

Martin Bochterle

**Methodenbausteine für die empirische
Modellierung des Verhaltens technischer
Systeme im Rahmen der Produktentwicklung
auf Basis linearer Regression**

Methods for empirical modeling of the behavior
of technical systems in the context of product
development based on linear regression

Band 154

Systeme ■ Methoden ■ Prozesse

Univ.-Prof. Dr.-Ing. Dr. h.c. A. Albers
Univ.-Prof. Dr.-Ing. S. Matthiesen
(Hrsg.)

Copyright: IPEK • Institut für Produktentwicklung, 2022
Karlsruher Institut für Technologie (KIT)
Die Forschungsuniversität in der Helmholtz-Gemeinschaft
Alle Rechte vorbehalten

Druck: Stolzenberger Druck und Werbung GmbH & Co. KG, Leimen
06224-7697915

ISSN 1615-8113

Methodenbausteine für die empirische Modellierung des Verhaltens technischer Systeme im Rahmen der Produktentwicklung auf Basis linearer Regression

Zur Erlangung des akademischen Grades eines
DOKTORS DER INGENIEURWISSENSCHAFTEN (Dr.-Ing.)
von der KIT-Fakultät für Maschinenbau des
Karlsruher Instituts für Technologie (KIT)

angenommene

DISSERTATION

von

Dipl.-Ing. Martin Bochterle

Tag der mündlichen Prüfung:	25.04.2022
Hauptreferent:	Univ.-Prof. Dr.-Ing. Dr. h. c. Albert Albers
Korreferent:	Univ.-Prof. Dr.-Ing. Marcus Geimer

Vorwort der Herausgeber

Wissen ist einer der entscheidenden Faktoren in den Volkswirtschaften unserer Zeit. Der Unternehmenserfolg wird mehr denn je davon abhängen, wie schnell ein Unternehmen neues Wissen aufnehmen, zugänglich machen und verwerten kann. Die Aufgabe eines Universitätsinstitutes ist es, hier einen wesentlichen Beitrag zu leisten. In den Forschungsarbeiten wird ständig Wissen generiert. Dieses kann aber nur wirksam und für die Gemeinschaft nutzbar werden, wenn es in geeigneter Form kommuniziert wird. Diese Schriftenreihe dient seit mehr als 20 Jahren als eine Plattform zum Transfer und macht damit das Wissenspotenzial aus aktuellen Forschungsarbeiten am IPEK - Institut für Produktentwicklung Karlsruhe* am Karlsruher Institut für Technologie (KIT) verfügbar. Die Forschung des IPEK ist dabei strukturiert in die Kategorien Systeme, Methoden und Prozesse, um so der Komplexität heutiger Produktentwicklung ganzheitlich gerecht zu werden. Erst die Verknüpfung dieser drei Kategorien ermöglicht die Synthese innovativer Systeme durch Nutzung neuester Methoden und Prozesse. Gleichzeitig werden durch die Systemsynthese die erforschten neuen Methoden und Prozesse validiert und deren Mehrwert für die Praxis abgesichert. Dieses Forschungskonzept prägt nicht nur das IPEK-Leitbild, sondern auch den Charakter dieser Schriftenreihe, da immer alle drei Kategorien und deren Wechselwirkungen berücksichtigt werden. Jeder Band setzt hier individuelle Schwerpunkte und adressiert dabei folgende Forschungsgebiete des IPEK:

- das Entwicklungs- und Innovationsmanagement,
- die Entwicklungs- und Konstruktionsmethodik,
- der Leichtbau von der Ebene des ganzen Systems bis hinunter zur Optimierung des Bauteils,
- die Validierung technischer Systeme auch unter Berücksichtigung der NVH Aspekte (Noise, Vibration, Harshness) mit dem Fokus auf Schwingungen und Akustik an Komponenten und in den Gesamtsystemen sowie deren subjektiver Beurteilung durch den Menschen,
- die Antriebssystemtechnik mit den Schwerpunkten komplette Antriebslösungen für Fahrzeuge und Maschinen,
- das Design, die Tribologie und Erprobung von Kupplungen und Bremsen sowie
- die Gerätetechnik mit dem Schwerpunkt auf Power-Tools.

Die Forschungsberichte stellen Ergebnisse unserer Forschung sowohl anderen Wissenschaftlern als auch den Unternehmen zu Verfügung um damit die Produktentwicklung in allen ihren Facetten mit innovativen Impulsen zu optimieren

Albert Albers und Sven Matthiesen

* Eh.: Institut für Maschinenkonstruktionslehre und Kraftfahrzeugbau, Universität Karlsruhe (TH)

Vorwort zu Band 154

Die zunehmende Komplexität moderner Produkte macht es zwingend erforderlich, in der Produktentwicklung mit neuen Ansätzen die Beschreibung des Verhaltens von Teilsystemen und Komponenten zu ermöglichen, um diese dann im Gesamtzusammenhang des Produktes modellieren zu können. Dabei stellen sich große Herausforderungen, da es bis heute praktisch nicht möglich ist, für viele Effekte stringente, physikalisch deduktive Modelle zur Verfügung zu haben. Auch wenn in den grundlagenwissenschaftlichen Arbeiten – zum Beispiel zum Thema der Tribologie – Ansätze zur Modellierung von Reibung bis in sehr große Detailtiefen vorangetrieben werden, sind diese bis heute nicht geeignet, um in der Produktsynthese die Reibungsphänomene in Wirkflächenpaaren durchgängig zu beschreiben. Hier arbeitet die Forschungsgruppe um ALBERS bereits seit vielen Jahren, um zu verbesserten Ansätzen zu kommen. Wichtig scheint es, auch neuere Ansätze einer Nutzung von Modellen, die durch empirische Beschreibungsansätze gewonnen werden, in die Produktentwicklungsprozesse zu integrieren und somit nutzbar zu machen. Auch hier hat es in den letzten Jahren erhebliche Fortschritte gegeben. Nachteilig ist bei den nichtparametrischen Regressionsmodellen, wie beispielsweise den Künstlichen neuronalen Netzen (KNN), dass hier keine deduktive physikalische Beschreibung des Verhaltens und der Wechselwirkungen mit Systemgrößen vorhanden ist. Lineare Regressionsmodelle basieren im Gegensatz zu den neuronalen Netzen auf Polynomen und erlauben so eine schnelle Erfassung und Interpretation des abgebildeten Verhaltens. Dies ist für die Produktentwicklung sehr wichtig, da im Syntheseprozess auch ein Verständnis der Wechselwirkungen in den Designräumen für den Konstrukteur bei der Gestalt-Neudefinition sehr hilfreich ist. Auf der anderen Seite haben lineare Regressionsmodelle in ihrem reinen Ansatz Probleme bei der Abbildung von Nichtlinearitäten. Hier sind Ansätze notwendig, wie diese Schwierigkeiten in der Praxis effektiv überwunden werden können und gleichzeitig auch Vorwissen aus der Produktentwicklung in die Modellbildung einfließen kann.

An dieser Stelle setzt die Arbeit von Herrn Dr.-Ing. Martin Bochterle an. Er hat sich zum Ziel gesetzt, durch eine konsequente Aufbereitung des im Bereich der Grundlagenforschung erarbeiteten Wissens zur Modellbildung für den Einsatz in den Syntheseprozessen in der Produktentwicklung entsprechende Vorgehensweisen und Methoden zu erstellen, die es dem Produktentwickelnden erlauben, effizient und effektiv eine entsprechende Modellbildung für die konkreten Herausforderungen in der Produktentwicklung zu erstellen. Im Rahmen einer mitarbeitenden Forschung im Kontext eines Unternehmens gelingt es Herrn Dr.-Ing. Martin Bochterle diesen Zusammenhang zwischen den theoretischen Grundlagen und der realen Produktentwicklung an konkreten Beispielen aufzuzeigen. Die entstandene Methode und der zugehörige Vorgehensprozess leisten einen wichtigen Beitrag für den Produktentwicklungsprozess in der Praxis der Unternehmen.

Kurzfassung

Die Bedeutung von Simulationsmodellen in der Produktentwicklung nimmt weiter stetig zu. Diese finden verschiedenste Anwendungen im Handlungssystem der Produktentwicklung und in zunehmendem Maße auch als Teil des Objektsystems, also des entwickelten Produkts selbst. Nicht immer lassen sich solche Simulationsmodelle auf Basis physikalischer Grundgleichungen entwickeln. In einem solchen Fall müssen die Modelle stattdessen auf empirischem Weg ermittelt werden.

Hierbei müssen auch die wirtschaftlichen Randbedingungen der Produktentwicklung eingehalten werden. Eine hohe Effizienz bei der Modellbildung ist also essentiell. Eine Methodik zur Minimierung des Versuchsaufwands bei gleichzeitiger Maximierung der erhaltenen Modellqualität ist die statistische Versuchsplanung. Diese lässt sich unterteilen in Verfahren zur Erstellung linearer und auch nichtparametrischer Regressionsmodelle. Nichtparametrische Regressionsmodelle, wie beispielsweise künstliche neuronale Netze, besitzen keine für Menschen interpretierbaren Parameter. Lineare Regressionsmodelle basieren im Gegensatz dazu auf Polynomen und erlauben so eine schnelle Erfassung und Interpretation des abgebildeten Verhaltens. Dies kann zu einem gesteigerten Systemverständnis führen, was in der Produktentwicklung von herausragender Bedeutung ist. Deshalb erscheint ihr Einsatz hier sehr attraktiv.

Lineare Regressionsmodelle haben in ihrer Grundform allerdings eine beschränkte Abbildungsleistung von Nichtlinearitäten. Ein generelles Problem in der statistischen Versuchsplanung ist zudem, dass die übliche vollständige Randomisierung bei technischen Versuchsreihen oft deren Effizienz beeinträchtigt. Des Weiteren stellt sich bei der Erstellung empirischer Modelle oft die Frage, wie Vorwissen über die inneren Zusammenhänge des abgebildeten Systems in das Modell einfließen kann. Um lineare Regressionsmodelle für Simulationszwecke in der Produktentwicklung optimal nutzen zu können, müssen diese Probleme gelöst werden. Dies ist das Ziel dieser Arbeit.

Hierzu werden die genannten Probleme zunächst analysiert und Randbedingungen für deren Lösung im Umfeld der Produktentwicklung ermittelt. Anschließend werden geeignete Lösungsansätze aus der Literatur ermittelt. Diese umfassen hybride Grey-Box-Modelle, semiparametrische Regressionsmodelle in Form von Varying-Coefficient-Modellen und Verfahren zur Erstellung und Auswertung von Split-Plot-Versuchsreihen. Der Großteil dieser Verfahren wird aktuell hauptsächlich in Disziplinen abseits der Produktentwicklung eingesetzt. In dieser Arbeit wird ihre Anpassung an die Randbedingungen der Produktentwicklung beschrieben und ihr Einsatz sowie die damit erzielbaren Ergebnisse anhand von zwei Praxisbeispielen demonstriert. In Summe entsteht eine Methode, mit deren Hilfe auch komplexe Eigenschaften technischer Systeme effizient modelliert werden können, während die gute Verständlichkeit linearer Regressionsmodelle erhalten bleibt.

Abstract

The importance of simulation models in product development is steadily increasing. Models find various applications in the operating system of product development and increasingly also as part of the system of objects, i.e. the developed product itself. Such simulation models cannot always be developed on the basis of physical equations. In such a case, empirical models must be used instead.

In doing so, the economic requirements of product development must also be kept. High efficiency in modelling is therefore essential. A methodology for minimizing the experimental effort while simultaneously maximizing the obtained model quality is design of experiments. It can be subdivided into methods for the creation of linear and non-parametric regression models. Nonparametric regression models, such as artificial neural networks, do not have parameters that can be interpreted by humans. In contrast, linear regression models are based on polynomials and thus allow a fast and easy interpretation of the mapped behavior. Thus, they can contribute to an increased understanding of the product, which is of outstanding importance in product development. Therefore, their use in this area seems very attractive.

However, linear regression models in their basic form have a limited capability of mapping nonlinearities. A general problem in design of experiments is that the usual complete randomization of technical experiments often reduces their efficiency. Furthermore, when creating empirical models, the question often arises as to how prior knowledge of the inner workings of the mapped system can be incorporated into the model. In order to use linear regression models for simulation purposes in product development, these problems must be solved. This is the goal of this thesis.

For this purpose, the mentioned problems are first analyzed and boundary conditions for their solution in the context of product development are determined. Subsequently, suitable solutions are determined from the literature. These include hybrid grey box models, semi-parametric regression models in the form of varying coefficient models and various methods for the design and analysis of split-plot experiments. The majority of these methods are currently mainly used in disciplines outside of product development. This thesis describes their adaptation to the boundary conditions of product development and demonstrates their application and the results achievable with them using two practical examples. Overall, a method is created which allows to model even complex properties of technical systems efficiently, while maintaining the good comprehensibility of linear regression models.

Danksagung

Die vorliegende Forschungsarbeit entstand im Zuge eines Kooperationsprojektes während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am IPEK – Institut für Produktentwicklung des Karlsruher Instituts für Technologie (KIT) und als Forschungsingenieur bei der Festo SE & Co. KG. An dieser Stelle möchte ich mich bei allen Menschen bedanken, die mich bei der Anfertigung der Dissertation und darüber hinaus unterstützt haben.

Mein besonderer Dank gilt Herrn Prof. Dr.-Ing. Dr. h.c. Albert Albers, der mein Forschungsvorhaben als Doktorvater im Rahmen von zahlreichen wissenschaftlichen Gesprächen mit stets wertvollen Impulsen geprägt hat. Für die Übernahme des Korreferates bedanke ich mich sehr herzlich bei Herrn Prof. Dr.-Ing. Marcus Geimer. Herrn Prof. Dr.-Ing. habil. Thomas Böhlke möchte ich für die Übernahme des Prüfungsvorsitzes danken.

Ein großer Dank gilt meinem Vorgesetzten Dr. Wolfgang Gauchel bei der Festo SE & Co. KG. Durch viele spannende Diskussionen und Anregungen, sowie das Gewähren von benötigten Freiräumen hat er entscheidend zur Fertigstellung dieser Arbeit beigetragen.

Beim gesamten IPEK-Team bedanke ich mich für die tolle Arbeitsatmosphäre und die ausdauernde Unterstützung. Besonders bedanken möchte ich mich bei Markus Spädinger, Sebastian Schmidt, Benoit Lorentz und Petra Müller.

Zuletzt gilt mein größter Dank meiner Familie, ganz besonders meiner Frau Dana und meinen Eltern, für ihre Unterstützung, Geduld und den mir stets gegebenen Rückhalt.

April 2022

Martin Bochterle

Inhaltsverzeichnis

Kurzfassung	i
Abstract	iii
Danksagung	v
Abbildungsverzeichnis	xi
Tabellenverzeichnis	xv
Abkürzungsverzeichnis	xvii
Formelzeichen	xix
1 Einleitung	1
2 Grundlagen und Stand der Forschung	5
2.1 Produktentstehung.....	5
2.1.1 PGE – Modell der Produktgenerationsentwicklung nach Albers ..	6
2.1.2 Das Integrierte Produktentstehungsmodell - iPeM.....	7
2.1.3 Problemlösungsprozess SPALTEN nach Albers.....	12
2.2 Simulationsmodelle in der Produktentwicklung.....	13
2.2.1 Simulationsmodelle als Teil des Handlungssystems.....	13
2.2.2 Simulationsmodelle als Teil des Ziel- und Objektsystems.....	18
2.3 Modellbildung.....	19
2.3.1 Contact and Channel-Ansatz.....	20
2.3.2 White-Box-Modelle.....	21
2.3.3 Black-Box-Modelle.....	22
2.3.4 Grey-Box-Modelle.....	23
2.4 Grundlagen der statistischen Versuchsplanung.....	24
2.4.1 Begriffsdefinitionen.....	25
2.4.2 Grundprinzipien für valide Versuchsergebnisse.....	30
2.5 Statistische Versuchsplanung für lineare Regressionsmodelle.....	32
2.5.1 Versuchspläne.....	33
2.5.2 Das klassische lineare Modell.....	47
2.5.3 Das allgemeine lineare Modell.....	60
2.5.4 Das gemischte lineare Modell.....	64
2.5.5 Split-Plot-Versuche.....	69
2.6 Statistische Versuchsplanung für nichtparametrische Regressionsmodelle.....	75
2.6.1 Versuchspläne.....	76
2.6.2 Nichtparametrische Regressionsmodelle.....	79
2.6.3 Semiparametrische Regressionsmodelle.....	84

3	Motivation und Zielsetzung der Arbeit.....	87
4	Forschungsdesign	93
5	Praxisbeispiel: pneumatisch angetriebene Prozessarmaturen	95
5.1	Simulationsmodelle pneumatisch angetriebener Prozessarmaturen	96
5.2	Kugelhähne.....	97
5.3	Pneumatische Schwenkantriebe.....	100
6	Integration von physikalischem Vorwissen in empirische Modelle	103
6.1	Spezielle Anforderungen in der Produktentwicklung	105
6.2	Lösungsansatz.....	106
6.3	Anwendung des Lösungsansatzes in der Produktentwicklung	107
6.3.1	Grundlegende Vorgehensweise	108
6.3.2	Paralleler Ansatz hybrider Grey-Box-Modelle.....	110
6.3.3	Serieller Ansatz hybrider Grey-Box-Modelle.....	113
6.4	Demonstration am Beispiel der Schwenkantriebe	116
6.4.1	Modellierung der Kinematik und der Druckkräfte.....	116
6.4.2	White-Box Reibungsmodellierung	118
6.4.3	Grey-Box Reibungsmodellierung	121
6.4.4	Ergebnisse	124
6.5	Zwischenfazit.....	131
7	Abbildung von Nichtlinearitäten	133
7.1	Spezielle Anforderungen in der Produktentwicklung	134
7.2	Lösungsansatz.....	135
7.3	Anwendung des Lösungsansatzes in der Produktentwicklung	136
7.3.1	Auswahl der linear und nichtparametrisch abzubildenden Faktoren.....	137
7.3.2	Auswahl eines geeigneten Regressionsmodells	137
7.3.3	Aufbau der Versuchsreihe.....	139
7.3.4	Ermittlung des semiparametrischen Regressionsmodells	143
7.4	Demonstration am Beispiel von Kugelhähnen.....	143
7.4.1	Modellierung der Kinematik und der Strömungskräfte.....	144
7.4.2	White-Box Reibungsmodellierung	145
7.4.3	Reibungsmodellierung durch ein semiparametrisches Regressionsmodell.....	148
7.4.4	Ergebnisse	151
7.5	Zwischenfazit.....	161
8	Steigerung der Effizienz von Versuchsreihen.....	163
8.1	Spezielle Anforderungen in der Produktentwicklung	164
8.2	Lösungsansatz.....	166

8.3	Anwendung von Split-Plot-Versuchsreihen in der Produktentwicklung	166
8.3.1	Empfehlungen für Split-Plot-Versuchsreihen mit bis zu drei Randomisierungsebenen.....	167
8.3.2	Empfehlungen für Split-Plot-Versuchsreihen mit mehr als drei Randomisierungsebenen.....	169
8.4	Demonstration am Beispiel der Kugelhähne.....	172
8.4.1	Auswahl und Aufbau des Versuchsplans.....	172
8.4.2	Aufbau des gemischten linearen Modells und Regression.....	173
8.4.3	Ergebnisse.....	176
8.5	Zwischenfazit.....	178
9	Zusammenfassung und Ausblick.....	181
9.1	Zusammenfassung.....	181
9.2	Ausblick.....	183
	Literaturverzeichnis.....	I
	Anhang.....	XIII

Abbildungsverzeichnis

Abbildung 2.1:	Eine Ebene des integrierten Produktentstehungsmodells – iPeM.....	9
Abbildung 2.2:	Das iPeM im Kontext der Produktgenerationsentwicklung	11
Abbildung 2.3:	Problemlösungsprozess SPALTEN.....	12
Abbildung 2.4:	Dimensionen im IPEK-XiL-Ansatz nach Albers	17
Abbildung 2.5:	Das IPEK XiL-Framework am Beispiel der Fahrzeugentwicklung	18
Abbildung 2.6:	C&C ² -A am Beispiel eines Zahnradpaares	21
Abbildung 2.7:	Paralleler und serieller Ansatz hybrider Grey-Box-Modelle	23
Abbildung 2.8:	Darstellung einer reinen Wechselwirkung	27
Abbildung 2.9:	Lack-of-fit	29
Abbildung 2.10:	Over-fit	29
Abbildung 2.11:	Verminderung des Fehlereinflusses von Trends durch Randomisierung	31
Abbildung 2.12:	Beispiele für die Abbildungsleistung mehrdimensionaler Polynome	35
Abbildung 2.13:	Vergleich zwischen orthogonalen und kollinearen Versuchsplänen.	36
Abbildung 2.14:	Darstellung eines konventionellen One-factor-at-a-time-Versuchsplans.....	37
Abbildung 2.15:	Darstellung eines vollständig faktoriellen 2 ³ Versuchsplans	38
Abbildung 2.16:	Darstellung eines fraktionell faktoriellen 2 ³⁻¹ Versuchsplans....	41
Abbildung 2.17:	Darstellung eines vollständig faktoriellen 3 ³ Versuchsplans	42
Abbildung 2.18:	Darstellung eines fraktionell faktoriellen 3 ³⁻¹ Versuchsplans....	43
Abbildung 2.19:	Darstellung eines Box-Behnken-Versuchsplans.....	44
Abbildung 2.20:	Darstellung eines zentral zusammengesetzten Versuchsplans	45
Abbildung 2.21:	Residuendiagramme: Residuen über der Zielgröße	57
Abbildung 2.22:	Residuendiagramme: Residuen über der Versuchsreihenfolge	58
Abbildung 2.23:	Residuendiagramme: Residuen über einem Faktor	58
Abbildung 2.24:	Residuendiagramme: Residuen über verzögerte Residuen	59

Abbildung 2.25:	Residuendiagramme: Residuen im Wahrscheinlichkeitsnetz ..	59
Abbildung 2.26:	TP-Spline mit zwei Knoten.....	81
Abbildung 2.27:	Schematische Darstellung eines neuronalen Netzes	83
Abbildung 4.1:	Entwicklungspfade entwicklungsmethodischer Unterstützung	94
Abbildung 5.1:	Pneumatisch angetriebener Kugelhahn mit Stellungsregler....	96
Abbildung 5.2:	Aufbau eines Kugelhahns	98
Abbildung 5.3:	C&C ² -A Modellierung eines Kugelhahns.....	99
Abbildung 5.4:	Gängige Bauweisen pneumatischer Schwenkantriebe	101
Abbildung 5.5:	C&C ² -A Modellierung eines Zahnstange-Ritzel-Antriebs	102
Abbildung 6.1:	Kräfte und Momente im Schwenkantrieb	117
Abbildung 6.2:	Reibmomentenverläufe nach verschiedenen Stillstandszeiten	125
Abbildung 6.3:	Regressionskoeffizienten für das Haftreibmoment.....	127
Abbildung 6.4:	Regressionskoeffizienten für das Gleitreibmoment.....	128
Abbildung 6.5:	Regressionskoeffizienten für die Zeitkonstante.....	129
Abbildung 6.6:	Regressionskoeffizienten für die charakteristische Wegstrecke	130
Abbildung 6.7:	Residuendiagramme der vier Black-Box-Teilmodelle.....	132
Abbildung 7.1:	Gekreuzter Versuchsplan für semiparametrische Regressionsmodelle	142
Abbildung 7.2:	An der Ventilkugel angreifende Kräfte	145
Abbildung 7.3:	Wirkläche zwischen Primärdichtung und Ventilkugel	146
Abbildung 7.4:	Mittleres Reibmoment.....	152
Abbildung 7.5:	Einfluss der Druckdifferenz	154
Abbildung 7.6:	Einfluss der Temperatur	155
Abbildung 7.7:	Wechselwirkung zwischen Druckdifferenz und Temperatur..	156
Abbildung 7.8:	Quadratischer Einfluss der Temperatur	157
Abbildung 7.9:	Einfluss der Öffnungsgeschwindigkeit	158
Abbildung 7.10:	Quadratischer Einfluss der Öffnungsgeschwindigkeit	158
Abbildung 7.11:	Wechselwirkung zwischen Öffnungsgeschwindigkeit und Druck	159
Abbildung 7.12:	Einfluss des Verschleißes.....	160
Abbildung 7.13:	Quadratischer Einfluss des Verschleißes	160

Abbildung 8.1:	Zeitaufwand für die Durchführung der drei Versuchsreihen...	177
Abbildung 8.2:	Vergleich der Schätzwerte mit den Messwerten des Testdatensatzes	177
Abbildung 8.3:	Bestimmtheitsmaße der drei Modelle für den Testdatensatz .	178
Abbildung A.1:	Quadratischer Einfluss der Druckdifferenz	XVI
Abbildung A.2:	Wechselwirkung zwischen Geschwindigkeit und Temperatur	XVI
Abbildung A.3:	Wechselwirkung zwischen Geschwindigkeit und Schaltspielzahl	XVII
Abbildung A.4:	Wechselwirkung zwischen Temperatur und Schaltspielzahl.	XVII
Abbildung A.5:	Wechselwirkung zwischen Druck und Schaltspielzahl.....	XVIII

Tabellenverzeichnis

Tabelle 2.1:	Beispielhafte statistische Auswertung einer Versuchsreihe in Tabellenform	53
Tabelle 4.1:	Fünf Kategorien von entwicklungsmethodischer Forschung nach Cantamessa	93
Tabelle 6.1:	Parameterwerte des Reibungsmodells in Abhängigkeit der Black-Box-Faktoren.....	126
Tabelle 8.1:	Equivalent-Estimation-Versuchsplan inklusive der Versuchsergebnisse.....	174
Tabelle A.1:	Regressionsmodell für das Haftreibmoment.....	XIII
Tabelle A.2:	Regressionsmodell für das Gleitreibmoment	XIII
Tabelle A.3:	Regressionsmodell für die Zeitkonstante.....	XIV
Tabelle A.4:	Regressionsmodell für die charakteristische Wegstrecke.....	XIV
Tabelle A.5:	Zur Untersuchung des Schwenkantriebs eingesetzter Versuchsplan	XV

Abkürzungsverzeichnis

C	Connector
C&C ² -A	Contact and Channel-Ansatz
DoE	Statistische Versuchsplanung (Design of Experiments)
EED	Equivalent-Estimation-Design
ETC-Faktor	Easy-To-Change-Faktor
FE-Modell	Finite-Elemente-Modell
GLS	Gewichtete Methode der kleinsten Fehlerquadrate (Generalized Least Squares)
HiL	Hardware-in-the-Loop
HTC-Faktor	Hard-To-Change-Faktor
IC	Informationscheck
IPEK	Institut für Produktentwicklung am KIT
iPeM	Integriertes Produktentstehungsmodell
KIS	Kontinuierlicher Ideen Speicher
KIT	Karlsruher Institut für Technologie
KNN	Künstliches neuronales Netz
LSS	Leitstützstruktur
MARS	Multivariate Adaptive Regression Splines
MiL	Model-in-the-Loop
MKM	Münchener Produktkonkretisierungsmodell
ML	Maximum-Likelihood

MLP	Multilayer Perceptron
MPC	Modellbasierte Regelung (Model Predictive Control)
OLS	Gewöhnliche Methode der kleinsten Fehlerquadrate (Ordinary Least Squares)
PLT	Problemlösungsteam
PTFE	Polytetrafluorethylen
RCP	Rapid-Control-Prototyping
REML	Restricted-Maximum-Likelihood
SiL	Software-in-the-Loop
SP	Sub-Plot
WFP	Wirkflächenpaar
WK	Wirkkontakt
WP	Whole-Plot
XiL	X-in-the-Loop

Formelzeichen

Besondere Schreibweisen am Beispiel einer Variablen a

a	Skalare Größe
\mathbf{a}	Vektor oder Matrix
\mathbf{a}'	Transponierte Matrix von \mathbf{a}
\hat{a}	Schätzwert für a
\bar{a}	Mittelwert von a

Formelzeichen

\mathbf{b}	(Regressions-)Koeffizienten der Effekte aller zufälligen Faktoren
c	frei wählbare Konstante
CI	Konfidenzintervall
$Cov()$	Kovarianz(-matrix)
$E()$	Erwartungswert einer Größe
\mathbf{G}	Kovarianzmatrix der zufälligen Effekte
i	Laufindex für den aktuell betrachteten Versuch oder das aktuell betrachtete Cluster einer Versuchsreihe
\mathbf{I}	Einheitsmatrix
\mathbf{J}	Blockdiagonalmatrix
k	Knoten eines Splines
L	Likelihoodfunktion
m	Anzahl der Cluster in einer Versuchsreihe
n	Anzahl der Versuchswiederholungen
N	Gesamtzahl der Versuche in einem Versuchsplan
N_v	Anzahl der Versuche im vollständig faktoriellen Teil eines zentral zusammengesetzten Versuchsplans
p	Anzahl der Faktoren eines linearen Modells

$p_{\text{statistic}}$	p-Wert
PI	Prognoseintervall
$t_{\text{statistic}}$	t-Wert
R^2	Bestimmtheitsmaß
\bar{R}^2	Korrigiertes Bestimmtheitsmaß
R	Kovarianzmatrix der Residuen in der marginalen Formulierung des gemischten linearen Modells
R_m	Universelle Gaskonstante
x	Einstellung eines Faktors
x	Einstellungen aller Faktoren für einen Versuch
X	Designmatrix (Enthält die Einstellungen aller festen Faktoren für alle Versuche einer Versuchsreihe)
y	Zielgröße
y	Ergebnisvektor (Werte der Zielgröße für alle Versuche einer Versuchsreihe)
Z	Designmatrix der zufälligen Effekte (Enthält die Einstellungen aller zufälligen Faktoren für alle Versuche einer Versuchsreihe)
α	Distanz der Sternpunkte vom Zentrum eines zentral zusammengesetzten Versuchsplans
β	(Regressions-)Koeffizient des Effekts eines festen Faktors
β	(Regressions-)Koeffizienten der Effekte aller festen Faktoren
δ	Whole-Plot Fehlerterm
ε	Residuum
ε	Residuenvektor (Residuen aller Versuche einer Versuchsreihe)
σ	Standardabweichung
σ^2	Varianz
Σ	Kovarianzmatrix der Residuen
Ω	Kovarianzmatrix der Residuen normiert auf σ^2

1 Einleitung

Der Anteil elektronischer Komponenten und zugehöriger Software steigt in modernen Produkten stetig an. So liegt beispielsweise der Wertanteil der Elektronik eines modernen Fahrzeugs mit Verbrennungsmotor bei ungefähr 16%. Im Jahr 2025 wird dieser Anteil in einem batterieelektrisch angetriebenen Fahrzeug bei etwa 35% liegen. Ähnliche Steigerungen lassen sich im Bereich der Softwareentwicklungskosten beobachten (Meissner et al., 2020, S. 10, 14). Durch diese Aufwände werden immer mehr und immer bessere Komfort-, Steuer-, Regel- und Diagnosefunktionen ermöglicht. Damit geht allerdings auch eine steigende Komplexität der zu entwickelnden Produkte einher.

