportlern wird nachfolgend analysiert, ob der verbleibende Parameter aus einer der zwei Varianten des einparametrischen Faltungsmodells unmittelbar mit der empirisch ermittelten Leistungsfähigkeit der Probanden zusammenhängt. Die Daten wurden in »Datensatz 3: Daten von Hobby-Sportlern« (Abschnitt 4.3) beschrieben.

Es wird sich zeigen, dass das identifizierte einparametrische Faltungsmodell mit freiem Parameter a_3 eine sehr hohe Ähnlichkeit und Korrelation zur Blutlaktatkonzentration an der geschätzten IAS aufweist. Zu den anderen Schwellwerten scheint ein hoher linearer Zusammenhang zum Inversen des Parameters vorzuliegen. Dieser Eindruck wird durch die Trendanalyse zusätzlich bestärkt. Bei dem einparametrischen Faltungsmodell mit freiem Parameter a_2 zeigen sich hingegen kaum lineare Zusammenhänge.

7.2.1. Experimenteller Aufbau und Methoden

Zur Analyse, in wie weit es möglich ist, über die Modellierung der Herzfrequenz Rückschlüsse auf die Leistungsentwicklung einer Person ziehen zu können, werden verschiedene Experimente durchgeführt. Die Modellierung der Herzfrequenz erfolgt dabei mit dem zuvor identifizierten einparametrischen Faltungsmodell aus Abschnitt 6.3 zur »Konstruktion eines einparametrischen Faltungsmodells: Studiendaten von Hobby-Sportlern«. Dazu werden die Trainingsdaten je Athlet auf zwei unterschiedliche Arten betrachtet.

Im Gegensatz zu den Analysen aus Abschnitt 7.1: »Analyse zur Abschätzung der Leistungsveränderung auf Basis von Herzfrequenz-Simulationen: Profi-Sportler« ist hierbei aufgrund von fehlenden Trainingsdaten keine Durchführung einer virtuellen Leistungsdiagnostik möglich. Anstelle eines einzelnen empirischen Leistungswertes wie zuvor werden nachfolgend drei unterschiedliche, aus der Laktatdiagnostik abgeleitete Leistungskennzahlen im zeitlichen Verlauf betrachtet, sodass sich die Analyse teilweise von der zuvor in Abschnitt 7.1 durchgeführten Untersuchung unterscheidet.

Abbildung 7.6 veranschaulicht den Ablauf des durchgeführten Prozesses: Die vorliegenden HF-Daten werden zur Kalibrierung des parameterreduzierten Faltungsmodells genutzt. Dabei wird jeder Datensatz einer Trainingseinheit einzeln und unabhängig von anderen Trainingseinheiten betrachtet. Auf diese Weise kann zu jeder Trainingseinheit ein Parameter identifiziert und im späteren Verlauf der Analyse mit der empirischen Leistung verglichen werden.

7. Nutzen des parameterreduzierten Faltungsmodells bei der Leistungsbeurteilung

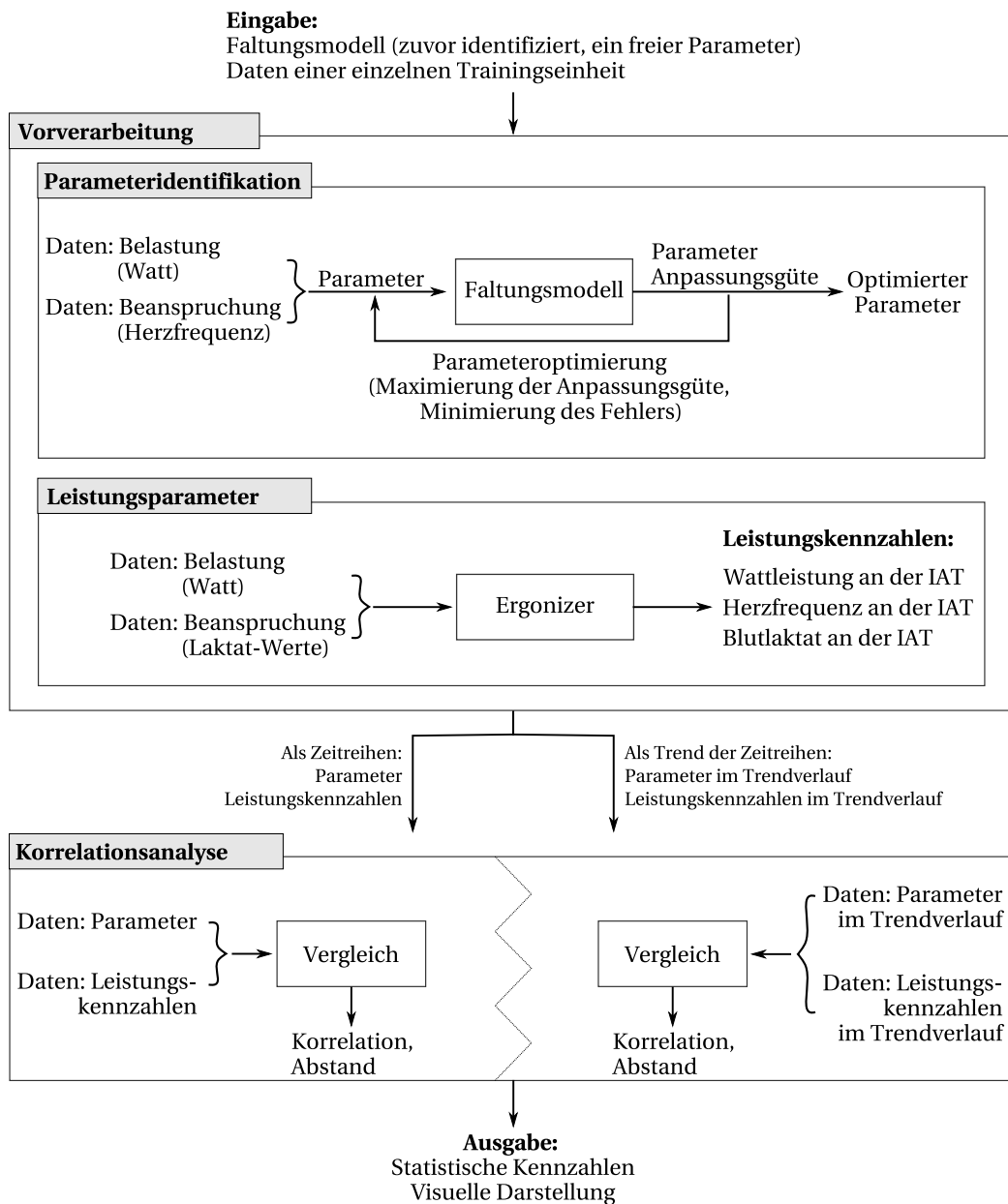


Abbildung 7.6.: Vorgehensweise zur Korrelationsanalyse zwischen der Leistung und dem Parametern der HF-Modellierung bei Hobby-Athleten

Parallel dazu werden Leistungswerte der Probanden anhand der Laktatdiagnostik betrachtet. Über den *Ergonizer* werden für jede Ergometer-Trainingseinheit die Wattleistungswerte, die Laktatschwellenwerte und die Herzfrequenz an der geschätzten IAS bestimmt. Die *Ergonizer*-Auswertung der einzelnen Protokolle findet sich im Anhang (Abbildung A.7, Abbildung A.8 und Abbildung A.9). Als Leistungswert werden alle drei Leistungsschwellenwerte betrachtet – d.h. die Wattleistung an der IAS, die Herzfrequenz an der IAS und die Blutlaktatkonzentration an der IAS. Der zuvor identifizierte freie Parameter des Faltungsmodells wird anschließend in Hinblick auf seine Korrelation zur Leistungsentwicklung anhand der

drei Leistungsschwellenwerte der Probanden hin analysiert. Dabei werden die zwei identifizierten einparametrischen Faltungsmodelle getrennt voneinander untersucht. Wie auch in »Unterabschnitt 7.1.1« zur »Analyse zur Abschätzung der Leistungsveränderung auf Basis von Herzfrequenz-Simulationen: Profi-Sportler« werden die beiden Zeitreihen dafür zunächst linear auf das Intervall $[0,1; 1,1]$ normalisiert. Eine Trendbereinigung findet erneut nicht statt, damit der Verlauf der Leistungsentwicklung anstelle von Tagesschwankungen betrachtet werden kann. Bei der Berechnung der Korrelationskoeffizienten ist zu berücksichtigen, dass für jeden der drei Probanden maximal sechs Werte zur Korrelationsanalyse zur Verfügung stehen. Die Berechnung des ICC ist daher nicht möglich. Durch die geringe Anzahl an Datenpunkten kann die Analyse dieses Datensatzes lediglich einen ersten Eindruck davon vermitteln, welche Möglichkeiten sich für die Betrachtung der Leistungsentwicklung bei Hobby-Athlet*innen und hinsichtlich unterschiedlicher Leistungsmaße ergeben können. Dies ist bei der Interpretation der Ergebnisse zu berücksichtigen.

7.2.2. Ergebnisse

Für jeden der drei Probanden lassen sich basierend auf den vorhandenen Daten drei unterschiedliche Korrelationsanalysen durchführen: Die Analyse der Korrelation zwischen dem jeweiligen Parameter des Faltungsmodells und jeweils einem der drei ermittelten Leistungsschwellenwerte. Für den RMSE ist zu beachten, dass der RMSE bei maximalem Abstand beider Zeitverläufe zueinander auf Grund der durchgeführten Normalisierung exakt 1 beträgt und der angegebene Wert demzufolge auch als prozentualer Fehler interpretiert werden kann.

Einparametrisches Faltungsmodell, Variante 1: a_2 verbleibt frei

Die visuellen Ergebnisse aus dem Vergleich des Modellparameters mit den drei Schwellenwerten an der geschätzten IAT eines jeden Probanden im zeitlichen Verlauf zeigt Abbildung 7.7 bzw. Abbildung 7.8 für die Trendanalyse. Der WAPE wird hier aufgrund der geringen Anzahl an Datenpunkten anstelle des MAPE betrachtet (vgl. Definition 2.1 in Unterabschnitt 2.1.1: »Zeitreihenanalyse und statistische Parameter«).

Eine Übersicht der statistischen Kennzahlen (WAPE, RMSE, Pearson's r) ist in Tabelle 7.3 und für die Trendanalyse in Tabelle 7.4 dargestellt.

Bei allen drei Probanden und allen drei Schwellenwerten an der geschätzten IAS sind die Fehlerwerte (WAPE und RMSE) tendenziell hoch. Es liegt lediglich ein Fall vor, in dem der WAPE einen Wert von 30% unterschreitet (Blutlaktatkonzentration an der IAS für Proband 2 im Trendverlauf mit 22,58%). Der RMSE liegt ebenfalls nur für diesen einen Fall bei unter einem Fehlerwert von 0,2, was etwa 20% entspricht².

²Aufgrund der Normalisierung beträgt der RMSE bei maximalem Abstand beider Zeitverläufe zueinander exakt 1.

7. Nutzen des parameterreduzierten Faltungmodells bei der Leistungsbeurteilung

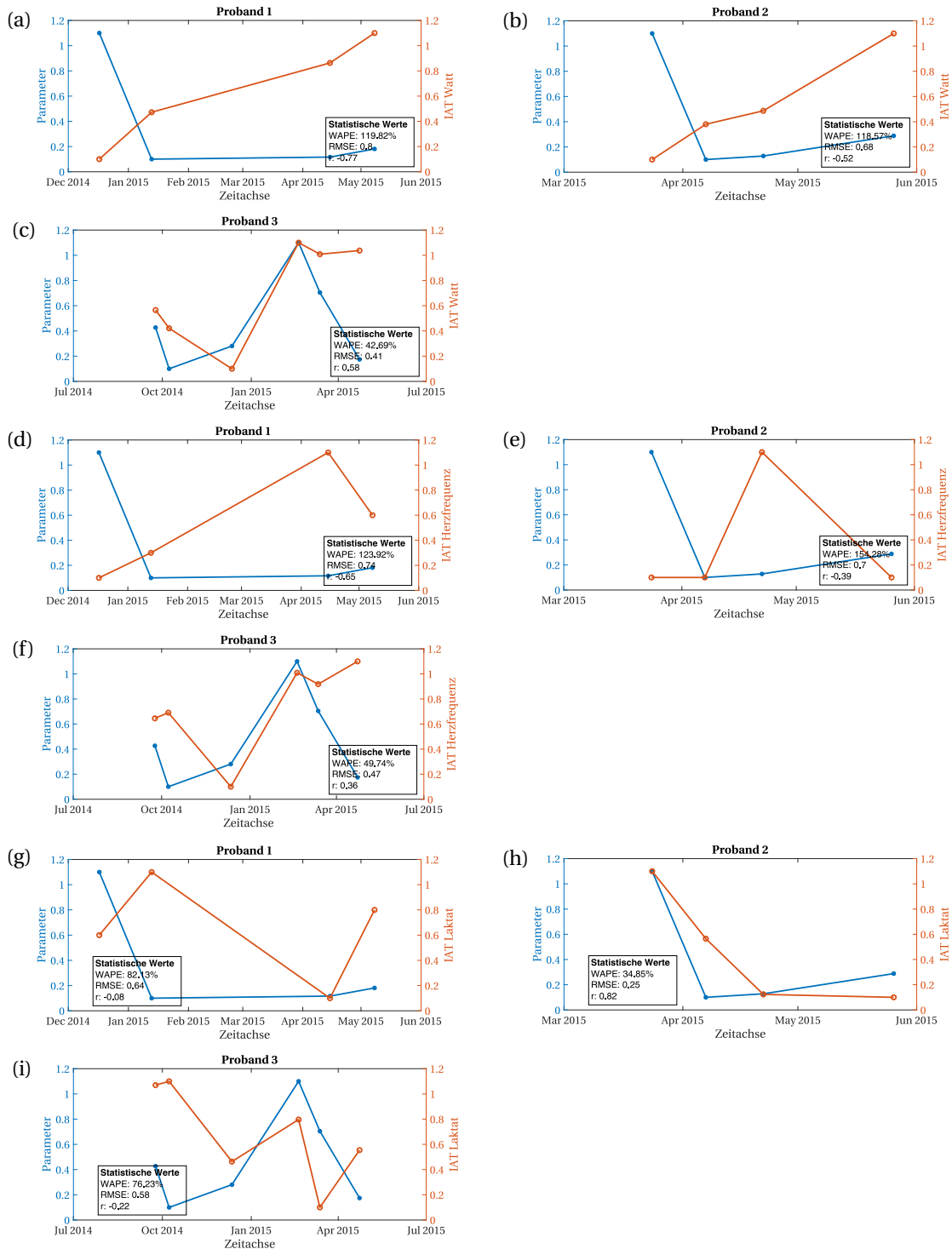


Abbildung 7.7.: Vergleich: Parameter a_2 des Faltungsmodells und Leistung anhand unterschiedlicher Schwellwerte bezüglich der geschätzten IAS für alle drei Probanden. (a)-(c): Vergleich mit der Wattleistung an der IAS, (d)-(f): Vergleich mit der Herzfrequenz an der IAS, (g)-(i): Vergleich mit der Blutlaktatkonzentration an der IAS

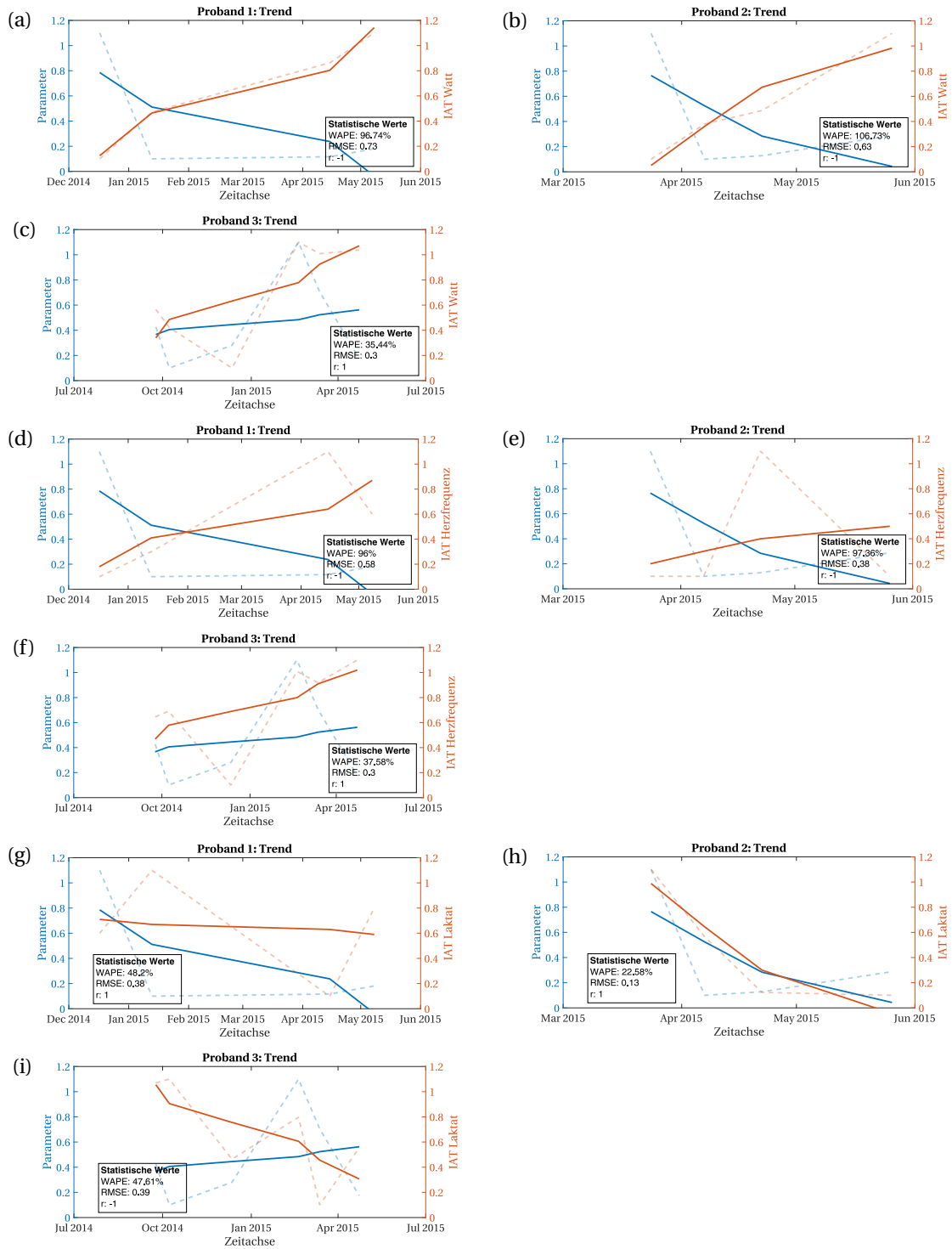


Abbildung 7.8.: Vergleich: Parameter a_2 des Faltungsmodells und Leistung anhand unterschiedlicher Schwellenwerte bezüglich der geschätzten IAS für alle drei Probanden. (a)-(c): Vergleich mit der Wattleistung an der IAS, (d)-(f): Vergleich mit der Herzfrequenz an der IAS, (g)-(i): Vergleich mit der Blutlaktatkonzentration an der IAS