Um trotz einer höheren Komplexität zuverlässige Produkte entwickeln zu können, muss der Aktivität der Validierung in der Produktentwicklung ein sehr hoher Stellenwert zukommen. Nach Albers (2010a, Hypothese drei) ist die Validierung die zentrale Aktivität der Produktentwicklung. Um eine deutlich gezieltere, frühere und umfangreichere Validierung zu ermöglichen, werden mechatronische Systeme mehr und mehr modellbasiert entwickelt.¹ Der grundlegende Gedanke hinter dieser Vorgehensweise ist, dass die Ursache einer Fehlfunktion bei der Validierung eines vollständigen mechatronischen Systems nur sehr schwer identifiziert werden kann. Dennoch muss eine Validierung stets am Gesamtsystem erfolgen, da nur dieses die kundenrelevante Funktion erfüllt (Albers et al., 2014a).

Um dieses Spannungsfeld zu lösen, werden bei der modellbasierten Entwicklung, parallel zur Entwicklung der physischen Komponenten des Produkts, alle funktionsrelevanten Komponenten des Systems in Simulationsmodellen abgebildet. Anschließend können beliebige Teilsysteme eines Produkts durch virtuelle Modelle ersetzt werden, die dann mittels spezieller Prüfstände mit physischen Teilsystemen desselben Produkts interagieren können. Diese Vorgehensweise bietet mehrere Vorteile: Da die inneren Zustände von Simulationsmodellen deutlich einfacher zu beobachten sind, als die von realen Systemen, wird die Eingrenzung von Fehlerquellen bei der Validierung deutlich vereinfacht. Zudem können in der Simulation Experimente nachgestellt werden, deren Durchführung in der Realität entweder zu teuer oder zu gefährlich wäre. Eine besonders flexible Beschreibung dieser Vorgehensweise liefert der IPEK-X-in-the-Loop-(IPEK-XiL)-Ansatz nach Albers². Um die Vorteile dieser Vorgehensweise erreichen zu können, werden jedoch geeignete Simulationsmodelle benötigt. Aufgrund der Interaktion mit physischen Teilsystemen

¹ Siehe Abschnitt 2.2.1.1 für Details

² Siehe Abschnitt 2.2.1.2

müssen Simulationsmodelle für den Einsatz in XiL-Systemen in Echtzeit lauffähig sein. Die Modelle dürfen also nur begrenzten Rechenaufwand benötigen.

Neben dem Bedarf von Simulationsmodellen zur Steigerung der Effizienz und Zuverlässigkeit der Produktentwicklung nimmt auch der Bedarf an direkt auf dem entwickelten Produkt lauffähigen Modellen stetig zu. Momentan entwickelt sich hierbei beispielsweise der Bereich der modellbasierten Regelungen (MPC) sehr schnell (siehe z.B. Raković & Levine, 2019). Aber auch für andere Bereiche, wie für Condition Monitoring Algorithmen besteht ein stets wachsender Bedarf an Simulationsmodellen, die direkt auf dem entwickelten Produkt lauffähig sind. Unabhängig vom Anwendungsfall ist aus Gründen der Kosten- und Energieeffizienz die verfügbare Rechenleistung und Speicherkapazität für solche Simulationsmodelle meist stark eingeschränkt.

In Summe gibt es in der modernen Produktentwicklung einen sehr großen Bedarf an einfachen Simulationsmodellen, die schnell berechnet werden können und wenig Speicherplatz benötigen. Klassischerweise werden diese in der Produktentwicklung basierend auf Gleichungen aufgebaut, die physikalische Gesetzmäßigkeiten beschreiben.³ Es gibt jedoch auch Fälle, in denen dies nicht möglich ist. In vielen Fällen sind physikalisch basierte Modelle zu komplex. So benötigen beispielsweise Berechnungsverfahren für kontinuumsmechanische Vorgänge mit räumlich diskretisierten Modellen für die oben beschriebenen Anwendungsfälle im Allgemeinen eine viel zu hohe Rechenleistung. Systembestandteile, deren Verhalten von mehrdimensionalen Verformungen von Festkörpern oder Fließvorgängen von Flüssigkeiten abhängt, müssen deshalb für den oben beschriebenen Einsatz in der Produktentwicklung auf anderem Wege modelliert werden. Ebenso kommt es vor, dass es nicht möglich oder zu aufwendig wäre, ein auf physikalischen Grundgleichungen basierendes Modell zu erstellen.

In solchen Fällen kann das Verhalten des betreffenden Teilsystems experimentell untersucht und in geeignete empirische Modelle abgelegt werden. Eine mögliche Vorgehensweise hierzu liefert die statistische Versuchsplanung (DoE)⁴. Deren Ziel ist, mit möglichst geringem Untersuchungsaufwand möglichst genaue und valide Informationen über ein untersuchtes System zu erhalten.

Statistische Versuchsplanung wird in der Produktentwicklung bereits für die Erstellung und/oder Parametrierung von Simulationsmodellen eingesetzt. Es gibt jedoch einige Einschränkungen, die den möglichen Einsatzbereich hierbei eng begrenzen.

³ Siehe Abschnitt 2.3.2

⁴ Abgeleitet von der englischen Bezeichnung „Design of Experiments“

Für andere Wissenschaftszweige, wie beispielsweise die Medizin, Biologie oder Soziologie wurden in den letzten Jahrzehnten viele neue statistische Methoden entwickelt, die die Fähigkeiten und die Effizienz der statistischen Versuchsplanung erweitern und verbessern. In dieser Arbeit wird untersucht, inwieweit solche Methoden eingesetzt werden können, um einen breiteren und effizienteren Einsatz der statistischen Versuchsplanung für die Modellerstellung in der Produktentwicklung zu ermöglichen.

Nachfolgend werden in Kapitel 2 die Einsatzgebiete von Simulationsmodellen in der modernen Produktentwicklung vorgestellt, sowie die Grundlagen bis hin zum Stand der Forschung der für diese Arbeit relevanten statistischen Methoden zusammengefasst.

Daraus wird in Kapitel 3 die Motivation und Zielsetzung für diese Arbeit abgeleitet. Die Zielsetzung wird dabei in drei Teilziele gegliedert. In Kapitel 4 folgt die Beschreibung der methodischen Vorgehensweise zur Erreichung dieser Ziele. Die hierzu entwickelten Vorgehensweisen werden zur besseren Verständlichkeit anhand eines Praxisbeispiels demonstriert. Dieses Praxisbeispiel wird in Kapitel 5 erläutert. Anschließend wird in den Kapiteln 6 bis 8 die Erreichung jeweils eines Teilziels abgeleitet und vorgestellt. In Kapitel 9 werden abschließend die in dieser Arbeit erzielten Ergebnisse zusammengefasst und ein Ausblick auf mögliche daran anknüpfende Arbeiten gegeben.

2 Grundlagen und Stand der Forschung

Die in dieser Arbeit vorgestellten statistischen Methoden werden im Kontext der Modellbildung in Produktentwicklungsprozessen betrachtet. Um die daraus entstehenden Randbedingungen ableiten und die Einsatzgebiete der Methoden in der Produktentwicklung aufzeigen zu können, wird in Abschnitt 2.1 zunächst ein moderner Produktentwicklungsprozess anhand des integrierten Produktentstehungsmodells (iPeM) im Kontext der Produktgenerationsentwicklung beschrieben. Darauf basierend wird in Abschnitt 2.2 die Entwicklung und Nutzung von Modellen in der Produktentwicklung erläutert. In Abschnitt 2.3 findet sich ein kurzer Überblick über die grundlegenden Möglichkeiten zur Modellbildung. Eine dieser Möglichkeiten ist die rein empirische Modellbildung, zu deren Unterstützung statistische Versuchsplanung eingesetzt werden kann.

In den Abschnitten 2.4 bis 2.6 wird eine Einführung in den Stand der Forschung der statistischen Versuchsplanung gegeben. Diese ist darauf ausgerichtet, Ingenieuren einen verständlichen Einstieg in die Thematik zu ermöglichen. Dazu werden zunächst in Abschnitt 2.4 die notwendigen statistischen Fachbegriffe definiert und die Grundprinzipien für die Erzielung valider Versuchsergebnisse erläutert. In den Abschnitten 2.5 und 2.6 wird ein Überblick über den Stand der Forschung der statistischen Versuchsplanung für lineare sowie nichtparametrische Regressionsmodelle dargestellt. In beiden Abschnitten werden sowohl die zur Datengenerierung verwendbaren Versuchspläne, als auch die Regressionsverfahren zur Erzeugung der auf die Daten angepassten Modelle beschrieben. Da ein Großteil dieser Arbeit sich mit linearen Regressionsmodellen auseinandersetzt, wird der diesbezügliche Stand der Forschung auch deutlich detaillierter dargestellt als der von nichtparametrischen Regressionsmodellen.

2.1 Produktentstehung

Das Ziel der Vorstellung der Produktentstehungsprozesse in diesem Abschnitt ist, Einsatzmöglichkeiten der in dieser Arbeit vorgestellten statistischen Methoden in der modernen Produktentstehung aufzuzeigen und einzuordnen, sowie die daraus entstehenden Randbedingungen abzuleiten. Dafür wird ein Produktentstehungsmodell benötigt, das die gesamte Produktentstehung umspannt und aus dem sich die Randbedingungen für den Einsatz der statistischen Methoden ableiten lassen.

Es gibt mehrere Modelle, die den Produktentstehungsprozess in zeitlich und inhaltlich trennbare Phasen gliedern. Bekannte Vertreter sind die VDI-Richtlinie 2221, das V-Modell nach VDI 2206 oder der Stage-Gate-Prozess nach Cooper (1990).

Zudem existieren ganzheitliche Modelle, die versuchen, die Phasen und Aktivitäten des Produktentstehungsprozesses in mehreren Auflösungsgraden zu beleuchten und dabei auch die unterschiedlichen Stakeholder (z.B. Entwicklung, Produktion, Management) zu berücksichtigen. Dazu zählen das Münchener Produktkonkretisierungsmodell (MKM, Ponn & Lindemann, 2011), das 3-Zyklen-Modell der Produktentstehung (Gausemeier et al., 2009) oder das integrierte Produktentstehungsmodell – iPeM (Albers & Meboldt, 2007).

Der Sachverhalt, dass industrielle Produkte in Produktgenerationen entwickelt werden, hat großen Einfluss auf die Einsatzmöglichkeiten der vorgestellten statistischen Methoden. Um die Randbedingungen für deren Einsatz umfassend klären zu können, muss das verwendete Produktentstehungsmodell dies umfassend berücksichtigen. Das integrierte Produktentstehungsmodell im Kontext des Modells der PGE - Produktgenerationsentwicklung ist das einzige Produktentstehungsmodell, das diese Forderung erfüllt. Aus diesem Grund erfolgt die Beschreibung von Produktentstehungsprozessen nachfolgend anhand von diesem Modell.

2.1.1 PGE – Modell der Produktgenerationsentwicklung nach Albers

Nach Albers et al. (2019) bestehen bei allen Produktentwicklungen *Referenzsysteme*, die als Basis für die Entwicklung neuer *Produktgenerationen* dienen. Referenzsysteme können sowohl Zwischenstände aus Forschung und Entwicklung als auch fertiggestellte Referenzprodukte oder Teilsysteme daraus beinhalten. Die Funktions- und Baustruktur der Referenzsystemelemente kann teilweise in neue Produktgenerationen übernommen werden und ist zudem Ausgangspunkt für Variationen. Dieser Sachverhalt wird im *PGE – Modell der Produktgenerationsentwicklung nach Albers* aufgenommen.

„Es postuliert, dass jede Produktentwicklung auf der Basis der Elemente eines zuvor definierten Referenzsystems durch gezielte Variationen in der Synthese abgebildet werden kann. Dieser Grundsatz hat weitreichende Auswirkungen. Er ist bei konsequenter Anwendung ein Schlüssel zu einer effizienteren, effektiveren, aber auch sichereren Produktentwicklung. Dabei gilt der in der PGE beschriebene Zusammenhang sowohl auf der Ebene von Produktarchitekturen auf Gesamtsystemebene als auch bis hinunter zu den Teilsystemen, beziehungsweise letztendlich den Komponenten in den Teilsystemen.“

A. Albers (persönl. Mitteilung, 31.08.2021)

Die in dieser Definition beschriebenen *Teilsysteme* und ihre Komponenten können dabei sowohl Elemente des Zielsystems¹ (z.B. Produktprofil, Anforderungen und Grenzwerte), als auch des Objektsystems (z.B. Lösungsprinzipien und Gestalt) sein (Klingler, 2017, S. 13). Die drei Variationsarten, mit denen Teilsysteme angepasst werden können, sind nach Albers et al. (2015, S. 4 f.):

- Die *Prinzipvariation*. Bei ihr wird das grundlegende Lösungsprinzip des Teilsystems verändert.
- Die *Gestaltvariation*. Bei ihr wird das Lösungsprinzip des Teilsystems beibehalten, aber die Gestalt maßgeblich verändert.
- Die *Übernahmevariation*. Bei ihr werden Prinzip und Gestalt des Teilsystems weitestgehend beibehalten. Lediglich an den Schnittstellen zu anderen Teilsystemen werden möglichst geringe Anpassungen vorgenommen.

Die Wahl der Anteile entscheidet auch über die Innovationsfähigkeit sowie das technische und wirtschaftliche Risiko des entwickelten Produkts. Eine angepasste Wahl der Variationsarten für die verschiedenen Teilsysteme ist deshalb entscheidend. (Albers et al., 2017b, S. 18)

Zur Unterscheidung der drei Variationsarten kann der Contact and Channel-Ansatz (C&C²-A) Indizien liefern (Albers et al., 2016b; Bursac, 2016, S. 39)²:

- Das Hinzufügen oder Entfernen von Wirkflächenpaaren indiziert eine Prinzipvariation.
- Veränderungen von Wirkflächenpaaren und Leitstützstrukturen indizieren eine Gestaltvariation.
- Veränderungen an Connectoren indizieren eine Übernahmevariation.

2.1.2 Das Integrierte Produktentstehungsmodell - iPeM

Die zentrale Grundlage für das Integrierte Produktentstehungsmodell (iPeM) bildet das von Ropohl (2009) entwickelte *ZHO-Modell*. In diesem wird die Produktentwicklung als soziotechnisches System, bestehend aus drei wesentlichen Teilsystemen, beschrieben. Dabei steht *Z* für *Zielsystem*, *H* für *Handlungssystem* und *O* für *Objektsystem*. Das Zielsystem enthält die Entwicklungsziele, an denen sich die Ent-

¹ Eine Definition von Ziel- und Objektsystem folgt in Abschnitt 2.1.2

² Siehe Abschnitt 2.3.1 für Details. Dort folgen auch die Definitionen von Wirkflächenpaaren, Leitstützstrukturen und Connectoren.

wickler orientieren. Die im Handlungssystem durchgeführten Aktivitäten zur Erreichung der im Zielsystem definierten Ziele führen zum Objektsystem, das die Ergebnisse der Produktentwicklung enthält.

Albers und Mebold (2007) sowie Albers und Braun (2011a) bauen darauf ihr *integriertes Produktentstehungsmodell (iPeM)* auf. Dieses beschreibt Produktentwicklung als Transformierung des Zielsystems in ein korrespondierendes Objektsystem durch das Handlungssystem.

„Based on systems theory, product engineering can be described as the transfer of an (initially vague) system of objectives into a concrete system of objects by an operating system.“

(Albers, 2010a, Hypothese zwei)

Das iPeM ist im Gegensatz zu klassischen Vorgehens- und Prozessmodellen ein Metamodell, das die individuelle Abbildung des Prozessverlaufs jeder Produktentwicklung erlaubt (Meboldt, 2008). Eines der Ziele des iPeM ist es, den kreativen Prozess der Produktentwicklung durch eine strukturierte Vorgehensweise zu unterstützen. Es kann zudem dabei helfen, die Lücke zwischen Entwicklung und Management sowie Controlling zu schließen, indem es ein konsistentes, für alle Unternehmensbereiche geeignetes Produktentstehungsmodell darstellt (Albers et al., 2010; Albers & Braun, 2011b).

Das Handlungssystem ist im iPeM ein soziotechnisches System, das Aktivitäten, Methoden, Prozesse und die dazugehörigen Ressourcen beinhaltet. Wie in Abbildung 2.1 gezeigt, wird das Handlungssystem im iPeM in drei Bereiche gegliedert: Die *Aktivitätenmatrix*, das *Ressourcen-System* und das *Phasenmodell*.

Die beiden Dimensionen der **Aktivitätenmatrix** beinhalten die Aktivitäten der Produktentstehung (*Makro-Aktivitäten*, senkrecht in Abbildung 2.1) sowie die Aktivitäten der Problemlösung (*Mikro-Aktivitäten*, waagrecht in Abbildung 2.1). Die Mikro-Aktivitäten werden aus dem Problemlösungsprozess *SPALTEN* (siehe Abschnitt 2.1.3) abgeleitet. Die Makro-Aktivitäten lassen sich in die beiden Aktivitäten-Cluster *Basisaktivitäten* (oben in Abbildung 2.1) und *Kernaktivitäten* (unten in Abbildung 2.1) unterteilen. Die immer wiederkehrenden Basisaktivitäten beinhalten organisatorische Aktivitäten sowie die Aktivität *Validieren und Verifizieren*. Diese ist laut Albers (2010a, Hypothese drei) die zentrale Aktivität im Produktentstehungsprozess.

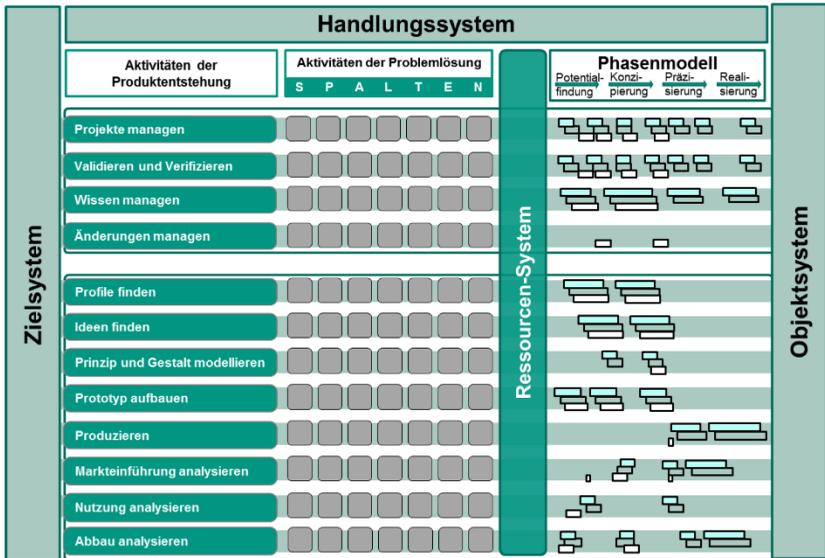


Abbildung 2.1: Eine Ebene des integrierten Produktentstehungsmodells – iPeM im Kontext der Produktgenerationsentwicklung (Albers et al., 2016d)³

Die Kernaktivitäten beinhalten die auf das Produkt ausgerichteten schöpferischen Aktivitäten der Produktentwicklung vom Finden eines Produktprofils bis zum Analysieren des Abbaus des Produkts. Sie orientieren sich an den Elementen des Produktlebenszyklus, sind aber an keine zeitliche oder logische Abfolge gebunden. Beschreibungen aller Makro-Aktivitäten finden sich beispielsweise in Albers et al. (2016d) sowie in Albers und Braun (2011a).

Das **Ressourcen-System** beinhaltet für die Produktentwicklung verfügbare Mitarbeiter, Arbeitsmittel (Prüfstände, Werkzeuge, Maschinen...), Informationen und das Budget.

Das **Phasenmodell** projiziert die Aktivitäten der Aktivitätenmatrix auf einen Zeitstrahl. So werden die Zeitdauer sowie die zeitliche Abfolge der einzelnen Aktivitäten

³ Diese Abbildung stellt einen Ausschnitt des iPeM im Kontext der PGE gemäß Abbildung 2.2 dar.

erkennbar. Dabei existieren drei verschiedene Modellierungen: Der *Referenzprozess* ist eine allgemein gehaltene Vorlage für den Projektverlauf, die aus Erfahrungen vorhergehender Produktentwicklungen oder branchenüblichen Standards abgeleitet werden kann (Albers et al., 2012; Albers & Deigendesch, 2010). Darauf aufbauend wird im *SOLL-Prozess* der spezifische Soll-Prozess für das aktuelle Projekt definiert, der als Projektplan dient. Im realen Projektverlauf ergeben sich meist Abweichungen von der ursprünglichen Planung. Diese sollten während des Projekts aufgezeichnet werden. Der reale Projektverlauf wird im *IST-Prozess* abgebildet. Durch einen Abgleich zwischen den drei Prozessen zum Projektende können der Referenzprozess sowie die SOLL-Prozesse zukünftiger Produktentwicklungen optimiert werden (Albers et al. 2019b, Albers & Braun, 2011a; Braun, 2013; Meboldt, 2008).

Im Kontext der Produktgenerationsentwicklung⁴ wird die Grundstruktur des iPeM aus Abbildung 2.1 auf mehrere Ebenen projiziert, die die Entwicklung verschiedener Produktgenerationen und die damit verknüpften Prozesse im entwickelnden Unternehmen abbilden. Die so entstehende Erweiterung des iPeM gemäß Abbildung 2.2 besitzt für jede entwickelte Produktgeneration eine eigene *Produktgenerationsebene*. Dabei bezeichnet g_n die aktuell in Entwicklung befindliche Produktgeneration und g_{n+1} deren Nachfolgenergeneration. In Abbildung 2.2 nicht dargestellt ist die Vorgängergeneration g_{n-1} des aktuell entwickelten Produkts. Zudem gibt es eine *Validierungssystemebene* und eine *Produktionssystemebene*. Diese beinhalten die Entwicklungen technischer Einrichtungen zur Validierung (Prüfstände, Testverfahren, ...) und Produktion (Maschinen, Herstellungsprozesse, Montagelinien, ...) des eigentlichen Produkts, die als eigene Produktentwicklungsprozesse angesehen werden können. Diese stehen in Wechselwirkung mit dem eigentlich entwickelten Produkt und müssen deshalb bei dessen Entwicklung berücksichtigt werden. Dasselbe gilt für die Entstehung und Weiterentwicklung von Leitlinien der Unternehmensstrategie, die in der Strategieebene abgebildet sind.

Alle Ebenen des iPeM im Kontext der PGE besitzen dieselbe Struktur. Sie teilen sich dabei ein gemeinsames Ressourcensystem, das die Entwicklungsressourcen des Unternehmens beinhaltet. Auch die Ziele der verschiedenen Unternehmensbereiche und –prozesse, die durch die verschiedenen Ebenen symbolisiert werden, sind in einem gemeinsamen Zielsystem zusammengefasst.

⁴ Siehe Abschnitt 2.1.1

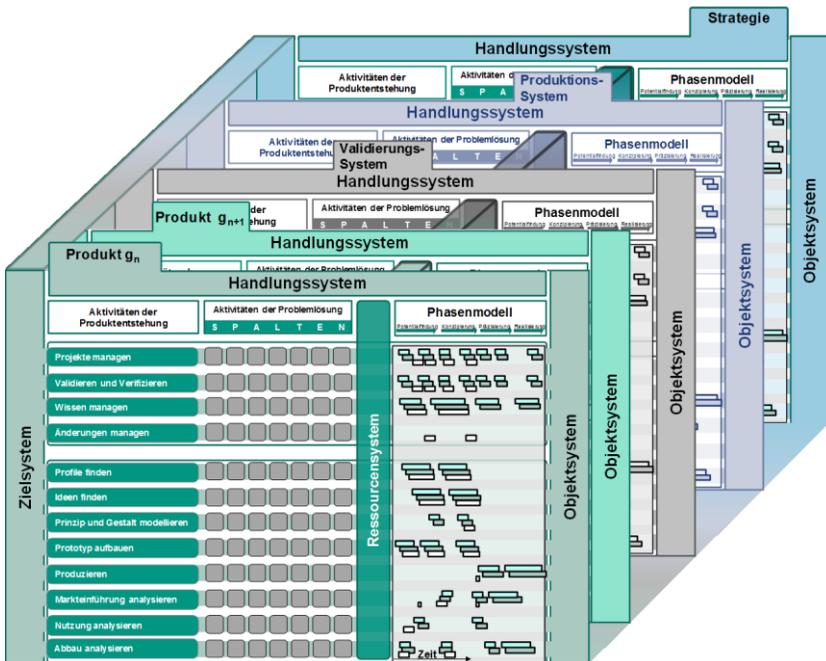


Abbildung 2.2: Das iPeM im Kontext der Produktgenerationsentwicklung (Albers et al., 2016d)

Dem gegenüber stehen für alle Ebenen separate Objektsysteme. Diese lassen sich nicht in ein gemeinsames Objektsystem zusammenfassen, da jede Ebene eigene Entwicklungsergebnisse besitzt, die sich von denen anderer Ebenen stark unterscheiden. Ebenso besitzt jede Ebene eine eigene Aktivitätenmatrix und ein eigenes Phasenmodell. Manche Elemente des iPeM existieren zwar separat für jede Ebene, sind aber dennoch zwischen den Ebenen verknüpft. So kann beispielsweise ein Prüfstand als Objekt aus dem Objektsystem der Validierungssystemebene zu einem Element des Handlungssystems der Produktebene der nächsten Produktgeneration werden (Albers et al., 2016d, S. 5).

2.1.3 Problemlösungsprozess SPALTEN nach Albers

Der Problemlösungsprozess SPALTEN nach Albers (Albers et al., 2005; Albers & Meboldt, 2006, siehe Abbildung 2.3) kann zur Lösung von Problemen verschiedenster Bereiche und Komplexitäten eingesetzt werden. Das Einsatzgebiet ist nicht auf die Produktentwicklung beschränkt (Saak, 2007, S. 34 f.).