7. Nutzen des parameterreduzierten Faltungsmodells bei der Leistungsbeurteilung

Tabelle 7.3.: Drei statistische Maße der drei Probanden P1, P2, P3 mit den drei unterschiedlichen Schwellwerten für das parameterreduzierte Faltungsmodell mit a_2 als freiem Parameter

Methode/Maßzahl	WAPE [%]			RMSE			r		
	P1	P2	P3	P1	P2	P3	P1	P2	P3
IAT: Watt	119,82	118,57	42,69	0,8	0,68	0,41	-0,77	-0,52	0,58
IAT: Herzfrequenz	123,92	154,28	49,74	0,74	0,7	0,47	-0,65	-0,39	0,36
IAT: Laktat	82,13	34,85	76,23	0,64	0,25	0,58	-0,08	-0,82	-0,22

Tabelle 7.4.: Drei statistische Maße der drei Probanden P1, P2, P3 mit den drei unterschiedlichen Schwellwerten für das parameterreduzierte Faltungsmodell mit a_2 als freiem Parameter im Trendverlauf

Methode/Maßzahl	WAPE [%]			RMSE			r		
	P1	P2	P3	P1	P2	P3	P1	P2	P3
IAT: Watt	96,74	106,73	35,44	0,73	0,63	0,3	-1	-1	1
IAT: Herzfrequenz	96,0	97,36	37,58	0,58	0,38	0,30	-1	-1	1
IAT: Laktat	48,2	22,58	47,61	0,38	0,13	0,39	1	1	-1

Grundsätzlich ließe sich der Fehler durch die Invertierung des Parameters (d.h. für einen Parameterwert a wird stattdessen $\frac{1}{a}$ betrachtet) teilweise senken, was durch die negativen Korrelationskoeffizienten erkennbar ist. Allerdings führt eine Invertierung hierbei nicht für alle Probanden gleichermaßen zu einer positiven Korrelation: Bei der Betrachtung des Parameters im zeitlichen Verlauf (ohne Invertierung) liegt ausschließlich im Vergleich mit der Blutlaktatkonzentration für alle drei Probanden eine negativer Korrelationskoeffizient vor. Bei den anderen beiden Schwellwerten (Wattleistung an der IAS und Herzfrequenz an der IAS) liegen für die Probanden 1 und 2 jeweils negative Korrelationskoeffizienten vor, für Proband 2 jedoch nicht. In der Trendanalyse sind die Korrelationskoeffizienten für die drei Probanden bei keinem der betrachteten Schwellwerte gleichermaßen positiv oder negativ. Für die Möglichkeit einer späteren Interpretation des Parameters ist eine Invertierung nur dann zweckmäßig, wenn sie für alle Probanden gleichermaßen durchführbar ist und zu einem positiven Korrelationskoeffizienten führt. Da dies hier nicht der Fall ist, wird auf die Analyse der invertierten Parameter im zeitlichem Verlauf verzichtet. Die starke Abhängigkeit der positiven bzw. negativen Korrelationskoeffizienten von den einzelnen Probanden wird innerhalb der Interpretation diskutiert.

Innerhalb der visuellen Darstellung fällt auf, dass der identifizierte Parameter bei Proband 1 vom maximal angenommenen Wert im ersten Ergometertest unmittelbar auf den minimal angenommenen Wert in Test 2 und 3 fällt. Auch bei Proband 2 fällt der Wert des Parameters von seinem Maximum im ersten Ergometertest unmittelbar auf sein Minimum im zweiten Ergometertest. Lediglich bei Proband 3 lässt sich ein weniger extremer Verlauf des Parameters über die Zeit erkennen. Dabei ist zu berücksichtigen, dass für die Probanden 1 und 2

lediglich vier Messzeitpunkte zur Verfügung stehen, während die Analyse für Proband 3 an sechs Messzeitpunkten stattfindet.

Die Ergebnisse dieses 1P-Modells weisen damit nicht auf einen eindeutigen Zusammenhang zu den einzelnen Leistungsmesswerten hin.

Einparameteriges Faltungsmodell, Variante 2: a_3 verbleibt frei

Die visuellen Ergebnisse aus dem Vergleich des Modellparameters mit den drei Schwellwerten an der geschätzten IAT eines jeden Probanden im zeitlichen Verlauf zeigt Abbildung 7.9 bzw. Abbildung 7.10 für die Trendanalyse. Der WAPE wird hier aufgrund der geringen Anzahl an Datenpunkten anstelle des MAPE betrachtet.

Bei dieser Modellvariante des einparametrischen Faltungsmodells fällt die Korrelation je nach zum Vergleich verwendeten Leistungsschwellwert recht unterschiedlich aus: Bei der Betrachtung der Wattleistung an der geschätzten IAT liegt für alle drei Probanden mit $r > 0,5$ bereits ein deutlicher linearer Zusammenhang vor. Für die Probanden 2 und 3 deutet der Korrelationskoeffizient mit $r > 0,9$ sogar auf einen sehr hohen bis nahezu perfekten linearen Zusammenhang hin. Der Eindruck einer starken Ähnlichkeit wird dabei auch vom WAPE mit Werten $< 20\%$ für die Probanden 2 und 3 und zumindest $< 50\%$ für Proband 1 und dem RMSE ($< 0,2$ für P2 und P3) gestützt. Auch visuell lässt sich die Ähnlichkeit zwischen dem Parameter und der Wattleistung an der geschätzten IAS im zeitlichen Verlauf besonders für die Probanden 2 und 3 in Abbildung 7.9 (b) bzw. (c) gut erkennen – die beiden Zeitreihen verlaufen für beide Probanden nahezu parallel und in einem ähnlichen (normierten) Wertebereich. Dabei ist zu beachten, dass der Parameter für eine bessere Vergleichbarkeit zum Schwellwert für alle drei Probanden invertiert werden konnte.

Auch im Vergleich mit der Herzfrequenz an der geschätzten IAS wurde der Parameter zur Analyse invertiert. Die Fehlerwerte fallen in diesem Vergleich insbesondere für die Probanden 1 und 2 mit Fehlerwerten von über 80% im WAPE und über 0,45 im RMSE deutlich höher aus als bei der Wattleistung. Ein linearer Zusammenhang lässt sich über den Korrelationskoeffizienten mit $r < 0,2$ für diese beiden Probanden nicht feststellen. Lediglich für Proband 3 scheint ein deutlicher linearer Zusammenhang ($r = 0,76$ und Fehlerwerte $< 30\%$) vorzuliegen. Dieser Eindruck lässt sich für alle drei Probanden visuell bestätigen (Abbildung 7.9 (d)-(f)).

Beim Vergleich des Modellparameters mit der Blutlaktatkonzentration an der geschätzten IAS zeigt sich für Proband 3 ein deutlicher und für Proband 2 ein sehr hoher linearer Zusammenhang ($r > 0,5$ bzw. $r > 0,8$), wenngleich die Fehlerwerte bei rund 24–51% liegen. Visuell ist der hohe lineare Zusammenhang besonders für Proband 2 (Abbildung 7.9 (h)) deutlich erkennbar.

7. Nutzen des parameterreduzierten Faltungsmodells bei der Leistungsbeurteilung

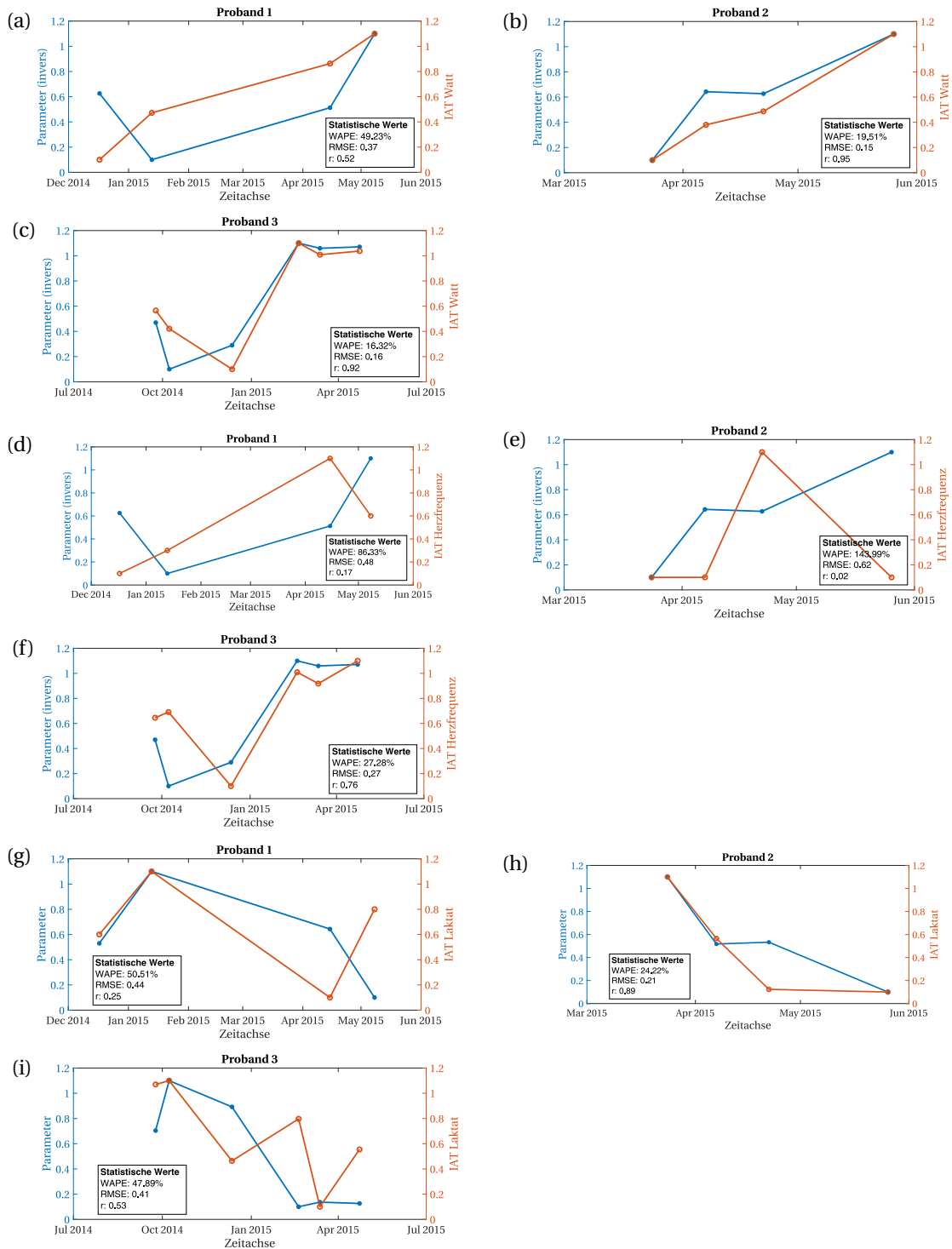


Abbildung 7.9.: Vergleich: Parameter a_2 des Faltungsmodells und Leistung anhand unterschiedlicher Schwellwerte bezüglich der geschätzten IAS für alle drei Probanden. (a)-(c): Vergleich mit der Wattleistung an der IAS, (d)-(f): Vergleich mit der Herzfrequenz an der IAS, (g)-(i): Vergleich mit der Blutlaktatkonzentration an der IAS

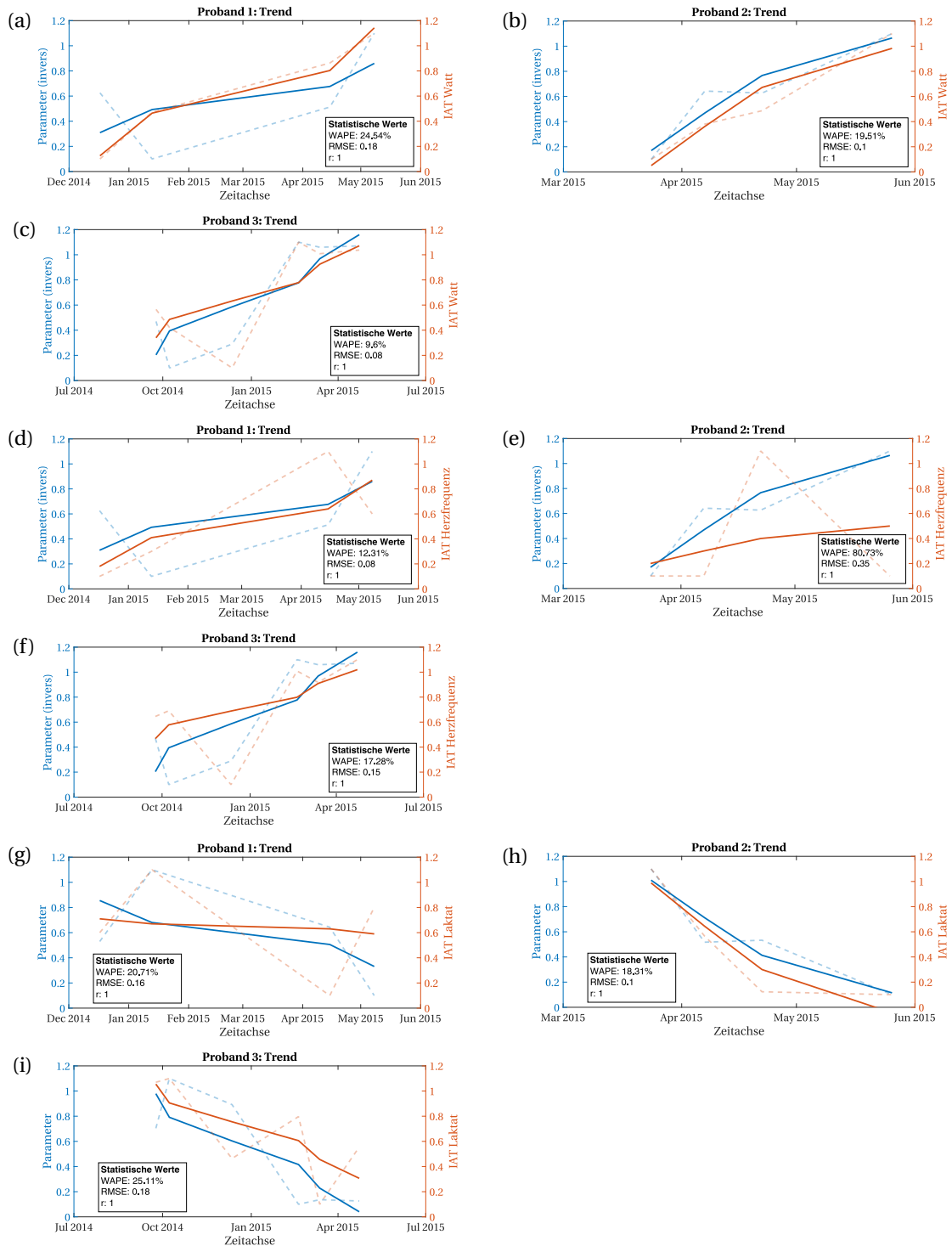


Abbildung 7.10.: Vergleich: Parameter a_2 des Faltungsmodells und Leistung anhand unterschiedlicher Schwellwerte bezüglich der geschätzten IAS für alle drei Probanden. (a)-(c): Vergleich mit der Wattleistung an der IAS, (d)-(f): Vergleich mit der Herzfrequenz an der IAS, (g)-(i): Vergleich mit der Blutlaktatkonzentration an der IAS

Eine Übersicht der statistischen Kennzahlen (WAPE, RMSE, Pearson's r) ist in Tabelle 7.5 und für die Trendanalyse in Tabelle 7.6 dargestellt.

Tabelle 7.5.: Drei statistische Maße (gerundet) der drei Probanden P1, P2, P3 mit den drei unterschiedlichen Schwellwerten für das parameterreduzierte Faltungsmodell mit a_3 als freiem Parameter

Methode/Maßzahl	WAPE [%]			RMSE			r		
	P1	P2	P3	P1	P2	P3	P1	P2	P3
IAT: Watt	49,23	19,51	16,32	0,37	0,15	0,16	0,52	0,95	0,92
IAT: Herzfrequenz	86,33	143,99	27,28	0,48	0,62	0,27	0,17	0,02	0,76
IAT: Laktat	50,51	24,22	47,89	0,44	0,21	0,41	0,25	0,89	0,53

Tabelle 7.6.: Drei statistische Maße (gerundet) der drei Probanden P1, P2, P3 mit den drei unterschiedlichen Schwellwerten für das parameterreduzierte Faltungsmodell mit a_3 als freiem Parameter im Trendverlauf

Methode/Maßzahl	WAPE [%]			RMSE			r		
	P1	P2	P3	P1	P2	P3	P1	P2	P3
IAT: Watt	24,54	19,51	9,6	0,18	0,1	0,08	1	1	1
IAT: Herzfrequenz	12,31	80,73	17,28	0,08	0,35	0,15	1	1	1
IAT: Laktat	20,71	18,31	25,11	0,16	0,1	0,18	1	1	1

Die Trendanalyse verstärkt den bereits vorhandenen Eindruck für die einzelnen Probanden und Schwellwerte: Für die Probanden 2 und 3 liegt beim Vergleich mit der Wattleistung eine nahezu perfekte Korrelation des Trendverlaufs mit minimalen Abstandswerten ($RMSE \leq 0,1$) vor, während der invertierte Parameter für Proband 1 am ehesten mit der Herzfrequenz an der geschätzten IAS zu korrelieren scheint ($RMSE < 0,1$). Beim Vergleich mit der Blutlaktatkonzentration verläuft der Trend des Parameters für alle drei Probanden jeweils ähnlich zu den individuellen Laktatwerten mit Fehlerwerten von rund 20 – 25% im WAPE bzw. RMSE-Werten von 0,1 bis $< 0,2$. Die Trendverläufe sind dabei vor allem für die Probanden 2 und 3 nahezu parallel mit minimaler Abweichung in der Steigung.

Insgesamt weisen die Ergebnisse dieses 1P-Faltungsmodells darauf hin, dass die Modellvariante mit freiem Parameter a_3 geeignet erscheint, um Rückschlüsse auf die Leistungsentwicklung eines Hobby-Athleten ziehen zu können.

7.2.3. Interpretation und Diskussion

Für die beiden betrachteten einparametrischen Faltungsmodell-Varianten zeigen sich grundlegend unterschiedliche Verläufe des freien Parameters. Entsprechend unterschiedlich ist die Korrelation zu den Leistungsschwellwerten ausgefallen: Während für den Verlauf des freien Parameters a_2 keine zuverlässige geeignete Korrelation mit einem der Schwellwerte

identifiziert werden konnte, zeigt sich der freie Parameter a_3 als durchaus für alle Probanden gut geeignetes Leistungsmaß.

Problematisch bei der Modellvariante mit freiem Parameter a_2 ist dabei neben der geringen Korrelation insbesondere, dass der Korrelationskoeffizient r je nach Proband für einen spezifischen Leistungsschwellwert positiv oder negativ ausfällt. Eine strukturelle Vereinheitlichung durch die Betrachtung des invertierten Parameters ist damit nicht global für alle Datensätze und Probanden möglich. Ein solches Modellverhalten lässt sich weder physiologisch noch phänomenologisch geeignet interpretieren.