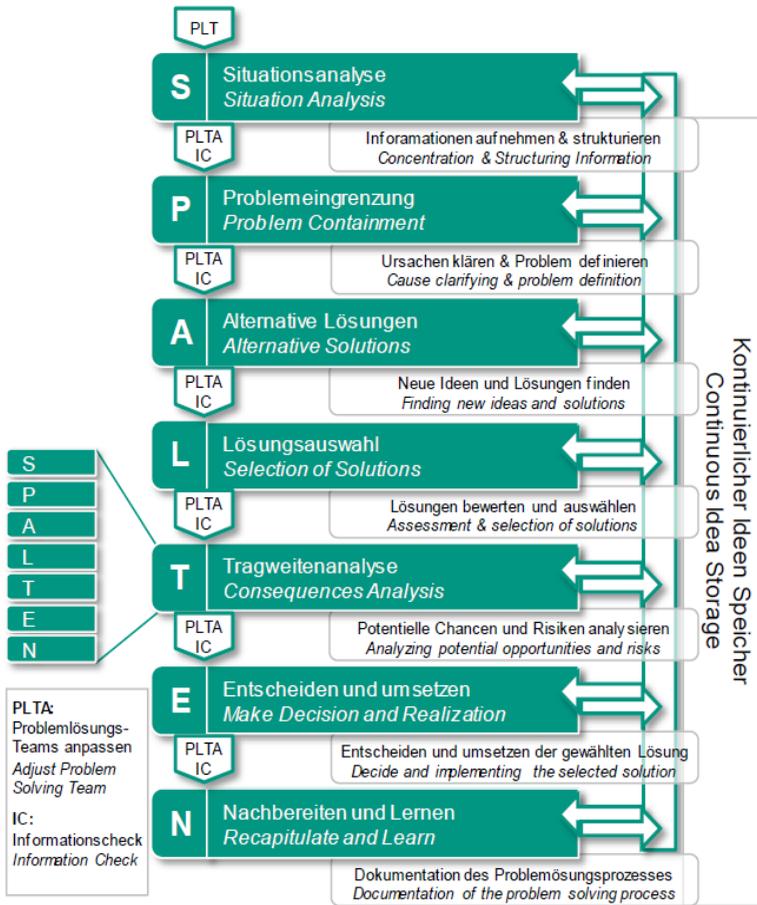


Abbildung 2.3: Problemlösungsprozess SPALTEN (Albers et al., 2016c)

SPALTEN beschreibt als Akronym die sieben Schritte aus denen der Problemlösungsprozess besteht: die *Situationsanalyse*, die *Problemeingrenzung*, die *alternative Lösungssuche*, die *Lösungsauswahl*, die *Tragweitenanalyse*, das *Entscheiden und Umsetzen* und das *Nachbereiten und Lernen*.

Vor jedem Schritt wird überprüft, ob das Problemlösungsteam (PLT) den aktuellen Anforderungen entspricht. Falls nicht, muss es neu zusammengesetzt werden. Nach jedem Schritt wird in einem Informationscheck (IC) geprüft, ob die aktuelle Informationsbasis ausreicht und auch entsprechend genutzt wird. Im Kontinuierlichen Ideen Speicher (KIS) werden aufkommende Ideen festgehalten, die zum aktuellen Schritt des Problemlösungsprozesses nicht relevant sind. Dies hilft einerseits bei der Fokussierung auf den jeweils aktuellen Schritt und andererseits dabei, keine wertvollen Ansätze zu vergessen. SPALTEN ist ein fraktaler Prozess, da die Lösung jedes einzelnen Schritts durch einen eigenen SPALTEN-Prozess erfolgen kann. Im iPeM wird SPALTEN zur Überführung von Elementen des Zielsystems in Elemente des Objektsystems eingesetzt. Dazu kann jede Makro-Aktivität der Aktivitätenmatrix in die sieben Schritte des SPALTEN-Prozesses untergliedert werden.

Neben dem SPALTEN-Prozess gibt es auch andere weit verbreitete Problemlösungsprozesse. Dazu zählen der Problemlösungszyklus der Systemtechnik nach Daenzer und Huber (1997, S. 96), sowie das Münchner Vorgehensmodell nach Lindemann (2009, Kap. 3.2.7).

2.2 Simulationsmodelle in der Produktentwicklung

In modernen Produktentstehungsprozessen nehmen Simulationsmodelle eine immer wichtiger werdende Position ein. Sie können dabei sowohl Teil des Handlungssystems, als auch des Ziel- und Objektsystems sein. Ihre dortigen Anwendungen werden nachfolgend beschrieben.

2.2.1 Simulationsmodelle als Teil des Handlungssystems

Die Komplexität moderner Produkte nimmt vor allem durch den steigenden Anteil mechatronischer Komponenten stark zu. Zur Funktionserfüllung müssen Mechanik, Elektronik und Software jeder Komponente fehlerfrei interagieren. In komplexeren Produkten wie beispielsweise Kraffahrzeugen oder auch verfahrenstechnischen Produktionsanlagen⁵ werden zudem mehrere Komponenten zu Subsystemen und mehrere Subsysteme zu Gesamtsystemen zusammengesetzt. In solchen *Systems-*

⁵ Deren pneumatisch angetriebene Prozessarmaturen liefern das Praxisbeispiel dieser Arbeit (siehe Kapitel 5)

of-Systems⁶ muss neben der Funktionserfüllung jeder Komponente auch das fehlerfreie Zusammenspiel zwischen allen Komponenten und Subsystemen (ggf. in mehreren Ebenen) gewährleistet sein.

Um die daraus entstehende Komplexität handhabbar zu machen, müssen bereits früh in der Produktentwicklung auf Komponenten- und Subsystemebene Validierungsmaßnahmen gestartet werden. Laut Albers (2010a, Hypothese drei) ist die Validierung die zentrale Aktivität in der Produktentwicklung. Validierung überprüft, ob der wirkliche Bedarf des Kunden durch das entwickelte Produkt erfüllt wird. Da der Bedarf nur durch ein vollständiges Produkt erfüllt werden kann, muss Validierung stets am Gesamtsystem erfolgen, auch wenn nur ein Teilsystem davon im aktuellen Produktentwicklungsprozess entwickelt wird. (Albers et al., 2014a)

In frühen Phasen der Produktentwicklung⁷ sind allerdings meist noch nicht alle Teile eines Gesamtsystems verfügbar. Zudem ist eine Validierung am realen Gesamtsystem oft extrem aufwendig. In solchen Fällen ist es sinnvoll, Teile des Gesamtsystems durch Modelle zu ersetzen. So können in jeder Phase der Produktentwicklung effiziente Validierungsmaßnahmen durchgeführt werden. Der Einsatz von Modellen hat zudem den Vorteil, dass sich innere Zustände von Simulationsmodellen deutlich einfacher erfassen lassen, als die von realen technischen (Sub-) Systemen. Deshalb ist mit dieser Vorgehensweise das Systemverhalten für den Entwickler deutlich einfacher nachvollziehbar. Bei Abweichungen vom gewünschten Verhalten kann so schneller und treffsicherer nach Lösungsansätzen gesucht werden.

2.2.1.1 Modellbasierte Funktionsentwicklung

Beispielsweise bei der Entwicklung *eingebetteter Systeme* („embedded systems“) oder auch bei Anlagensteuerungen wird die Strecke (also das zu steuernde bzw. zu regelnde System) aus den oben genannten Gründen oft in einem *Streckenmodell* abgebildet. So bekommen die Entwickler eine Möglichkeit, die Funktion ihrer Algorithmen in jedem beliebigen Stadium direkt an ihrem Arbeitsplatz kontinuierlich zu validieren. Die Programmierung der Algorithmen erfolgt dabei oft nicht mehr direkt in einer auf der Zielplattform (also der Hardware des Steuergeräts) lauffähigen Programmiersprache. Stattdessen wird ein *Funktionsmodell* der Algorithmen aus vorgefertigten Funktionsbausteinen in einer Entwicklungsumgebung wie beispielsweise MATLAB Simulink® erstellt. Dies erlaubt eine sehr schnelle und zuverlässige Synthese der gewünschten Funktion. Anschließend kann das Funktionsmodell durch automatische Codegenerierung in die Programmiersprache der Zielplattform

⁶ Eine Definition liefert u.a. Maier (1998)

⁷ Für eine formale Definition siehe Albers et al. (2017), Abschnitt 4

übertragen werden. Wenn in einem Entwicklungsprozess eingebetteter Systeme sowohl die Algorithmen, als auch die zugehörige Strecke zunächst in Form von Modellen vorliegen, dann wird von *modellbasierter Funktionsentwicklung* gesprochen (Arenz & Potrykus, 2010).

Neben dem geringen Aufwand und der hohen Zuverlässigkeit bietet die modellbasierte Funktionsentwicklung durch die Möglichkeit der automatischen Codegenerierung eine hohe Flexibilität für Änderungen der Hardware der Zielplattform. Dies ist vor allem im Kontext der PGE – Modell der Produktgenerationsentwicklung nach Albers interessant: Wenn für eine neue Produktgeneration des zu entwickelnden Steuergeräts die Rechenleistung der bisherigen Hardware nicht mehr ausreicht, kann das Funktionsmodell mit geringem Aufwand auf andere Hardware portiert werden, auch wenn diese eine andere Programmiersprache erfordert. Der größte Vorteil der modellbasierten Funktionsentwicklung im Vergleich zur konventionellen Funktionsentwicklung sind jedoch die stark verbesserten Möglichkeiten zu kontinuierlichen Validierung.⁸

- Ein erster Validierungszyklus kann bei der modellbasierten Funktionsentwicklung rein simulativ durchgeführt werden, indem das Funktionsmodell mit dem Streckenmodell gekoppelt wird. Dies wird *Model-in-the-Loop (MiL)* genannt.
- Wird das Funktionsmodell stattdessen über geeignete Hardware⁹ mit einer realen Strecke gekoppelt, wird von *Rapid-Control-Prototyping (RCP)* gesprochen.
- Wenn das Funktionsmodell z.B. durch automatische Codegenerierung in eine Programmiersprache überführt worden ist, die auf der gewünschten Zielplattform lauffähig ist, kann anschließend auch der so erzeugte Code mit dem Streckenmodell gekoppelt werden. Dies wird als *Software-in-the-Loop (SiL)* bezeichnet.
- Nach Aufspielen des Codes auf die Zielplattform kann das vollständige Steuergerät mit dem Streckenmodell gekoppelt werden, wodurch ein *Hardware-in-the-Loop (HiL)* Test entsteht. Für diesen wird, ähnlich wie beim RCP, geeignete Hardware benötigt, um virtuelle und reale Systembestandteile koppeln zu können.
- Im letzten Validierungsschritt kann abschließend das reale Steuergerät mit der realen Strecke in einem *Systemtest* untersucht werden.

⁸ Die Definitionen der nachfolgenden Liste entstammen sinngemäß (Isermann, 2014, Kap. 6.11)

⁹ beispielsweise Systeme von dSpace oder National Instruments

Bei der konventionellen Steuergeräteentwicklung ist, im Gegensatz zur modellbasierten Funktionsentwicklung, die erste und einzige Validierungsmöglichkeit der Systemtest. Zu diesem Zeitpunkt ist allerdings das validierte System bereits so komplex, dass eine Lokalisierung von Fehlerursachen sehr schwierig wird. Als Fehlerquellen müssen die Soft- und Hardware des Steuergeräts genau wie Probleme mit der Strecke in Betracht gezogen werden. Zudem ist die Behebung von Fehlern in einem so späten Stadium auch sehr aufwendig. Aufgrund der viel früheren und umfangreicheren Validierungsmöglichkeiten erlaubt die modellbasierte Funktionsentwicklung eine schnellere und zuverlässige Entwicklung von eingebetteten Systemen.

2.2.1.2 Der IPEK-X-in-the-Loop-(IPEK-XiL)-Ansatz nach Albers

Das *IPEK-X-in-the-Loop-(IPEK-XiL)-Ansatz nach Albers* überträgt den Ansatz der modellbasierten Funktionsentwicklung mit den damit verknüpften Begriffen MiL, SiL, HiL und RCP von der Entwicklung einzelner Steuergeräte auf die Entwicklung beliebiger und beliebig komplexer technischer Systeme.

Das zu entwickelnde (Teil-)System wird im IPEK-XiL als *System-in-Development* bezeichnet. Im Gegensatz zur modellbasierten Funktionsentwicklung kann dies im IPEK-XiL durchaus auch ein rein mechanisches (Teil-)System sein. Die mit dem System-in-Development verbundenen (Teil-)Systeme werden im IPEK-XiL als *Connected Systems* bezeichnet. Im IPEK-XiL können beliebige Elemente des System-in-Development physisch und beliebige andere Elemente virtuell oder in Mischformen vorliegen. Dasselbe gilt für die Elemente der Connected Systems. So kann beispielsweise für die Entwicklung eines E-Motors mit angeschlossenem Getriebe, welcher Teil eines Gesamtfahrzeugs werden soll, der zur Validierung eingesetzte IPEK-XiL-Ansatz wie in Abbildung 2.4 dargestellt ausgebildet sein.

Nicht bei jeder Validierungsaktivität steht ausschließlich das im aktuellen Produktentwicklungsprozess entwickelte System-in-Development im Fokus. Deshalb wird im IPEK-XiL für das untersuchte (Teil-)System der weiter gefasste Begriff *System-under-Investigation* eingeführt. Das System-under-Investigation kann dabei, wie in Abbildung 2.5 dargestellt, unterschiedliche Anteile des Gesamtsystems beinhalten. Das Spektrum reicht dabei vom vollständigen Gesamtsystem (System-in-the-Loop) über Teilsysteme (Subsystem-in-the-Loop) bis hin zu funktionserfüllenden Komponenten auf Wirkflächenpaarebene (WSP-in-the-Loop).

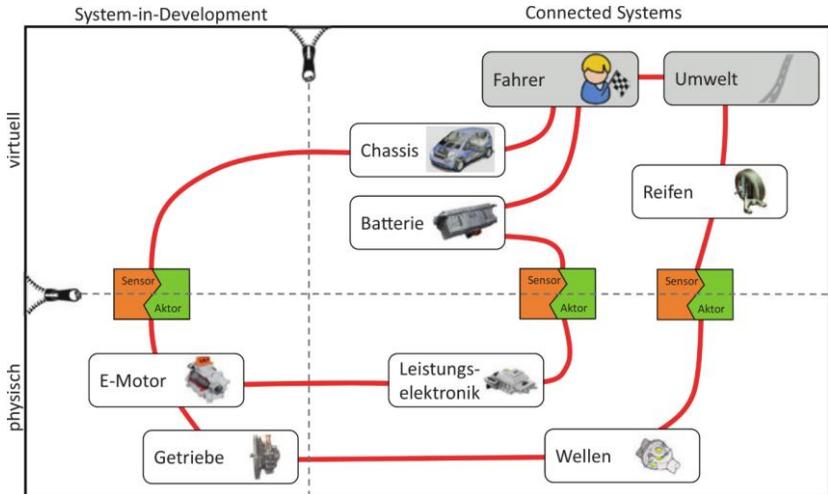


Abbildung 2.4: Dimensionen im IPEK-XiL-Ansatz nach Albers (Albers et al., 2016a)

Da Validierung stets am Gesamtsystem erfolgen muss¹⁰, werden die nicht im System-under-Investigation enthaltenen Komponenten des Gesamtsystems in einem Rest-System-Modell abgebildet. Verbundene Systeme (Connected Systems), die in der realen Nutzung in Wechselwirkung mit dem Gesamtsystem stehen, werden im IPEK XiL-Framework ebenfalls mit modelliert. Diese schließen auch den Anwender (Fahrer in Abbildung 2.5) sowie die Umwelt des Systems mit ein. Für eine zuverlässige Validierung ist es zudem unerlässlich, dass praxisnahe Testfälle generiert werden, mit denen das Systemverhalten des System-under-Investigation untersucht werden kann. Die verschiedenen Testfälle können dabei das Verhalten aller im IPEK XiL-Framework enthaltenen Systeme beeinflussen.

¹⁰ vgl. S. 14

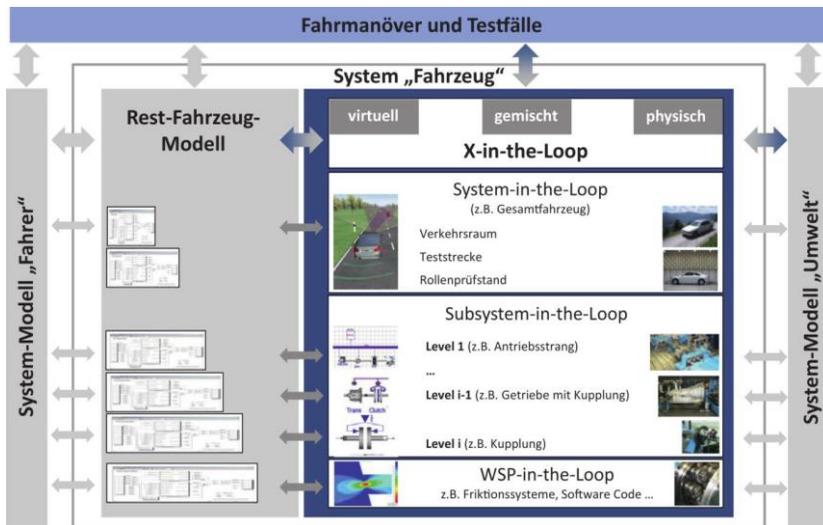


Abbildung 2.5: Das IPEK XiL-Framekwork am Beispiel der Fahrzeugentwicklung (Albers et al., 2016a)

2.2.2 Simulationsmodelle als Teil des Ziel- und Objektsystems

Neben der vorhergehend beschriebenen Nutzung von Modellen als Teil des Handlungssystems einer Produktentwicklung kommt es auch immer öfter vor, dass Modelle Teil des Ziel- und Objektsystems eines entwickelten Produkts werden. Hierbei lassen sich zwei Fälle unterscheiden:

Einerseits können die Modelle in das entwickelte Produkt selbst integriert werden. Ein solcher Anwendungsfall ist z.B. bei Produkten gegeben, die modellprädiktive Regelungen (Model Predictive Control) beinhalten. Bei diesen dient ein Modell des zu regelnden Prozesses dazu, das zukünftige Verhalten vorherzusagen und so den optimalen Verlauf der Stellgröße berechnen zu können, um ein gewünschtes Systemverhalten zu erzielen. Ein zweiter möglicher Anwendungsfall sind Systeme, deren Zustand per Condition Monitoring überwacht werden soll, um Ausfallrisiken vorab und im laufenden Betrieb erkennen zu können. Hierbei untersucht ein Algorithmus, ob die aktuell erfassten Sensordaten auf aufkommende Probleme schließen lassen. Eine Möglichkeit, solche Algorithmen zu erstellen, besteht darin, das normale Systemverhalten unter den gegebenen Betriebsbedingungen von einem

Modell berechnen zu lassen und Abweichungen von diesem Verhalten zu detektieren.

Andererseits gibt es für immer mehr Produkte die Anforderung, dass neben dem Produkt selbst auch Verhaltensmodelle an den Kunden mitgeliefert werden. Dies gilt unter anderem für Komponenten und Subsysteme der Automatisierungstechnik. In diesem Umfeld wird mit Hilfe der Verhaltensmodelle der Komponenten und Subsysteme ein digitales Abbild neu aufzubauender Produktionsanlagen erstellt. Anhand dieses digitalen Abbilds kann dann eine virtuelle Inbetriebnahme¹¹ erfolgen. Dabei wird die Programmierung der Steuerungstechnik der Anlage anhand des virtuellen Abbilds der Anlage getestet. Der Hersteller der Produktionsanlagen nutzt die Komponentenmodelle also, wie in Abschnitt 2.2.1 beschrieben, zur Validierung der Software seiner Steuerungstechnik als Teil seines Handlungssystems. Für den Komponentenhersteller sind die Verhaltensmodelle hingegen eine Forderung im Zielsystem und damit später auch Teil des Objektsystems seiner zu entwickelnden Produkte.

2.3 Modellbildung

Im vorhergehenden Abschnitt wurde dargestellt, an welchen Stellen und zu welchen Zwecken Simulationsmodelle in der Produktentwicklung eingesetzt werden. Auf eine Definition des Begriffs des *Modells* wurde bisher verzichtet, da jeder Leser bereits ein intuitives Verständnis dieses Begriffs besitzt. Im nun folgenden Abschnitt wird eine formale Definition des Modellbegriffs nachgereicht. Aus dieser Definition ergeben sich verschiedene Möglichkeiten zur Erstellung von Modellen. Diese werden auf die Definition folgend aufgezeigt und erläutert.

Nach Zingel (2013) ist ein Modell „[...] *eine abstrahierte Beschreibung der Realität*“.

Stachowiak (1973, S. 131 ff.) definiert Modelle anhand von drei Hauptmerkmalen:

- Das *Abbildungsmerkmal* besagt, dass Modelle stets Repräsentationen eines Originals sind.
- Gemäß dem *Verkürzungsmerkmal* erfassen Modelle nicht alle Attribute des Originals, sondern nur die, die den Erstellern und Nutzern des Modells als relevant erscheinen.
- Das *pragmatische Merkmal* definiert, dass Modelle stets einem bestimmten Zweck dienen.

¹¹ Siehe (Breitbach, 2015; Strahilov, 2015)

Für die Modellbildung folgert Zeller (2012, S.11) deshalb treffend:

„Die Kunst der Modellbildung besteht darin, die betrachtete Realität soweit aufgrund der Problemstellung zulässig zu vereinfachen, ohne die interessierenden Sachverhalte unzulässig zu verfälschen.“

Der Ersteller des Modells hat also die Aufgabe, die Eigenschaften des Originals (Abbildungsmerkmal) gemäß dem Anwendungszweck des Modells (pragmatisches Merkmal) in wichtige und unwichtige Eigenschaften zu unterteilen. Ausschließlich die für den Anwendungszweck wichtigen Eigenschaften sollte er im Modell abbilden (Verkürzungsmerkmal). Damit er das Original in für das Modell wichtige und unwichtige Anteile unterteilen kann, muss er das Original zunächst in seine Elemente zerlegen. Eine Methode, die ihn dabei unterstützen kann, ist der Contact and Channel-Ansatz, der in Abschnitt 2.3.1 vorgestellt wird.

Um das Verhalten der so ermittelten Elemente abzubilden, stehen dem Modellersteller zwei verschiedene Ansätze zur Verfügung. Einerseits kann er versuchen, die inneren Zusammenhänge des Originals zu verstehen und die relevanten Zusammenhänge auf Basis physikalischer Gesetzmäßigkeiten in Gleichungen zu fassen. Dies führt zu White-Box-Modellen, die in Abschnitt 2.3.2 beschrieben werden. Andererseits kann er auch das Original durch verschiedene Stimuli anregen, dessen Reaktionen beobachten und dann den relevanten Teil durch approximierende Gleichungen beschreiben. Dies führt zu Black-Box-Modellen, die in Abschnitt 2.3.3 erläutert werden. Die Kombination von White-Box- und Black-Box-Anteilen in einem Modell führt zu Grey-Box-Modellen. Diese werden in Abschnitt 2.3.4 vorgestellt.

2.3.1 Contact and Channel-Ansatz

Der Contact and Channel-Ansatz C&C²-A dient der Modellierung von Zusammenhängen zwischen Gestalt und Funktion eines technischen Produkts. Im C&C²-A existieren drei Grundelemente:

- *Connectoren (C)* verbinden das betrachtete (Teil-)System mit seiner Umwelt. In Abbildung 2.6 werden Teile eines Zahnradpaares betrachtet. Die Connectoren C1 und C2 verbinden das modellierte Teilsystem mit den Wellen der Zahnräder, welche in diesem Beispiel außerhalb des Betrachtungsraumes liegen.
- Gemäß der Grundhypothese 1 des C&C²-A braucht Funktion Wechselwirkung zwischen verschiedenen Systemelementen. Diese entsteht in *Wirkflächenpaaren (WFP)*. Im Beispiel aus Abbildung 2.6 sind dies die Zahnflanken eines Zahnradpaares. Zur Funktionserfüllung werden stets

mindestens zwei Wirkflächenpaare benötigt. Der Teil des Wirkflächenpaares, der aktuell in Kontakt steht, wird als *Wirkkontakt (WK)* bezeichnet.

- Zwischen den Wirkflächenpaaren und Connectoren des betrachteten (Teil-)Systems muss zur Funktionserfüllung Energie, Materie oder Information übertragen werden. Die Strukturelemente entlang derer dies geschieht werden als *Leitstützstrukturen (LSS)* bezeichnet. In Abbildung 2.6 sind dies LSS1 und LSS2, entlang welcher der Kraftfluss zwischen WFP1 und den Connectoren C2 und C1 verläuft. Allgemein können durch LSS neben Kräften auch Energie, Information oder Stoffe geleitet werden.

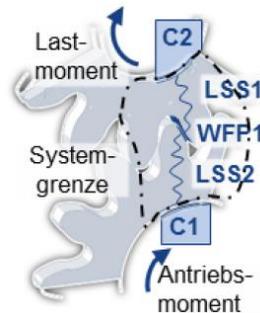


Abbildung 2.6: C&C²-A am Beispiel eines Zahnradpaares (Matthiesen et al., 2018)

Neben diesen Grundelementen existieren im C&C²-A auch für die Funktionserfüllung nicht entscheidende Nebenelemente und Strukturelemente, die die Modellbestandteile in Beziehung zueinander setzen. Einen detaillierteren Einblick geben beispielsweise Matthiesen et al. (2018) oder Freudenmann (2014).

2.3.2 White-Box-Modelle

White-Box-Modelle beschreiben das Verhalten eines Systems basierend auf physikalischen Grundgleichungen. Zur Erstellung eines White-Box-Modells müssen die inneren Zusammenhänge des abzubildenden Systems vom Modellersteller analysiert, verstanden und in mathematischen Gleichungen abgebildet werden.

Nach Meywerk (2007, S. 17) stellt die Erstellung von White-Box-Modellen eine sehr anspruchsvolle Aufgabe dar, da zu ihrer Erledigung „das Wissen über eine große Vielzahl physikalischer Gesetzmäßigkeiten und deren Umsetzung in numerisch ein-

fach zu lösende Gleichungen“ benötigt wird. Dieses Wissen wird für die am häufigsten in technischen Systemen vorkommenden Gesetzmäßigkeiten im Rahmen des Ingenieursstudiums vermittelt. Es umfasst verschiedene Fachgebiete, von der Festkörpermechanik über die Thermodynamik und Elektrotechnik bis hin zur Strömungsmechanik. Obwohl die Erstellung von White-Box-Modellen eine sehr komplexe Aufgabe ist, kann deshalb davon ausgegangen werden, dass der Leser dieser Arbeit das Wissen und die Fähigkeiten besitzt, um diese Aufgabe prinzipiell lösen zu können. Auf weitere Ausführungen dazu wird deshalb an dieser Stelle verzichtet. Eine methodische Unterstützung zur Vorgehensweise wird beispielsweise von Panreck (2002) und Nollau (2009) gegeben.

2.3.3 Black-Box-Modelle

Wenn die Erstellung von White-Box-Modellen zu aufwendig oder unmöglich ist, oder zu Modellen zu hoher Komplexität führen würde, kann das Verhalten des abzubildenden Systems auch experimentell ermittelt werden. Dies führt zu Black-Box-Modellen. Bei diesen wird die Reaktion des Systemverhaltens auf definierte Anregungen ermittelt. Solche Modelle beschreiben also einen empirisch ermittelten Zusammenhang zwischen Ein- und Ausgangssignalen eines Systems, der ohne Rückschlüsse auf die inneren Zusammenhänge des betreffenden Systems auskommt.

Eine häufig in der Regelungstechnik eingesetzte Vorgehensweise zur Erstellung von Black-Box-Modellen dynamischer Systeme ist die *Systemidentifikation*. Dabei wird das System mit einem breitbandigen, meist klar definierten Eingangssignal angeregt. Die Systemantwort, also das auf das Eingangssignal folgende Ausgangssignal des Systems, wird aufgezeichnet. Aus den Zusammenhängen zwischen Anregung und Systemantwort wird anschließend mittels geeigneter Identifikationsverfahren auf eine Übertragungsfunktion geschlossen, die als Black-Box-Modell des untersuchten Systems dient.¹² Mit dieser Vorgehensweise wird also das dynamische (zeitabhängige) Verhalten eines Systems abgebildet. Einflüsse weiterer Faktoren werden hierbei üblicherweise nicht berücksichtigt.

Wie die Einflüsse mehrerer Faktoren auf das stationäre Verhalten eines untersuchten Systems möglichst effizient abgebildet werden können, ist Themengebiet der statistischen Versuchsplanung. Ein Überblick über die darin zusammengefassten Methoden zum Planen und Auswerten von Versuchsreihen zum Zweck der Modellbildung wird in den Abschnitten 2.4 bis 2.6 gegeben.

¹² Für Details siehe z.B. (Bohn & Unbehauen, 2016)

2.3.4 Grey-Box-Modelle

Wenn ein Modell Black-Box-Anteile mit White-Box-Anteilen kombiniert, wird von Grey-Box-Modellen gesprochen. Zum Aufbau von Grey-Box-Modellen existieren mehrere mögliche Vorgehensweisen. Die nachfolgend dargestellte Einteilung basiert auf der Arbeit von Sohlberg und Jacobsen (2008).

- Die *Black-Box-Identifikation mit Randbedingungen* stellt eine Gruppe von Grey-Box-Modellen dar, bei denen der Black-Box-Anteil den White-Box-Anteil deutlich überwiegt. Der White-Box-Anteil besteht lediglich darin, dass physikalisches Grundlagewissen dazu genutzt wird, Randbedingungen für Parameter des Black-Box-Modells zu ermitteln. Beispiele für solche Modelle liefern beispielsweise Tulleken (1991) oder Murakami und Seborg (2000).
- Auch bei *semi-physikalischer Modellbildung* stellt der Black-Box-Anteil den größten Anteil am Gesamtmodell. White-Box-Modelle werden hier dazu genutzt, gezielt Faktortransformationen (vgl. S.49) zur Linearisierung der Zusammenhänge im empirischen Teilmodell durchzuführen. Beispiele hierfür geben beispielsweise Kroll (2000) oder Rogers et al. (2017).
- Bei *Hybrid-Modellbildung* beinhaltet das Gesamtmodell separate White- und Black-Box-Teilmodelle. Sohlberg und Jacobsen unterscheiden hierbei zwischen einem seriellen Ansatz und einem parallelen Ansatz. Beim seriellen Ansatz wird die Ausgangsgröße des Black-Box-Modells als zusätzliche Eingangsgröße des White-Box-Modells genutzt. Das White-Box-Modell liefert dann die Ausgangsgröße des Gesamtmodells. Beim parallelen Ansatz gibt es keine Interaktion zwischen beiden Teilmodellen. Die Ausgangsgröße des Gesamtmodells wird durch Addition der Ausgangsgrößen beider Teilmodelle gebildet. Beide Ansätze sind zur Verdeutlichung in Abbildung 2.7 dargestellt.

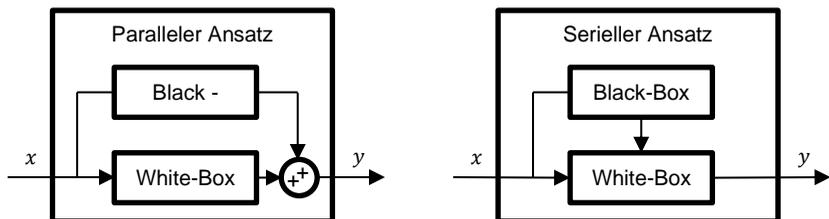


Abbildung 2.7: Paralleler und serieller Ansatz hybrider Grey-Box-Modelle nach Sohlberg & Jacobsen (2008)

2.4 Grundlagen der statistischen Versuchsplanung

In einer Versuchsreihe wird das Verhalten eines Systems in diskreten Zuständen experimentell untersucht. Die statistische Versuchsplanung ist eine Methode zur effizienten Planung und Auswertung von Versuchsreihen (Siebertz et al., 2010, S. 1). Das Ziel der statistischen Versuchsplanung ist es also, eine Versuchsreihe so zu gestalten und auszuwerten, dass mit möglichst geringem Aufwand möglichst valide, umfassende und genaue Ergebnisse erzielt werden. Das für das Verständnis dieser Methode benötigte Fachvokabular wird in Abschnitt 2.4.1 definiert.

Zur effizienten Gestaltung und Auswertung von Versuchsreihen müssen zwei grundlegende Schwierigkeiten von experimentellen Untersuchungen gelöst werden. Zum einen ist jeder Versuch und jede Messung unvermeidbaren Fehlereinflüssen unterworfen. Damit diese Fehlereinflüsse die Validität und Genauigkeit der Versuchsergebnisse möglichst wenig beeinträchtigen, müssen bereits bei der Versuchsplanung einige Grundprinzipien eingehalten werden. Diese werden in Abschnitt 2.4.2 beschrieben.

Die andere Schwierigkeit entsteht, wenn ein System durch mehrere Faktoren¹³ beeinflusst wird. Hier ist die Reaktion auf die Änderung eines Parameters oft abhängig von der Einstellung mindestens eines weiteren Parameters. Dies hat zur Folge, dass die Parameter nicht unabhängig voneinander untersucht werden können, wenn man eine umfassende Beschreibung des untersuchten Systems erhalten möchte. Es gibt also bei einer Anzahl von n Parametern einen n -dimensionalen Parameterraum, der untersucht werden muss. Dieser Parameterraum wird Versuchsraum genannt. Wenn nur ein diskreter Zustand pro Versuch untersucht werden kann, stellt sich die Frage, wie man die Versuchspunkte im Versuchsraum anordnet, um mit möglichst wenigen Versuchen eine möglichst umfassende und genaue Beschreibung des Systems im interessierenden Bereich zu erhalten. Die Antwort auf diese Frage hängt von der Art des Modells ab, das an die Versuchsdaten angepasst werden soll.

In der klassischen statistischen Versuchsplanung, die in Abschnitt 2.5 beschrieben wird, werden als Regressionsmodelle mehrdimensionale Polynome eingesetzt. Die dabei eingesetzten Versuchspläne sind so gestaltet, dass sich mehrdimensionale Polynome möglichst effizient an die Versuchsdaten anpassen lassen. Diese Art von Versuchsplanung und –auswertung ist ein Teilgebiet der Statistik. Sie wurde und

¹³ Weiter oben wurden diese noch mit dem Begriff „Eingangsgröße“ aus der Regelungstechnik bezeichnet. Im Bereich der statistischen Versuchsplanung ist dagegen die Bezeichnung „Faktor“ üblich. (Siehe Abschnitt 2.4.1.)

wird dazu entwickelt, reale Experimente in der Industrie und verschiedenen Wissenschaftszweigen möglichst effizient und valide zu gestalten. Anwendungen reichen von den Werkstoffwissenschaften über die Medizin bis hin zur Soziologie. Da Störgrößen und ihre Effekte in solchen Anwendungen stets präsent sind, kommt ihrer Erfassung und Bewertung eine hohe Bedeutung zu. Die in Abschnitt 2.5 beschriebenen Versuchspläne und Auswertungsmethoden sind darauf ausgerichtet, dies zu ermöglichen. Die größte Schwäche solcher Versuchspläne und der mit ihnen erstellten polynomialen Modelle ist ihr begrenztes Auflösungsvermögen. Üblicherweise werden Polynome maximal zweiten Grades zur Abbildung des untersuchten Systemverhaltens verwendet. Damit lassen sich weder sprunghafte Änderungen noch lokale Details des Systemverhaltens erfassen. Aus diesem Grund werden diese Methoden im Ingenieursumfeld bisher meist nicht zur Modellbildung, sondern als reines Optimierungstool verwendet¹⁴, beispielsweise für Produktionsprozesse, bei denen die optimale Kombination mehrerer Fertigungsparameter gefunden werden muss.

Neben der oben beschriebenen klassischen statistischen Versuchsplanung gibt es Versuchspläne sowie Regressionsmodelle und –verfahren, die eine deutlich höhere Auflösung des untersuchten Systemverhaltens ermöglichen. Diese werden in Abschnitt 2.6 beschrieben.

2.4.1 Begriffsdefinitionen

Ein sehr frühes Buch über die statistische Versuchsplanung schrieb Fisher (1935). Darin prägte er einen Großteil des Vokabulars, das auch heute noch in der statistischen Versuchsplanung gebräuchlich und auch für das Verständnis der vorliegenden Arbeit notwendig ist. Die nachfolgende Liste ist ergänzt um einige grundlegende Begriffe aus der mathematischen Statistik. Die darin verwendeten Definitionen sind sinngemäß aus Kleppmann (2011), Siebertz et al. (2010) und Czado & Schmidt (2011) entnommen. Auf die Einführung von Begriffen und Konzepten, die zum Verständnis dieser Arbeit nicht unbedingt notwendig sind, wird bewusst verzichtet.

- Die *Grundgesamtheit* beschreibt die gesamte untersuchte Einheit, deren Eigenschaften in einem bestimmten Umfang abgebildet werden sollen.
- Da die Grundgesamtheit meist sehr groß oder ein Kontinuum mit unendlich vielen Zuständen ist, kann zu ihrer Untersuchung im Normalfall nur eine ausgewählte Teilmenge, die *Stichprobe*, untersucht werden.
- Der *Versuchsumfang* ist die Gesamtzahl der Versuche in einem Versuchsplan.

¹⁴ Siehe z.B. (English et al., 2012; Jensen & Kowalski, 2012; Potcner & Kowalski, 2004; Vining, 2012; Vining et al., 2005)

- Oft ist das Verhalten der Grundgesamtheit nur in einem begrenzten Bereich von Interesse. Ebenso können Faktoren in vielen Fällen aus technischen Gründen nur in begrenzten Bereichen untersucht werden. Der Parameterraum, der durch die zu untersuchenden Faktoren in ihren zu untersuchenden Bereichen aufgespannt wird, wird *Versuchsraum* oder *Faktorraum* genannt.
- Eingangsgrößen des zu untersuchenden Systems werden *Parameter* oder *Einflussgrößen* genannt. In dieser Arbeit wird ausschließlich der Begriff „Einflussgrößen“ verwendet um eine klare Abgrenzung zu Parametern einer mathematischen Gleichung zu ermöglichen.
- Ausgangsgrößen des zu untersuchenden Systems, die vom Modell in Abhängigkeit der Eingangsgrößen berechnet werden sollen, werden als *Qualitätsmerkmale* oder *Zielgrößen* bezeichnet.
- Einflussgrößen lassen sich aufteilen in *Steuergrößen* (einstellbare Einflussgrößen) und *Störgrößen*.
- Die Steuergrößen, deren Effekte auf die Zielgröße im Versuchsplan untersucht werden sollen, heißen *Faktoren*.
- Faktoren können *quantitativ* (z.B. Temperatur oder Kraft) oder *qualitativ* (z.B. ja/nein oder rot/gelb/grün) sein.
- Die in einer Versuchsreihe untersuchten Werte eines Faktors heißen *Stufen* oder *Level*.
- Ein *Effekt* ist die Änderung der Zielgröße, die durch die Änderung der Stufe eines Faktors verursacht wird.
- Wenn der Effekt eines Faktors von der Einstellung eines anderen Faktors abhängt, nennt man dies eine *Wechselwirkung* bzw. einen *Wechselwirkungseffekt*. Zur Veranschaulichung, welche Auswirkung eine Wechselwirkung auf das Modellverhalten besitzt, ist in Abbildung 2.8 der Effekt einer Wechselwirkung zwischen zwei Faktoren x_1 und x_2 dargestellt. Mathematisch wird eine solche Wechselwirkung durch den Term $\beta_{12}x_1x_2$ beschrieben. Der Koeffizient β_{12} besitzt für die dargestellte Abhängigkeit einen positiven Wert. Entlang der Koordinatenachsen ist keine Änderung der Zielgröße zu erkennen. Mit steigenden Werten des Faktors x_1 ergibt sich hingegen eine steigende, positive und immer linear bleibende Abhängigkeit zwischen der Zielgröße y und dem Faktor x_2 . Ebenso ergibt sich mit steigenden Werten des Faktors x_2 eine größer werdende, ebenfalls positive sowie linear bleibende Abhängigkeit zwischen der Zielgröße und dem Faktor x_1 . Eine Wechselwirkung zwischen zwei Faktoren x_1 und x_2 beschreibt also immer sowohl die Änderung des Zusammenhangs zwischen x_1 und y in Abhängigkeit von x_2 , als auch umgekehrt die Änderung des Zusammenhangs zwischen x_2 und y in Abhängigkeit von x_1 .

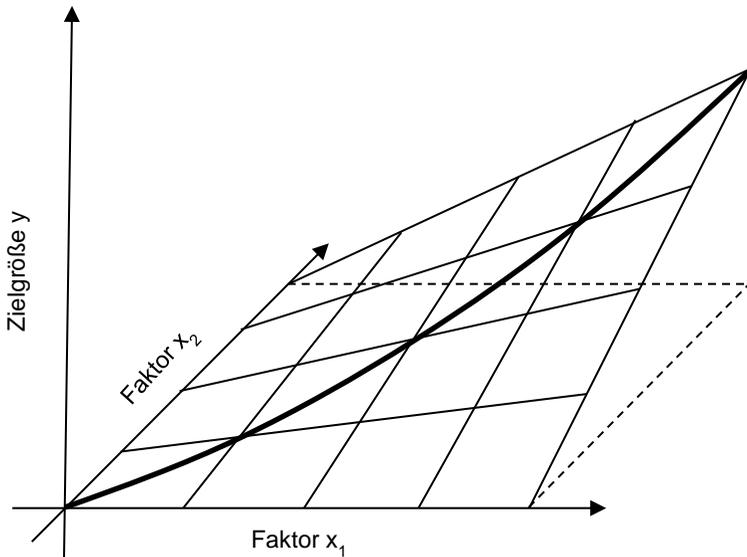


Abbildung 2.8: Darstellung einer reinen Wechselwirkung

Eine Dreifachwechselwirkung $\beta_{123}x_1x_2x_3$ der Faktoren x_1 , x_2 und x_3 kann analog dazu als Änderung der Zweifachwechselwirkung $\beta_{12}x_1x_2$ in Richtung des Faktors x_3 angesehen werden. Diese ist wiederum mit der Änderung der Wechselwirkung $\beta_{23}x_2x_3$ in Richtung des Faktors x_1 sowie der Änderung der Wechselwirkung $\beta_{13}x_1x_3$ in Richtung des Faktors x_2 identisch. Obwohl die Fläche in Abbildung 2.8 parallel zu den Koordinatenachsen stets linear ist, ist sie in Richtung der Diagonalen parabelförmig gekrümmt. Dies kommt daher, dass in den Wechselwirkungsterm das Produkt aus x_1 und x_2 eingeht. Entlang der eingezeichneten Diagonalen gilt $x_1 = x_2$. Deshalb wächst der Term $\beta_{12}x_1x_2$ in dieser Richtung quadratisch an. Wechselwirkungen kann man sich also auch als das Verhalten der Zielgröße in Richtung der Diagonalen zwischen den zugehörigen Koordinatenachsen vorstellen. Die Form dieser Abhängigkeit besitzt denselben Grad wie die Wechselwirkung. Sie ist also beispielsweise für eine Zweifachwechselwirkung quadratisch und für eine Dreifachwechselwirkung kubisch.

- Der *Erwartungswert* einer Größe ist der reale Wert oder Mittelwert dieser Größe in der Grundgesamtheit.
- Eine *Schätzfunktion* (*Schätzer*) dient dazu, aus den Daten einer Stichprobe einen *Schätzwert* für einen gesuchten Erwartungswert zu ermitteln.

- Der *Freiheitsgrad* zur Berechnung eines Schätzwerts ist die Anzahl der unabhängigen Beobachtungswerte, die zur Berechnung des Schätzwerts zur Verfügung stehen.
- Die *Varianz* ist die mittlere quadratische Abweichung einer Größe von ihrem Mittelwert.
- Die *Standardabweichung* ist die Quadratwurzel aus der Varianz.
- Die *Kovarianz* beschreibt, wie stark sich zwei Merkmale gemeinsam ändern. Sie ist positiv, wenn sich die Merkmale gleichläufig zusammen ändern, und negativ, wenn sie sich gegenläufig ändern. Eine Kovarianz von null bedeutet, dass sich kein Zusammenhang zwischen den Merkmalen erkennen lässt.
- Um die Kovarianz unabhängig von der Varianz der einzelnen Merkmale darzustellen, wird sie durch die Wurzel des Produkts der Varianzen der beiden Merkmale geteilt. So ergibt sich der *Korrelationskoeffizient*. Er beschreibt maßstabsunabhängig, wie stark zwei Merkmale miteinander korrelieren. Er kann Werte zwischen plus eins und minus eins annehmen. Ein Wert von eins bedeutet einen positiven linearen Zusammenhang, null bedeutet keinen Zusammenhang und minus eins bedeutet einen negativen linearen Zusammenhang.
- Der *Vertrauensbereich* ist ein Wertebereich, innerhalb dessen sich der Erwartungswert einer untersuchten Größe mit einer bestimmten Wahrscheinlichkeit (*Vertrauensniveau*) befindet. Durch einen Vertrauensbereich wird ein nicht exakt bestimmbarer Erwartungswert eingegrenzt. Ein möglichst kleiner bzw. enger Vertrauensbereich ist deshalb wünschenswert. Je schlechter die Datenbasis für die Errechnung des Schätzwerts für die gesuchte Größe ist (wenige Datensätze oder große Streuung), desto größer wird der Vertrauensbereich. Wenn ein hohes Vertrauensniveau gewünscht wird, bedeutet dies, dass der Erwartungswert mit hoher Wahrscheinlichkeit im Vertrauensbereich enthalten sein muss. Deshalb vergrößert ein hohes Vertrauensniveau den Vertrauensbereich.
- Wenn ein Polynom auf die Ergebnisse einer Versuchsreihe angepasst wird, dann muss dessen Grad geeignet gewählt werden. Existieren in den Versuchsergebnissen starke Nichtlinearitäten und der Grad des Polynoms wird zu niedrig gewählt, dann kann das Polynom den Nichtlinearitäten nicht ausreichend folgen. In einem solchen Fall wird von *Lack-of-fit* gesprochen. Ein Beispiel hierfür ist in Abbildung 2.9 dargestellt.

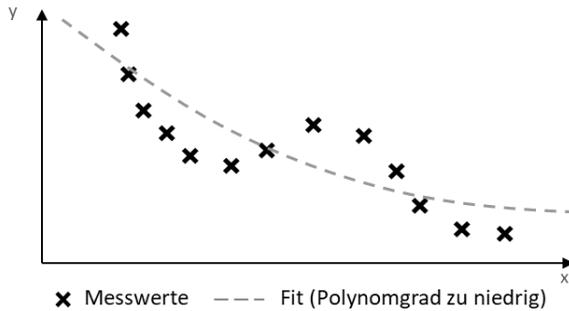


Abbildung 2.9: Lack-of-fit¹⁵

- Umgekehrt sollte der Grad des Polynoms allerdings auch nicht zu hoch gewählt werden, da es sonst zu einem *Over-fit* kommen kann. Dies bedeutet, dass das Polynom zwar so an die Versuchsergebnisse angepasst wird, dass es die Versuchspunkte optimal trifft. Zwischen den Versuchspunkten und vor allem am Rand des Untersuchungsbereiches kann es dagegen ein Verhalten aufweisen, das sich nicht durch die Versuchsergebnisse stützen lässt. Ein Beispiel hierfür ist in Abbildung 2.10 dargestellt.

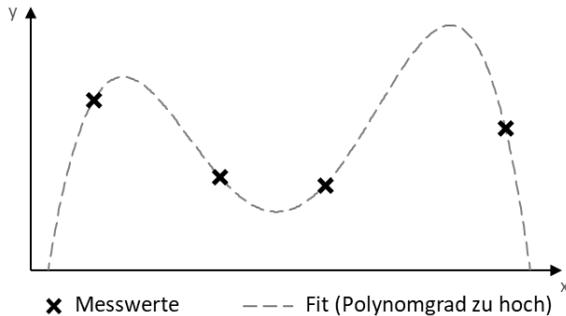


Abbildung 2.10: Over-fit

¹⁵ Vgl. beispielsweise (Sharmara et al., 2014, Fig. 3) oder (Shao et al., 2017, Fig. 2)

2.4.2 Grundprinzipien für valide Versuchsergebnisse

Bei jeder Versuchsdurchführung gibt es Fehlereinflüsse, die die Versuchsergebnisse verfälschen: Die Einstellung der untersuchten Einflussfaktoren ist nie vollkommen genau, es gibt meist Störgrößen, deren Einfluss nicht unmittelbar bestimmt werden kann, und auch bei der Messung der Zielgröße lässt sich keine absolute Genauigkeit erreichen. All diese Fehlereinflüsse führen dazu, dass ein Versuchsergebnis den untersuchten Zusammenhang nie vollständig richtig beschreiben kann. Aus diesem Grund wird eine Vorgehensweise benötigt, die sicherstellt, dass die Fehlereinflüsse die gewonnene Beschreibung des untersuchten Systems so wenig wie möglich verfälschen. Zusätzlich zur Minimierung der Fehlereinflüsse sollten diese auch quantifiziert werden. Nur so kann die Validität der aus der Versuchsreihe abgeleiteten Modelle bewertet werden. Beide Ziele können durch Beachtung der nachfolgend beschriebenen Grundprinzipien bei der Versuchsplanung erreicht werden.

2.4.2.1 Randomisierung

In real durchgeführten Versuchen gibt es oft Störgrößen oder Messungenauigkeiten, die sich über die Zeit nur langsam ändern (Trends). Zudem kann es vorkommen, dass Versuchsergebnisse durch vorhergehende Versuche beeinflusst werden. Beides hat eine systematische Verfälschung der Versuchsergebnisse zur Folge, wenn Faktoren im Versuchsplan in steigender oder fallender Abfolge variiert werden. Eine Verfälschung der Versuchsergebnisse durch einen fallenden Trend, bei Änderung eines Faktors in steigender Reihenfolge, ist in Abbildung 2.11 links dargestellt. Trotz Verfälschung weisen die so erzielten Versuchsergebnisse keine erhöhte Streuung auf, anhand derer auf das Vorhandensein von relevanten Störgrößen geschlossen werden könnte. Man erhält also verfälschte Versuchsergebnisse, ohne es zu bemerken. Auch statistische Bewertungsgrößen¹⁶ werden für eine solche Versuchsreihe nicht korrekt bestimmt. Dies zeigt sich beispielsweise in Abbildung 2.11 links daran, dass die reale Abhängigkeit weit außerhalb des ermittelten Vertrauensbereichs liegt. Der Vertrauensbereich wird hier viel zu klein geschätzt.

Um die Entstehung dieser Probleme zu verhindern, müssen alle Versuche in zufälliger Reihenfolge durchgeführt werden. Dieses Prinzip wird Randomisierung genannt. Es stellt sicher, dass auch der Einfluss von zeitlich veränderlichen Störgrößen als statistisch regellose Verfälschung in die Versuchsergebnisse einfließt (siehe Abbildung 2.11, rechts). Wenn eine Ausgleichsgerade oder eine andere Ansatzfunktion niedrigen Grades durch die Ergebnisse einer randomisierten Versuchsreihe ge-

¹⁶ Vgl. Abschnitt 2.5.2.2

legt wird, dann wird diese durch die Störgröße im Mittel nicht beeinflusst. Die Verfälschung der auf die Versuchsergebnisse angepassten Funktion durch eine trendförmige Störung ist so also sehr viel geringer als ohne Randomisierung. Zudem lässt sich durch die nun erkennbare Streuung der Versuchsergebnisse auch darauf schließen, dass es vermutlich relevante Störgrößen bei der Versuchsdurchführung gegeben hat. Dadurch lassen sich auch Bewertungsgrößen, wie der Vertrauensbereich, korrekt ermitteln. (Siebertz et al., 2010, Kap. 4.3.1)

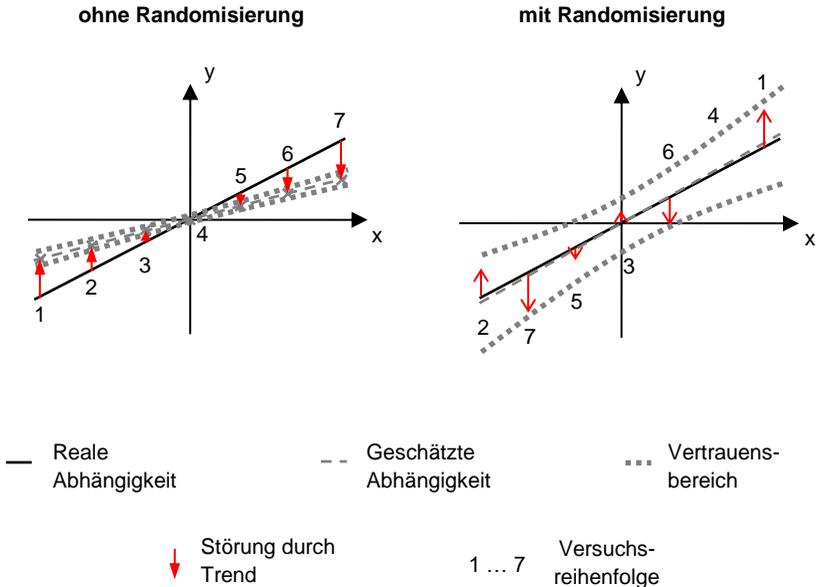


Abbildung 2.11: Verminderung des Fehlereinflusses von Trends durch Randomisierung

2.4.2.2 Wiederholung

Wenn aus mehreren streuenden Versuchsergebnissen ein Mittelwert gebildet wird, dann unterliegt auch dieser bei Wiederholung der Versuche einer Streuung. Die Varianz der Streuung dieses Mittelwerts ist umgekehrt proportional zur Anzahl der Versuche über die gemittelt wurde (Kleppmann, 2011, S. 73). Die Streuung in den Versuchsergebnissen lässt sich also durch Mittelwertbildung aus mehreren Versuchswiederholungen reduzieren, wodurch die Genauigkeit der Versuchsergebnisse gesteigert wird. Zudem lässt sich durch mehrere Versuchswiederholungen

erst bestimmen, wie groß die Streuung zwischen den einzelnen Versuchsergebnissen ist. Als Maß für die Streuung kann ein Schätzwert $\hat{\sigma}^2$ für die Varianz σ^2 nach Gleichung 2.1 bestimmt werden. Darin ist y_i das Ergebnis der i -ten Versuchswiederholung und \bar{y} der Mittelwert aller n Versuchsergebnisse.

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \quad 2.1$$

Dabei ist es wichtig, den kompletten Versuch inklusive der erneuten Einstellung aller Faktoren zu wiederholen. Ansonsten würde der Anteil der Streuung, der durch ungenaue Einstellung der Einflussfaktoren entsteht, nicht erfasst werden. Durch Vergleich der ermittelten Gesamtstreuung mit den gemessenen Effekten lässt sich bewerten, wie wahrscheinlich es ist, dass die Effekte real existieren und nicht nur durch statistische Streuung verursacht wurden. (Kowalski et al., 2007)

2.4.2.3 Blockbildung

Wenn bei größeren Versuchsreihen bekannt ist, dass eine Störgröße sich schleichend oder sprunghaft ändert, dann kann die Versuchsreihe in mehrere Blöcke unterteilt werden. Dadurch wird der Einfluss dieser Störgröße auf die Versuchsergebnisse reduziert. Die Versuchsreihe sollte so unterteilt werden, dass die durchschnittliche Einstellung jedes Faktors innerhalb jedes Blocks identisch ist. Die Effekte der sich ändernden Störgröße können dann bestimmt werden, indem die Mittelwerte der Zielgröße innerhalb jedes Blocks berechnet und anschließend miteinander verglichen werden. Durch Elimination der Störgrößeneffekte aus der Modellgleichung kann die statistische Streuung, und damit der Vertrauensbereich des Modells, verkleinert werden. (Kleppmann, 2011, S. 28)

2.5 Statistische Versuchsplanung für lineare Regressionsmodelle

Die nachfolgend dargestellten Versuchspläne und dazugehörigen Auswerteverfahren sind darauf ausgerichtet, *lineare Modelle* möglichst effizient an das gemessene Systemverhalten anpassen zu können. Linear bedeutet hierbei lediglich, dass die bei der Auswertung zu bestimmenden Koeffizienten linear in das Modell eingehen müssen. Die Einflüsse der Faktoren selbst können auch nichtlinear sein.¹⁷ Meist werden als Modelltyp mehrdimensionale Polynome eingesetzt. In Abschnitt 2.5.1

¹⁷ Vgl. Abschnitt 2.5.2

werden zunächst die gängigsten Versuchspläne für diese Aufgabe vorgestellt. Die dazugehörenden Modelle und Auswerteverfahren werden in den Abschnitten 2.5.2 bis 2.5.4 erläutert. In vielen Fällen kann die Effizienz der Versuchspläne aus Abschnitt 2.5.1 weiter gesteigert werden, wenn von der vollständigen Randomisierung¹⁸ abgewichen wird. Welche Vorteile dies bringt und wie eine solche Versuchsreihe strukturiert und ausgewertet werden muss, um dennoch valide Ergebnisse zu erzielen, wird in Abschnitt 2.5.5 dargestellt.