Anders verhält es sich bei der zweiten untersuchten Modellvariante mit freiem Parameter a_3 . Der Korrelationskoeffizient ist über alle Probanden hinweg für einen spezifischen Leistungsschwellwert entweder positiv oder negativ – ein wechselndes Verhalten je nach Proband konnte nicht beobachtet werden. Auch phänomenologisch lässt sich dieser Modellparameter gut interpretieren: Als additiver Parameter in der Herzfrequenzsimulation stellt a_3 einen »Startwert« oder eine »untere Schwelle« für die simulierte Herzfrequenz dar. Physiologisch ist bekannt, dass die Ruhe-Herzfrequenz bei steigender Leistungsfähigkeit absinkt (Haber, 2009, S. 89). Durch die Ökonomisierung der Herzleistung sinkt gleichzeitig auch die Herzfrequenz unter (submaximaler) Belastung (Haber, 2009, S. 89). Es erscheint daher durchaus logisch, dass der freie Parameter a_3 als »untere Schwelle« der simulierten HF durch eine Leistungssteigerung ebenfalls absinkt. Interessant ist dabei, dass sich der Modellparameter im zeitlichen Verlauf invers zur Herzfrequenz an der geschätzten IAS und invers zur Wattleistung an der geschätzten IAS verhält, jedoch unmittelbar mit der Blutlaktatkonzentration an der geschätzten IAS korreliert. Physiologisch lässt sich dieses Verhalten wie folgt erklären:

1. Wie bereits erwähnt hat eine Leistungssteigerung ein Absinken der Herzfrequenz zur Folge;
2. Die IAS ist der Gleichgewichtszustand im Körper, bei dem die Menge an produziertem Laktat der Kapazität des Laktatabbaus entspricht;
3. Eine Veränderung der Blutlaktatkonzentration durch eine Veränderung der Leistungsfähigkeit geht damit direkt mit einer Verschiebung der IAS einher;
4. Eine Leistungssteigerung führt dazu, dass bei gleicher Wattleistung weniger Laktat produziert wird – die IAS verschiebt sich »nach hinten«;
5. Durch die Verschiebung der IAS kann die Leistungssteigerung zu höheren Werte der HF und der Wattleistung an der IAS führen.

Eine Leistungssteigerung führt also dazu, dass durch die Verschiebung der IAS die Blutlaktatkonzentration an der (geschätzten) IAS absinkt, wohingegen die Herzfrequenz und die Wattleistung an der (geschätzten) IAS ansteigen. Der freie Parameter a_3 entspricht wie bereits dargestellt einer »unteren Schwelle« der (simulierten) Herzfrequenz. Damit müsste der Wert von a_3 im Rahmen einer Leistungssteigerung über die Zeit absinken. Exakt dieses Ver-

hältnis lässt sich in der vorangegangenen Analyse beobachten: Es ergibt sich eine direkte Korrelation zwischen dem Verlauf des Modellparameters und der Blutlaktatkonzentration, sowie eine inverser linearer Zusammenhang zwischen dem Verlauf des Modellparameters und der Herzfrequenz bzw. der Wattleistung an der (geschätzten) IAS.

Insgesamt hat sich für jeden der drei Probanden eine hohe Korrelation des Modellparameters zu mindestens einem der Schwellwerte gezeigt. Die Verläufe der einzelnen Trends stimmen mit einer Ausnahme (P2 im Trendverlauf der HF) sogar bei allen Probanden für alle Schwellwerte weitgehend überein. Problematisch ist dabei jedoch die sehr geringe Anzahl an Datenpunkten, die für die Auswertung betrachtet werden konnte. Die im Rahmen dieser Analyse verwendeten Datensätze enthalten einerseits maximal sechs Leistungstest, andererseits liegen keine ausreichenden Informationen zum Training außerhalb der Tests vor. Dadurch war die Möglichkeit einer Analyse zwischen einem verbleibenden Parameter des Faltungsmodells und den betrachteten Leistungsschwellwerten vergleichsweise eingeschränkt. Auch die Durchführung einer virtuellen Leistungsdiagnostik wie in Abschnitt 7.1 war aufgrund der fehlenden Trainingsdaten nicht möglich.

Korrelation und Datenmenge

Der Korrelationskoeffizient deutet je nach betrachtetem Schwellenwert auf einen teilweise deutlichen linearen Zusammenhang zwischen der empirischen Leistung und dem Parameter des (einparametrischen) Faltungsmodells hin. Dabei ist zu beachten, dass die Datenmenge als sehr gering einzustufen ist – eine Korrelation konnte pro Proband über lediglich vier bis sechs Datenpunkte berechnet werden.

Diese geringe Menge an Daten hat gleich zwei Effekte unmittelbar zur Folge:

1. Es ist unklar, wie zufällig der ähnliche Verlauf beider Zeitreihen möglicherweise ist oder ob der postulierte lineare Zusammenhang auch über längere Zeiträume Bestand hat;
2. Bereits ein einzelner »Ausreißer« in einer der Zeitreihen übt einen großen Effekt auf den Gesamteindruck und die berechnete Korrelation aus.

Zudem ist fraglich, wie stark der Parameter des Faltungsmodells realistisch betrachtet maximal mit einem der Leistungsschwellwerte korrelieren *kann*. Abbildung 7.11 zeigt anstelle eines Vergleichs mit dem Parameter des Faltungsmodells den Vergleich zweier Leistungsschwellwerte für jeden Probanden. Beim Vergleich der Wattleistung an der geschätzten IAT und der Blutlaktatkonzentration an der geschätzten IAT zeigt sich im zeitlichen Verlauf ein ähnlich »durchwachsenes« Bild wie teilweise bei der Analyse mit dem Modellparameter.

Tabelle 7.7 und Tabelle 7.8 fassen die statistischen Kennzahlen der drei Schwellwerte untereinander bzw. ihres Trendverlaufs zusammen.

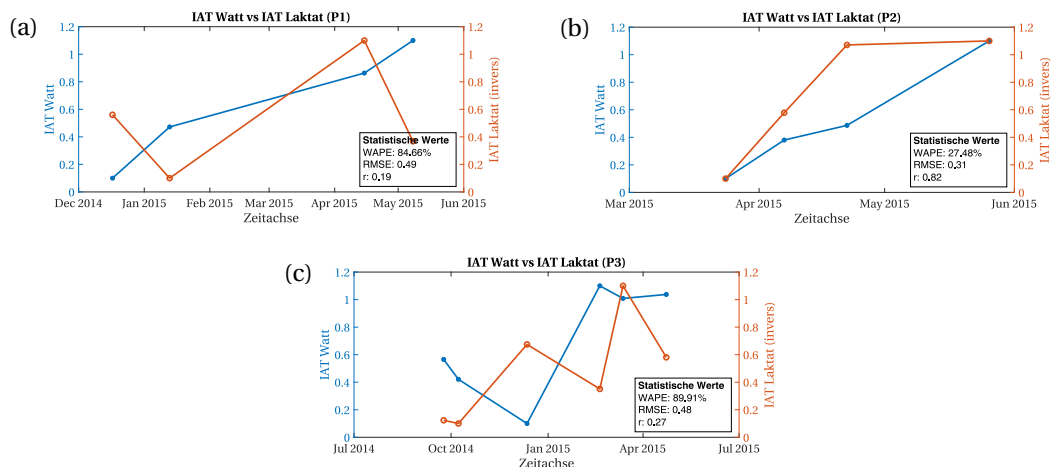


Abbildung 7.11.: Vergleich: Wattleistung an der geschätzten IAS und Blutlaktatkonzentration an der geschätzten IAS

Maßzahl	WAPE [%]			RMSE			<i>r</i>		
	P1	P2	P3	P1	P2	P3	P1	P2	P3
Watt vs. HF	43,29	135,24	13,3	0,29	0,6	0,13	0,75	-0,05	0,94
Watt vs. 1/Laktat	84,66	27,48	89,91	0,49	0,31	0,48	0,19	0,82	0,27
1/Laktat vs. HF	42,51	107,68	68,25	0,28	0,55	0,53	0,72	0,51	0,06

Tabelle 7.7.: Drei statistische Maße der drei Probanden P1, P2, P3 mit den drei unterschiedlichen Schwellwerten für die drei unterschiedlichen Schwellwerte an der geschätzten IAT untereinander

Maßzahl	WAPE [%]			RMSE			<i>r</i>		
	P1	P2	P3	P1	P2	P3	P1	P2	P3
Watt vs. HF	25,94	68,95	8,17	0,16	0,29	0,07	1	1	1
Watt vs. Laktat	55,81	27,48	44,45	0,35	0,2	0,22	1	1	1
Laktat vs. HF	35,77	105,18	34,36	0,21	0,46	0,26	1	1	1

Tabelle 7.8.: Drei statistische Maße der drei Probanden P1, P2, P3 mit den drei unterschiedlichen Schwellwerten für die drei unterschiedlichen Schwellwerte an der geschätzten IAT untereinander im Trendverlauf

Allein die große Schwankungsbreite innerhalb der Korrelationen der einzelnen Leistungsschwellwerte verdeutlicht, dass auch der Modellparameter mit maximal einem der Leistungsschwellwerte geeignet korrelieren kann, nicht jedoch mit mehreren gleichzeitig. Dabei ist fraglich, welcher Leistungsschwellwert besonders geeignet ist, ob der Zusammenhang zwischen dem Parameter und einem der Leistungsschwellwerte grundsätzlich personenabhängig ist, und wie valide die Leistungsschwellwerte an sich sind. Dabei ist zu berücksichtigen, dass auch diese Werte bereits auf einer Modellierung beruhen. Alle drei Leistungsschwellwerte basieren auf der Modellierung des Ergonizers. Dabei wird die IAS anhand der gemessenen Laktatwerte geschätzt. Anschließend wird der jeweilige Schwellwert an der ge-

geschätzten IAS über eine Spline-Interpolation berechnet. Demzufolge liegen den betrachteten Schwellwerten mehrere Fehlerquellen zugrunde:

1. Messungenauigkeiten bei der Laktatmessung;
2. Schwankungen der Blutlaktatkonzentration durch die Tagesform und äußere Einflussfaktoren (z.B. Ernährung, vorherige Belastung);
3. Modellierungsungenauigkeiten durch die Schätzung und Interpolation des Ergonizers.

Ergänzend zur Korrelation: Pearson's r und ICC

Wie auch in der vorangegangenen Studie in Abschnitt 7.1 ist der Korrelationskoeffizient nach Pearson bei Zeitreihen mit Vorsicht zu betrachten. Da wie mehrfach erläutert bei der Analyse der Leistungsentwicklung explizit der Trend von Interesse ist, wurde auch im Rahmen dieser Analyse auf eine Trendbereinigung verzichtet und explizit eine Analyse des Trendverlaufs durchgeführt. In den durchgeführten Analysen zum Trendverlauf ergibt die Berechnung des r -Werts erneut stets einen Wert von 1, wobei sich auch hier erneut lediglich feststellen lässt, dass beide Zeitreihen – die Zeitreihe des Modellparameters und die der empirischen Leistungsdaten – im Zeitverlauf dem gleichen linearen Trend in Form eines streng monotonen Anstiegs oder Abfalls folgen.

Eine Berechnung der ICC konnte im Rahmen dieser empirischen Studie nicht durchgeführt werden: Die ICC kann erst für Datenreihen mit mindestens 10 Messpunkten zuverlässig bestimmt werden. Diese Menge an Datenpunkten lag pro Proband jedoch nicht vor, sodass auf eine Berechnung der ICC verzichtet werden musste.

Aussagekraft der Abstandsmaße (WAPE, RMSE)

Wie auch in der vorangegangenen Studie in Abschnitt 7.1 sind für den WAPE und den RMSE als Fehlermaße bei der Interpretation der Ergebnisse die gleichen zwei Aspekte zu berücksichtigen:

1. Die betrachteten Daten wurden normiert;
2. Während die empirische Leistung (ohne Normierung) in kJ/kg angegeben werden kann, sind die Parameter abstrakter Natur und dimensionslos.

Der WAPE wurde wie eingangs erläutert auf Grund der sehr geringen Menge an Datenpunkten anstelle des MAPE verwendet, kann aber ansonsten analog interpretiert werden.

Die Art Normierung kann bereits Einfluss darauf haben, wie nah die empirische und die simulierte Kurve beieinander liegen. Aus diesem Grund wurden beide Zeitreihen in den durchgeführten Experimenten ohne nähere Betrachtung linear auf das Intervall $[0,1; 1,1]$

normiert. Der Wert 0 wurde dabei nicht verwendet, um eine adäquate Umrechnung der Parameter zu gewährleisten. Letztendlich spielt die genaue Positionierung der Zeitreihen zueinander jedoch auch nur eine untergeordnete Rolle: Ziel der Untersuchung war es, die Ähnlichkeit im Verlauf zwischen den beiden Zeitreihen zu analysieren. Dabei spielen weder die spezifische Skalierung, noch die Abstände untereinander eine wesentliche Rolle. Zentral bei der Betrachtung des Verlaufs sind die grundlegenden Trendentwicklungen und ob es mit Hilfe des Parameters möglich ist, einen Leistungszuwachs und eine Leistungsabnahme tendenziell über die Zeit voneinander unterscheiden zu können. In diesem Sinne ist auch das Fehlen einer Dimensionsgröße bei dem freien Modellparameter kein Hindernis: Der Vergleich, ob und inwieweit sich der Parameter über die Zeit ähnlich zu empirischen Leistungswerten entwickelt, ist nicht von einer spezifischen Einheit des Parameters abhängig und kann ebenso gut abstrakt betrachtet werden.

Umgekehrt folgt daraus jedoch auch, dass die Aussagekraft der exakten Fehlerwerte durch die Normierung nicht als allgemeingültig angenommen werden kann. Die Fehlermaße können als Vergleichswerte lediglich in einem ebenso aufgebauten Setting verwendet werden. Für die hier durchgeführten Vergleiche unterschiedlicher empirischer Leistungswerte und zur besseren Einschätzung der Ähnlichkeit zwischen den beiden Zeitreihen erscheinen sie grundsätzlich geeignet.

Modellauswahl

Ein weiterer zentraler Aspekt der Validität der Analyse betrifft die Modellauswahl. Für das parameterreduzierte Faltungsmodell konnten zwei unterschiedliche Modellvarianten (a_2 bzw. a_3 als freier Parameter) identifiziert werden. Die Analyse zur Parameterreduktion hat dabei auch gezeigt, dass die exakte Wahl von beispielsweise a_1 in einem spezifischen Wertebereich keinen Einfluss auf das Ergebnis einer Fitting- oder Vorhersagesimulation mit dem parameterreduzierten Faltungsmodell nimmt. Entsprechend könnten ebenso gut noch weitere Modellvariationen mit anderen konstant gewählten Parameterwerten betrachtet werden. Es hat sich jedoch gezeigt, dass leichte Veränderungen der konstant gewählten Parameter keinen Einfluss auf den grundlegenden Verlauf des freien Parameters haben – während der exakte Wert des freien Parameters durch unterschiedliche Werte von a_1 leicht variiert, bleibt das Verhältnis der Parameter im zeitlichen Verlauf bestehen. Die hier durchgeführten Analysen sind damit unabhängig von der spezifischen Wahl des Wertes von a_1 .

Abbildung 7.12 zeigt den Verlauf des freien Parameters a_2 in Modellvariante 1 (Abbildungsteil (a)) bzw. den Verlauf des freien Parameters a_3 in Modellvariante 2 (Abbildungsteil (b)) exemplarisch für vier verschiedene Werte von a_1 anhand von Proband 1.

Gleiches gilt für die unterschiedlichen möglichen Werte von a_2 in Modellvariante 2 (freier Parameter a_3), wie sich in Abbildung 7.12 (c) erkennen lässt: Auch hier kamen während der Parameterreduktion unterschiedliche Werte für a_2 in Frage. Wird a_2 auf einen anderen Wert

7. Nutzen des parameterreduzierten Faltungsmodells bei der Leistungsbeurteilung

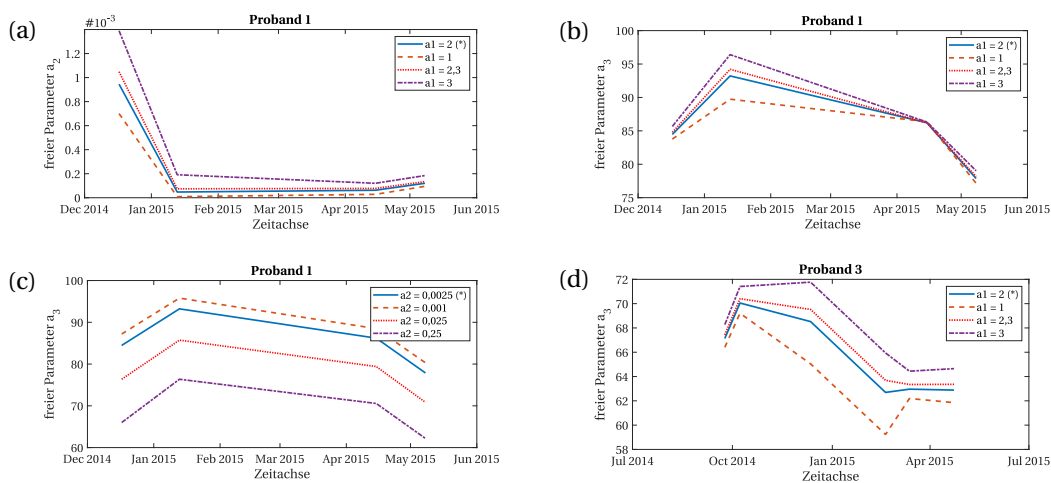


Abbildung 7.12.: Vergleich: Parameterverlauf bei unterschiedlichen konstanten Parameterwerte für a_1 bzw. a_2

fixiert, ändert sich jedoch lediglich die Skalierung des Parameters, nicht sein zeitlicher Verlauf.

Lediglich bei Proband 3 gibt es bei der Wahl von $a_1 = 1$ in beiden Modellvarianten eine leichte Abweichung im vorletzten Wert des freien Parameters, exemplarisch für Modellvariante 2 dargestellt in Abbildung 7.12 (d). Allerdings hatte sich bereits innerhalb von Abschnitt 6.3: »Konstruktion eines einparametrischen Faltungsmodells: Studiendaten von Hobby-Sportlern« gezeigt, dass die Fehlerwerte von $a_1 = 1$ in der Vorhersage signifikant schlechter ausfallen als für die meisten anderen konstanten Werte für a_1 . Die Abweichung im Verlauf des freien Parameters für $a_1 = 1$ bei Proband 3 kann daher vernachlässigt werden. Die restlichen Vergleiche der unterschiedlichen konstanten Werte für die anderen Probanden sind im Anhang in Abbildung A.13 (S. 261) dargestellt.

Auf eine ausführliche Analyse derartiger Modellvarianten mit verschiedenen Werten von a_1 bzw. a_2 kann folglich verzichtet werden.

7.2.4. Fazit zur Studie

Für das Faltungsmodell konnten zuvor innerhalb von Abschnitt 6.3, »Konstruktion eines einparametrischen Faltungsmodells: Studiendaten von Hobby-Sportlern«, zwei unterschiedliche einparametrische Varianten identifiziert werden: Ein Modell, in dem der multiplikative Parameter a_2 zur Individualisierung frei bleibt, und eine Modellvariante mit freiem additiven Parameter a_3 . Im Rahmen dieser empirischen Studie wurde der jeweilige Modellparameter im zeitlichen Verlauf für drei Probanden mit drei unterschiedlichen Leistungsschwellwerten verglichen: Der Herzfrequenz, der Wattleistung und der Blutlaktatkonzentration an der geschätzten IAS.