2.5.1 Versuchspläne

Das Ziel der in diesem Abschnitt beschriebenen Versuchspläne ist, eine möglichst genaue und umfassende Abbildung des Systemverhaltens auf Polynombasis mit einem möglichst geringen Versuchsaufwand zu erzielen. Bereits die Wahl des an die Versuchsergebnisse anzupassenden Polynoms entscheidet über die erreichbare Effizienz. Ein geeignetes Polynom sollte üblicherweise in der Lage sein, den räumlichen Verlauf der Zielgröße im Faktorraum innerhalb des interessierenden Bereichs möglichst überall gleich detailliert abzubilden. Für ein mehrdimensionales Polynom bedeutet dies, dass es meist keinen Sinn macht, Terme höheren Grades für manche Faktoren oder Wechselwirkungen zu bestimmen, während noch nicht alle Terme der niedrigeren Grade bestimmt sind (Kleppmann, 2011, S. 216).

Ein Beispiel für ein wenig sinnvolles zweidimensionales Polynom ist deshalb Gleichung 2.2. Diese beschreibt die Effekte der beiden Faktoren x_1 und x_2 mit Termen bis zum jeweils vierten Grad. Entlang der Koordinatenachsen ermöglicht dieses Polynom also eine verhältnismäßig genaue Abbildung beliebigen Verhaltens. Da keine Wechselwirkungen enthalten sind, kann das Verhalten entlang der Diagonalen der x_1 - x_2 -Ebene durch dieses Polynom nicht weiter beeinflusst werden. Ein Beispiel für den abbildbaren Verlauf der Zielgröße durch eine solche Modellgleichung ist in Abbildung 2.12 a dargestellt. Durch die fehlenden Wechselwirkungen besitzen alle Linien der Gitterdarstellung der Funktionsfläche, die in derselben Richtung verlaufen, dieselbe Form. Der Einfluss des Faktors x_1 ist also für alle Werte von x_2 identisch. Dasselbe gilt für den Einfluss von x_2 , der für alle Werte von x_1 identisch ist.

$$\begin{aligned}
 y = & \beta_0 + \beta_1 x_1 + \beta_{11} x_1^2 + \beta_{111} x_1^3 + \beta_{1111} x_1^4 \\
 & + \beta_2 x_2 + \beta_{22} x_2^2 + \beta_{222} x_2^3 + \beta_{2222} x_2^4
 \end{aligned}
 \tag{2.2}$$

¹⁸ Siehe Abschnitt 2.4.2.1

Den umgekehrten Fall stellt Gleichung 2.3 dar. Hier sind Wechselwirkungen bis zum vierten Grad enthalten, während Effekte der Faktoren nicht abgebildet werden. Damit kann der Verlauf der Zielgröße entlang der Diagonalen in Abbildung 2.12 b sehr genau abgebildet werden. Entlang der Koordinatenachsen (für $x_1 = 0$ und $x_2 = 0$) besitzt y hingegen den konstanten Wert β_0 . Auch die Abbildungsleistung dieses Polynoms ist also stark abhängig von der Lage in der x_1 - x_2 -Ebene.

$$y = \beta_0 + \beta_{12}x_1x_2 + \beta_{112}x_1^2x_2 + \beta_{122}x_1x_2^2 + \beta_{1122}x_1^2x_2^2 \quad 2.3$$

Sinnvoll ist hingegen das Polynom aus Gleichung 2.4, das alle Terme bis zum vierten Grad enthält. Es ermöglicht damit eine ähnlich gute Abbildungsleistung entlang der Koordinatenachsen und der Diagonalen. Damit können nahezu beliebig gekrümmte Flächen angenähert werden. Eine exemplarische grafische Darstellung einer solchen Funktion ist in Abbildung 2.12 c gegeben.

$$\begin{aligned} y = & \beta_0 + \beta_1x_1 + \beta_{11}x_1^2 + \beta_{111}x_1^3 + \beta_{1111}x_1^4 \\ & + \beta_2x_2 + \beta_{22}x_2^2 + \beta_{222}x_2^3 + \beta_{2222}x_2^4 \\ & + \beta_{12}x_1x_2 + \beta_{112}x_1^2x_2 + \beta_{122}x_1x_2^2 + \beta_{1122}x_1^2x_2^2 \end{aligned} \quad 2.4$$

Welche Terme eines Polynoms mit einem Versuchsplan bestimmbar sind, entscheidet sich wie folgt: Im Versuchsplan müssen für jeden zu bestimmenden Term Versuche mit unterschiedlicher Einstellung des enthaltenen Faktors vorkommen. Der Effekt eines Faktors, dessen Einstellung im Versuchsplan nicht variiert wird, kann nicht bestimmt werden. Damit die Effekte mehrerer Faktoren möglichst effizient in einem Versuchsplan bestimmt werden können, sollten die Einstellungslisten¹⁹ der Faktoren zudem orthogonal zueinander sein. Dies bedeutet, dass eine Änderung der Einstellung eines Faktors im Mittel keinen Einfluss auf die Einstellung eines anderen Faktors hat. Ein orthogonaler Versuchsplan ist in Abbildung 2.13 links dargestellt.

¹⁹ Mit „Einstellungsliste“ ist hier ein Spaltenvektor gemeint, der alle Einstellungen des Faktors in der Reihenfolge der Versuchsdurchführung enthält. Jede Zeile des Vektors entspricht einem Versuch in der Versuchsreihe.

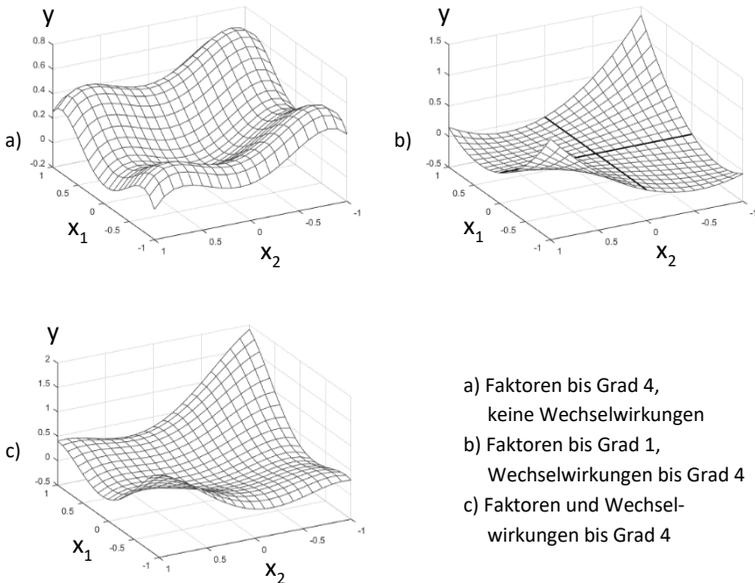


Abbildung 2.12: Beispiele für die Abbildungsleistung mehrdimensionaler Polynome

Wenn die Einstellungslisten zweier Faktoren nicht orthogonal zueinander sind, dann werden die Effekte dieser Faktoren miteinander vermengt. Dies bedeutet, dass bei der Auswertung nicht vollständig geklärt werden kann, von welchem Faktor der gemessene Effekt stammt. Dann kann beispielsweise ein geringer Effekt eines Faktors durch einen großen Effekt eines anderen Faktors überdeckt werden. Dadurch werden die Vertrauensbereiche der nicht orthogonalen Faktoren größer. Im Extremfall sind die Einstellungslisten zweier Faktoren kollinear, das heißt, dass sie durch Multiplikation mit einer Konstanten ineinander überführt werden können. Ein kollinear Versuchsplan ist in Abbildung 2.13 rechts dargestellt. Bei einem solchen Versuchsplan sind die Effekte der beiden Faktoren x_1 und x_2 in den Versuchsergebnissen vollständig und untrennbar miteinander vermengt. Gute Versuchspläne sind also bezüglich der Einstellungslisten aller zu bestimmender Faktoren möglichst orthogonal.

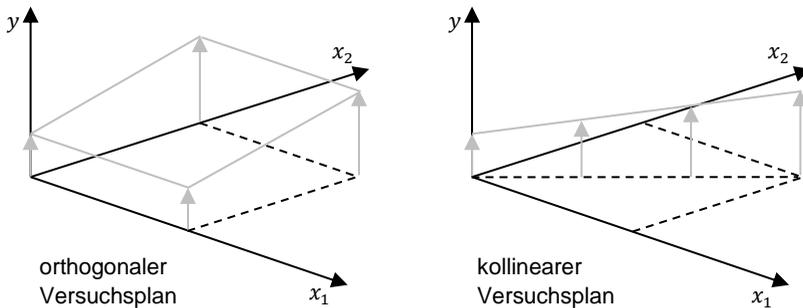


Abbildung 2.13: Vergleich zwischen orthogonalen und kollinearen Versuchsplänen. Beim kollinearen Versuchsplan kann nicht unterschieden werden, von welchem der beiden Faktoren x_1 und x_2 der gemessene Effekt verursacht wurde. Vgl. (Kleppmann, 2011, Bild 10-11).

Mathematisch formuliert sind die Einstellungslisten zweier Faktoren genau dann orthogonal, wenn ihre Kovarianz null ist. Die Kovarianz zwischen zwei Faktoren x_1 und x_2 kann nach Gleichung 2.5 berechnet werden, in der N die Gesamtzahl der Versuche im Versuchsplan beschreibt (Czado & Schmidt, 2011, S. 135).

$$s_{x_1 x_2} = \frac{1}{N} \sum_{i=1}^N (x_{1,i} - \bar{x}_1)(x_{2,i} - \bar{x}_2) \quad 2.5$$

Nachfolgend werden in den Abschnitten 2.5.1.1 bis 2.5.1.8 die gängigsten Versuchsplantypen dargestellt. Die Informationen hierzu sind sinngemäß aus Lehrbüchern (Kleppmann, 2011; Montgomery, 2009; Siebertz et al., 2010) zur statistischen Versuchsplanung entnommen. Zu beachten ist, dass die in den Tabellen dieses Kapitels dargestellten Versuchspläne für eine bessere Übersichtlichkeit in einer initialen Version, also noch vor der Anwendung der Prinzipien aus Abschnitt 2.4.2, dargestellt sind.

2.5.1.1 One-factor-at-a-time

In One-factor-at-a-time-Versuchsreihen wird der Effekt jedes Faktors in mehreren Stufen, aber dabei nur für eine einzige Einstellungskombination der restlichen Faktoren untersucht. Deshalb kann hier nicht überprüft werden, ob der Effekt eines Faktors selbst von der Einstellung von einem oder mehreren anderen Faktoren abhängt.

Es können also keine Wechselwirkungen abgebildet werden. In Abbildung 2.14 ist ein Versuchsplan dargestellt, der auf diese Weise für drei Faktoren erstellt wurde, die in jeweils fünf Stufen untersucht werden.

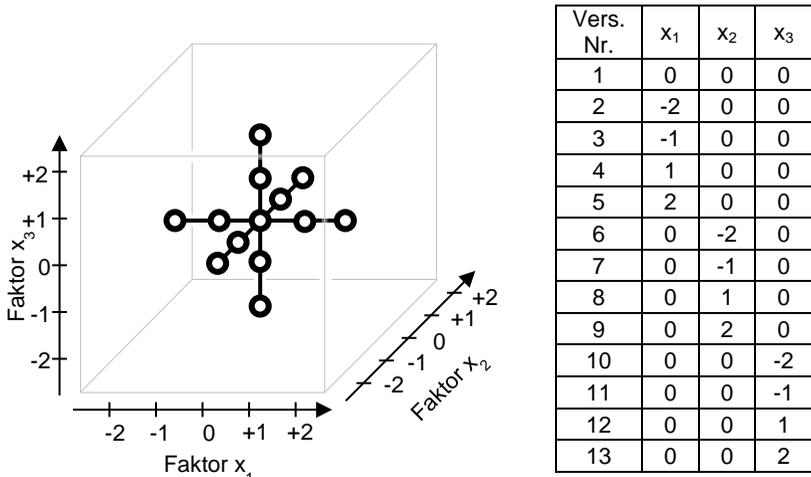


Abbildung 2.14: Darstellung eines konventionellen One-factor-at-a-time-Versuchsplans Darstellungsweise in Anlehnung an (Kleppmann, 2011, S.108)

Da alle Faktoren in jeweils fünf Stufen untersucht werden, kann für alle Faktoren ein Polynom vierten Grades an die Versuchsergebnisse angepasst werden. Wechselwirkungen können hingegen nicht bestimmt werden, da die Effekte jedes Faktors nur für eine Einstellungskombination der anderen Faktoren untersucht werden. Aus den Versuchsergebnissen kann ein mathematisches Modell, wie es in Gleichung 2.6 gezeigt ist, erstellt werden.

$$\begin{aligned}
 y = & \beta_0 + \beta_1 x_1 + \beta_{11} x_1^2 + \beta_{111} x_1^3 + \beta_{1111} x_1^4 \\
 & + \beta_2 x_2 + \beta_{22} x_2^2 + \beta_{222} x_2^3 + \beta_{2222} x_2^4 \\
 & + \beta_3 x_3 + \beta_{33} x_3^2 + \beta_{333} x_3^3 + \beta_{3333} x_3^4
 \end{aligned} \tag{2.6}$$

Hierbei ist es weder konsequent, noch effizient, für die Faktoren Terme bis zum vierten Grad zu bestimmen, während Wechselwirkungen selbst zweiten Grades vernachlässigt werden.²⁰

2.5.1.2 Zweistufige vollständig faktorielle Versuchspläne

In vollständig faktoriellen Versuchsplänen wird jede mögliche Stufenkombination aller Faktoren untersucht. Ein nach dieser Methode erstellter Versuchsplan besitzt im zweistufigen Fall, bei einer Anzahl von p Faktoren, einen Versuchsumfang von 2^p Versuchen (Kleppmann, 2011, S. 107). Beispielhaft ist ein solcher Versuchsplan für $p = 3$ Faktoren in Abbildung 2.15 dargestellt.

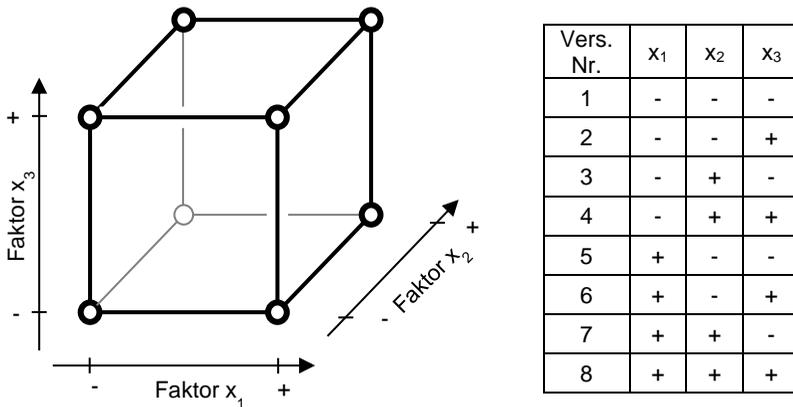


Abbildung 2.15: Darstellung eines vollständig faktoriellen 2^3 Versuchsplans in Anlehnung an (Kleppmann, 2011, S.108)

In einem zweistufigen vollständig faktoriellen Versuchsplan können die Effekte jedes Faktors in einem linearen Modell abgebildet werden. Im Gegensatz zum One-factor-at-a-time-Versuchsplan wird der Effekt jedes Faktors für jede mögliche Kombination der anderen Faktoren untersucht. Ob und wie der Effekt eines Faktors von der Einstellung der restlichen Faktoren abhängt, kann deshalb hier ermittelt werden. An die

²⁰ Vgl. Einleitung Abschnitt 2.5.1, vor allem Gleichung 2.2

Ergebnisse eines vollständig faktoriellen Versuchsplans wird üblicherweise eine Modellgleichung angepasst, wie sie in Gleichung 2.7 für den Versuchsplan aus Abbildung 2.15 dargestellt ist.

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_{12}x_1x_2 + \beta_{13}x_1x_3 + \beta_{23}x_2x_3 + \beta_{123}x_1x_2x_3 \quad 2.7$$

Hierbei beschreibt β_0 den Mittelwert aller durchgeführten Versuche und somit einen Schätzwert für die Zielgröße im Zentrum des Würfels aus Abbildung 2.15. β_1 , β_2 und β_3 sind, wie schon im vorhergehenden Abschnitt, die Koeffizienten der Effekte der Faktoren x_1 , x_2 und x_3 . Die Koeffizienten β_{12} , β_{13} und β_{23} beschreiben die Stärke der Wechselwirkungen zwischen jeweils zwei dieser Faktoren. Dementsprechend ist der Koeffizient β_{123} ein Maß für die Dreifachwechselwirkung zwischen allen drei Faktoren. Die Einstellungslisten sind für alle Faktoren dieses Modells im zweistufigen vollständig faktoriellen Versuchsplan orthogonal.

Trotz seiner Vorzüge gegenüber One-factor-at-a-time-Versuchsreihen besitzt der zweistufige vollständig faktorielle Versuchsplan auch Nachteile: Während die Effekte aller Faktoren nur durch eine lineare Gleichung abgebildet werden, sind Wechselwirkungen mindestens zweiten, dritten oder noch höheren Grades aus den Versuchsergebnissen ermittelbar. Hier wird also das Verhalten entlang der Diagonalen genauer abgebildet, als das entlang der Koordinatenachsen.²¹ Je mehr Faktoren in der Versuchsreihe untersucht werden sollen, desto größer wird dieses Missverhältnis, da die Anzahl der untersuchten Wechselwirkungen auch deutlich höheren Grades hier sehr schnell steigt. So wird beispielsweise durch einen vollständig faktoriellen Versuchsplan mit drei Faktoren nur eine Wechselwirkung dritten oder höheren Grades bestimmt. Bei fünf Faktoren sind es bereits 16 solcher Wechselwirkungen und bei 10 Faktoren sind es 968. Der größte Teil der $2^{10} = 1024$ Versuche²² eines zweistufigen vollständig faktoriellen Versuchsplans mit zehn Faktoren wird also dazu verwendet, Wechselwirkungsterme dritten oder höheren Grades zu bestimmen. Für die Effekte der einzelnen Faktoren werden hingegen nur Terme ersten Grades bestimmt. Vor allem für höhere Anzahlen von Faktoren werden also auch vollständig faktorielle Versuchspläne ineffizient.

2.5.1.3 Zweistufige fraktionell faktorielle Versuchspläne

Um die geringe Effizienz vollständig faktorieller Versuchspläne, vor allem für größere Anzahlen von Faktoren, verbessern zu können, wurden fraktionell faktorielle

²¹ Vgl. Einleitung Abschnitt 2.5.1, vor allem Gleichung 2.3

²² bzw. der durch die Versuche generierten Freiheitsgrade

Versuchspläne entwickelt. Diese beinhalten eine gezielt ausgewählte Teilmenge der Versuche von vollständig faktoriellen Versuchsplänen. Die Auswahl wird dabei so getroffen, dass vor allem die Einstellungslisten von Wechselwirkungen höheren Grades untereinander nicht mehr orthogonal, sondern meist vollständig vermengt sind. Diese Wechselwirkungseffekte können dann zwar nicht mehr zuverlässig bestimmt werden, der Versuchsumfang lässt sich so aber für höhere Faktoranzahlen sehr stark senken.

Es gibt verschiedene Konstruktionsmethoden, die zu fraktionell faktoriellen Versuchsplänen unterschiedlicher Auflösung führen. Diese werden ausführlich beispielsweise von Kleppmann (2011, Kap. 8) und Siebertz et al. (2010, Kap. 2.2) erläutert. Ihr Auflösungsvermögen wird in römischen Ziffern angegeben. Es ergibt sich als Summe der Grade von untereinander vermengten Effekten. So sind beispielsweise bei einem fraktionell faktoriellen Versuchsplan mit Auflösung IV sowohl Zweifachwechselwirkungen untereinander ($2+2=4$), als auch die Effekte linearer Faktoren mit Dreifachwechselwirkungen ($1+3=4$) vermengt. Je größer die Zahl für das Auflösungsvermögen eines fraktionell faktoriellen Versuchsplans ist, desto höher ist der Grad der Wechselwirkungen, die vermengungsfrei bestimmt werden können. Allerdings steigt auch die Anzahl der Versuche in der Versuchsreihe mit steigendem Grad stark an (Kleppmann, 2011, S. 137 ff.). Üblich sind fraktionell faktorielle Versuchspläne mit Graden zwischen III und VI.

Fraktionell faktorielle Versuchspläne eignen sich aufgrund ihres geringen Versuchsumfangs besonders für Untersuchungen, bei denen Versuche sehr teuer oder aufwendig sind. Zudem können sie auch eingesetzt werden, um bei vielen in Frage kommenden Einflussgrößen in einer Voruntersuchung diejenigen mit den größten Effekten zu ermitteln. Für detailliertere Untersuchungen von nichtlinearen Zusammenhängen sind sie aber nicht geeignet, da sie nur ein lineares Beschreibungsmodell des untersuchten Systemverhaltens liefern. Ein fraktionell faktorieller Versuchsplan mit Auflösung III für drei Faktoren ist in Abbildung 2.16 dargestellt.

Da bei Versuchsplänen der Auflösung III bereits Zweifachwechselwirkungen mit Effekten linearer Faktoren vermengt sind, können hier keine Wechselwirkungen bestimmt werden. Als Modellgleichung ergibt sich für den dargestellten Versuchsplan somit Gleichung 2.8.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \tag{2.8}$$

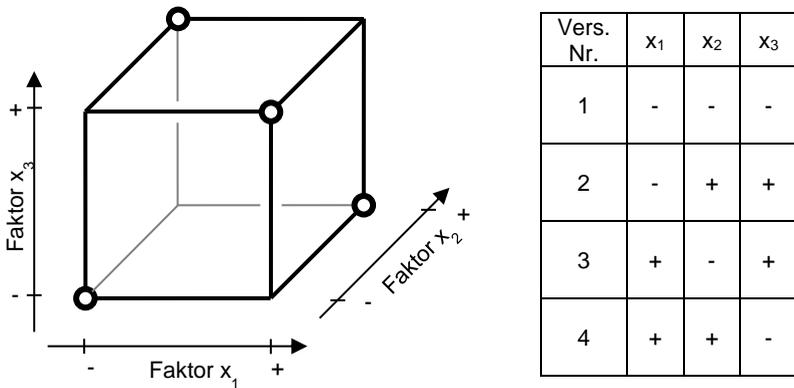


Abbildung 2.16: Darstellung eines fraktionell faktoriellen 2^{3-1} Versuchsplans
Darstellungsweise in Anlehnung an (Kleppmann, 2011, S.108)

2.5.1.4 Dreistufige vollständig faktorielle Versuchspläne

Wenn durch eine Versuchsreihe nichtlineare Zusammenhänge untersucht und abgebildet werden sollen, dann müssen die untersuchten Faktoren in mindestens drei Stufen variiert werden. Die einfachste Möglichkeit einen solchen Versuchsplan zu erstellen besteht in der Erweiterung des zweistufigen vollständig faktoriellen Versuchsplans um eine weitere Faktorstufe. Diese zusätzliche Faktorstufe wird im Versuchsplan durch eine „0“ gekennzeichnet und befindet sich für jeden Faktor in der Mitte zwischen den Faktorstufen „+“ und „-“. Für drei zu untersuchende Faktoren ist ein solcher Versuchsplan in Abbildung 2.17 dargestellt. Mit ihm können auch quadratische Effekte abgebildet werden. Allerdings löst ein solcher Versuchsplan auch Wechselwirkungen zwischen quadratischen Effekten auf. Dies verschärft das Problem des zweistufigen vollständig faktoriellen Versuchsplans mit dem überhöhten Auflösungsvermögen der Wechselwirkungen noch deutlich. So können bereits bei drei untersuchten Faktoren Wechselwirkungen sechsten Grades abgebildet werden, während dies für die Faktoren selbst nur mit einem Polynom zweiten Grades möglich ist. Dies ist inkonsequent und führt zu einem unverhältnismäßig großen Versuchsumfang. Der Einsatz eines solchen Versuchsplans ist deshalb normalerweise nicht sinnvoll. (Kleppmann, 2011, S. 213 f.)

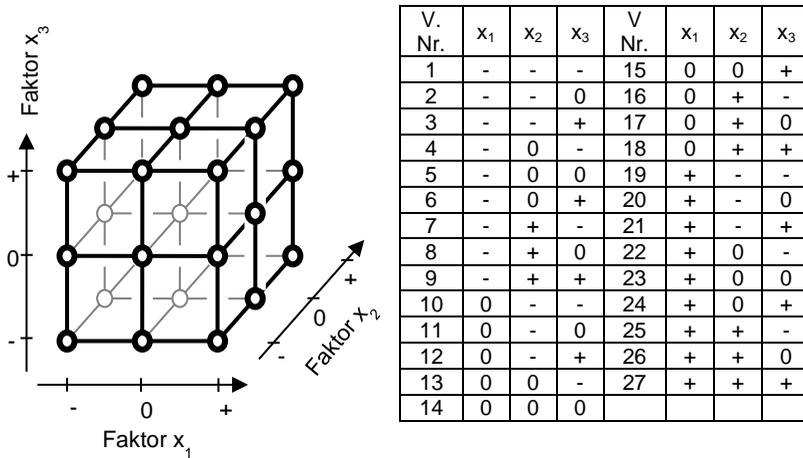


Abbildung 2.17: Darstellung eines vollständig faktoriellen 3³ Versuchsplans
Darstellungsweise in Anlehnung an (Kleppmann, 2011, S.108)

Auf die Ergebnisse des Versuchsplans aus Abbildung 2.17 kann eine Modellgleichung der Form von Gleichung 2.9 angepasst werden.

$$\begin{aligned}
 y = & \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_{11}x_1^2 + \beta_{22}x_2^2 + \beta_{33}x_3^2 + \beta_{12}x_1x_2 \\
 & + \beta_{13}x_1x_3 + \beta_{23}x_2x_3 + \beta_{112}x_1^2x_2 + \beta_{113}x_1^2x_3 \\
 & + \beta_{122}x_1x_2^2 + \beta_{133}x_1x_3^2 + \beta_{223}x_2^2x_3 + \beta_{233}x_2x_3^2 \\
 & + \beta_{123}x_1x_2x_3 + \beta_{1122}x_1^2x_2^2 + \beta_{1133}x_1^2x_3^2 + \beta_{2233}x_2^2x_3^2 \quad 2.9 \\
 & + \beta_{1123}x_1^2x_2x_3 + \beta_{1223}x_1x_2^2x_3 + \beta_{1233}x_1x_2x_3^2 \\
 & + \beta_{11223}x_1^2x_2^2x_3 + \beta_{11233}x_1^2x_2x_3^2 + \beta_{12233}x_1x_2^2x_3^2 \\
 & + \beta_{112233}x_1^2x_2^2x_3^2
 \end{aligned}$$

2.5.1.5 Dreistufige fraktionell faktorielle Versuchspläne

Wie schon beim zweistufigen vollständig faktoriellen Versuchsplan gibt es auch beim dreistufigen vollständig faktoriellen Versuchsplan die Möglichkeit, nur eine gezielt getroffene Auswahl der Versuche durchzuführen, um den Versuchsumfang zu

reduzieren. Ein solcher Versuchsplan ist für drei Faktoren in Abbildung 2.18 dargestellt. Solche Versuchspläne besitzen allerdings meist den Nachteil, dass sie quadratische Effekte auflösen können, aber Zweifachwechselwirkungen nicht orthogonal untereinander und zu anderen Effekten sind. Sie sind hier so stark mit anderen Effekten vermengt, dass sie faktisch nicht bestimmt werden können. Als Modellgleichung ergibt sich deshalb Gleichung 2.10.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 \quad 2.10$$

Bei diesen Versuchsplänen ist also das Auflösungsvermögen von Wechselwirkungen deutlich geringer als das der restlichen Faktoren. Aus diesem Grund ist ihr Einsatz nach Box & Wilson (1951) und Kleppmann (2011, S. 236–239) nur in Sonderfällen sinnvoll.