Es hat sich gezeigt, dass die Modellvariante mit freiem Parameter a_2 keine interpretierbaren Rückschlüsse auf die Leistungsfähigkeit zulässt, wohingegen die Modellvariante mit freiem Parameter a_3 sehr gut zur Einschätzung von Leistungsveränderungen geeignet erscheint. Für diese Modellvarianten konnten für zwei der Probanden im Vergleich mit mindestens einem Leistungsschwellwert stets ein mindestens deutlicher, teilweise sogar starker linearer Zusammenhang anhand des Korrelationskoeffizienten und der visuellen Betrachtung identifiziert werden. In der Trendanalyse zeigten sich bei acht von neun Auswertungen hohe lineare Zusammenhänge zwischen dem Modellparameter und den unterschiedlichen Leistungsschwellwerten.

Trotz der geringen Anzahl an Datenpunkten liegt der Schluss nahe, dass das einparametrische Faltungsmodell mit freiem Parameter a_3 grundsätzlich Rückschlüsse auf die Leistungsfähigkeit in der Radsportpraxis zulässt.

7.3. Zusammenfassung und Fazit

Ziel dieses Kapitels war es, die dritte, zentrale Fragestellung zu untersuchen:

Fragestellung 3:

Trägt der aus der Herzfrequenzmodellierung stammenden Parameter des parameterreduzierten Faltungsmodells basierend auf alltäglichem Training zur Gewinnung von Rückschlüssen auf die Leistungsentwicklung einer Athletin oder eines Athleten im Ausdauersport bei?

Zur Analyse dieser Fragestellung wurden zwei große Vorarbeiten verwendet: Das zuvor in Kapitel 5: »Das Faltungsmodell: Ein neuer Ansatz zur Simulation und Vorhersage der Herzfrequenz im Training« entwickelte und evaluierte Faltungsmodell, sowie die Weiterentwicklung zu einem einparametrischen Faltungsmodell aus Kapitel 6: »Verbesserung des Faltungsmodells: Parameterreduktion«. Darauf basierend wurde in zwei empirischen Studien der freie Parameter einer auf den jeweiligen Datensatz abgestimmten innerhalb von Kapitel 6 identifizierten Modellvariante des Faltungsmodells im zeitlichen Verlauf näher untersucht und mit empirischen Messergebnissen zur Leistungsentwicklung verglichen.

In beiden empirischen Studien hat sich gezeigt, dass das einparametrische Faltungsmodell mit freiem additiven Parameter a_3 großes Potenzial hat, um Rückschlüsse auf die Leistungsfähigkeit im Radsport zu erlauben.

Innerhalb der ersten Studie hat sich gezeigt, dass je nach verwendeten Daten der Einfluss tagesaktueller Schwankungen und externer Faktoren auf die Herzfrequenz sehr groß zu sein scheint. Wurden diese Einflüsse durch die Parameterkalibrierung auf mehreren Datensätzen (im Rahmen einer virtuellen Leistungsdiagnostik) oder durch eine Analyse des Trend-

verlaufs abgeschwächt, zeigte sich bei zwei von vier Probanden ein deutlicher linearer Zusammenhang des Parameters zu den empirischen Leistungswerten. Für die anderen beiden Probanden wurden mögliche Ursachen identifiziert, die das Ergebnis der Untersuchungen negativ beeinträchtigt haben könnten.

In der zweiten Studie hat sich in der Betrachtung unterschiedlicher Leistungsschwellwerte in acht von neun Fällen ein deutlicher Zusammenhang zwischen dem Parameter und der empirischen Leistung im zeitlichen Verlauf gezeigt. Das Verhalten des Modellparameters konnte zudem phänomenologisch und physiologisch im Kontext der Leistungsentwicklung interpretiert werden.

Die durchgeführten Experimente haben damit gezeigt, dass letztlich auch Fragestellung 3 positiv beantwortet werden kann: Der aus der Herzfrequenzmodellierung stammenden Parameter des parameterreduzierten Faltungsmodells trägt basierend auf alltäglichem Training zur Gewinnung von Rückschlüssen auf die Leistungsentwicklung einer Athletin oder eines Athleten im Ausdauersport bei.

Zusammenfassung und Ausblick



Abstract.

Die Inhalte und Ergebnisse der einzelnen Kapitel werden abschließend kurz zusammengefasst. Basierend auf den Antworten auf die zentrale Fragestellung wird ein Ausblick gegeben, inwieweit eine weiterführende Validierung der Ergebnisse zweckdienlich erscheint und wozu das 1P-Faltungsmodell in der Sportpraxis verwendet werden könnte.

Die übergeordnete Zielsetzung der vorliegenden Arbeit war die Analyse, inwieweit es mithilfe eines möglichst einfachen mathematisch-phänomenologischen Modells möglich ist, basierend auf regulären Trainingseinheiten und der Herzfrequenzmessung Rückschlüsse auf die Leistungsentwicklung einer Sportlerin oder eines Sportlers zu ziehen. Dazu wurden zunächst Grundlagen aus den Bereichen der Physiologie, Trainingswissenschaft, Modelltheorie und Statistik innerhalb der Kapitel 1 und 2 aufbereitet, die für die Evaluation und Interpretation der empirischen Studien notwendig waren. Darauf basierend konnten drei Fragestellungen abgeleitet werden (Kapitel 3), mithilfe derer die Zielsetzung unterteilt und in drei unmittelbar aufeinander aufbauende Themenschwerpunkte gegliedert werden konnte. Diese drei Themenschwerpunkte bildeten den Kern der Arbeit im empirischen Teil.

Nach einer Vorstellung der verwendeten Datensätze (Kapitel 4) wurde in Kapitel 5 in einem ersten Schritt mit dem *Faltungsmodell* ein neues Modell zur Simulation und Vorhersage der Herzfrequenz während des Radtrainings entwickelt. Es existieren zwar in der Literatur bereits zahlreiche ähnliche Modelle, diese werden üblicherweise jedoch nur zur Simulation der Herzfrequenz oder der Vorhersage weniger Sekunden (oft im Rahmen einer PID-Regelung) verwendet. Eine frühere Studie hat gezeigt, dass manche dieser Modelle zwar grundsätzlich auch zur Vorhersage vollständiger Trainingseinheiten verwendet werden können, in der Regel jedoch mit einer Vielzahl von freien Parametern kalibriert werden müssen. Je mehr Parameter ein Modell hat, desto mehr Daten werden zur Kalibrierung benötigt – und desto eher neigt das Modell dazu, durch die Kalibrierung an die dort verwendeten Daten »überangepasst« (engl. *overfitting*) zu werden. Auch eine physiologische oder phänomenologische Interpretation der Modellparameter ist in diesen Modellen kaum möglich, sodass auf Basis einzelner Parameter in der Regel keine Rückschlüsse auf die Leistungsfähigkeit einer Person gezogen werden können.

Im Rahmen einer ersten empirischen Studie konnte das Faltungsmodell validiert und mit den Ergebnissen eines Wiener-Modells aus der Literatur verglichen werden. Dabei hat sich gezeigt, dass sich die Ergebnisse des Faltungsmodells und des Wiener-Modells in der Simulation und der Vorhersage nicht signifikant voneinander unterscheiden. Im Gegensatz zu dem in der Literatur verwendeten »Black Box«-Wiener-Modell hat das Faltungsmodell den Vorteil, dass es mit lediglich vier freien Parametern individualisiert werden und zudem phänomenologisch interpretiert werden kann. Eine mathematisch-strukturelle Analyse hat ergeben, dass das Faltungsmodell ebenfalls ein Wiener-Modell ist – während die anderen betrachteten Differentialgleichungsmodelle zur Klasse der Hammerstein-Modelle gehören.

In einem zweiten Schritt wurde das Faltungsmodell innerhalb von Kapitel 6 weiter optimiert, um eine mögliche Überanpassung an die Kalibrierungsdaten zu verringern und eine wechselseitige Beeinflussung verschiedener Parameter zu reduzieren. Innerhalb von zwei empirischen Studien mit unterschiedlichen Probandendaten konnte gezeigt werden, dass sich die Anzahl der Parameter über die dort präsentierte Vorgehensweise ohne signifikanten Genauigkeitsverlust in der Vorhersage deutlich reduzieren lässt. In beiden Studien konnte mindes-

tens eine Variante des Faltungsmodells mit nur noch einem freien Parameter identifiziert werden. In einer der beiden Studien wurden die Daten von Profi-Athleten verwendet. Für diese konnte zusätzlich gezeigt werden, dass mit dem einparametrischen Faltungsmodell die Herzfrequenz auch auf regulären Outdoor-Trainingsdaten valide simuliert und prognostiziert werden kann. Innerhalb der anderen Studie konnten zwei Modellvarianten mit nur noch einem freien Parameter identifiziert werden.

Abschließend wurden die identifizierten Faltungsmodelle mit einem freien Parameter verwendet, um einen möglichen Zusammenhang zwischen dem freien Parameter und der Leistungsentwicklung über die Zeit zu untersuchen. Das zuvor entwickelte Faltungsmodell wurde damit auf einen neuen Anwendungsbereich übertragen. Analog zur Analyse der zweiten Fragestellung wurden auch hier zwei empirische Studien mit unterschiedlichen Probandendaten durchgeführt. Dabei war vor allem der Trendverlauf über die Zeit von Interesse. Leistungsschwankungen über die einzelnen Tage hingegen hatten für die betrachtete Fragestellung keine Relevanz. Aus diesem Grund wurde auf eine Trendbereinigung verzichtet. Die Zeitverläufe wurden sowohl ohne Trendbereinigung, als auch speziell hinsichtlich ihres Trendverlaufs analysiert. Für die beiden Studien in Kapitel 7 wurde der Modellparameter mit unterschiedlichen Leistungsparametern verglichen: Während in der ersten empirischen Studie zu dieser Fragestellung (Abschnitt 7.1) die kumulierte Wattleistung bis zum Testabbruch zur Verfügung stand, wurden in der zweiten Studie (Abschnitt 7.2) drei verschiedene Leistungsschwellwerte basierend auf der gemessenen Blutlaktatkonzentration betrachtet. Innerhalb der ersten Studie wurde ergänzend eine »virtuelle Leistungsdiagnostik« betrachtet, indem reguläre Outdoor-Trainingseinheiten zur Kalibrierung genutzt und der so identifizierte Parameter auf die zeitlich anschließende Ergometer-Einheit zur Validierung angewendet wurde.

Insgesamt hat sich gezeigt, dass das Faltungsmodell mit freiem additiven Parameter im Trendverlauf Rückschlüsse auf die Leistungsfähigkeit im Radsport erlaubt. Während der Einfluss tagesaktueller Schwankungen und externer Faktoren auf die Herzfrequenz anscheinend sehr groß ist, zeigten die Analysen auf den Trendverläufen wie auch basierend auf mehreren Kalibrierungsdaten (im Rahmen der virtuellen Leistungsdiagnostik) bei zwei von vier Probanden einen deutlichen linearen Zusammenhang des Parameters zu den empirischen Leistungswerten in der ersten Studie. In der zweiten Studie konnte in acht von neun Fällen ein deutlicher Zusammenhang zwischen dem Parameter und der empirischen Leistung im zeitlichen Verlauf identifiziert werden. Zusätzlich war es möglich, das Verhalten des Modellparameters phänomenologisch und physiologisch im Kontext der Leistungsentwicklung zu interpretieren.

Es konnte damit schließlich gezeigt werden, dass das Faltungsmodell in seiner Variante mit dem additiven Herzfrequenzparameter als einzigem freien Parameter verwendet werden kann, um Rückschlüsse auf die Leistungsfähigkeit im Radsport zu ziehen. Grundsätzlich ist es möglich, das Modell dazu auf regulären Outdoor-Trainingseinheiten zu kalibrieren,

sodass möglicherweise auf eine spezifische Leistungsdiagnostik verzichtet werden kann, sofern es lediglich um eine Kontrolle der Tendenz innerhalb der Leistungsentwicklung geht. Es ist denkbar, dass auch andere Modelle der Herzfrequenzsimulation dazu geeignet sind, Rückschlüsse auf die Leistungsfähigkeit zu ermöglichen. Das Faltungsmodell hat jedoch den Vorteil, dass derartige Zusammenhänge durch die Betrachtung des einen freien Parameters leicht untersucht werden und phänomenologisch interpretiert werden können.

Die Ergebnisse dieser Arbeit insbesondere hinsichtlich möglicher Rückschlüsse auf die Leistungsentwicklung sollten in künftigen Forschungsarbeiten basierend auf größeren Datensätzen verifiziert werden. Gerade hinsichtlich der sehr guten Korrelationsergebnisse zwischen dem Parameter des 1P-Faltungsmodells und den Leistungsschwellwerten aus der Laktatdiagnostik (Abschnitt 7.2) ist es notwendig, diesen Zusammenhang näher und auf einer größeren Datenbasis zu untersuchen. Wenn die hier dargestellten Forschungsergebnisse dabei verifiziert werden können, könnte das 1P-Faltungsmodell grundsätzlich zur Überprüfung der Leistungsentwicklung basierend auf regulären Trainingseinheiten und ohne Durchführung einer spezifischen Leistungsdiagnostik herangezogen werden. Ob die Qualität des Parameters im Zeitverlauf im Leistungssportbereich ausreicht oder eher eine Unterstützung im Breitensport darstellt, müsste in diesem Kontext ebenfalls evaluiert werden.

Literaturverzeichnis

- Achten, J. und Jeukendrup, A. E. (2003). Heart rate monitoring. *Sports medicine*, 33(7):517–538.
- Allen, H. und Coggan, A. (2012). *Wattmessung im Radsport und Triathlon*. Spomedis.
- Amrhein, V., Greenland, S., und McShane, B. (2019). Scientists rise up against statistical significance.
- Amrhein, V., Korner-Nievergelt, F., und Roth, T. (2017). The earth is flat ($p > 0.05$): significance thresholds and the crisis of unreplicable research. *PeerJ*, 5:e3544.
- Andreacci, J. L., Lemura, L. M., Cohen, S. L., Urbansky, E. A., Chelland, S. A., und Duvillard, S. P. v. (2002). The effects of frequency of encouragement on performance during maximal exercise testing. *Journal of sports sciences*, 20(4):345–352.
- Anscombe, F. J. (1973). Graphs in statistical analysis. *The american statistician*, 27(1):17–21.
- Argha, A., Ye, L., Su, S. W., Nguyen, H., und Celler, B. G. (2016). Real-time modelling of heart rate response during exercise using a novel constrained parameter estimation method. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, Seiten 2680–2683. IEEE.
- Asteroth, A. (2000). *Effiziente Identifikation parametrisierter Kreislaufmodelle (Efficient identification of parameterized circulatory models)*. Dissertation, Universitäts- und Landesbibliothek Bonn.
- Åstrand, P.-O., Rodahl, K., Dahl, H. A., und Strømme, S. B. (1986). *Textbook of work physiology*. USA: Mc Graw-Hill.
- Baig, D.-e.-Z., Su, H., Cheng, T. M., Savkin, A. V., Su, S. W., und Celler, B. G. (2010). Modeling of human Heart Rate response during walking, cycling and rowing. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, Seiten 2553–2556. IEEE.
- Becque, M., Katch, V., Marks, C., und Dyer, R. (1993). Reliability and within subject variability of VE, VO₂, heart rate and blood pressure during submaximum cycle ergometry. *International journal of sports medicine*, 14(04):220–223.
- Bellenger, C. R., Fuller, J. T., Thomson, R. L., Davison, K., Robertson, E. Y., und Buckley, J. D. (2016). Monitoring athletic training status through autonomic heart rate regulation: a systematic review and meta-analysis. *Sports medicine*, 46(10):1461–1486.

- Bennett, F., Christini, D., Ahmed, H., Lutchen, K., Hausdorff, J., und Oriol, N. (1993). Time series modeling of heart rate dynamics. In *Computers in Cardiology 1993, Proceedings.*, Seiten 273–276. IEEE.
- Berend, J., Brammeier, M., Jones, N., Holliman, S., und Hackney, A. (1994). Effect of the menstrual cycle phase and diet on blood lactate responses to exercise. *Biology of sport*, 11(4):241.
- Billings, S. und Fakhouri, S. (1977). Identification of nonlinear systems using the Wiener model. *Electronics letters*, 13(17):502–504.
- Billman, G. E. (2011). Heart rate variability—a historical perspective. *Frontiers in physiology*, 2:86.
- Bischoff, D. (2002). *Die Bestimmung des Sauerstoffverbrauchs bei Patienten mit Sepsis: ein Methodenvergleich zwischen der Berechnung mit der Fickschen Formel und der indirekten Kalorimetrie*. Dissertation, Freie Universität Berlin.
- Bland, J. M. und Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*, 327(8476):307–310.
- Bodner, M. E. und Rhodes, E. C. (2000). A review of the concept of the heart rate deflection point. *Sports Medicine*, 30(1):31–46.
- Bonate, P. L. (2011). *Pharmacokinetic-pharmacodynamic modeling and simulation*. Springer.
- Boos, D. D. und Stefanski, L. A. (2011). P-value precision and reproducibility. *The American Statistician*, 65(4):213–221.
- Borresen, J. und Lambert, M. I. (2008). Autonomic control of heart rate during and after exercise. *Sports medicine*, 38(8):633–646.
- Borresen, J. und Lambert, M. I. (2009). The quantification of training load, the training response and the effect on performance. *Sports medicine*, 39(9):779–795.
- Borszcz, F. K., Tramontin, A. F., de Souza, K. M., Carminatti, L. J., und Costa, V. P. (2018). Physiological correlations with short, medium, and long cycling time-trial performance. *Research quarterly for exercise and sport*, 89(1):120–125.
- Boucsein, W. (2000). *Engineering psychophysiology: issues and applications*. CRC Press.
- Brisswalter, J. und Legros, P. (1994). Daily stability in energy cost of running, respiratory parameters and stride rate among well-trained middle distance runners. *International journal of sports medicine*, 15(05):238–241.
- Brooke, J. D. und Hamley, E. J. (1972). The heart-rate - physical work curve analysis for the prediction of exhausting work ability. *Medicine and Science in Sports*, 4(1):23–26.
- Buchheit, M., Chivot, A., Parouty, J., Mercier, D., Al Haddad, H., Laursen, P., und Ahmaidi, S. (2010). Monitoring endurance running performance using cardiac parasympathetic function. *European journal of applied physiology*, 108(6):1153–1167.
- Buchheit, M., Millet, G. P., Parisy, A., Pourchez, S., Laursen, P. B., und Ahmaidi, S. (2008). Supramaximal training and postexercise parasympathetic reactivation in adolescents. *Medicine and science in sports and exercise*, 40(2):362–371.

- Buchheit, M., Simpson, B. M., Schmidt, W. F., Aughey, R. J., Soria, R., Hunt, R. A., Garvican-Lewis, L. A., Pyne, D. B., Gore, C. J., und Bourdon, P. C. (2013). Predicting sickness during a 2-week soccer camp at 3600 m (ISA3600). *British Journal of Sports Medicine*, 47(Suppl 1):i124–i127.
- Buchman, T. G., Stein, P. K., und Goldstein, B. (2002). Heart rate variability in critical illness and critical care. *Current opinion in critical care*, 8(4):311–315.
- Bunc, V., Heller, J., und Leso, J. (1988). Kinetics of heart rate responses to exercise. *Journal of sports sciences*, 6(1):39–48.
- Busso, T., Denis, C., Bonnefoy, R., Geysant, A., und Lacour, J.-R. (1997). Modeling of adaptations to physical training by using a recursive least squares algorithm. *Journal of applied physiology*, 82(5):1685–1693.
- Calvert, T. W., Banister, E. W., Savage, M. V., und Bach, T. (1976). A systems model of the effects of training on physical performance. *IEEE Transactions on Systems, Man and Cybernetics*, 6(2):94–102.
- Carter, J. B., Banister, E. W., und Blaber, A. P. (2003). Effect of endurance exercise on autonomic control of heart rate. *Sports medicine*, 33(1):33–46.
- Cheng, T. M., Savkin, A. V., Celler, B. G., Su, S. W., Wang, L., u a. (2008). Nonlinear modeling and control of human heart rate response during exercise with various work load intensities. *IEEE Transactions on biomedical engineering*, 55(11):2499–2508.
- Cheng, T. M., Savkin, A. V., Celler, B. G., Wang, L., und Su, S. W. (2007a). A nonlinear dynamic model for heart rate response to treadmill walking exercise. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, Seiten 2988–2991. IEEE.
- Cheng, T. M., Savkin, A. V., Celler, B. G., Wang, L., und Su, S. W. (2007b). A nonlinear dynamic model for heart rate response to treadmill walking exercise. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, Seiten 2988–2991. IEEE.
- Chinarro, D. (2014). *System Engineering Applied to Fuenmayor Karst Aquifer (San Julián de Banzo, Huesca) and Collins Glacier (King George Island, Antarctica)*. Springer.
- Christini, D. J., Bennett, F., Lutchen, K., Ahmed, H., Hausdorff, J., und Oriol, N. (1995). Application of linear and nonlinear time series modeling to heart rate dynamics analysis. *IEEE Transactions on biomedical engineering*, 42(4):411–415.
- Clénin, G. (2019). Leistungsdiagnostik im Ausdauersport–anaerobe Schwelle, VO₂max, aerobe Kapazität - wohin geht die Reise? *Swiss Sports & Exercise Medicine*, 67(1):6–14.
- Colquhoun, D. (2014). An investigation of the false discovery rate and the misinterpretation of p-values. *Royal Society open science*, 1(3):140216.
- Conconi, F., Ferrari, M., Ziglio, P. G., Droghetti, P., und Codeca, L. (1982). Determination of the anaerobic threshold by a noninvasive field test in runners. *Journal of Applied Physiology: Respiratory, Environmental and Exercise Physiology*, 52(4):869–873.

- Cumming, G. (2008). Replication and p intervals: p values predict the future only vaguely, but confidence intervals do much better. *Perspectives on psychological science*, 3(4):286–300.
- Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.
- Daanen, H. A., Lamberts, R. P., Kallen, V. L., Jin, A., und Van Meeteren, N. L. (2012). A systematic review on heart-rate recovery to monitor changes in training status in athletes. *International journal of sports physiology and performance*, 7(3):251–260.
- Davis, J. A. und Convertino, V. A. (1975). A comparison of heart rate methods for predicting endurance training intensity. *Medicine and Science in Sports*, 7(4):295–298.
- de Smet, D., Francaux, M., Hendrickx, J. M., und Verleysen, M. (2016). Heart rate modelling as a potential physical fitness assessment for runners and cyclists. In *Proceedings of the Machine Learning and Data Mining for Sports Analytics Workshop at ECML/PPKD. Riva del Garda, Italy*.
- Dickhuth, H.-H., Huonker, M., Münzel, T., Drexler, H., Berg, A., und Keul, J. (1991). Individual anaerobic threshold for evaluation of competitive athletes and patients with left ventricular dysfunction. In *Advances in ergometry*, Seiten 173–179. Springer.
- Dickhuth, H.-H., Mayer, F., Röcker, K., und Berg, A. (2007). *Sportmedizin für Ärzte*. Köln: Deutscher Ärzteverlag.
- Dörr, C. (2010). *Untersuchung der Validität verschiedener Laktatschwellenkonzepte an Ausdauersportlern*. Dissertation, Universitätsbibliothek Giessen.
- Dur-e Zehra Baig, H. S., Cheng, T. M., Savkin, A. V., Su, S. W., und Celler, B. G. (2010). Modeling of Human Heart Rate Response during Walking, Cycling and Rowing. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, Seiten 2553–2556. IEEE.
- El-Amrawy, F. und Nounou, M. I. (2015). Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial? *Healthcare informatics research*, 21(4):315–320.
- Endler, S. (2013). *Adaptation of the meta model PerPot to endurance running for optimization of training and competition*. Dissertation, Johannes Gutenberg University Mainz.
- Endler, S. und Friedrich, O. K. (2016). Determining reference speeds and heart rates for endurance running prescription in altitude profile. In Lamas, L. und Russomanno, T., Herausgeber, *Proceedings of the 2016 International Association of Computer Science in Sport (IACSS) Conference*, Seiten 96 – 101, Brasilia, Brazil.
- Endler, S., Hoffmann, S., Sterzing, B., Simon, P., und Pfeiffer, M. (2017). The PerPot simulated anaerobic threshold : a comparison to typical lactate-based thresholds. *International journal of human movement and sports sciences*, 5(1):9 – 15.
- Eskinat, E., Johnson, S. H., und Luyben, W. L. (1991). Use of Hammerstein models in identification of nonlinear systems. *AIChE Journal*, 37(2):255–268.

- Esmaeili, A. und Ibeas, A. (2016). Particle Swarm Optimization modelling of the heart rate response in treadmill exercise. In *System Theory, Control and Computing (ICSTCC), 2016 20th International Conference on*, Seiten 613–618. IEEE.
- Fairbairn, M. S., Blackie, S. P., McElvaney, N. G., Wiggs, B. R., Pare, P. D., und Parady, R. L. (1994). Prediction of heart rate and oxygen uptake during incremental and maximal exercise in healthy adults. *Chest*, 105(5):1365–1369.
- Feliu, J., Ventura, J., Segura, R., Rodas, G., Riera, J., Estruch, A., Zamora, A., und Capdevila, L. (1999). Differences between lactate concentration of samples from ear lobe and the finger tip. *J Physiol Biochem*, 55(4):333–339.
- Füller, M., Meenakshi Sundaram, A., Ludwig, M., Asteroth, A., und Prassler, E. (2015). Modeling and Predicting the Human Heart Rate During Running Exercise. In *Information and Communication Technologies for Ageing Well and e-Health*, Band 578, Seiten 106–125. Springer International Publishing.
- Gans, D. (1981). Use of a preliminary test in comparing two sample means: Use of a preliminary test. *Communications in Statistics-Simulation and Computation*, 10(2):163–174.
- Gillinov, S., Etiwy, M., Wang, R., Blackburn, G., Phelan, D., Gillinov, A. M., Houghtaling, P., Javadikasgari, H., und Desai, M. Y. (2017). Variable Accuracy of Wearable Heart Rate Monitors during Aerobic Exercise. *Medicine and science in sports and exercise*, 49(8):1697.
- Gilman, M. und Wells, C. (1993). The use of heart rates to monitor exercise intensity in relation to metabolic variables. *International journal of sports medicine*, 14(06):339–344.
- Gilman, M. B. (1996). The use of heart rate to monitor the intensity of endurance training. *Sports Medicine*, 21(2):73–79.
- Girard, C., Ibeas, A., Vilanova, R., und Esmaeili, A. (2016). Robust discrete-time linear control of heart rate during treadmill exercise. In *Electrical Engineering (ICEE), 2016 24th Iranian Conference on*, Seiten 1113–1118. IEEE.
- Gnehm, P., Reichenbach, S., Altpeter, E., Widmer, H., und Hoppeler, H. (1997). Influence of different racing positions on metabolic cost in elite cyclists. *Medicine and science in sports and exercise*, 29(6):818–823.
- Gonzalez, A. A., Bertschinger, R., Brosda, F., Dahmen, T., Thumm, P., und Saupe, D. (2015). Modeling Oxygen Dynamics under Variable Work Rate. In *Proceedings of the 3rd International Congress on Sport Sciences Research and Technology Support*, Seiten 198–207.
- Gonzalez, A. A., Bertschinger, R., und Saupe, D. (2016). Modeling VO₂ and VCO₂ with Hammerstein-Wiener Models. In *Proceedings of the 4th International Congress on Sport Sciences Research and Technology Support (icSPORTS 2016)*.
- Graf, C. und Höher, J. (2009). *Fachlexikon Sportmedizin: Bewegung, Fitness und Ernährung von A-Z*. Deutscher Ärzteverlag.
- Griffin, M. P., O’Shea, T. M., Bissonette, E. A., Harrell, F. E., Lake, D. E., und Moorman, J. R. (2003). Abnormal heart rate characteristics preceding neonatal sepsis and sepsis-like illness. *Pediatric research*, 53(6):920–926.

- Grodins, F. S. (1959). Integrative Cardiovascular Physiology: A Mathematical Synthesis of Cardiac and Blood Vessel Hemodynamics. *The Quarterly Review of Biology*, 34(2):93–116. PMID: 13675404.
- Gundlach, H. (1980). *Zu den Strukturmerkmalen der Leistungsfähigkeit, der Wettkampfleistung und des Trainingsinhaltes in den Schnellkraft- und Ausdauersportarten*. Dissertation, Leipzig: DHfK.
- Haber, P. (2009). *Leitfaden zur medizinischen Trainingsberatung – Rehabilitation bis Leistungssport*, Band 3. Springer-Verlag.
- Hagg, A., Asteroth, A., Rasche, C., Bach, K., und Pfeiffer, M. (2021). *Künstliche Intelligenz für den Spitzensport im Spannungsfeld zwischen Big und Small Data (KISs-BiS)*, Band 3 in 2021. SPORTVERLAG Strauß, 1. Auflage. Schriftenreihe des Bundesinstituts für Sportwissenschaft.
- Hartmann, R., Schmidt, F. M., Sander, C., und Hegerl, U. (2019). Heart rate variability as indicator of clinical state in depression. *Frontiers in psychiatry*, 9:735.
- Heaps, C. L., Gonzalez-Alonso, J., und Coyle, F. F. (1994). Hypohydration Causes Cardiovascular Drift Without Reducing Blood Volume. *International Journal of Sports Medicine*, 15(2):74–79.
- Heck, H. und Schulz, H. (2002). Methoden der anaeroben Leistungsdiagnostik. *Deutsche Zeitschrift für Sportmedizin*, 53(7):8.
- Hintze, J. L. und Nelson, R. D. (1998). Violin plots: a box plot-density trace synergism. *The American Statistician*, 52(2):181–184.
- Hirsch, J. A. und Bishop, B. (1981). Respiratory sinus arrhythmia in humans: how breathing pattern modulates heart rate. *American Journal of Physiology-Heart and Circulatory Physiology*, 241(4):H620–H629.
- Hoffmann, K. und Wiemeyer, J. (2017). Statistical Models for Predicting Short-Term HR Responses to Submaximal Interval Exercise. In *International Symposium on Computer Science in Sport*, Seiten 57–68. Springer.
- Hohmann, A., Lames, M., und Letzelter, M. (2007). *Einführung in die Trainingswissenschaft (4. Aufl.)*. Limpert Wiebelsheim.
- Hohmann, A., Lames, M., Letzelter, M., und Pfeiffer, M. (2020). *Einführung in die Trainingswissenschaft. 7., überarb. und erw.* Limpert.
- Hollman, W. und Hettinger, T. (2000). Sports medicine. *Stuttgart, Germany*.
- Hollmann, W. und Mader, A. (2000). Grenzen der Leistungsfähigkeit des Menschen aus physiologischer Sicht. *Leistungssport 30 (1)*, Seiten 11–17.
- Hotehama, M., Lindsay, S. T., und Takemori, T. (2003). Simulation of Living Behavior. In *SICE Annual Conference Program and Abstracts SICE Annual Conference 2003*, Seiten 12–12. The Society of Instrument and Control Engineers.
- Hottenrott, K. und Neumann, G. (2016). *Trainingswissenschaft: Ein Lehrbuch in 14 Lektionen*, Band 7. Meyer & Meyer Verlag.

- Hung, H. J., O'Neill, R. T., Bauer, P., und Kohne, K. (1997). The behavior of the p-value when the alternative hypothesis is true. *Biometrics*, Seiten 11–22.
- Hunt, K. J., Fankhauser, S. E., und Saengsuwan, J. (2015). Identification of heart rate dynamics during moderate-to-vigorous treadmill exercise. *Biomedical engineering online*, 14(1):117.
- Hunt, K. J. und Hunt, A. J. (2016). Feedback control of heart rate during outdoor running: A smartphone implementation. *Biomedical Signal Processing and Control*, 26:90–97.
- Hunter, I. W. und Korenberg, M. (1986). The identification of nonlinear biological systems: Wiener and Hammerstein cascade models. *Biological cybernetics*, 55(2):135–144.
- Ioannidis, J. P. (2019a). The importance of predefined rules and prespecified statistical analyses: do not abandon significance. *Jama*, 321(21):2067–2068.
- Ioannidis, J. P. (2019b). Retiring statistical significance would give bias a free pass. *Nature*, 567(7749):461–462.
- Jang, D.-G., Ko, B.-H., Sunoo, S., Nam, S.-S., Park, H.-Y., und Bae, S.-K. (2016). A preliminary study of a running speed based heart rate prediction during an incremental treadmill exercise. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, Seiten 5323–5326. IEEE.
- Javorka, M., Zila, I., Balharek, T., und Javorka, K. (2002). Heart rate recovery after exercise: relations to heart rate variability and complexity. *Brazilian Journal of Medical and Biological Research*, 35(8):991–1000.
- Jeukendrup, A. und Diemen, A. V. (1998). Heart rate monitoring during training and competition in cyclists. *Journal of Sports Sciences*, 16(sup1):91–99.
- Jones, A. M. und Doust, J. H. (1995). Lack of reliability in Conconi's heart rate deflection point. *International journal of sports medicine*, 16(08):541–544.
- Karemaker, J. M. (2009). Counterpoint: respiratory sinus arrhythmia is due to the baroreflex mechanism. *Journal of applied physiology*, 106(5):1742–1743.
- Karjalainen, J. und Viitasalo, M. (1986). Fever and cardiac rhythm. *Archives of internal medicine*, 146(6):1169–1171.
- Kindermann, W. (2004). Anaerobe Schwelle. *Deutsche Zeitschrift für Sportmedizin*, 55(6):161–162.
- Knottnerus, J. A. und Tugwell, P. (2020). Thresholds and innovation: discussion on statistical significance. *Journal of clinical epidemiology*, 118:A5–A7.
- Koenig, A. C., Somaini, L., Pulfer, M., Holenstein, T., Omlin, X., Wieser, M., und Riener, R. (2009). Model-based Heart rate prediction during Lokomat walking. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, Seiten 1758–1761. IEEE.
- Korner-Nievergelt, F. und Hüppop, O. (2016). Kurze Einführung in Bayes-Statistik mit R für Ornithologen. *Vogelwarte*, 54:181–194.
- Lambert, M., Mbambo, Z., und Gibson, A. S. C. (1998). Heart rate during training and competition for longdistance running. *Journal of sports sciences*, 16(sup1):85–90.

- Lamberts, R., Swart, J., Capostagno, B., Noakes, T., und Lambert, M. (2010). Heart rate recovery as a guide to monitor fatigue and predict changes in performance parameters. *Scandinavian journal of medicine & science in sports*, 20(3):449–457.
- Lamberts, R. P. (2009). The development of an evidence-based submaximal cycle test designed to monitor and predict cycling performance. *The Lamberts and Lambert Submaximal Cycle Test (LSCT)*. Enschede, The Netherlands: Ipskamp Drukkers.
- Lamberts, R. P. und Lambert, M. I. (2009). Day-to-day variation in heart rate at different levels of submaximal exertion: implications for monitoring training. *The Journal of Strength & Conditioning Research*, 23(3):1005–1010.
- Lamberts, R. P., Swart, J., Noakes, T. D., und Lambert, M. I. (2011). A novel submaximal cycle test to monitor fatigue and predict cycling performance. *British Journal of Sports Medicine*, 45(10):797–804.
- Lazzeroni, L., Lu, Y., und Belitskaya-Levy, I. (2014). P-values in genomics: apparent precision masks high uncertainty. *Molecular psychiatry*, 19(12):1336–1340.
- Le, A., Jaitner, T., Tobias, F., und Litz, L. (2008). A dynamic heart rate prediction model for training optimization in cycling. *The Engineering of Sport 7*, Seiten 425–433.
- Le, A., Jaitner, T., Tobias, F., und Litz, L. (2009). A dynamic heart rate prediction model for training optimization in cycling. In *The Engineering of Sport 7*, Seiten 425–433. Springer.
- Lee, R.-G., Hsiao, C.-C., Chen, C.-Y., und Lin, R. (2015). Heart Rate Monitoring Systems in Groups for Assessment of Cardiorespiratory Fitness Analysis. In *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on*, Seiten 1145–1150. IEEE.
- Leicht, A. S., Hirning, D. A., und Allen, G. D. (2003). Heart rate variability and endogenous sex hormones during the menstrual cycle in young women. *Experimental physiology*, 88(3):441–446.
- Lim, K., Yang, S.-H., und Shim, E. (2012). Systemic modelling of human bioenergetics and blood circulation. *IET systems biology*, 6(5):187–195.
- Londeree, B. R. und Moeschberger, M. L. (1982). Effect of age and other factors on maximal heart rate. *Research quarterly for exercise and sport*, 53(4):297–304.
- Lucía, A., Hoyos, J., Pérez, M., und Chicharro, J. L. (2000). Heart rate and performance parameters in elite cyclists: a longitudinal study. *Medicine and science in sports and exercise*, 32(10):1777–1782.
- Ludwig, M., Grohgan, H. G., und Asteroth, A. (2016). A Convolution Model for Heart Rate Prediction in Physical Exercise. In *Proceedings of the 4th International Congress on Sport Sciences Research and Technology Support (icSPORTS 2016)*, Seiten 157–164.
- Ludwig, M., Grohgan, H. G., und Asteroth, A. (2019). A Convolution Model for Prediction of Physiological Responses to Physical Exercises. In *Sport Science Research and Technology Support Information (icSPORTS 2016/2017)*, Band 975 in *Communications in Computer and Information Science*, Seiten 18–35. Springer Nature Switzerland AG.