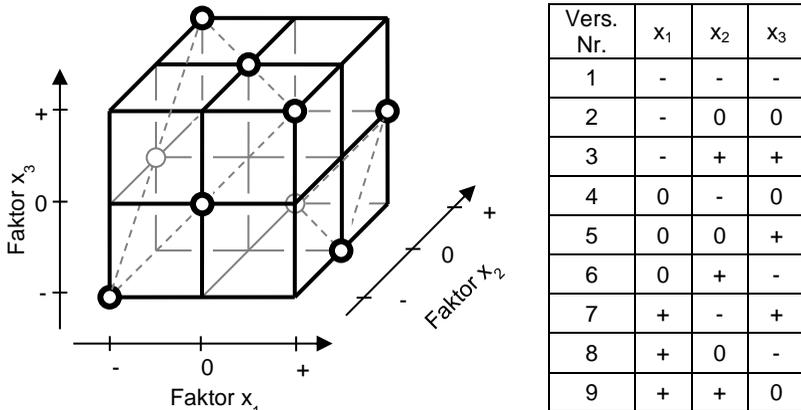


Abbildung 2.18: Darstellung eines fraktionell faktoriellen 3^{3-1} Versuchsplans
Darstellungsweise in Anlehnung an (Kleppmann, 2011, S.108)

2.5.1.6 Box-Behnken-Versuchspläne

Ähnlich zu dreistufigen fraktionell faktoriellen Versuchsplänen bestehen Box-Behnken-Versuchspläne aus einer Auswahl von Versuchen eines dreistufigen vollständig faktoriellen Versuchsplans. Hier ist die Auswahl aber im Gegensatz zu den lateinischen Quadraten so getroffen, dass Zweifachwechselwirkungen vermengungsfrei bestimmt werden können. Als Modellgleichung ergibt sich somit Gleichung 2.11.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 \tag{2.11}$$

Allerdings sind hier die quadratischen Effekte nicht vollständig orthogonal zueinander (Kleppmann, 2011, S. 215). Da die Abweichung von der Orthogonalität nur gering ist, können diese Effekte hier bestimmt werden. Dennoch verbreitert diese Abweichung die Vertrauensbereiche der für diese Effekte ermittelten Regressionskoeffizienten. Ein Box-Behnken-Versuchsplan für drei Faktoren ist in Abbildung 2.19 dargestellt. Er besteht aus dem Zentrumsunkt des dreistufigen vollständig faktoriellen Versuchsplans, sowie aus allen Punkten, die auf den Mitten der Würfelkanten liegen.

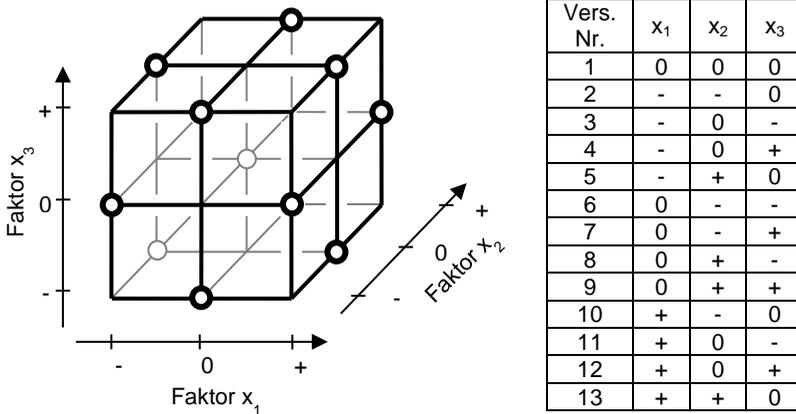


Abbildung 2.19: Darstellung eines Box-Behnken-Versuchsplans Darstellungsweise in Anlehnung an (Kleppmann, 2011, S.108)

2.5.1.7 Zentral zusammengesetzte Versuchspläne

Ein zentral zusammengesetzter Versuchsplan für drei Faktoren ist in Abbildung 2.20 dargestellt. Er besteht aus einem zweistufigen vollständig faktoriellen Versuchsplan, der um einen Zentrumsunkt und sogenannte „Sternpunkte“ erweitert ist. Der Zentrumsunkt befindet sich in der Mitte dieses Würfels und besitzt in der Tabelle die

Versuchsnummer eins. Der vollständig faktorielle Teil bildet die würfelförmige Struktur in Abbildung 2.20 und besteht aus den Versuchen zwei bis neun in der zugehörigen Tabelle. Die Sternpunkte befinden sich mittig über den Würfelflächen im Abstand α vom Zentrumspunkt. In der Tabelle haben sie die Nummern zehn bis fünfzehn. Der Wert von α kann frei gewählt werden, hat aber Auswirkungen auf die Eigenschaften des Versuchsplans. Für zentral zusammengesetzte Versuchspläne wird üblicherweise dieselbe Art von Modellgleichung (Gleichung 2.11) verwendet wie bei Box-Behnken-Plänen.

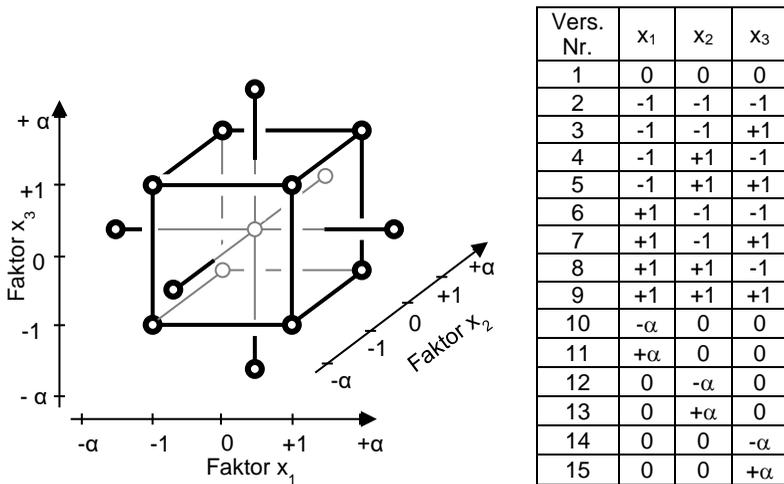


Abbildung 2.20: Darstellung eines zentral zusammengesetzten Versuchsplans
Darstellungsweise in Anlehnung an (Kleppmann, 2011, S.108)

Wenn α so gewählt wird, dass Gleichung 2.12 erfüllt ist, dann sind im Versuchsplan alle lineare und quadratische Effekte sowie Zweifachwechselwirkungen orthogonal zueinander und können vermengungsfrei bestimmt werden (Kleppmann, 2011, S. 200). In dieser Gleichung beschreibt N die Gesamtzahl der Versuche im Versuchsplan und N_v die Zahl der Versuche im vollständig faktoriellen Teil des Versuchsplans. Für das Beispiel in Abbildung 2.20 gilt also $N = 15$ und $N_v = 8$.

$$\alpha^2 = \frac{1}{2}(\sqrt{N N_v} - N_v) \quad 2.12$$

Wird α dagegen so gewählt, dass Gleichung 2.13 erfüllt ist, dann wird eine Eigenschaft erreicht, die *Drehbarkeit* genannt wird. Diese sagt aus, dass die Vorhersagegenauigkeit des mit dem Versuchsplan ermittelten Modells in alle Richtungen gleich gut ist. Die Breite des Vertrauensbereichs eines Schätzwerts an einer beliebigen Stelle des Faktorraums ist dann nur vom Abstand vom Zentrum des Versuchsplans abhängig und nicht von der Richtung.

$$\alpha^2 = \sqrt{N_v} \quad 2.13$$

Man kann durch Wiederholung des Zentrumsversuchs ($x_1 = x_2 = x_3 = 0$) den Versuchsumfang N so beeinflussen, dass das α aus Gleichung 2.12 näherungsweise dem α aus Gleichung 2.13 entspricht. Dann erhält man einen Versuchsplan der nahezu orthogonal und vollständig drehbar ist. Im Beispiel aus Abbildung 2.20 sind dazu neun Durchführungen des Zentrumsversuchs notwendig. Daraus resultiert dann ein Wert von etwa 1,682 für α .

Der zentral zusammengesetzte Versuchsplan ist der einzige bisher beschriebene Versuchsplan, bei dem alle Faktoren auf fünf Stufen untersucht werden. Da ein Polynom zweiten Grades, das üblicherweise für diesen Versuchsplan als Modellgleichung eingesetzt wird, damit überbestimmt ist, lässt sich überprüfen, ob Lack-of-fit²³ vorliegt. Eine mögliche Vorgehensweise dazu wird in Abschnitt 2.5.2.3 beschrieben. Insgesamt ergibt sich so ein größerer Versuchsumfang als bei Box-Behnken-Versuchsplänen, dafür erhält man aber ein genaueres Modell höherer Informationsdichte, das auch Lack-of-fit-Tests ermöglicht.

Eine Abwandlung von zentral zusammengesetzten Versuchsplänen stellen kleine zusammengesetzte Pläne nach Draper & Lin (1990) dar. Bei diesen Versuchsplänen wird der vollständig faktorielle Teil des zentral zusammengesetzten Versuchsplans durch einen entsprechenden fraktionell faktoriellen Teil ersetzt. So kann der Versuchsumfang reduziert werden, was allerdings Vermengungen zwischen den Effekten höheren Grades zur Folge hat.

2.5.1.8 Optimale Versuchspläne

Unter optimalen Versuchsplänen werden Versuchspläne verstanden, bei denen die Versuchspunkte mittels eines Optimierungsalgorithmus ausgewählt werden. Dazu wird zunächst die Modellgleichung festgelegt, für die der Versuchsplan optimiert

²³ Grad des Polynoms ist zu klein, um stark nichtlinearem Verhalten folgen zu können. Vgl. Abbildung 2.9

werden soll. Anschließend muss dem Algorithmus eine Kandidatenliste, also eine Menge an möglichen Versuchspunkten, vorgegeben werden. Ebenso muss die Anzahl an Versuchspunkten festgelegt werden, die der Algorithmus aus dieser Kandidatenliste auswählen soll. Diese Auswahl trifft der Algorithmus so, dass sie ein Optimalitätskriterium bestmöglich erfüllt. Das am häufigsten verwendete Kriterium ist hierbei die sogenannte D-Optimalität. Als zu optimierender Kennwert dient hier das Produkt der Vertrauensbereichsbreiten aller Regressionskoeffizienten. Der Algorithmus wählt hier also diejenige Untermenge aus den Versuchskandidaten aus, die dieses Produkt minimiert.

Unabhängig vom gewählten Optimalitätskriterium ist der erhaltene Versuchsplan immer auf die vorab festgelegte Modellgleichung optimiert. Ob diese für den Anwendungsfall geeignet ist, oder Lack-of-fit oder Over-fit²⁴ vorliegt, lässt sich aus den Versuchsergebnissen deshalb nur schwer erkennen. Zudem wird üblicherweise spezielle Software benötigt, die den Optimierungsalgorithmus ausführen kann. Die große Stärke optimaler Versuchspläne ist hingegen ihre Flexibilität: Dadurch, dass eine beliebige Menge an möglichen Versuchspunkten als Kandidatenliste vorgegeben werden kann, lässt sich ein solcher Versuchsplan für beliebig geformte Versuchsräume erstellen. Im Gegensatz dazu können die bisher vorgestellten Versuchspläne nur näherungsweise (hyper-) quaderförmige Faktorräume gut ausfüllen und damit umfassend untersuchen. (Kleppmann, 2011, S. 217)

2.5.2 Das klassische lineare Modell

Es existieren drei unterschiedliche Typen linearer Modelle, die sich in ihrer Fähigkeit der Modellierung streuender (bzw. zufälliger) Einflüsse unterscheiden. Zufällige Einflüsse bestehen oft aus Fehlereinflüssen auf die Versuchsergebnisse. Dazu zählen Messfehler, Störgrößeneinflüsse und fehlerhafte Faktoreinstellungen. Das *klassische lineare Modell*, das in diesem Abschnitt beschrieben wird, unterliegt der Annahme, dass alle Einzelversuche der Versuchsreihe denselben zufälligen Einflüssen unterworfen sind.

Damit diese Annahme erfüllt ist, müssen alle oben genannten Fehlereinflüsse für jeden Einzelversuch dieselbe Varianz aufweisen. Dies trifft bereits nicht mehr zu, wenn eine Pause bei der Versuchsdurchführung eingelegt wurde. Es muss nämlich davon ausgegangen werden, dass Störgrößeneinflüsse (wie beispielsweise die Luftfeuchtigkeit oder die Raumtemperatur) sich über längere Pausen hinweg stärker verändern als zwischen zwei direkt aufeinander folgenden Versuchen. Spätestens wenn in einer Versuchsreihe von der vollständigen Randomisierung abgewichen

²⁴ Grad des Polynoms ist zu groß, sodass es stark schwingt. Vgl. Abbildung 2.10

wird²⁵, ist die Annahme konstanter Fehlervarianz über die Versuchsreihe hinweg nicht mehr haltbar. Für solche Fälle muss das eingesetzte Modell einen komplexeren Fehlerterm erlauben, um dennoch valide Ergebnisse zu erhalten. Diese Forderung erfüllt das *allgemeine lineare Modell* aus Abschnitt 2.5.3.

In manchen Fällen umfassen zufällige Einflüsse nicht nur ungewollte Fehlereinflüsse. Im Praxisbeispiel dieser Arbeit²⁶ wird u.a. untersucht, welchen Einfluss verschiedene Faktoren auf das Reibmoment von Kugelhähnen besitzen. Durch Produktionstoleranzen kann dieses Reibmoment zwischen unterschiedlichen Exemplaren deutlich schwanken. Doch nicht nur der Mittelwert des Reibmoments unterliegt solchen Schwankungen. Auch der Einfluss einzelner Faktoren kann für unterschiedliche Exemplare verschieden sein. Falls solche Schwankungen von Interesse sind, muss ein lineares Modell eingesetzt werden, das neben den *festen Effekten* der Faktoren auch umfangreiche *zufällige Effekte* beschreiben kann. Modelle, die beide Arten von Effekten beinhalten, werden *gemischte lineare Modelle* genannt. Dieser Modelltyp wird in Abschnitt 2.5.4 vorgestellt.

Das klassische lineare Modell ist in Gleichung 2.14 für eine Anzahl von p Faktoren formuliert. Darin ist y_i der im i -ten Versuch einer Versuchsreihe bestimmte Wert der Zielgröße. $x_{1,i}$ bis $x_{p,i}$ sind die Einstellungen der p Faktoren im jeweiligen Versuch. β_0 bis β_p sind die *Regressionskoeffizienten*. Dabei ist β_0 der Mittelwert der Zielgröße über alle Versuche und wird *Interzeptparameter* genannt. β_1 bis β_p quantifizieren die Effekte der Faktoren x_1 bis x_p . Zur Abbildung streuender Einflüsse, sowie von Passungsgenauigkeiten des Modells, ist ein Fehlerterm ε_i in der Modellgleichung enthalten. Dieser beschreibt den Anteil des untersuchten Systemverhaltens, der nicht durch das ermittelte Modell für die p Faktoren erklärt werden kann. (Czado & Schmidt, 2011, S. 193 f.)

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \dots + \beta_p x_{p,i} + \varepsilon_i \quad 2.14$$

Wenn nicht ein einzelner Versuch betrachtet wird, sondern eine komplette Versuchsreihe, dann ergibt sich für N in der Versuchsreihe durchgeführte Versuche das Gleichungssystem 2.15.

²⁵ Vgl. Abschnitt 2.5.5

²⁶ Siehe Kapitel 5

$$\begin{aligned}
 y_1 &= \beta_0 + \beta_1 x_{1,1} + \beta_2 x_{2,1} + \dots + \beta_p x_{p,1} + \varepsilon_1 \\
 y_2 &= \beta_0 + \beta_1 x_{1,2} + \beta_2 x_{2,2} + \dots + \beta_p x_{p,2} + \varepsilon_2 \\
 &\vdots \\
 y_N &= \beta_0 + \beta_1 x_{1,N} + \beta_2 x_{2,N} + \dots + \beta_p x_{p,N} + \varepsilon_N
 \end{aligned}
 \tag{2.15}$$

In Matrixschreibweise ist es in Gleichung 2.16 dargestellt. Darin wird \mathbf{X} als *Designmatrix* bezeichnet (Czado & Schmidt, 2011, S. 195). Sie beinhaltet alle Faktoreinstellungen des Versuchsplans.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{1,1} & x_{2,1} & \dots & x_{p,1} \\ 1 & x_{1,2} & x_{2,2} & \dots & x_{p,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1,N} & x_{2,N} & \dots & x_{p,N} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix}
 \tag{2.16}$$

Bei der Auswertung einer Versuchsreihe sind die Einstellungen der untersuchten Faktoren für jeden Versuch bekannt. Sie werden also als Zahlenwerte in \mathbf{X} eingesetzt. Das Gleichungssystem zur Bestimmung von $\boldsymbol{\beta}$ bleibt also linear, auch wenn nichtlineare Effekte in der Designmatrix enthalten sind. Aus diesem Grund bezieht sich der Begriff *lineares* Modell ausschließlich auf die Regressionskoeffizienten und nicht auf die Faktoren dieser Gleichung. Als Faktoren können auch transformierte Größen, wie quadratische Effekte oder Wechselwirkungen, eingesetzt werden (z.B. $x_3 = x_1 x_2$ oder $x_3 = x_1^2$) um ein mehrdimensionales Polynom als Modellgleichung verwenden zu können. Transformierte Größen werden hier als eigenständige Faktoren behandelt. Somit sind alle Modellgleichungen der in Abschnitt 2.5 beschriebenen Versuchspläne als lineare Modelle aufzufassen. Faktortransformationen können zudem dazu genutzt werden, nichtlineare Einflüsse auf das zu modellierende Verhalten zu linearisieren. Wenn also bekannt ist, dass der Faktor x_3 einen exponentiellen Einfluss auf die Zielgröße besitzt, dann kann stattdessen eine Transformation $x_3^* = \ln(x_3)$ als Faktor eingesetzt werden.

Kategoriale Einflussfaktoren können im klassischen linearen Modell und ebenso in den später beschriebenen linearen Modelltypen abgebildet werden. Dazu wird eine Ausprägung des kategorialen Faktors als Referenz festgelegt. Für jede andere abzubildende Ausprägung wird jeweils eine sogenannte *Dummyvariable* eingesetzt. Diese bekommt für Versuche mit der zugehörigen Ausprägung des Faktors den Wert eins zugewiesen und für alle anderen Versuche den Wert null. (Wentura & Pospeschill, 2015, S. 82 f.)

Das in den Gleichungen 2.14 bis 2.16 dargestellte lineare Modell heißt klassisches lineares Modell, wenn folgende Annahmen gelten (Fahrmeir et al., 2009, S. 61 f.):

- Die Designmatrix X besitzt vollen Spaltenrang. Es lässt sich also keine Spalte von X als Linearkombination anderer Spalten darstellen. Dies bedeutet, dass keine Kollinearität im Versuchsplan auftritt.²⁷ Die in Abschnitt 2.5 für die verschiedenen Versuchspläne angegebenen Modellgleichungen sind alle so gewählt, dass diese Annahme zutrifft.
- Der Erwartungswert der Modellfehler ist null. Das heißt, der Mittelwert aller Fehler über unendlich viele Versuche hinweg ist null.

$$E(\varepsilon) = 0 \qquad 2.17$$

- Die Kovarianzmatrix Σ der Residuen ε entspricht dem Produkt einer konstanten Varianz σ^2 mit der Einheitsmatrix I . Dies bedeutet einerseits, dass die Fehler *homoskedastisch* sind. Ihre Varianz ist also für alle Versuche konstant. Zum anderen ist die Kovarianz zwischen verschiedenen Versuchen immer null. Alle Werte von ε sind also vollkommen unabhängig voneinander und es kann beispielsweise nicht aus dem Wert ε_i eines Versuches auf den Wert ε_{i+1} des nachfolgenden Versuches geschlossen werden.

$$Cov(\varepsilon) = \Sigma = E(\varepsilon\varepsilon') = \sigma^2 I \qquad 2.18$$

Das klassische lineare Modell ist sehr restriktiv bezüglich des Fehlerterms ε . Dadurch ist die Bestimmung der Regressionskoeffizienten und die anschließende statistische Bewertung im Vergleich zu den nachfolgend vorgestellten Modellen einfach. Zur Auswertung kann die gewöhnliche Methode der kleinsten Fehlerquadrate eingesetzt werden.

Nachfolgend wird zunächst in Abschnitt 2.5.2.1 vorgestellt, wie die Koeffizienten des klassischen linearen Modells ermittelt werden können. Anschließend wird in Abschnitt 2.5.2.2 beschrieben, wie das Modell und dessen Vorhersagegenauigkeit statistisch bewertet werden können. In Abschnitt 2.5.2.3 wird dann beleuchtet, wie die Einhaltung der Annahmen des klassischen linearen Modells überprüft werden kann.

²⁷ Vgl. S. 35

2.5.2.1 Parameterschätzung mit der gewöhnlichen Methode der kleinsten Fehlerquadrate

Bei der *gewöhnlichen Methode der kleinsten Fehlerquadrate (OLS)*²⁸ werden die Schätzwerte $\hat{\beta}$ für die Modellkoeffizienten β so bestimmt, dass die Summe der Quadrate der im Gleichungssystem enthaltenen Residuen minimiert wird. Dazu ist kein Optimierungsalgorithmus notwendig. Stattdessen kann die Summe der Fehlerquadrate nach Gleichung 2.19 ausmultipliziert und anschließend nach β abgeleitet werden. Dadurch ergibt sich Gleichung 2.20. (Fahrmeir et al., 2009, S. 91)

$$\varepsilon' \varepsilon = (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta) \quad 2.19$$

$$\frac{\partial \varepsilon' \varepsilon}{\partial \beta} = -2\mathbf{X}'\mathbf{y} + 2(\mathbf{X}'\mathbf{X})\beta \quad 2.20$$

Anschließendes Nullsetzen dieser Ableitung führt zu Gleichung 2.21, da $\mathbf{X}'\mathbf{X}$ positiv definit und damit invertierbar ist (Czado & Schmidt, 2011, S. 203; Fahrmeir et al., 2009, S. 92). Mittels dieser Gleichung lassen sich Schätzwerte für die Koeffizienten $\hat{\beta}_{OLS}$ analytisch bestimmen.

$$\hat{\beta}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad 2.21$$

Wenn die so ermittelten Koeffizienten in Gleichung 2.22 eingesetzt werden, erhält man einen Schätzwert \hat{y}_0 für das Systemverhalten mit den Faktoreinstellungen x_0 .

$$\hat{y}_0 = x_0 \hat{\beta}_{OLS} \quad 2.22$$

$\hat{\beta}_{OLS}$ ist *erwartungstreu*²⁹, unabhängig davon, ob die Störgröße ε normalverteilt ist oder einer anderen Verteilungsklasse angehört. Auch wenn die Bedingung $Cov(\varepsilon) = \sigma^2 \mathbf{I}$ verletzt wird, bleibt $\hat{\beta}_{OLS}$ erwartungstreu. Dann wird die Schätzung allerdings ineffizient. (Fahrmeir et al., 2009, S. 101, 124)

²⁸ Aus dem Englischen für *ordinary least squares*

²⁹ Das heißt $\hat{\beta}_{OLS} \rightarrow \beta$ für $N \rightarrow \infty$

2.5.2.2 Statistische Bewertung

Um keine falschen Schlüsse aus dem ermittelten Modell zu ziehen, ist eine statistische Bewertung des Modells notwendig. Die wichtigsten Bewertungsgrößen und -verfahren werden nachfolgend kurz vorgestellt.

2.5.2.2.1 Bewertung des Modells

Eine anschauliche Möglichkeit, die Güte der Anpassung der Modellgleichung an die Versuchsdaten zu beurteilen, stellt das *Bestimmtheitsmaß* R^2 dar. Es wird gemäß Gleichung 2.23 berechnet. (Kleppmann, 2011, S. 178)

$$R^2 = \frac{Q_{Regression}}{Q_{Gesamt}} \quad \text{mit } Q_{Regression} = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2, \quad Q_{Gesamt} = \sum_{i=1}^N (y_i - \bar{y})^2 \quad 2.23$$

Darin beschreibt Q_{Gesamt} die Summe der quadratischen Abweichungen der einzelnen Messwerte y_i von ihrem Durchschnitt \bar{y} . $Q_{Regression}$ bezeichnet die Summe der quadratischen Abweichungen der Regressionswerte \hat{y}_i vom Durchschnitt. Somit kann das Bestimmtheitsmaß R^2 als derjenige Anteil der gesamten Varianz der Messwerte gesehen werden, der durch die Regressionsgleichung erklärt wird. R^2 kann Werte zwischen null und eins annehmen. Ein Wert nahe eins deutet dabei auf eine gute Anpassung der Modellgleichung an das gemessene Verhalten hin. Werte, die stark von eins abweichen, können dagegen mehrere Ursachen haben. Diese sind (Kleppmann, 2011, S. 180):

- Lack-of-fit³⁰
- eine geringe Abhängigkeit der Zielgröße von den untersuchten Faktoren
- eine große Zufallsstreuung in den Versuchsergebnissen.

Während das Bestimmtheitsmaß sehr anschaulich ist, eignet es sich weniger gut für den Vergleich zwischen verschiedenen Modellen. Dies liegt daran, dass R^2 bei der Aufnahme neuer Einflussgrößen in ein Modell stets zunimmt, auch wenn deren Effekte statistisch nicht signifikant sind. Für einen Modellvergleich sind Bewertungskriterien wie das korrigierte Bestimmtheitsmaß \bar{R}^2 , das Akaike Informationskriterium AIC oder das Bayesianische Informationskriterium BIC besser geeignet.³¹

³⁰ Siehe S. 28

³¹ Für Details siehe (Fahrmeir et al., 2009, S. 159 ff.)

2.5.2.2 Bewertung des Einflusses der einzelnen Faktoren

Zusätzlich zur Bewertung des gesamten Modells ist meist auch eine Bewertung des Einflusses der einzelnen Faktoren von Interesse. Diese wird meist in Tabellenform angegeben. Ein Beispiel hierfür ist in Tabelle 2.1 dargestellt. Auch Softwarepakete für statistische Auswertungen geben ihre Ergebnisse oft in dieser Form aus.

Tabelle 2.1: Beispielhafte statistische Auswertung einer Versuchsreihe in Tabellenform

Variable	Koeffizient	Std.-Abw.	t-Wert	p-Wert	95% Konfidenzintervall	
Konstante	15,21	0,12	126,75	< 0,001	14,97	15,45
Temperatur	6,95	0,14	49,64	< 0,001	6,67	7,23
Druck	0,09	0,14	0,64	0,311	-0,19	0,37
Geschwindigkeit	-2,19	0,14	15,64	< 0,001	1,91	2,47

In dieser Tabelle werden nach den Variablennamen die zugehörigen, gemäß Abschnitt 2.5.2.1 geschätzten Koeffizienten $\hat{\beta}$ angegeben. Um aus den Werten dieser Koeffizienten die Effektgröße des zugehörigen Faktors einfach ablesen zu können, werden bei der Regression üblicherweise normierte Faktoreinstellungen verwendet. Das heißt, dass jeder Faktor in seiner maximalen Einstellung den normierten Wert 1 und in seiner minimalen Einstellung den normierten Wert -1 zugewiesen bekommt. Dadurch beschreibt nach der Regression der Koeffizient jedes Faktors den maximalen Effekt, den dieser Faktor während der Versuchsreihe hatte. Und zwar unabhängig davon, welche Einheit die Einstellungen des Faktors gewählt wurden. Im Beispiel aus Tabelle 2.1 führte also die größte untersuchte Einstellung der Temperatur zu einer 6,95 erhöhten Zielgröße, während die niedrigste Einstellung zu einer um 6,95 verringerten Zielgröße führte. Entsprechend führte die maximale Einstellung der Geschwindigkeit zu einer um 2,19 verringerten Zielgröße und die minimale Einstellung zu einer um 2,19 erhöhten Zielgröße. Nach diesem Schema lassen sich die Einflüsse aller Faktoren intuitiv und einfach erfassen und vergleichen.