- Ludwig, M., Hoffmann, K., Endler, S., Asteroth, A., und Wiemeyer, J. (2018). Measurement, Prediction, and Control of Individual Heart Rate Responses to Exercise - Basics and Options for Wearable Devices. *Front. Physiol.*, 9(778).
- Ludwig, M., Sundaram, A. M., Füller, M., Asteroth, A., und Prassler, E. (2015). On modeling the cardiovascular system and predicting the human heart rate under strain. In *Proceedings of the 1st International Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AgeingWell)*, Seiten 106 – 117.
- MacKay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge university press.
- Mackenzie, B. (2005). Performance evaluation tests. *London: Electric World plc*, 24(25):57–158.
- Mader, A., Liesen, H., Philippi, H., Rost, R., Schuerch, P., und Hollmann, W. (1976). Zur beurteilung der sportartspezifischen Ausdauerleistungsfähigkeit im Labor. *Sportarzt und Sportmedizin*, 27:80–88.
- Markworth, P. (2010). *Sportmedizin: physiologische Grundlagen*. Nikol Verlag, 24. Auflage.
- Marx, A. J., Porcari, J. P., Doberstein, S., Arney, B. E., Bramwell, S., und Foster, C. (2018). The accuracy of heart rate based zone training using predicted versus measured maximal heart rate. *International Journal of Research in Exercise Physiology*, 13(1):21–28.
- Matejka, J. und Fitzmaurice, G. (2017). Same stats, different graphs: generating datasets with varied appearance and identical statistics through simulated annealing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Seiten 1290–1294. ACM.
- Mazzoleni, M. J., Battaglini, C. L., Martin, K. J., Coffman, E. M., und Mann, B. P. (2016). Modeling and predicting heart rate dynamics across a broad range of transient exercise intensities during cycling. *Sports Engineering*, 19(2):117–127.
- McLeod, J. (1966). PHYSBE... a physiological simulation benchmark experiment. *Transactions of the Society for Computer Simulation*, 7(6):324–329.
- Medbø, J. I., Mamen, A., Holt Olsen, O., und Evertsen, E. (2000). Examination of four different instruments for measuring blood lactate concentration. *Scandinavian journal of clinical and laboratory investigation*, 60(5):367–380.
- Mohammad, S., Guerra, T. M., Grobois, J. M., und Hecquet, B. (2011a). Heart rate control during cycling exercise using Takagi-Sugeno models. In *18th IFAC World Congress, Milano (Italy)*, Seiten 12783–12788.
- Mohammad, S., Guerra, T. M., GROBOIS, J. M., und Hecquet, B. (2011b). Heart rate control during cycling exercise using takagi-sugeno models. *IFAC Proceedings Volumes*, 44(1):12783–12788.
- Moser, B. K. und Stevens, G. R. (1992). Homogeneity of variance in the two-sample means test. *The American Statistician*, 46(1):19–21.
- Mutijarsa, K., Ichwan, M., und Utami, D. B. (2016). Heart rate prediction based on cycling cadence using feedforward neural network. In *Computer, Control, Informatics and its*

- Applications (IC3INA), 2016 International Conference on*, Seiten 72–76. IEEE.
- Narendra, K. und Gallman, P. (1966). An iterative method for the identification of non-linear systems using a Hammerstein model. *IEEE Transactions on Automatic control*, 11(3):546–550.
- Nes, B., Janszky, I., Wisløff, U., Støylen, A., und Karlsen, T. (2013). Age-predicted maximal heart rate in healthy subjects: The HUNT Fitness Study. *Scandinavian journal of medicine & science in sports*, 23(6):697–704.
- Neumann, G. und Schüler, K.-P. (1994). *Sportmedizinische Funktionsdiagnostik*. J. A. Barth, Leipzig, 2. Auflage.
- No, M. und Kwak, H.-B. (2016). Effects of environmental temperature on physiological responses during submaximal and maximal exercises in soccer players. *Integrative medicine research*, 5(3):216–222.
- Ortlieb, C. P., Dresky, C. v., Gasser, I., und Günzel, S. (2013). *Mathematische Modellierung*. Springer.
- Paradiso, M., Pietrosanti, S., Scalzi, S., Tomei, P., und Verrelli, C. M. (2013). Experimental heart rate regulation in cycle-ergometer exercises. *IEEE Transactions on biomedical engineering*, 60(1):135–139.
- Parak, J. und Korhonen, I. (2014). Evaluation of wearable consumer heart rate monitors based on photoplethysmography. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, Seiten 3670–3673. IEEE.
- Paton, C. D. und Hopkins, W. G. (2001). Tests of cycling performance. *Sports medicine*, 31(7):489–496.
- Perl, J. (2004). PerPot - a meta-model and software tool for analysis and optimisation of load-performance-interaction. *International Journal of Performance Analysis of Sport*, 4(2):61–73.
- Pfeiffer, M. (2020). *Trainingswissenschaft in 60 Minuten*. UVK Verlag.
- Piepoli, M., Sleight, P., Leuzzi, S., Valle, F., Spadacini, G., Passino, C., Johnston, J., und Bernardi, L. (1997). Origin of respiratory sinus arrhythmia in conscious humans: an important role for arterial carotid baroreceptors. *Circulation*, 95(7):1813–1821.
- Rasch, D., Kubinger, K. D., und Moder, K. (2011). The two-sample t test: pre-testing its assumptions does not pay off. *Statistical papers*, 52(1):219–231.
- Richards, J. E. (1980). The statistical analysis of heart rate: A review emphasizing infancy data. *Psychophysiology*, 17(2):153–166.
- Röcker, K. (2013). Die sportmedizinische Laktatdiagnostik: Technische Rahmenbedingungen und Einsatzbereiche. *Deutsche Zeitschrift für Sportmedizin*, 64(12).
- Röcker, K. und Dickhuth, H.-H. (2001). Praxis der Laktatmessung. *Deutsche Zeitschrift für Sportmedizin*, 52(1):33–34.
- Roecker, K., Schotte, O., Niess, A. M., Horstmann, T., und Dickhuth, H.-H. (1998). Predicting competition performance in long-distance running by means of a treadmill test. *Medicine and science in sports and exercise*, 30:1552–1557.

- Rojas, R. (2013). *Theorie der neuronalen Netze: eine systematische Einführung*. Springer-Verlag.
- Ruxton, G. D. (2006). The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test. *Behavioral Ecology*, 17(4):688–690.
- Scalzi, S., Tomei, P., und Verrelli, C. M. (2012). Nonlinear control techniques for the heart rate regulation in treadmill exercises. *IEEE Transactions on Biomedical Engineering*, 59(3):599–603.
- Scharhag-Rosenberger, F. (2010). Spiroergometrie zur Ausdauerleistungsdiagnostik. *Deutsche Zeitschrift für Sportmedizin*, 61(6):146–147.
- Scharhag-Rosenberger, F. und Schommer, K. (2013). Die Spiroergometrie in der Sportmedizin. *Deutsche Zeitschrift für Sportmedizin*, 64(12):362–366.
- Schnabel, G., Harre, D., und Borde, A., Herausgeber (1994). *Trainingswissenschaft. Leistung–Training–Wettkampf*, Band 19972. Sportverlag, Berlin.
- Schrot, C. (2015). *Optimierung der Trainingssteuerung im Straßenradsport mittels simulationsgestützter Trainingswirkungsanalyse mit dem Performance-Potential-Metamodell*. Dissertation, Universität Bayreuth.
- Schwegler, J. S. und Lucius, R. (2011). *Der Mensch: Anatomie und Physiologie*. Thieme.
- She, J., Yokota, S., und Du, E. Y. (2013). Automatic heart-rate-based selection of pedal load and control system for electric cart. *Mechatronics*, 23(3):279–288.
- Sinclair, J., Hingston, P., Masek, M., und Nosaka, K. (2009). Using a virtual body to aid in exergaming system development. *IEEE Computer Graphics and Applications*, 29(2):39–48.
- Sornanathan, L. und Khalil, I. (2010). Fitness monitoring system based on heart rate and SpO2 level. In *Information Technology and Applications in Biomedicine (ITAB), 2010 10th IEEE International Conference on*, Seiten 1–5. IEEE.
- Stevinson, C. D. und Biddle, S. J. H. (1998). Cognitive orientations in marathon running and hitting the wall. *British Journal of Sports Medicine*, 32:229–235.
- Stirling, J. R., Zakyntinaki, M., Refoyo, I., und Sampedro, J. (2008). A model of heart rate kinetics in response to exercise. *Journal of Nonlinear Mathematical Physics*, 15(sup3):426–436.
- Stork, M., Novak, J., und Zeman, V. (2017). Modelling of Heart Rate During Exercise. In *MEASUREMENT 2017, Proceedings of the 11th International Conference*, Seiten 251–254.
- Su, S. W., Huang, S., Wang, L., Celler, B. G., Savkin, A. V., Guo, Y., und Cheng, T. (2007a). Non-parametric Hammerstein model based model predictive control for heart rate regulation. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, Seiten 2984–2987. IEEE.
- Su, S. W., Huang, S., Wang, L., Celler, B. G., Savkin, A. V., Guo, Y., und Cheng, T. M. (2010a). Optimizing heart rate regulation for safe exercise. *Annals of biomedical engineering*, 38(3):758–768.

- Su, S. W., Huang, S., Wang, L., Celler, B. G., Savkin, A. V., Guo, Y., und Cheng, T. M. (2010b). Optimizing heart rate regulation for safe exercise. *Annals of biomedical engineering*, 38(3):758–768.
- Su, S. W., Wang, L., Celler, B. G., Savkin, A. V., und Guo, Y. (2007b). Identification and control for heart rate regulation during treadmill exercise. *IEEE Transactions on biomedical engineering*, 54(7):1238–1246.
- Taelman, J., Vandeput, S., Spaepen, A., und Van Huffel, S. (2009). Influence of mental stress on heart rate and heart rate variability. In *4th European conference of the international federation for medical and biological engineering*, Seiten 1366–1369. Springer.
- Tanaka, H., Monahan, K. D., und Seals, D. R. (2001). Age-predicted maximal heart rate revisited. *Journal of the American College of Cardiology*, 37(1):153–156.
- Tanaka, M., Volle, M. A., Brisson, G. R., und Dion, M. (1979). Body temperature and heart rate relationships during submaximal bicycle ergometer exercises. *European journal of applied physiology and occupational physiology*, 42(4):263–270.
- ten Haaf, T., Foster, C., Meeusen, R., Roelands, B., Piacentini, M. F., van Staveren, S., Koenderman, L., und de Koning, J. J. (2019). Submaximal heart rate seems inadequate to prescribe and monitor intensified training. *European journal of sport science*, Seiten 1–10.
- Thayer, J. F., Åhs, F., Fredrikson, M., Sollers III, J. J., und Wager, T. D. (2012). A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. *Neuroscience & Biobehavioral Reviews*, 36(2):747–756.
- Tomasits, J. und Haber, P. (2011). *Leistungsphysiologie: Grundlagen für Trainer, Physiotherapeuten und Masseur*. Springer-Verlag.
- Ursino, M. (1998). Interaction between carotid baroregulation and the pulsating heart: a mathematical model. *American Journal of Physiology-Heart and Circulatory Physiology*, 275(5):H1733–H1747.
- Uth, N., Sørensen, H., Overgaard, K., und Pedersen, P. K. (2004). Estimation of VO₂max from the ratio between HR max and HR rest—the Heart Rate Ratio Method. *European journal of applied physiology*, 91(1):111–115.
- Vargas-Uricoechea, H., Bonelo-Perdomo, A., und Sierra-Torres, C. H. (2014). Effects of thyroid hormones on the heart. *Clínica e Investigación en Arteriosclerosis*, 26(6):296–309.
- Vautier, J., Vandewalle, H., und Monod, H. (1994). Prediction of exhaustion time from heart rate drift. *Archives internationales de physiologie, de biochimie et de biophysique*, 102(1):61–65.
- Wang, L., Su, S. W., Celler, B. G., Chan, G. S., Cheng, T. M., und Savkin, A. V. (2009). Assessing the human cardiovascular response to moderate exercise: feature extraction by support vector regression. *Physiological measurement*, 30(3):227.
- Wang, L., Su, S. W., Chan, G., Celler, B. G., Cheng, T. M., und Savkin, A. V. (2008). Nonlinear modeling of cardiovascular response to exercise. In *International Conference on Bio-inspired Systems and Signal Processing*, Seiten 40–46. Institute for Systems and

Technologies of Information.

- Wasserman, K. und McIlroy, M. B. (1964). Detecting the threshold of anaerobic metabolism in cardiac patients during exercise. *The American journal of cardiology*, 14(6):844–852.
- Weghorn, H. (2013). Application and UI Design for Ergonomic Heart Rate Monitoring in Endurance Sports: Realizing an Improved Tool for Health and Sports Activities on Base of Android Smartphone Programming and ANT+. In *International Congress on Sports Science Research and Technology Support*, Seiten 25–41. Springer.
- Weghorn, H. (2016). Experimental Investigation of the Usfulness of Bracelet Trackers in Sports and Health Monitoring - Critical Evaluation of a New Handheld Activity Monitoring Device Class. In *Proceedings of the 4th International Congress on Sport Sciences Research and Technology Support - Volume 1: icSPORTS*, Seiten 124–133. INSTICC, SciTePress.
- Weippert, M., Kumar, M., Kreuzfeld, S., Arndt, D., Rieger, A., und Stoll, R. (2010). Comparison of three mobile devices for measuring R–R intervals and heart rate variability: Polar S810i, Suunto t6 and an ambulatory ECG system. *European journal of applied physiology*, 109(4):779–786.
- Whittam, A., Clayton, R., Lord, S., McComb, J., und Murray, A. (1998). Computer modelling of heart rate and blood pressure. In *Computers in Cardiology 1998*, Seiten 149–152. IEEE.
- Wilmore, J. H. (1968). Influence of motivation on physical work capacity and performance. *Journal of applied physiology*, 24(4):459–463.
- Wolff, H.-G. und Bacher, J. (2010). Hauptkomponentenanalyse und explorative Faktorenanalyse. In *Handbuch der sozialwissenschaftlichen Datenanalyse*, Seiten 333–365. Springer.
- Xiao, F., Chen, Y., Yuchi, M., Ding, M., und Jo, J. (2010). Heart rate prediction model based on physical activities using evolutionary neural network. In *Genetic and Evolutionary Computing (ICGEC), 2010 Fourth International Conference on*, Seiten 198–201. IEEE.
- Xiao, F., Yuchi, M., Ding, M., und Jo, J. (2011). A research of heart rate prediction model based on evolutionary neural network. In *Intelligent Computation and Bio-Medical Instrumentation (ICBMI), 2011 International Conference on*, Seiten 304–307. IEEE.
- Xiao, F., Yuchi, M., Ding, M.-y., Jo, J., und Kim, J.-H. (2009). A multi-step heart rate prediction method based on physical activity using Adams-Bashforth technique. In *Computational Intelligence in Robotics and Automation (CIRA), 2009 IEEE International Symposium on*, Seiten 355–359. IEEE.
- Yang, Y., Ji, L., Wu, H., und Wu, J. (2012). An exercise-driven heart rate statistical process model. In *Information Fusion (FUSION), 2012 15th International Conference on*, Seiten 432–438. IEEE.
- Yuchi, M. und Jo, J. (2008). Heart rate prediction based on physical activity using feedforward neural network. In *Convergence and Hybrid Information Technology, 2008. ICHIT'08. International Conference on*, Seiten 344–350. IEEE.

- Zakynthinaki, M. S. (2015). Modelling heart rate kinetics. *PloS one*, 10(4):e0118263.
- Zakynthinaki, M. S. (2016). Simulating heart rate kinetics during incremental and interval training. *Biomedical Human Kinetics*, 8(1):144–152.
- Zimmer, P., Oberste, M., und Bloch, W. (2015). Einfluss von Sport auf das zentrale Nervensystem-Molekulare und zelluläre Wirkmechanismen. *Deutsche Zeitschrift für Sportmedizin*, 66(2).
- Zimmerman, D. W. (2004). A note on preliminary tests of equality of variances. *British Journal of Mathematical and Statistical Psychology*, 57(1):173–181.
- Zimmerman, D. W. und Zumbo, B. D. (1993). Rank transformations and the power of the Student t test and Welch t'test for non-normal populations with unequal variances. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 47(3):523.

Anhang

Definition 8.1: Grundlegende statistische Begriffe.

Sei n die Anzahl gemessener oder betrachteter Datenpunkte der Stichprobenmenge X und seien $x_1, x_2, \dots, x_n \in \mathbb{R}$ die einzelnen Messwerte. Die nachfolgend definierten Begriffe beziehen sich jeweils auf die empirische Statistik, die auf Messwerten (und nicht auf einer mathematischen Verteilung) basiert.

Mittelwert In der vorliegenden Arbeit bezeichnet der Mittelwert den klassischen Durchschnitt, also das arithmetische Mittel. Das arithmetische Mittel der gegebenen Werte $x_i, 1 \leq i \leq n$ ist definiert als

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Quantile (Empirische) Quantile sind Lagemaße einer Stichprobe. Ein p -Quantil gibt an, bei welchem Wert Q_p der Stichprobe $p\%$ der Werte unterhalb (und $1 - p\%$ oberhalb) von Q_p liegen. Für aufsteigend geordnete Messwerte $x_1 \leq x_2 \leq \dots \leq x_n$ ist das p -Quantil Q_p dieser Datenpunkte für $p \in (0,1)$ beschrieben als

$$Q_p = \begin{cases} \frac{1}{2}(x_{n \cdot p} + x_{n \cdot p + 1}), & \text{wenn } n \cdot p \text{ ganzzahlig} \\ x_{\lfloor n \cdot p + 1 \rfloor}, & \text{wenn } n \cdot p \text{ nicht ganzzahlig.} \end{cases}$$

Dabei bezeichnet $\lfloor x \rfloor$ die Abrundungsfunktion, die eine reelle Zahl x auf die nächstkleinere ganze Zahl abrundet.

Quartile Quartile sind die Quantile bei 25%, 50% und 75%. Dabei wird das Quartil bei 25% als »unteres Quartil« (Q_1) bezeichnet, das Quartil bei 50% definiert den »Median« (Q_2) und das Quartil bei 75% wird als »oberes Quartil« (Q_3) bezeichnet.

Median Der Median bezeichnet das 50%-Quantil.

Standardabweichung Die (empirische) Standardabweichung beschreibt ein Maß für die Streuung innerhalb einer Messreihe und gibt an, wie weit einzelne Werte durchschnittlich vom Mittelwert der Zahlen aus dieser Messreihe entfernt sind.

Dabei ist die Standardabweichung s (für eine Stichprobe) definiert als Wurzel aus der empirischen Varianz:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

IQA Der IQA beschreibt die Differenz aus oberem und unterem Quartil, also $IQA = Q_3 - Q_1 = Q_{0,75} - Q_{0,25}$. Er enthält 50 % der Messwerte.

Ausreißer Ein Wert innerhalb der Datenpunkte gilt als Ausreißer, wenn er mehr als das 1,5-fache des IQA oberhalb von Q_3 bzw. unterhalb von Q_1 liegt. Das heißt ein Wert x_{upper} ist oberer Ausreißer, wenn $x_{\text{upper}} \geq Q_3 + 1,5 \cdot IQA$ und ein Wert x_{lower} ist unterer Ausreißer, wenn $x_{\text{lower}} \leq Q_1 - 1,5 \cdot IQA$.