In der auf die Koeffizienten folgenden Spalte stehen Schätzwerte für die Standardabweichungen $\hat{\sigma}_{\hat{\beta}}$ dieser Koeffizienten. Sie sind ein Maß dafür, zu welchem Betrag der zugehörige Koeffizient potentiell durch Störgrößen und Messungenauigkeiten verfälscht ist. Sie lassen sich als die Quadratwurzel der diagonalen Elemente der Kovarianzmatrix für die Koeffizienten ermitteln (Kowalski et al., 2007, S. 13). Die Kovarianzmatrix für die Koeffizienten kann für klassische lineare Modelle nach Gleichung 2.24 berechnet werden (Fahrmeir et al., 2009, S. 101).

$$\text{Cov}(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \quad 2.24$$

Darin ist σ^2 die Varianz der Störgrößen. Für Versuchspläne, in denen ein oder mehrere Versuche mit identischen Einstellungen wiederholt wurden, kann $\hat{\sigma}^2$ direkt als die Varianz der Ergebnisse dieser Versuche nach Gleichung 2.1 bestimmt werden (Kowalski et al., 2007, S. 12). Ansonsten kann σ^2 für klassische lineare Modelle nach Gleichung 2.25 geschätzt werden, in der p die Anzahl der in β enthaltenen Elemente beschreibt (Fahrmeir et al., 2009, S. 102).³²

$$\hat{\sigma}^2 = \frac{1}{N - p} \hat{\varepsilon}'\hat{\varepsilon} \quad 2.25$$

$\hat{\varepsilon}$ ist ein Schätzwert für den Modellfehler ε und wird *Residuum* genannt. $\hat{\varepsilon}$ wird durch Einsetzen der Schätzwerte $\hat{\beta}$ in die Modellgleichung 2.16 und anschließendes Auflösen nach $\hat{\varepsilon}$ gemäß Gleichung 2.26 geschätzt.

$$\hat{\varepsilon} = \mathbf{X}\hat{\beta} - \mathbf{y} \quad 2.26$$

Damit Gleichung 2.25 lösbar ist, muss der Versuchsumfang N größer sein, als die Anzahl p der Koeffizienten, die in der Modellgleichung zu bestimmen sind. Der Term $N - p$ beschreibt die zur Bestimmung der Fehlervarianz übrigen Freiheitsgrade.

Wenn $\hat{\sigma}^2$ gemäß Gleichung 2.1 aus Versuchswiederholungen bestimmt wird, beinhaltet es den reinen Einfluss von Störgrößen und Messfehlern („reiner Fehler“). In Gleichung 2.25 wird $\hat{\sigma}^2$ aus den Residuen $\hat{\varepsilon}$ bestimmt. Residuen entstehen allerdings nicht nur durch Störgrößen und Messfehler, sondern auch durch Lack-of-fit. Dadurch ist der durch Gleichung 2.25 bestimmte Wert von $\hat{\sigma}^2$ größer als der Wert gemäß Gleichung 2.1.

Der *t-Wert* in Tabelle 2.1 ist der Quotient aus dem Schätzwert $\hat{\beta}_j$ für den jeweiligen Koeffizienten und seiner geschätzten Standardabweichung $\hat{\sigma}_{\beta_j}$ (Fahrmeir et al., 2009, S. 116).³³ Je größer sein Betrag ist, desto sicherer kann von einem real existierenden Effekt ausgegangen werden.

³² Vgl. Abschnitt 2.4.2.2

³³ siehe Gleichung 2.27

$$t_{\text{statistic}} = \frac{\hat{\beta}_j}{\hat{\sigma}_{\beta_j}} \quad 2.27$$

Ein besser interpretierbares Maß dafür, wie sicher ein ermittelter Effekt real existiert, liefert der *p*-Wert. Er beschreibt die Wahrscheinlichkeit, dass für einen Koeffizienten ein Betrag von mindestens der ermittelten Größe erreicht wird, obwohl die von ihm beschriebene Abhängigkeit real nicht existiert (Kleppmann, 2011, S. 310). Berechnet werden kann der *p*-Wert gemäß Gleichung 2.28, in der t_{N-p} die kumulative Verteilungsfunktion der Studentischen *t*-Verteilung für $N - p$ Freiheitsgrade darstellt.

$$p_{\text{statistic}} = 2 \cdot t_{N-p}(-|t_{\text{statistic}}|) \quad 2.28$$

Manchmal sind noch weitere statistische Kenngrößen für die Effekte der Faktoren in der Versuchsauswertung enthalten. In Tabelle 2.1 ist beispielsweise der Vertrauensbereich CI_{β_j} für ein Vertrauensniveau (siehe Abschnitt 2.1) von 95 % enthalten. Er wird hier als „95 % Konfidenzintervall“ bezeichnet. Der Erwartungswert für den zugehörigen Koeffizienten ist mit 95-prozentiger Wahrscheinlichkeit in diesem Intervall enthalten. Das Konfidenzintervall CI_{β_j} wird nach Gleichung 2.29 berechnet (Fahrmeir et al., 2009, S. 121). In dieser Gleichung ist $t_{N-p}^{-1}\left(1 - \frac{\alpha}{2}\right)$ der Wert der Quantilfunktion (Inverse der kumulativen Verteilungsfunktion) der Studentischen *t*-Verteilung an der Stelle $1 - \frac{\alpha}{2}$ für $N - p$ Freiheitsgrade.³⁴

$$CI_{\beta_j} = \left[\hat{\beta}_j - t_{N-p}^{-1}\left(1 - \frac{\alpha}{2}\right) \hat{\sigma}_{\beta_j}; \hat{\beta}_j + t_{N-p}\left(1 - \frac{\alpha}{2}\right) \hat{\sigma}_{\beta_j} \right] \quad 2.29$$

2.5.2.2.3 Ermittlung der Vorhersagegenauigkeit des Modells

Zur Vorhersagegenauigkeit des Modells gibt es zwei relevante Größen: das Konfidenzintervall und das Prognoseintervall. Das Konfidenzintervall CI_{y_0} beschreibt den Bereich, innerhalb dessen sich der Erwartungswert $E(y_0)$ der Zielgröße für die Faktoreinstellungskombination x_0 , mit dem gewählten Vertrauensniveau befindet. Das Konfidenzintervall beschreibt also die Unsicherheit der Schätzung des Erwartungswertes.

³⁴ Für Details siehe (Czado & Schmidt, 2011, S. 13 f.)

twerts von y_0 aufgrund der Ungenauigkeit der Schätzung von $\hat{\beta}$. Das Konfidenzintervall CI_{y_0} kann gemäß Gleichung 2.30 berechnet werden. Darin ist $\hat{\sigma}_{x_0}$ der Schätzwert für die Varianz der Schätzung der Zielgröße an der Stelle x_0 . (Fahrmeir et al., 2009, S. 123)

$$CI_{y_0} = \left[\mathbf{x}_0' \hat{\beta} - t_{N-p}^{-1} \left(1 - \frac{\alpha}{2} \right) \hat{\sigma}_{x_0}; \mathbf{x}_0' \hat{\beta} + t_{N-p}^{-1} \left(1 - \frac{\alpha}{2} \right) \hat{\sigma}_{x_0} \right] \quad 2.30$$

mit $\hat{\sigma}_{x_0} = \hat{\sigma} (\mathbf{x}_0' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_0)^{\frac{1}{2}}$

Das Prognoseintervall PI_{y_0} beschreibt das Intervall, innerhalb dessen sich ein gemessenes y_0 mit dem gewählten Vertrauensniveau befindet. In das Prognoseintervall fließt neben der Ungenauigkeit der Schätzung von $\hat{\beta}$ (und damit der Unsicherheit σ_{x_0} der Schätzung des Erwartungswertes der Zielgröße y an der Stelle x_0) auch die Streuung σ der Fehler ε mit ein. Es kann nach Gleichung 2.31 berechnet werden (Fahrmeir et al., 2009, S. 123).

$$PI_{y_0} = \left[\mathbf{x}_0' \hat{\beta} - t_{N-p}^{-1} \left(1 - \frac{\alpha}{2} \right) \sqrt{\hat{\sigma}_{x_0}^2 + \hat{\sigma}^2}; \mathbf{x}_0' \hat{\beta} + t_{N-p}^{-1} \left(1 - \frac{\alpha}{2} \right) \sqrt{\hat{\sigma}_{x_0}^2 + \hat{\sigma}^2} \right] \quad 2.31$$

2.5.2.3 Überprüfung der Modellannahmen

Die vorhergehend beschriebene statistische Bewertung ist nur gültig, wenn die Annahmen des klassischen linearen Modells³⁵ und zusätzlich die Annahme normalverteilter Fehler erfüllt sind.

Die Annahme, dass die Designmatrix \mathbf{X} vollen Spaltenrang besitzt, wird von jedem in Abschnitt 2.5 vorgestellten Versuchsplan mit der dazugehörigen Form der angegebenen Modellgleichung erfüllt. Falls ein anderer Versuchsplan eingesetzt wird, der diese Annahme nicht erfüllt, schlägt bereits die Bestimmung von $\hat{\beta}_{OLS}$ mittels Gleichung 2.21 fehl, da $\mathbf{X}' \mathbf{X}$ dann nicht invertierbar ist.

Die Annahme, dass der Störgrößeneinfluss im Mittel null beträgt, lässt sich bei der Auswertung der Versuchsreihe nicht nachweisen oder widerlegen. Dies ist allerdings auch nicht notwendig. Man muss sich lediglich bewusst machen, dass das

³⁵ Siehe S. 50

erstellte Modell und alle damit verbundenen Aussagen mit dieser Annahme verknüpft sind. Das erstellte Modell ist also nur für Umgebungsbedingungen gültig, bei denen die Störgrößen denselben Effekt besitzen, den sie auch bei der Versuchsreihe hatten, auf die das Modell angepasst wurde.

Zu überprüfen bleiben also die Annahmen der Homoskedastizität, Unabhängigkeit und Normalverteilung der Fehler ε . Da die realen Fehler ε nicht ermittelbar sind, wird die Überprüfung stattdessen anhand der Residuen $\hat{\varepsilon}$ durchgeführt. Dazu kommen Residuendiagramme zum Einsatz.

Die Homoskedastizität der Fehler kann mit Hilfe der Diagramme aus den nachfolgenden drei Abbildungen überprüft werden. In Abbildung 2.21 sind die Residuen über den Wert der geschätzten Zielgröße \hat{y} aufgetragen. Bei homoskedastischen Systemen sollte hier keine Abhängigkeit der Streubreite des Residuums von der Zielgröße erkennbar sein. Auch von der Versuchsreihenfolge sollte die Streubreite der Residuen unabhängig sein. Dies lässt sich mit den Residuendiagrammen aus Abbildung 2.22 überprüfen. Zudem sollte die Streubreite der Residuen auch unabhängig von der Einstellung der Faktoren sein (siehe Abbildung 2.23).

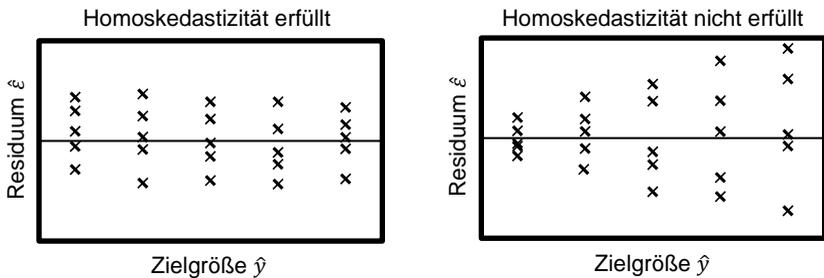


Abbildung 2.21: Residuendiagramme: Residuen über der Zielgröße
Vgl. (Kleppmann, 2011, Bild 10-6)

Mittels der Residuendiagramme aus Abbildung 2.22 lässt sich neben der Homoskedastizität der Residuen auch deren Unabhängigkeit von der Versuchsreihenfolge zeigen. Wenn sich hier ein Trend zeigt (wie im Diagramm rechts), dann hat sich der Einfluss einer Störgröße im Laufe der Versuchsreihe langsam geändert.

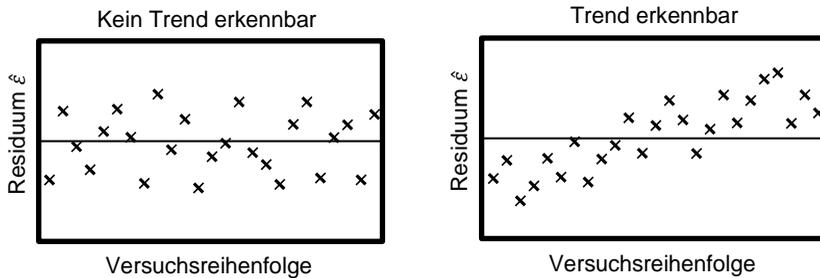


Abbildung 2.22: Residuendiagramme: Residuen über der Versuchsreihenfolge
Vgl. (Kleppmann, 2011, Bild 10-5)

Die Residuendiagramme aus Abbildung 2.23 erlauben neben des Nachweises der Homoskedastizität der Residuen auch die Überprüfung ihrer Unabhängigkeit von der Einstellung des im Diagramm enthaltenen Faktors. Wenn sich hier eine klare Abhängigkeit der Residuen von der Faktoreinstellung zeigt, kann dies indizieren, dass die Modellgleichung nicht zum modellierten Systemverhalten passt. Diese Diagramme können also als Lack-of-fit-Test eingesetzt werden.

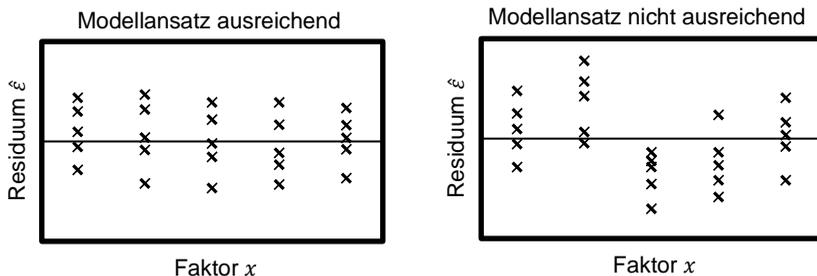


Abbildung 2.23: Residuendiagramme: Residuen über einem Faktor
Vgl. (Kleppmann, 2011, Bild 10-4)

Zum Nachweis der Unabhängigkeit der Residuen muss auch gezeigt werden, dass ein Versuch die Ergebnisse des nächsten Versuches nicht beeinflusst. Dazu werden

in den Diagrammen aus Abbildung 2.24 die Residuen aller Versuche über den Residuen der Vorgängerversuche aufgetragen. Wenn sich hier ein klarer Zusammenhang zeigt, ist die Annahme der Unabhängigkeit der Residuen widerlegt.

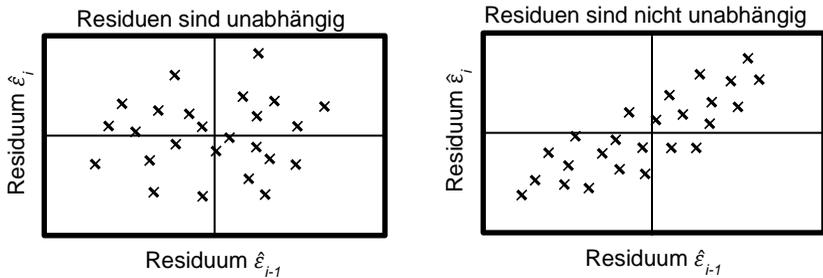


Abbildung 2.24: Residuendiagramme: Residuen über verzögerte Residuen
Vgl. (NIST/SEMATECH, 2012, Kap. 4.4.4.4)

Ob die Annahme normalverteilter Residuen erfüllt ist, kann mit den Diagrammen aus Abbildung 2.25 überprüft werden. Je näher die eingetragenen Punkte der eingezeichneten Gerade folgen, desto genauer entspricht die Verteilung der Residuen einer Normalverteilung.

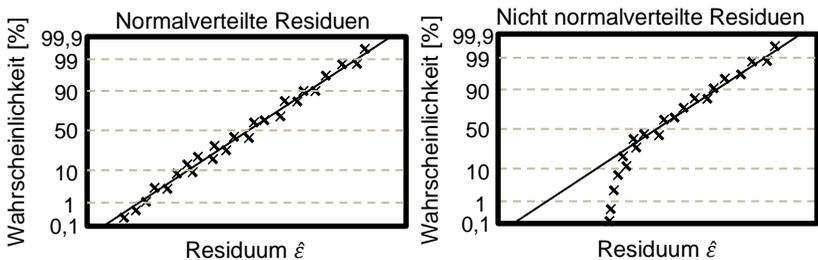


Abbildung 2.25: Residuendiagramme: Residuen im Wahrscheinlichkeitsnetz
Vgl. (Kleppmann, 2011, S.70)

2.5.3 Das allgemeine lineare Modell

Die Modellgleichung des allgemeinen linearen Modells entspricht dem klassischen linearen Modell aus Gleichung 2.16. Es unterscheidet sich lediglich in der Annahme über die Kovarianzmatrix der Residuen. Für das allgemeine lineare Modell gilt für $Cov(\boldsymbol{\varepsilon})$ Gleichung 2.32, in der $\boldsymbol{\Omega}$ eine beliebige positiv definite Matrix sein kann. Im Unterschied zum klassischen linearen Modell sind also heteroskedastische Residuen erlaubt, deren Varianzen nicht für alle Versuche identisch sind. Zudem dürfen auch die Kovarianzen zwischen zwei Versuchen ungleich null sein.

$$Cov(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Sigma} = \sigma^2\boldsymbol{\Omega} \quad 2.32$$

Im Vergleich zum klassischen linearen Modell besitzt das allgemeine lineare Modell somit eine deutlich weniger restriktive Annahme über den Aufbau der Kovarianzmatrix der Residuen. Dadurch kann das allgemeine lineare Modell auch angewendet werden, wenn über den Verlauf einer Versuchsreihe die Messgenauigkeit oder der Einfluss von Störgrößen nicht als konstant betrachtet werden kann.

Nachfolgend wird zunächst erklärt, wie die Koeffizienten des allgemeinen linearen Modells ermittelt werden können. Hierzu existieren zwei verschiedene Verfahren, die in den Abschnitten 2.5.3.1 und 2.5.3.2 vorgestellt werden. Anschließend wird in Abschnitt 2.5.3.3 beschrieben, wie das Modell und dessen Vorhersagegenauigkeit statistisch bewertet werden können. In Abschnitt 2.5.3.4 wird dann beleuchtet, wie die Einhaltung der Annahmen des allgemeinen linearen Modells überprüft werden kann.

2.5.3.1 Parameterschätzung mit der generalisierten Methode der kleinsten Fehlerquadrate

Wenn die Fehler $\boldsymbol{\varepsilon}$ heteroskedastisch oder korreliert sind, dann liefert die gewöhnliche Methode der kleinsten Fehlerquadrate (OLS)³⁶ zwar einen erwartungstreuen Schätzwert für die Koeffizienten, ist allerdings nicht effizient. Dies kommt daher, dass bei OLS die Summe der Fehlerquadrate ohne jegliche Gewichtung minimiert wird. Wenn allerdings die Fehlervarianz nicht für alle Versuche gleich groß ist, dann ist bei Beobachtungen mit hoher Fehlervarianz eine genaue Anpassung nicht so wertvoll wie bei Beobachtungen mit niedriger Fehlervarianz. Eine Methode zur

³⁶ Aus dem Englischen für *ordinary least squares*

Schätzung von β , die dies berücksichtigt, ist die gewichtete (oder auch: generalisierte) Methode der kleinsten Fehlerquadrate (GLS)³⁷. Bei GLS werden gemäß Gleichung 2.33 alle Beobachtungen durch die Varianz des zugehörigen Fehlerterms ε_i geteilt und so gewichtet.³⁸ Bei großer Fehlervarianz bekommt das zu minimierende Fehlerquadrat so eine kleine Gewichtung. (Fahrmeir et al., 2009, Kap. 3.4.2)

$$\hat{\beta}_{GLS} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y \quad 2.33$$

Die Schwierigkeit bei der Bestimmung von $\hat{\beta}_{GLS}$ nach dieser Gleichung ist, dass Ω bekannt sein muss. Falls Ω nicht bekannt ist, muss es ebenfalls geschätzt werden. Dazu existiert eine Vielzahl an Verfahren, von denen einige z.B. von Greene (2008) vorgestellt werden. Ein sehr häufig eingesetztes Verfahren ist die Maximum-Likelihood-Methode, mit der β , Ω und σ^2 gemeinsam geschätzt werden können.

2.5.3.2 Parameterschätzung mit der Maximum-Likelihood-Methode

Beim Maximum-Likelihood-Verfahren (ML-Verfahren) wird zunächst eine Likelihoodfunktion definiert, die die Wahrscheinlichkeit beschreibt, dass die in der Versuchsreihe beobachteten Werte der Zielgröße y für gegebene Werte der Regressionskoeffizienten β und der Residuenvarianz σ^2 erzielt werden. Unter der Annahme normalverteilter Störgrößen ist die Likelihoodfunktion L für das allgemeine lineare Modell gegeben durch Gleichung 2.34, in der N der Versuchsumfang ist.

$$L(\beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{N/2} |\Omega|^{1/2}} \exp\left(-\frac{1}{2\sigma^2} (y - X\beta)' \Omega^{-1} (y - X\beta)\right) \quad 2.34$$

Es sind nun die Parameter β , Ω und σ^2 gesucht, die diese Likelihoodfunktion maximieren, also die Wahrscheinlichkeit der beobachteten Werte der Zielgröße maximieren. Zunächst wird dazu die Likelihoodfunktion logarithmiert um die Suche nach einem Maximum mathematisch einfacher zu gestalten. So werden aus den Produkten, die in der Likelihoodfunktion enthalten sind, Summen. Da die Logarithmusfunktion für positive Werte streng monoton zunehmend ist und die Likelihoodfunktion stets positive Werte liefert, entspricht das Maximum der Likelihoodfunktion L auch dem Maximum der Log-Likelihoodfunktion l nach Gleichung 2.35.

³⁷ Aus dem Englischen für *generalized least squares*

³⁸ Siehe Gleichung 2.32 für den Zusammenhang zwischen ε und Ω

$$l(\boldsymbol{\beta}, \sigma^2) = -\frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\sigma^2) - \frac{1}{2} \log|\boldsymbol{\Omega}| - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Omega}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad 2.35$$

Da im Maximum von l die Ableitungen von l nach $\boldsymbol{\beta}$ und σ^2 null sein müssen, können $\boldsymbol{\beta}$ und σ^2 durch Ableiten von l nach diesen Größen und anschließendes Nullsetzen bestimmt werden. Hierzu muss allerdings $\boldsymbol{\Omega}$ bekannt sein. Die Ableitung von l nach $\boldsymbol{\beta}$ führt zu Gleichung 2.36. Die Ableitung von l nach σ^2 führt zu Gleichung 2.37.

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \frac{1}{\sigma^2} \mathbf{X}' \boldsymbol{\Omega}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad 2.36$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{2}{\sigma^4} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Omega}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad 2.37$$

Nullsetzen und Auflösen nach $\boldsymbol{\beta}$ bzw. σ^2 führt zu den ML-Schätzern $\hat{\boldsymbol{\beta}}_{ML}$ und $\hat{\sigma}_{ML}^2$ gemäß den Gleichungen 2.38 und 2.39 (Greene, 2008, S. 522).

$$\hat{\boldsymbol{\beta}}_{ML} = (\mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{y} \quad 2.38$$

$$\hat{\sigma}_{ML}^2 = \frac{\hat{\boldsymbol{\varepsilon}}' \boldsymbol{\Omega}^{-1} \hat{\boldsymbol{\varepsilon}}}{N} \quad 2.39$$

Der so entstehende ML-Schätzer $\hat{\boldsymbol{\beta}}_{ML}$ entspricht für normalverteilte Störgrößen dem GLS-Schätzer $\hat{\boldsymbol{\beta}}_{GLS}$ nach Gleichung 2.33. Der Schätzer $\hat{\sigma}_{ML}^2$ ist vor allem für kleine Versuchsumfänge n verzerrt und wird deshalb selten genutzt. Häufiger wird auf eine Variante der Maximum-Likelihood-Schätzung zurückgegriffen, die Restricted-Maximum-Likelihood (REML) genannt wird. Der REML-Schätzer $\hat{\sigma}_{REML}^2$ für σ^2 ist in Gleichung 2.40 angegeben, wobei p die Anzahl der Elemente von $\boldsymbol{\beta}$ ist (Fahrmeir et al., 2009, S. 127).³⁹

$$\hat{\sigma}_{REML}^2 = \frac{\hat{\boldsymbol{\varepsilon}}' \boldsymbol{\Omega}^{-1} \hat{\boldsymbol{\varepsilon}}}{N - p} \quad 2.40$$

³⁹, Für Details zu REML siehe (Verbeke & Molenberghs, 2001, S. 43 ff.)

Falls Ω nicht bekannt ist, muss es zusammen mit β und σ^2 geschätzt werden. Da hierzu im Allgemeinen keine analytische Lösung mehr existiert, kommen stattdessen meist iterative Optimierungsverfahren zur Maximierung von Gleichung 2.35 zum Einsatz. In den meisten Fällen schließen diese Verfahren Annahmen über den Aufbau (aber nicht über die enthaltenen Werte) von Ω mit ein, um die Anzahl der zu schätzenden Parameter klein zu halten. (Greene, 2008, S. 156 ff.)

2.5.3.3 Statistische Bewertung

Die statistische Bewertung allgemeiner linearer Modelle läuft analog zur in Abschnitt 2.5.2.2 beschriebenen Bewertung klassischer linearer Modelle ab. Lediglich die angepasste Kovarianz von ε muss berücksichtigt werden. So wird zur Schätzung von σ^2 der REML-Schätzer nach Gleichung 2.40 statt Gleichung 2.25 eingesetzt. Die Kovarianzmatrix $Cov(\hat{\beta})$ wird bestimmt nach Gleichung 2.41, in der σ^2 nach Gleichung 2.40 geschätzt werden sollte. (Fahrmeir et al., 2009, S. 127)

$$Cov(\hat{\beta}) = \sigma^2(X'\Omega X)^{-1} \quad 2.41$$

Auch für die Ermittlung der Vorhersagegenauigkeit des Modells muss die geänderte Form von Σ berücksichtigt werden. Wenn das allgemeine lineare Modell als marginales gemischtes lineares Modell betrachtet wird (siehe Abschnitt 2.5.4.1, vor allem Gleichung 2.45), kann zur Ermittlung der Vertrauens- und Vorhersagewerte die Vorgehensweise aus Abschnitt 2.5.4.2 verwendet werden. Eine ausführlichere und allgemeinere Betrachtung gibt Harville (1992, 2008).

2.5.3.4 Überprüfung der Modellannahmen

Zur Überprüfung der Modellannahmen des allgemeinen linearen Modells werden dieselben Methoden wie für das klassische lineare Modell verwendet (siehe Abschnitt 2.5.2.3). Die Residuen sollten allerdings aufgrund ihrer Heteroskedastizität vor der Darstellung in Residuendiagrammen in sogenannte *Pearson-Residuen* transformiert werden. Um auch die Normalverteilungsannahme überprüfen zu können, sind weitere Transformationen erforderlich. (Galecki, 2013, S. 136, 137, 190, 191)

2.5.4 Das gemischte lineare Modell

Das gemischte lineare Modell unterscheidet sich von den bisher vorgestellten Modelltypen dadurch, dass es neben den festen Effekten auch zufällige Effekte enthält. Zufällige Effekte beschreiben den Einfluss von Faktoren, deren Einstellung bei der Versuchsdurchführung nicht frei gewählt und oft auch nicht ermittelt werden kann. Die Einstellung dieser Faktoren ist also zufällig. Damit wird auch die Wirkung dieser Faktoren auf die Zielgröße als zufällige Schwankung sichtbar. Im Gegensatz zu festen Effekten wird für zufällige Effekte auch kein fester Regressionskoeffizient bestimmt. Stattdessen wird meist angenommen, dass der Regressionskoeffizient normalverteilt mit Mittelwert null streut (Galecki, 2013, S. 248). Zu dieser Normalverteilung wird dann die Standardabweichung bestimmt, die als Maß für den Einfluss des zufälligen Effekts verwendet wird.

Oft beschreiben zufällige Effekte den Einfluss von *Clustern*, die die Versuche einer Versuchsreihe in mehrere Gruppen einteilen. In solchen Fällen werden gemischte lineare Modelle auch *hierarchische lineare Modelle* oder *Mehrebenenmodelle* genannt. Häufig treten solche Cluster in Längs- und Querschnittstudien auf. In Längsschnittstudien werden mehrere Individuen wiederholt untersucht. Ein Beispiel hierfür ist die Bestimmung der Wirkung eines Medikaments. Das Medikament wird mehreren Personen verabreicht und die Wirkung für jeden Patienten im zeitlichen Verlauf bestimmt. Jeder Patient besitzt allerdings eigene Einflussgrößen (z.B. Zustand des Immunsystems) auf den zeitlichen Verlauf der Krankheit, die nicht frei festgelegt und auch nicht vollständig ermittelt werden können. Die Effekte dieser Einflussgrößen werden im gemischten linearen Modell als zufällige Effekte abgebildet.