Extremwerte Extremwerte sind ein Spezialfall der Ausreißer. Ein Wert innerhalb der Datenpunkte gilt als Extremwert, wenn er mehr als das dreifache des IQA oberhalb von Q_3 bzw. unterhalb von Q_1 liegt. Das heißt ein Wert x_{upEx} ist oberer Extremwert, wenn $x_{\text{upEx}} \geq Q_3 + 3 \cdot IQA$ und ein Wert x_{lowEx} ist unterer Extremwert, wenn $x_{\text{lowEx}} \leq Q_1 - 3 \cdot IQA$.

Ergänzend zu Abschnitt 6.2

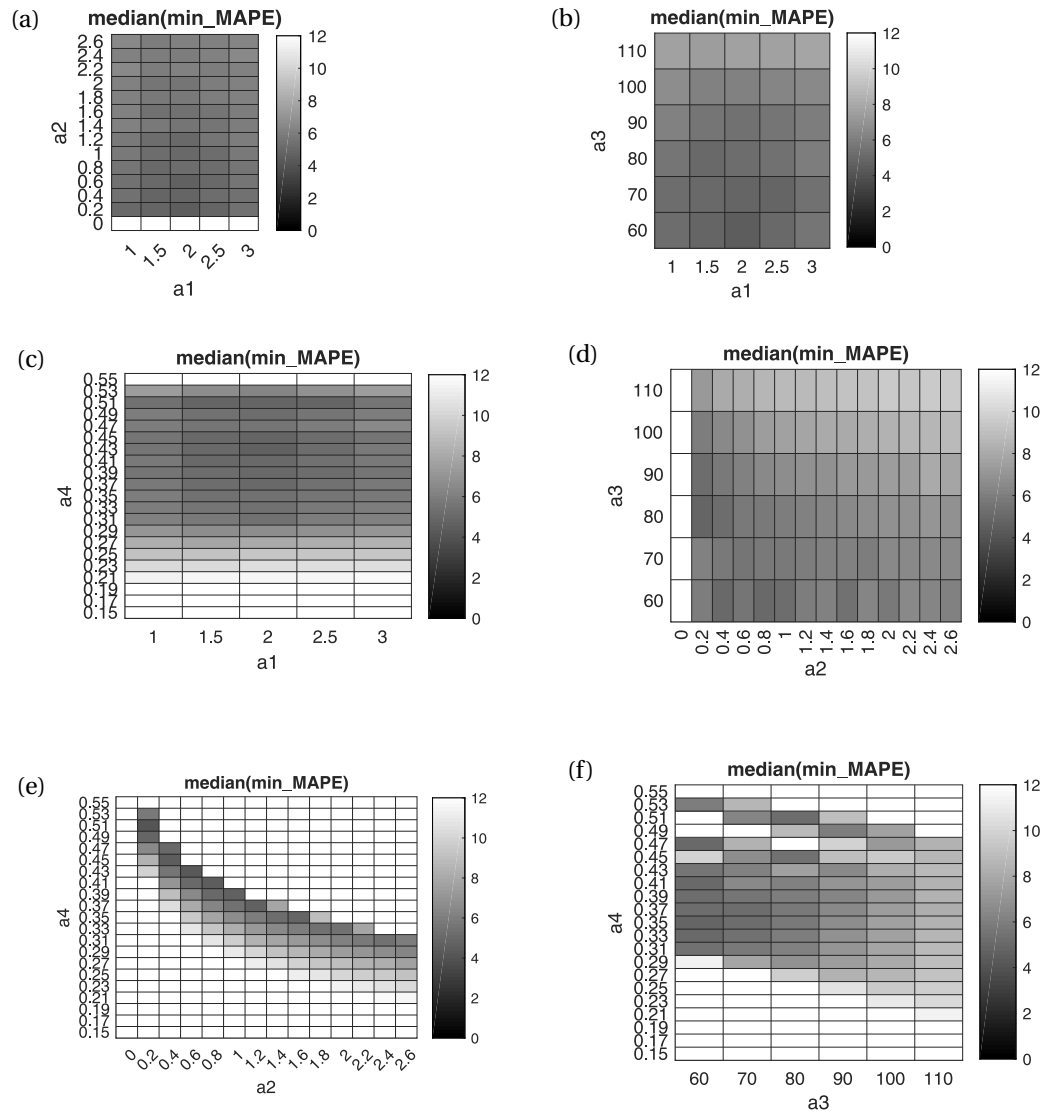


Abbildung A.1.: Heatmap-Darstellung für die Kombination aus je zwei Parametern. Für jedes Wertepaar aus der betrachteten Parameterkombination wird der Median der minimalen MAPE-Werte für alle Trainingseinheiten aller Probanden über die Einfärbung betrachtet

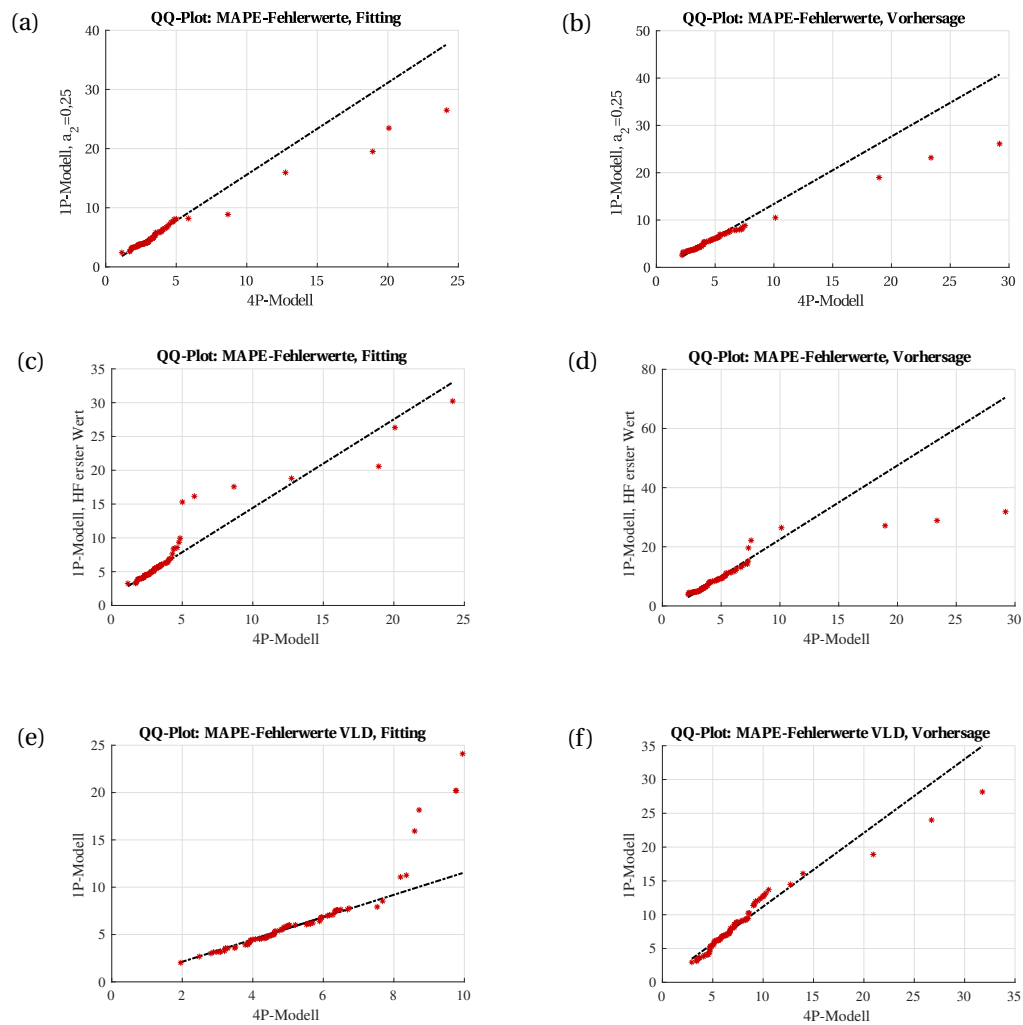


Abbildung A.2.: QQ-Plots zur Evaluation der Parameterreduktion im Vergleich des 4P- mit dem 1P-Faltungsmodell in unterschiedlichen Settings: (a) finale Parameter des 1P-Modells im Fitting, (b) finale Parameter des 1P-Modells in der Vorhersage, (c) zuvor betrachtetes 1P-Modell mit fixem Herzfrequenzparameter im Fitting, (d) zuvor betrachtetes 1P-Modell mit fixem Herzfrequenzparameter in der Vorhersage, (e) Ergebnis der virtuellen Leistungsdiagnostik im Fitting, (f) Ergebnis der virtuellen Leistungsdiagnostik in der Vorhersage

Ergänzend zu Abschnitt 6.3

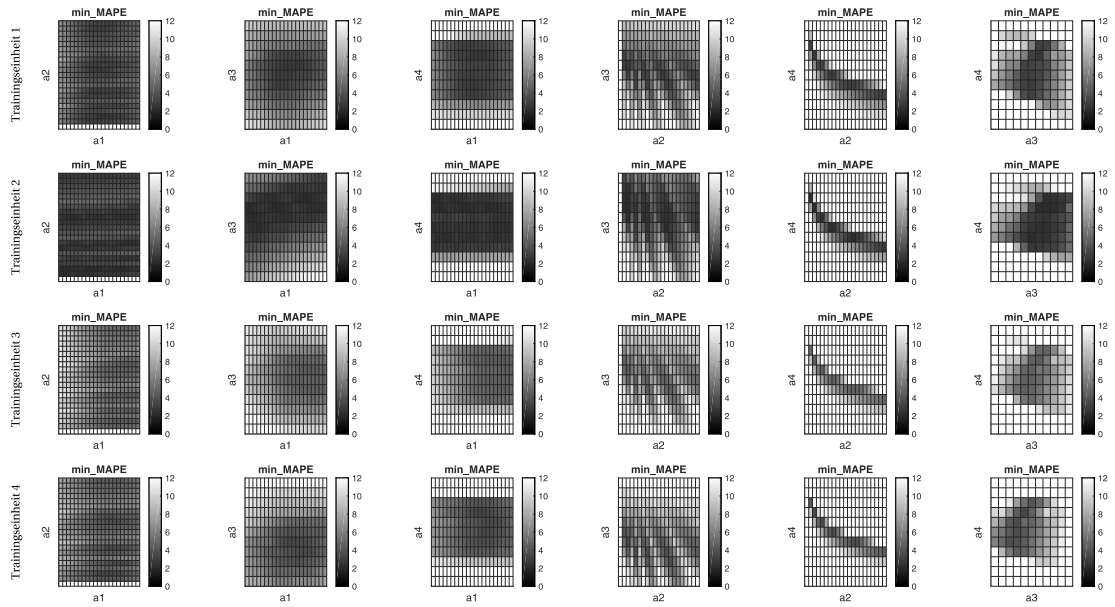


Abbildung A.3.: Heatmaps der einzelnen Trainingseinheiten von Proband 1 aus Studie 3

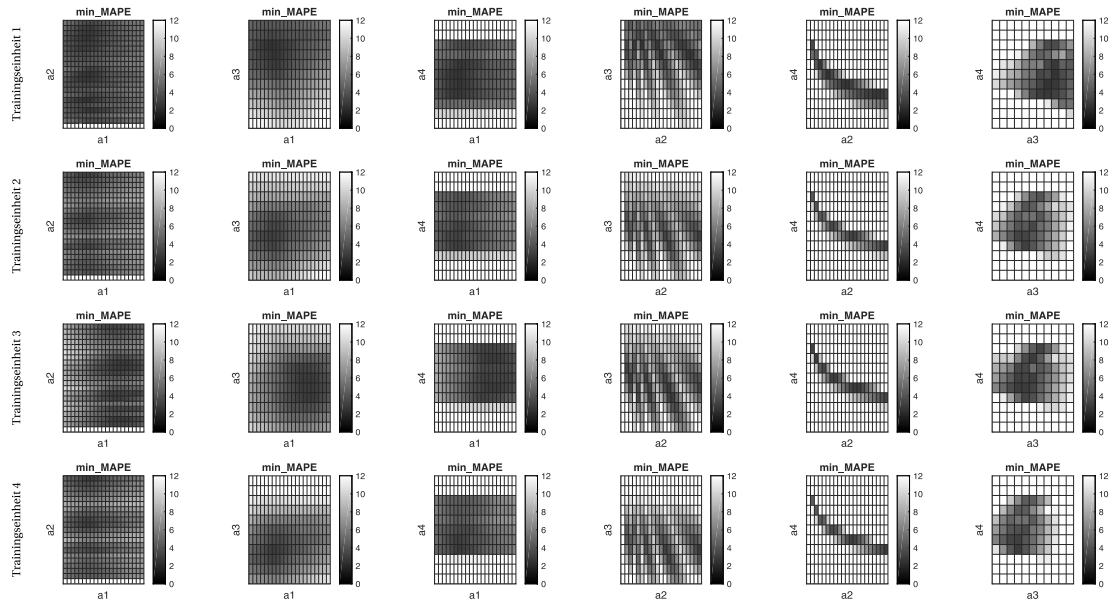


Abbildung A.4.: Heatmaps der einzelnen Trainingseinheiten von Proband 2 aus Studie 3

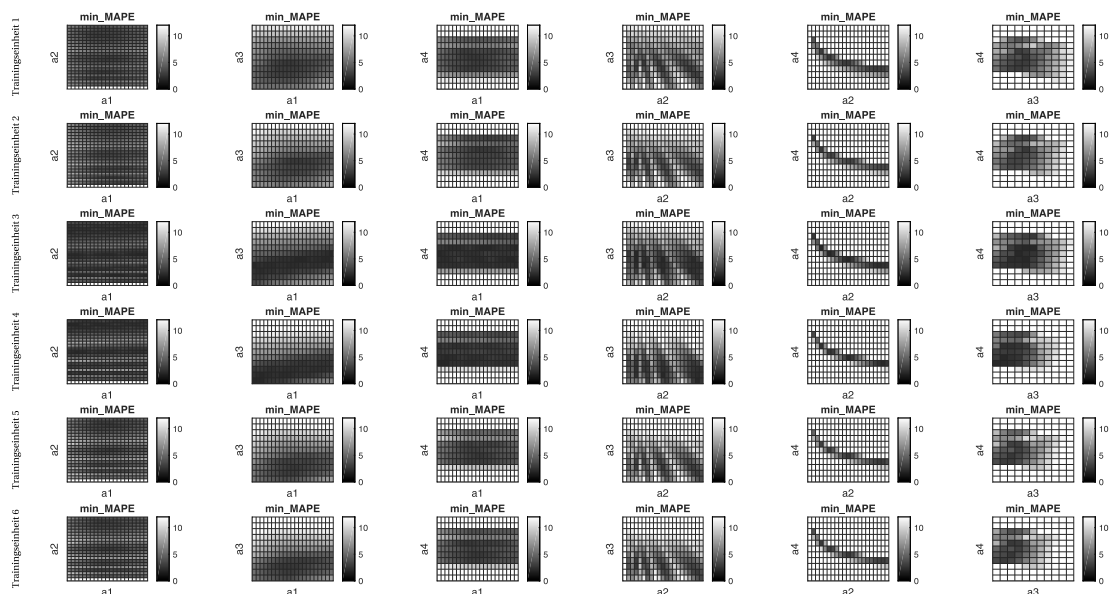


Abbildung A.5.: Heatmaps der einzelnen Trainingseinheiten von Proband 3 aus Studie 3

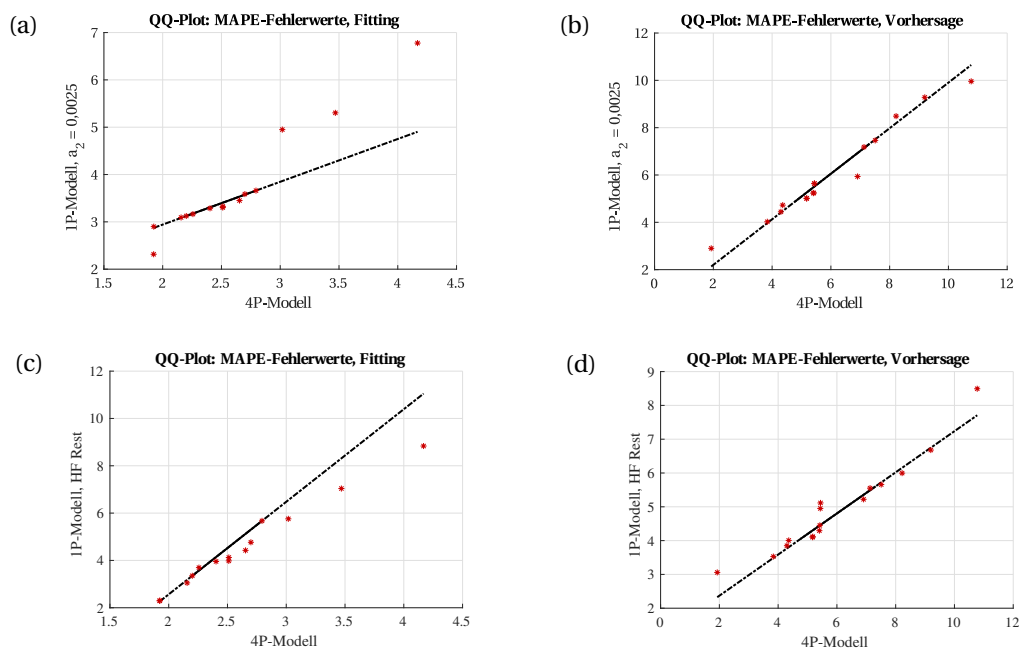


Abbildung A.6.: QQ-Plots zur Evaluation der Parameterreduktion im Vergleich des 4P- mit dem 1P-Faltungsmodell in unterschiedlichen Settings: (a) freier Parameter a_3 im Fitting, (b) freier Parameter a_3 in der Vorhersage, (c) freier Parameter a_2 im Fitting, (d) freier Parameter a_2 in der Vorhersage

Ergänzend zu Abschnitt 7.2

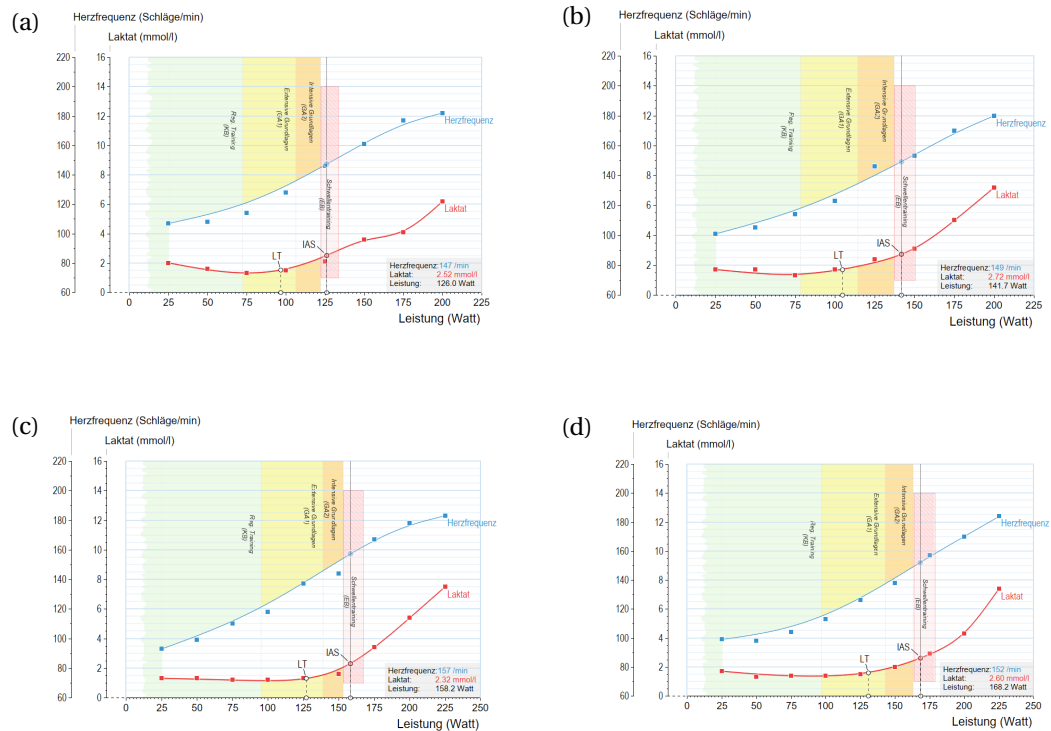


Abbildung A.7.: Ergebnis des Ergonizers für Proband 1. Aufgetragen sind die Wattleistung (X-Achse) gegen die Herzfrequenz und Blutlaktatkonzentration. Die rote Kurve zeigt die approximierte Kurve der Blutlaktatkonzentration, die blaue Kurve die der Herzfrequenzwerte. Empirisch gemessene Werte sind durch einen Punkt innerhalb der Kurve gekennzeichnet. Der grüne Bereich entspricht einem Regenerationstraining, der gelbe Bereich GA 1 und der orangene Bereich GA 2. Mit einem roten Balken markiert ist das »Schwellentraining«. LT und IAT sind durch einen Punkt hervorgehoben.

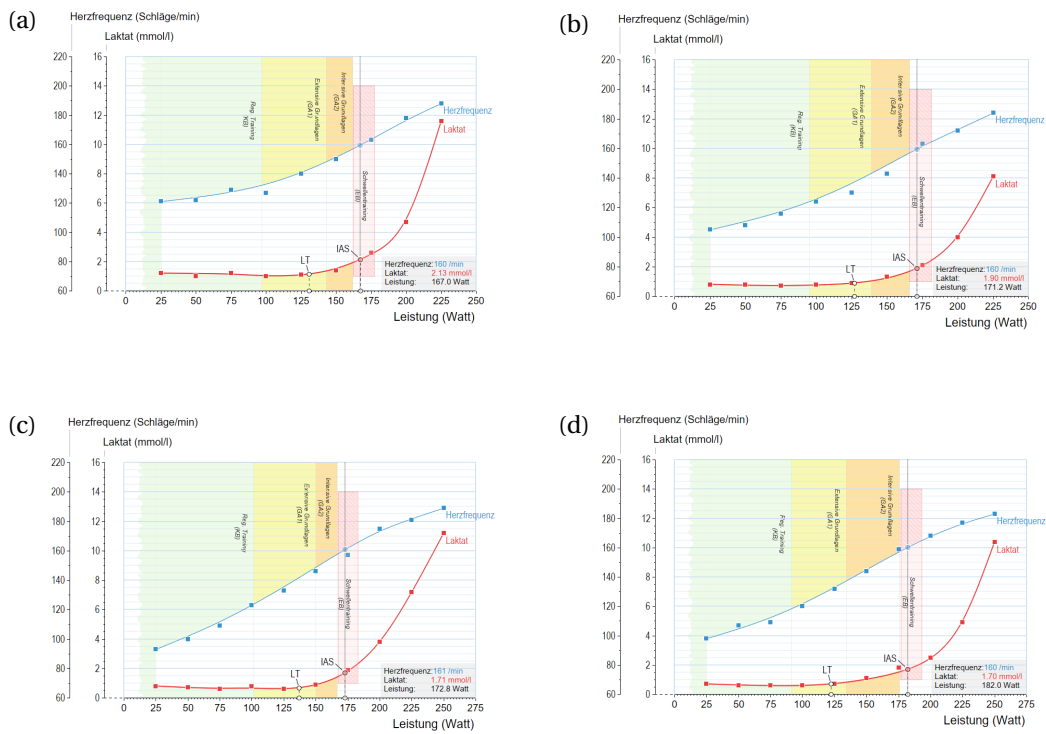


Abbildung A.8.: Ergebnis des Ergonizers für Proband 2. Aufgetragen sind die Wattleistung (X-Achse) gegen die Herzfrequenz und Blutlaktatkonzentration. Die rote Kurve zeigt die approximierte Kurve der Blutlaktatkonzentration, die blaue Kurve die der Herzfrequenzwerte. Empirisch gemessene Werte sind durch einen Punkt innerhalb der Kurve gekennzeichnet. Der grüne Bereich entspricht einem Regenerationstraining, der gelbe Bereich GA 1 und der orangene Bereich GA 2. Mit einem roten Balken markiert ist das »Schwellentraining«. LT und IAT sind durch einen Punkt hervorgehoben.

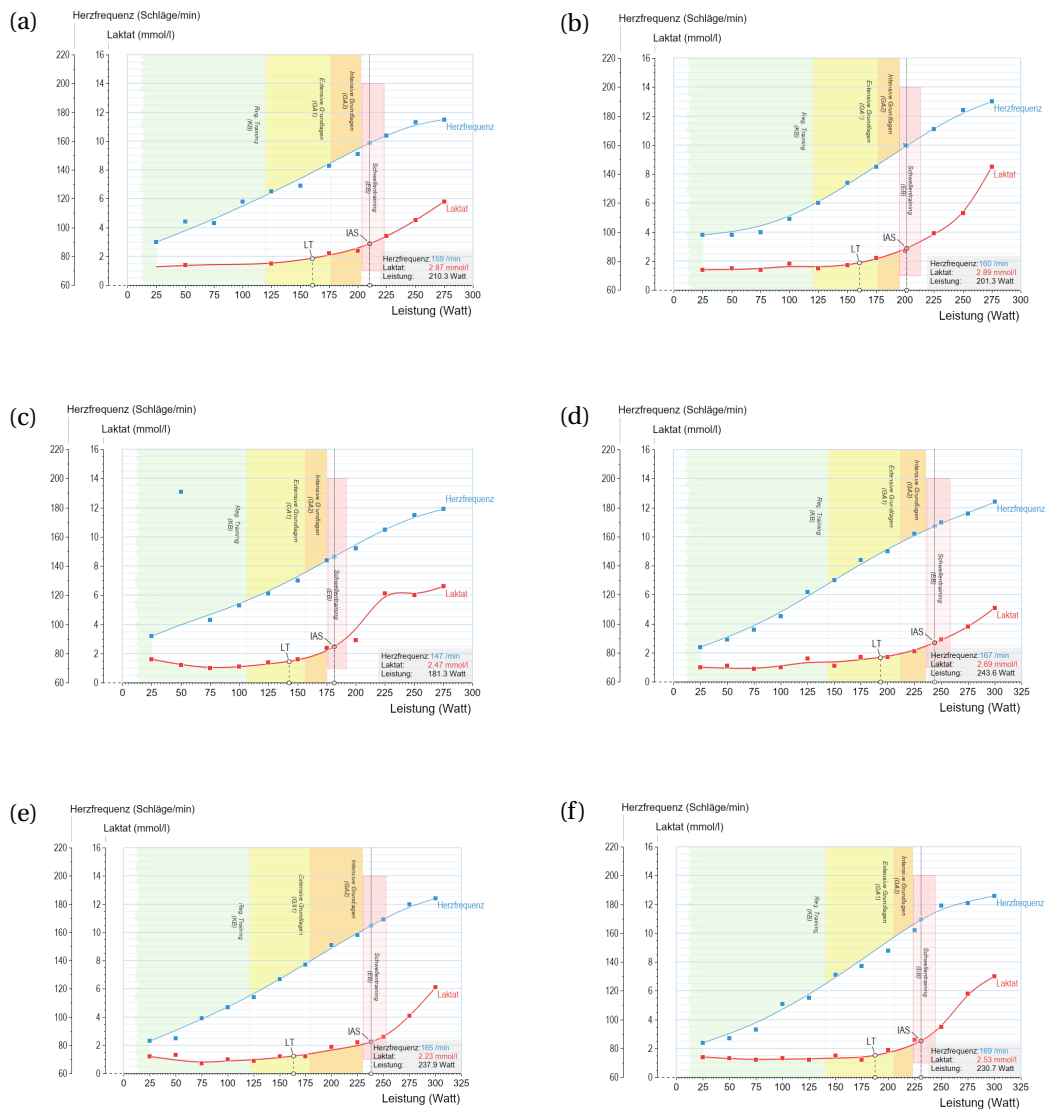


Abbildung A.9.: Ergebnis des Ergonizers für Proband 3. Aufgetragen sind die Wattleistung (X-Achse) gegen die Herzfrequenz und Blutlaktatkonzentration. Die rote Kurve zeigt die approximierte Kurve der Blutlaktatkonzentration, die blaue Kurve die der Herzfrequenzwerte. Empirisch gemessene Werte sind durch einen Punkt innerhalb der Kurve gekennzeichnet. Der grüne Bereich entspricht einem Regenerationstraining, der gelbe Bereich GA 1 und der orangene Bereich GA 2. Mit einem roten Balken markiert ist das »Schwellentraining«. LT und IAT sind durch einen Punkt hervorgehoben.

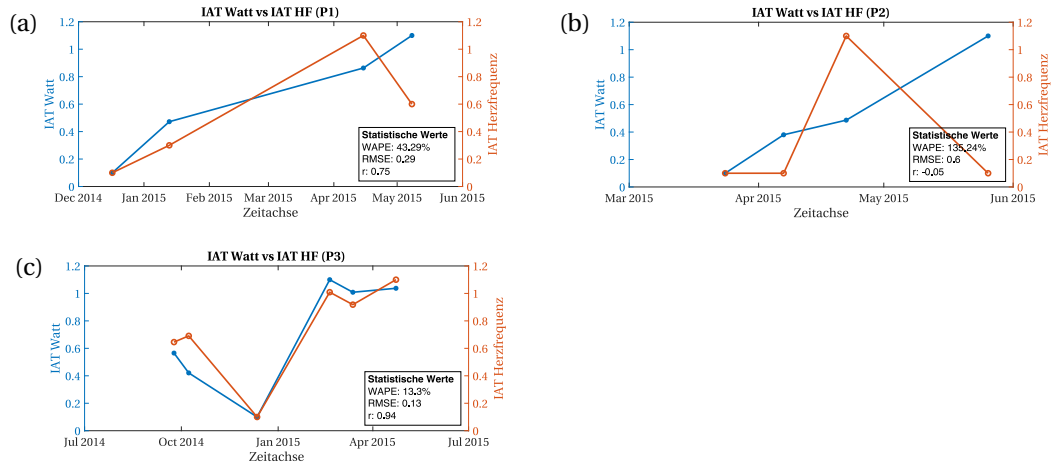


Abbildung A.10.: Vergleich: Wattleistung an der geschätzten IAS und Herzfrequenz an der geschätzten IAS der drei Probanden aus Studie 5

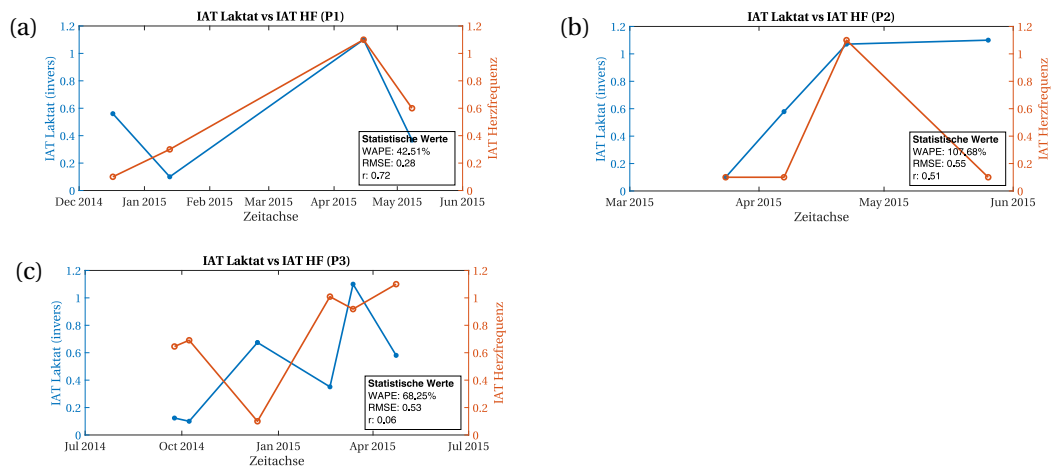


Abbildung A.11.: Vergleich: Blutlaktatkonzentration an der geschätzten IAS und Herzfrequenz an der geschätzten IAS der drei Probanden aus Studie 5

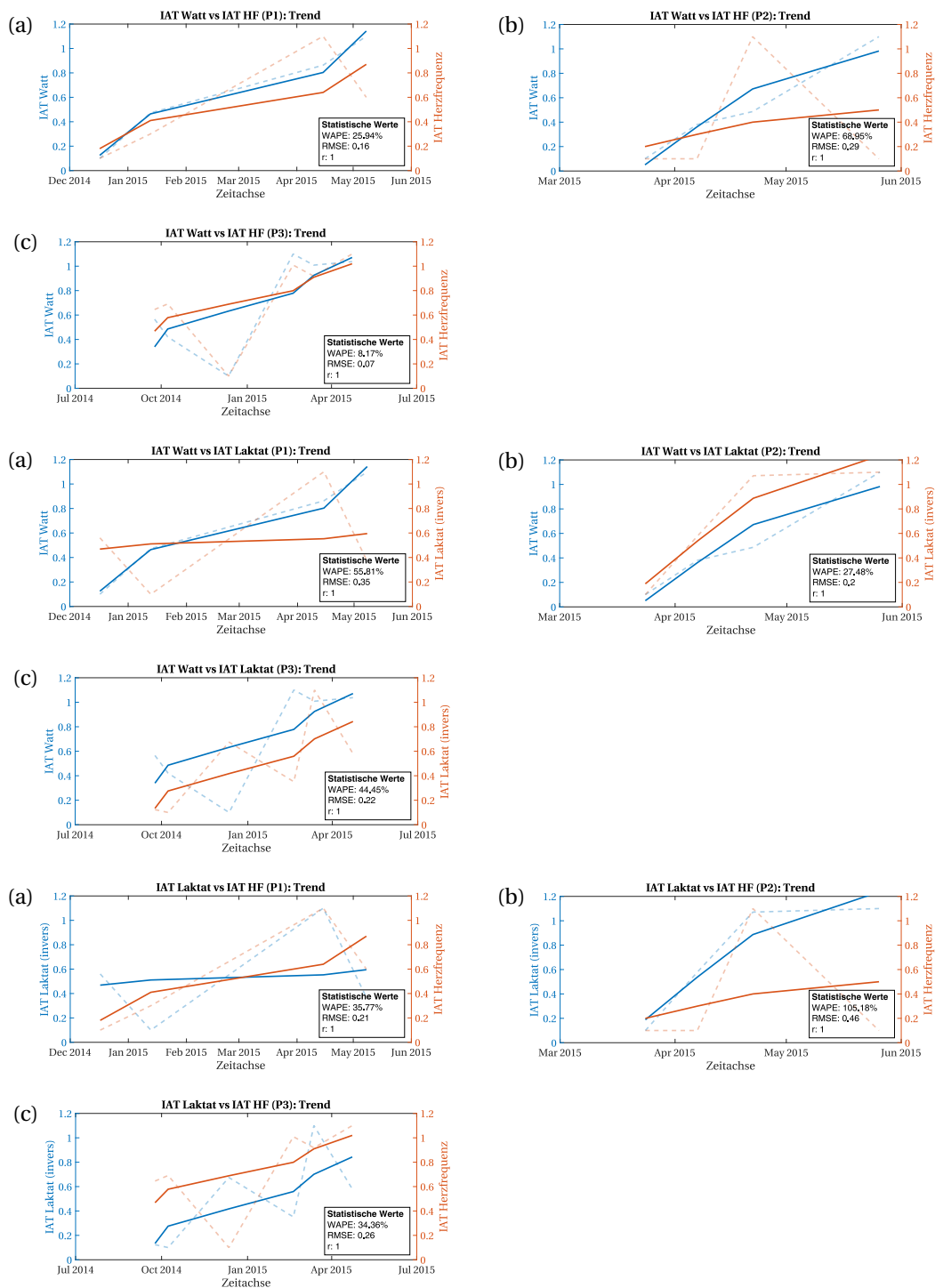


Abbildung A.12.: Vergleich: drei verschiedene Schwellwerte an der geschätzten IAS paarweise verglichen im Trendverlauf der drei Probanden aus Studie 5

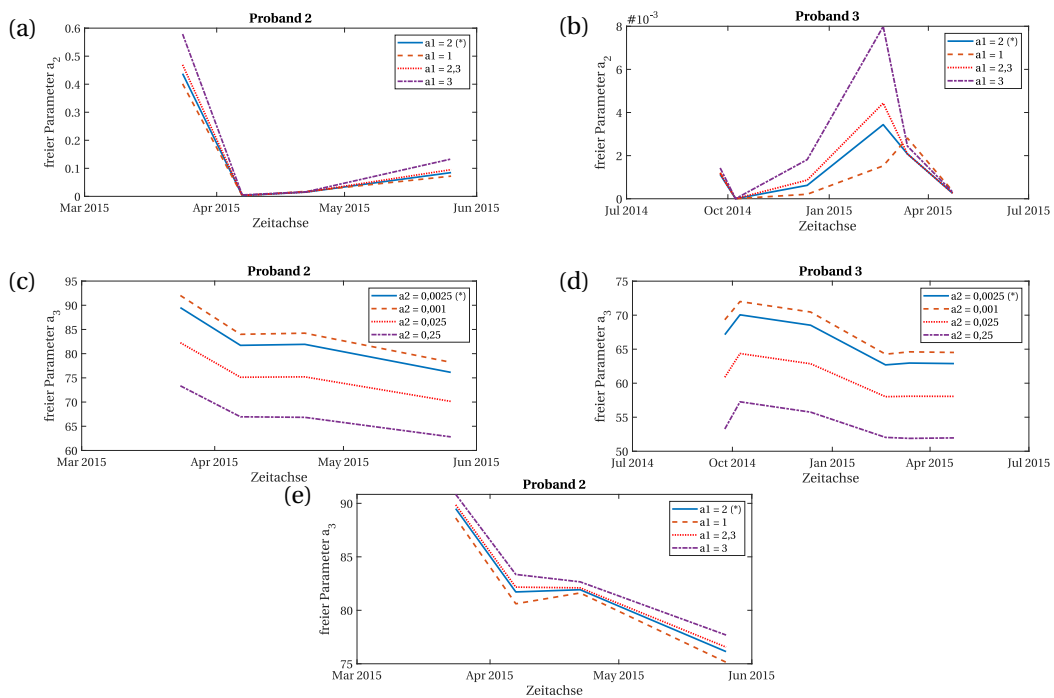


Abbildung A.13.: Vergleich: Parameterverlauf bei unterschiedlichen konstanten Parameterwerte für a_1 bzw. a_2 der drei Probanden aus Studie 5