In Querschnittstudien wird ein repräsentativer Querschnitt durch die Grundgesamtheit einmalig untersucht. Ein oft verwendetes Beispiel für Querschnittstudien sind Schulleistungsuntersuchungen, wie beispielsweise die PISA-Studien (siehe z.B. Raudenbush & Bryk, 2002). Dabei sind einzelne Schüler in Schulklassen und diese wiederum in Schulen, Landkreise, Bundesländer und Länder untergliedert. Jede Ebene dieser hierarchischen Struktur hat potentielle Einflüsse auf die erzielte Leistung. Diese Einflüsse können im gemischten linearen Modell ebenfalls als zufällige Effekte abgebildet werden. Die Modellgleichung des gemischten linearen Modells in Matrixform ist in Gleichung 2.42 angegeben. (Galecki, 2013, S. 248)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon} \qquad 2.42$$

Die im Vergleich zum klassischen oder allgemeinen linearen Modell neu hinzugekommenen zufälligen Effekte werden durch den Term $\mathbf{Z}\mathbf{b}$ modelliert. Darin ist \mathbf{Z} die

Designmatrix der zufälligen Effekte und ähnelt \mathbf{X} für die festen Effekte. \mathbf{b} ist der Vektor der Regressionskoeffizienten für die zufälligen Effekte.

Der Aufbau von \mathbf{Z} und \mathbf{b} ist komplexer als der von \mathbf{X} und $\boldsymbol{\beta}$, da neben den Effekten selbst auch die Cluster-Struktur des Experiments in \mathbf{Z} und \mathbf{b} abgebildet wird. Wenn es nur eine Ebene mit m unterschiedlichen Clustern im Experiment gibt, kann Gleichung 2.42 für das i -te Cluster umformuliert werden zu Gleichung 2.43, in der \mathbf{Z}_i und \mathbf{b}_i einen einfacheren Aufbau besitzen. (Fahrmeir et al., 2009, S. 259)

Innerhalb eines Clusters i sind die zufälligen Effekte des Clusters und somit die Regressionskoeffizienten \mathbf{b}_i dieser Effekte konstant. Der prinzipielle Aufbau von \mathbf{Z}_i und \mathbf{b}_i entspricht deshalb hier dem von \mathbf{X}_i und $\boldsymbol{\beta}$. Jede Spalte von \mathbf{Z}_i beinhaltet genau wie jede Spalte von \mathbf{X}_i die Einstellungen eines Faktors. In den meisten Fällen besteht \mathbf{Z}_i aus einer Teilmenge der Spalten von \mathbf{X}_i . Für die zu den Spalten von \mathbf{Z}_i gehörenden Faktoren erlaubt \mathbf{b}_i eine clusterspezifische Variation der Effekte, die durch $\boldsymbol{\beta}$ beschrieben werden. Innerhalb jedes Clusters ist \mathbf{b}_i konstant, für verschiedene Cluster allerdings unterschiedlich.

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, m$$

$$\text{mit } \mathbf{y}_i = \begin{pmatrix} y_{i,1} \\ y_{i,2} \\ \vdots \\ y_{i,n_i} \end{pmatrix}, \mathbf{X}_i = \begin{pmatrix} 1 & x_{i,1,1} & x_{i,2,1} & \dots & x_{i,k,1} \\ 1 & x_{i,1,2} & x_{i,2,2} & \dots & x_{i,k,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{i,1,n_i} & x_{i,2,n_i} & \dots & x_{i,k,n_i} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \quad 2.43$$

$$\mathbf{Z}_i = \begin{pmatrix} 1 & z_{i,1,1} & z_{i,2,1} & \dots & z_{i,l,1} \\ 1 & z_{i,1,2} & z_{i,2,2} & \dots & z_{i,l,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{i,1,n_i} & z_{i,2,n_i} & \dots & z_{i,l,n_i} \end{pmatrix}, \mathbf{b}_i = \begin{pmatrix} b_{i,0} \\ b_{i,1} \\ \vdots \\ b_{i,l} \end{pmatrix}, \boldsymbol{\varepsilon}_i = \begin{pmatrix} \varepsilon_{i,1} \\ \varepsilon_{i,2} \\ \vdots \\ \varepsilon_{i,n_i} \end{pmatrix}$$

Gleichung 2.44 beschreibt, wie sich die Matrizen für die gesamte Versuchsreihe aus den Einzelmatrizen für die einzelnen Cluster zusammensetzen. Die Matrizen \mathbf{y} , \mathbf{X} und $\boldsymbol{\varepsilon}$ werden als Spaltenvektoren aus den zugehörigen Matrizen der einzelnen Cluster zusammengesetzt. Da \mathbf{b}_i für die unterschiedlichen Cluster nicht identisch ist, wird auch \mathbf{b} als Spaltenvektor aus den Einzelvektoren \mathbf{b}_i zusammengesetzt. Die Matrix \mathbf{Z} setzt sich hingegen als Diagonalmatrix aus den Einzelmatrizen \mathbf{Z}_i zusammen. So wird erreicht, dass die clusterspezifischen Effekte \mathbf{b}_i auch nur im i -ten Cluster wirksam sind.

Im Gegensatz zum Vektor β der festen Effekte ist die Bestimmung von b nicht das Ziel der Auswertung. Stattdessen wird eine Kovarianzmatrix $G = cov(b)$ gesucht. Diese beschreibt die Varianzen und Kovarianzen der in b_i (Gleichung 2.43) enthaltenen zufälligen Effekte zwischen den $i = 1 \dots m$ Clustern.

$$\begin{aligned}
 y &= \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}, & X &= \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{pmatrix}, & Z &= \begin{pmatrix} Z_1 & 0 & \dots & 0 \\ 0 & Z_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & Z_m \end{pmatrix}, \\
 & & b &= \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}, & \varepsilon &= \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix}
 \end{aligned} \tag{2.44}$$

Das gemischte lineare Modell besitzt dieselben Annahmen wie das allgemeine lineare Modell. Zusätzlich wird meist angenommen, dass auch der Erwartungswert von b null ist, sowie dass b und ε normalverteilt sind. Die Normalverteilungsannahme ist dabei nicht generell für das gemischte lineare Modell notwendig, ermöglicht allerdings die Parameterschätzung mit ML- bzw. REML-Verfahren. (Fahrmeir et al., 2009, S. 258)

Nachfolgend wird zunächst in Abschnitt 2.5.4.1 vorgestellt, wie die Koeffizienten des gemischten linearen Modells ermittelt werden können. Anschließend wird in Abschnitt 2.5.4.2 beschrieben, wie das Modell und dessen Vorhersagegenauigkeit statistisch bewertet werden können. In Abschnitt 2.5.4.3 wird dann beleuchtet, wie die Einhaltung der Annahmen des gemischten linearen Modells überprüft werden kann.

2.5.4.1 Parameterschätzung

Die Parameterschätzung für das gemischte lineare Modell ist mit unterschiedlichen Verfahren möglich. Die am häufigsten eingesetzten Verfahren sind ML- und REML-Verfahren (Fahrmeir et al., 2009, S. 264). Diese werden in Abschnitt 2.5.3.2 vorgestellt. Zur Schätzung wird oft ein sogenanntes marginales Modell gemäß Gleichung 2.45 eingesetzt, bei dem die zufälligen Effekte mit den Residuen zu einem neuen Fehlerterm ε^* zusammengefasst werden. Die Kovarianzmatrix von ε^* ist durch Gleichung 2.46 gegeben, in der R die Kovarianzmatrix von ε darstellt. (Fahrmeir et al., 2009, S. 261)

$$y = X\beta + \varepsilon^* \quad \text{mit} \quad \varepsilon^* = Zb + \varepsilon \tag{2.45}$$

$$\text{Cov}(\boldsymbol{\varepsilon}^*) = \boldsymbol{\Sigma} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R} \quad 2.46$$

Zur Bestimmung der Modellparameter von gemischten linearen Modellen mit Hilfe der Maximum-Likelihood-Methode kommen üblicherweise iterative Optimierungsverfahren zum Einsatz. Mit diesen wird die Log-Likelihood gemäß Gleichung 2.35 maximiert. Um die im Vergleich zu allgemeinen linearen Modellen geänderte Modellgleichung zu berücksichtigen, muss dabei $\boldsymbol{\Omega}$ durch $\boldsymbol{\Sigma}$ gemäß Gleichung 2.32 ersetzt werden. Die große Anzahl der unbekannt Elemente in \mathbf{G} und \mathbf{R} führt allerdings dazu, dass Annahmen über den Aufbau dieser beiden Kovarianzmatrizen getroffen werden müssen. Ansonsten ist keine eindeutige Bestimmung möglich. Da die zufälligen Effekte bereits einen großen Teil der systematischen Varianz von $\boldsymbol{\varepsilon}^*$ erklären, kann beispielsweise oft $\mathbf{R} = \sigma^2\mathbf{I}$ angenommen werden.⁴⁰

Ein häufig eingesetztes iteratives Optimierungsverfahren zur Parameterbestimmung von gemischten linearen Modellen auf Basis von REML ist der Expectation-Maximization-Algorithmus (Galecki, 2013, S. 264).

2.5.4.2 Statistische Bewertung

Die Signifikanz fester Effekte im gemischten linearen Modell lässt sich analog zum allgemeinen linearen Modell bestimmen. Dazu muss in Gleichung 2.41 die Kovarianzmatrix $\boldsymbol{\Omega}$ durch $\boldsymbol{\Sigma}$ gemäß Gleichung 2.32 ersetzt werden (Fahrmeir et al., 2009, S. 269).

Für zufällige Effekte lässt sich die Signifikanz nicht in äquivalenter Weise bestimmen. Stattdessen werden hier oft Methoden eingesetzt, die die Qualität von Modellen mit dem zu testenden Term mit solchen Modellen vergleicht, die den jeweiligen Term nicht enthalten. Als Vergleichskriterium kann die Log-Likelihood entsprechend Gleichung 2.35 genutzt werden. Diese Art von Test wird Likelihood-Quotienten-Test genannt. Alternativ können Informationskriterien, wie das Informationskriterium nach Akaike *AIC* oder das Bayesianische Informationskriterium *BIC* zum Einsatz kommen. (Galecki, 2013, S. 268)

Bei der Ermittlung des Prognoseintervalls gibt es für gemischte lineare Modelle zwei unterschiedliche Anwendungsfälle. Einer besteht darin, dass die Vorhersagegenau-

⁴⁰ Dies wird z.B. üblicherweise für Split-Plot-Versuchsreihen (siehe Abschnitt 2.5.5) angenommen. Bei Split-Plot-Versuchsreihen werden zudem auch Annahmen über $\mathbf{Z}\mathbf{G}\mathbf{Z}'$ getroffen (siehe Gleichungen 2.50 und 2.51). Für Details zu weiteren möglichen Annahmen für \mathbf{R} und \mathbf{G} siehe Galecki (2013, S. 202–209)

igkeit für ein spezielles Cluster gesucht ist, für das die Parameter der zufälligen Effekte \mathbf{b} bekannt sind. In diesem Fall sind die „zufälligen“ Effekte nicht mehr variabel, sondern können den festen Effekten zugeordnet werden. So wird das gemischte lineare Modell zu einem klassischen linearen Modell und die Bestimmung der Vorhersagegenauigkeit kann analog zu Abschnitt 2.5.2.2 ablaufen.

Im „marginalen Fall“ (vergleiche „marginale Modell“ in Abschnitt 2.5.4.1) ist die Vorhersagegenauigkeit für ein bisher unbekanntes Cluster gesucht. Dadurch wirken sich die Varianzen der zufälligen Effekte \mathbf{b} auf das Konfidenz- und Vorhersageintervall aus.

Die Varianz einer Vorhersage des Modells beinhaltet deshalb drei Komponenten:

- Die Unsicherheit $\sigma_{x_0}^2$ der Schätzung des Erwartungswertes der Zielgröße y an der Stelle \mathbf{x}_0 aufgrund von Unsicherheiten der Schätzung von $\boldsymbol{\beta}$
- Die Streuung $\sigma_{z_0}^2$, die durch die zufälligen Effekte an der Stelle \mathbf{z}_0 hervorgerufen wird
- die Streuung σ der Fehler ε

Für das Vorhersageintervall ergibt sich deshalb Gleichung 2.47 (Albers, 2010b, S. 3).

$$PI_{y_0} = \left[\mathbf{x}_0' \hat{\boldsymbol{\beta}} - t_{N-p}^{-1} \left(1 - \frac{\alpha}{2}\right) \sqrt{\hat{\sigma}_{x_0}^2 + \hat{\sigma}_{z_0}^2 + \hat{\sigma}^2}; \mathbf{x}_0' \hat{\boldsymbol{\beta}} + t_{N-p}^{-1} \left(1 - \frac{\alpha}{2}\right) \sqrt{\hat{\sigma}_{x_0}^2 + \hat{\sigma}_{z_0}^2 + \hat{\sigma}^2} \right]$$

mit $\hat{\sigma}_{x_0} = \hat{\sigma}(\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0)^{\frac{1}{2}}$ 2.47

und $\hat{\sigma}_{z_0} = (\mathbf{z}_0' \mathbf{G} \mathbf{z}_0)^{\frac{1}{2}}$

Die Bestimmung des Vorhersageintervalls gemäß Gleichung 2.47 berücksichtigt keine Ungenauigkeiten in der Schätzung von $\sigma_{z_0}^2$. Wenn diese ebenfalls berücksichtigt werden sollen, müssen aufwendigere Verfahren zum Einsatz kommen. Zwei dieser Verfahren werden von Rhinehart (2016, Kap. 3.6.2.2 und 3.8) vorgestellt.

2.5.4.3 Überprüfung der Modellannahmen

Zusätzlich zu den Annahmen des allgemeinen linearen Modells besteht beim gemischten linearen Modell meist die Annahme normalverteilter zufälliger Effekte. Zur Überprüfung dieser Annahme können die zufälligen Effekte \mathbf{b} für das i -te Cluster nach Gleichung 2.48 geschätzt werden. Anschließend kann die Normalverteilung von $\hat{\mathbf{b}}_i$ anhand eines Wahrscheinlichkeitsnetzes (siehe Abbildung 2.25) überprüft werden.

$$\hat{\mathbf{b}}_i = \widehat{\mathbf{G}}\mathbf{Z}_i'\widehat{\boldsymbol{\Sigma}}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}) \quad 2.48$$

Die Verteilung von $\hat{\mathbf{b}}$ entspricht allerdings nicht notwendigerweise der Verteilung von \mathbf{b} . Deshalb sollte die Normalverteilungsannahme besser geprüft werden, indem ein Modell mit dieser Annahme mit einem Modell ohne dieser Annahme verglichen wird (Galecki, 2013, S. 264 f.).⁴¹

2.5.5 Split-Plot-Versuche

In vielen Versuchsreihen ist die Änderung der Einstellung von einigen Faktoren schwieriger oder aufwendiger, als die der restlichen Faktoren. So ist bei tribologischen Versuchsreihen beispielsweise der Wechsel eines Schmiermediums sehr aufwendig, da dafür alle mit ihm in Kontakt kommenden Systemkomponenten gereinigt und neu geschmiert werden müssen. Die Änderung von Kontaktdrücken oder Gleitgeschwindigkeiten ist im Vergleich dazu meist deutlich weniger aufwendig. Die schwierig zu ändernden Faktoren werden hard-to-change-Faktoren (HTC-Faktoren) genannt, während die anderen Faktoren easy-to-change-Faktoren (ETC-Faktoren) genannt werden. In vielen Fällen benötigt die Einstellungsänderung eines HTC-Faktors um mehrere Größenordnungen mehr Zeit, als die Durchführung eines Versuches, bei dem nur ETC-Faktoren neu eingestellt werden müssen. Um den Aufwand in solchen Versuchsreihen zu reduzieren, wird hier deshalb oft von der vollständigen Randomisierung (siehe Abschnitt 2.4.2.1) abgewichen. Die Reihenfolge der Versuchsdurchführung wird dann so umgestellt, dass alle Versuche mit derselben Einstellungskombination der HTC-Faktoren nacheinander durchgeführt werden. Eine solche Gruppe von aufeinanderfolgenden Versuchen mit konstanter Einstellung der HTC-Faktoren wird Whole-Plot (WP) genannt. Die Einzelversuche innerhalb eines Whole-Plots heißen Split-Plots oder Sub-Plots (SP). (Kowalski et al., 2007, S. 1, 4)

⁴¹ Für Details siehe (Verbeke & Molenberghs, 2001, S. 83–92)

Durch die Gruppierung der gesamten Versuchsreihe in WPs wird bei Split-Plot-Versuchsreihen eine heteroskedastische Fehlerstruktur erzeugt: Innerhalb eines WPs bleibt die Einstellung der HTC-Faktoren konstant. Dasselbe gilt dadurch auch für Ungenauigkeiten der HTC-Faktoreinstellungen, wodurch innerhalb eines WPs konstant bleibende Fehleranteile erzeugt werden. Zudem werden Versuche in einem WP meist in schneller Abfolge durchgeführt, während zwischen zwei WPs zeitaufwendige Einstellarbeiten der HTC-Faktoren für Verzögerungen sorgen. Deshalb sind auch die Einflüsse sich langsam ändernder Störgrößen innerhalb eines WPs relativ konstant, während sie zwischen verschiedenen WPs größere Unterschiede aufweisen. In Summe ist die Fehlervarianz innerhalb eines WPs deshalb meist deutlich kleiner als zwischen verschiedenen WPs. Aufgrund dieser heteroskedastischen Fehler können Split-Plot-Versuchsreihen nicht mit Hilfe des klassischen linearen Modells ausgewertet werden. Vining (2012, S. 480) und viele andere Autoren setzen deshalb ein allgemeines lineares Modell gemäß den Gleichungen 2.16 und 2.32 ein. Darin teilen sie den Fehler gemäß Gleichung 2.49 in zwei Komponenten δ und ε auf. ε beschreibt den „Sub-Plot-Fehler“. Es beinhaltet die Fehlervarianz innerhalb eines WPs. ε nimmt für jeden Versuch neue Werte an. Der „Whole-Plot Fehler“ δ beinhaltet hingegen die Fehlervarianz zwischen verschiedenen WPs und ist damit innerhalb eines WPs konstant.

$$y = X\beta + \delta + \varepsilon \tag{2.49}$$

Eine alternative, aber zu Gleichung 2.49 äquivalente Formulierung des Split-Plot-Modells verwenden beispielsweise Jones und Nachtsheim (2009). Sie drücken den Whole-Plot Fehler δ als zufälligen Effekt Zb aus und erhalten damit ein gemischtes lineares Modell gemäß Gleichung 2.42. Gleichung 2.49 kann als marginale Formulierung des gemischten linearen Modells gesehen werden (vgl. Gleichung 2.45). Beide Formulierungen sind also ineinander überführbar.

Da der Whole-Plot Fehler δ innerhalb eines WPs konstant ist, kann die Kovarianzmatrix Σ des Gesamtfehlers gemäß Gleichung 2.50 formuliert werden. In dieser Gleichung steht σ_δ^2 für die Varianz von δ und σ^2 für die Varianz von ε . (Kowalski et al. 2007, S. 12)

$$\Sigma = Cov(\delta + \varepsilon) = \sigma_\delta^2 J + \sigma^2 I \tag{2.50}$$

Darin ist I eine Einheitsmatrix und J eine Blockdiagonalmatrix gemäß Gleichung 2.51. n_1 bis n_m sind die Anzahlen der Versuche innerhalb der m WPs der Versuchsreihe.

$$J = \begin{pmatrix} \mathbf{1}_{n_1} \mathbf{1}_{n_1}' & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{1}_{n_m} \mathbf{1}_{n_m}' \end{pmatrix} \quad 2.51$$

Sowohl zur Schätzung von β mit Hilfe von GLS nach Gleichung 2.33, als auch zur Bewertung der statistischen Signifikanz der einzelnen Effekte gemäß Gleichung 2.41 muss Σ bekannt sein. Die darin unbekanntenen Varianzkomponenten σ_δ^2 und σ^2 können dabei auf verschiedene Arten bestimmt werden. Die am häufigsten eingesetzte Methode zur Schätzung der Varianzkomponenten ist REML⁴². REML kann unabhängig vom eingesetzten Versuchsplan verwendet werden. Alternative Ansätze liefern Weerakkody und Johnson (1992), Hasegawa et al. (2010), sowie Ikeda et al. (2014). Zudem ist auch ein „pure error“-Ansatz zur Bestimmung der Varianzkomponenten möglich. Dabei werden die Varianzen direkt aus den Unterschieden zwischen nominell identischen WPs und SPs abgeleitet. Dadurch fließen keine Lack-of-fit-Anteile⁴³ in die Varianzkomponenten ein. Voraussetzung hierfür ist allerdings, dass im Versuchsplan sowohl mehrere identische WPs, als auch WPs enthalten sind, innerhalb welchen SPs mit identischen Faktoreinstellungen wiederholt werden. (Kowalski et al., 2007, S. 12)

Auch wenn Split-Plot-Versuchspläne schon 1925 von Fisher (1925) vorgeschlagen worden sind, wurden erst kurz vor der Jahrtausendwende nennenswerte Weiterentwicklungen dieser Versuchspläne erreicht. Die aktuellen Forschungen zu diesem Thema lassen sich grob in drei Bereiche gliedern, die in den nachfolgenden Abschnitten beschrieben werden: Faktorielle Split-Plot-Versuchspläne (Abschnitt 2.5.5.1), Equivalent-Estimation-Versuchspläne (Abschnitt 2.5.5.2) und D-optimale Split-Plot-Versuchspläne (Abschnitt 2.5.5.3).

2.5.5.1 Faktorielle Split-Plot-Versuchspläne

Um vollständig faktorielle Versuchspläne zu Split-Plot-Versuchsplänen zu machen, wird zunächst ein vollfaktorieller Versuchsplan erstellt und randomisiert, der ausschließlich die HTC-Faktoren enthält. Anschließend werden anstelle der Einzelversuche dieses Versuchsplans vollständig faktorielle Versuchspläne eingesetzt, die die ETC-Faktoren enthalten. Diese werden separat randomisiert (Jones & Nachtsheim, 2009, S. 345). Die so erstellten Split-Plot-Versuchspläne werden *gekreuzte separat randomisierte* („crossed bi-randomized“) oder *kartesisches Produkt* („cartesian product“) Versuchspläne genannt (Vining, 2012, S. 482). Letsinger et al. (1996)

⁴² Siehe Abschnitt 2.5.3.2

⁴³ Siehe S. 28

konnten zeigen, dass für diese Art von Versuchsplänen GLS und OLS dieselben Schätzwerte für die Modellparameter liefern.⁴⁴

Fraktionell faktorielle Split-Plot-Versuchspläne können analog zu vollständig faktoriellen Split-Plot-Versuchsplänen generiert werden. Hierbei können für die HTC- und ETC-Faktoren allerdings Versuchspläne unterschiedlicher Auflösung verwendet werden. Zudem gibt es verschiedene Möglichkeiten, fraktionell faktorielle Versuchspläne einer bestimmten Auflösung zu generieren.⁴⁵ Methoden um mit den daraus resultierenden Variationsmöglichkeiten möglichst effiziente fraktionell faktorielle Split-Plot-Versuchspläne generieren zu können, sowie eine weitere Konstruktionsmöglichkeit namens „split plot confounding“, wurden von Huang et al. (1998) und Bingham & Sitter (1999) beschrieben. Die damit konstruierten Versuchspläne sind zweistufig und können deshalb nur lineare Abhängigkeiten zwischen der Zielgröße und den einzelnen Faktoren abbilden. Die Konstruktionsmethoden neigen dazu, einen Versuchsplan zu erzeugen, der eine große Anzahl von eher kleinen Whole-Plots besitzt (Wang, 2006). Da zwischen den Whole-Plots die HTC-Faktoreinstellungen geändert werden müssen, resultiert daraus oft ein hoher Aufwand zur Durchführung der Versuchsreihe, obwohl die Gesamtzahl der Versuche gering ist.

2.5.5.2 Equivalent-Estimation-Versuchspläne

Split-Plot-Versuchspläne für quadratische Modelle wurden erstmals von Letsinger et al. (1996) intensiver behandelt. Sie, sowie Draper und John (1998) und Trinca und Gilmour (2001) berücksichtigen die entstehende heteroskedastische Fehlerstruktur, indem sie REML und GLS zur Auswertung einsetzen. Vining et al. (2005) entwickelten im Gegensatz dazu erstmals Versuchspläne für quadratische Modelle, bei denen eine Auswertung mittels OLS dieselben Ergebnisse liefert wie GLS. Sie nennen diesen Versuchsplantyp Equivalent-Estimation-Design (EED). EEDs basieren auf zentral zusammengesetzten Versuchsplänen oder Box-Behnken-Versuchsplänen, die nach einem gezielten Schema umsortiert werden. Zusätzlich werden bestimmte Versuche wiederholt. Dass die Parameter β in EEDs mittels OLS geschätzt werden können, vereinfacht die Auswertung solcher Versuchspläne deutlich. Hierzu müssen lediglich y und X in Gleichung 2.21 eingesetzt werden. Dies erhöht auch die Auswahl an zur Auswertung geeigneter Software deutlich, da diese lediglich die Durchführung von Matrixoperationen ermöglichen muss. Bei Versuchsplänen, die gemäß dem Schema von Vining et al. erstellt werden, muss zur Bewertung der statistischen Signifikanz der Effekte allerdings weiterhin REML oder ein

⁴⁴ vgl. Abschnitt 2.5.5.2

⁴⁵ Es könne unterschiedliche „Generatoren“ eingesetzt werden. Siehe Kleppmann (2011, S. 126).

ähnliches Verfahren (zur hierbei notwendigen Bestimmung von δ und ϵ) eingesetzt werden.

Parker et al. (2007b) leiteten eine allgemeine Bedingung her, unter der ein Split-Plot-Versuchsplan die Eigenschaft der OLS-GLS-Äquivalenz besitzt. Diese ist durch 2.52 gegeben.

$$XK = JX \tag{2.52}$$

In Gleichung 2.52 ist K durch Gleichung 2.53 definiert.

$$K = (X'X)^{-1}X'JX \tag{2.53}$$

Diese Bedingung nutzen Parker et al., um eine alternative Konstruktionsmethode für EEDs zu erstellen, die die Anzahl der WPs im Versuchsplan minimiert. Ebenfalls auf der Äquivalenzbedingung 2.52 aufbauend, entwickelten Vining und Kowalski (2008) eine weitere Konstruktionsmethode. Diese führt zu Versuchsplänen, die einerseits WPs enthalten, die ausschließlich aus Wiederholungen eines Versuchs bestehen. Andererseits beinhalten die mit ihr erstellten Versuchspläne auch Wiederholungen von vollständigen WPs. Dies ermöglicht die direkte Bestimmung von σ und σ_δ gemäß den Gleichungen 2.54 und 2.55. Diese Konstruktionsmethode erlaubt damit erstmals eine vollständige Auswertung der Versuchsreihe, inklusive aller statistischer Bewertungsgrößen aus Abschnitt 2.5.2.2, ohne den Einsatz von REML oder ähnlichen Techniken.

Gleichung 2.54 dient bei einer solchen Versuchsreihe dazu, einen Schätzwert $\hat{\sigma}^2$ für die Fehlervarianz der Einzelversuche zu ermitteln. Dazu werden die WPs verwendet, in welchen ausschließlich Versuchswiederholungen enthalten sind. Die Anzahl solcher WPs wird durch m_s beschrieben. Die Varianz der Versuchsergebnisse innerhalb dieser WPs wird mit S_i^2 bezeichnet.

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{m_s} S_i^2}{m_s} \tag{2.54}$$

Zur Ermittlung eines Schätzwerts für die WP-Fehlervarianz werden die WPs herangezogen, die identisch zueinander sind. Die Schätzung basiert hier auf der Varianz S_{WP}^2 der Mittelwerte der Versuchsergebnisse dieser WPs. Da diese Mittelwerte auch

durch die Fehlervarianz $\hat{\sigma}^2$ der Einzelversuche beeinflusst werden, wird $\widehat{\sigma}_\delta^2$ zudem um deren Einfluss korrigiert. n ist dabei die Anzahl der Einzelversuche in jedem WP.

$$\widehat{\sigma}_\delta^2 = S_{WP}^2 - \frac{1}{n} \hat{\sigma}^2 \quad 2.55$$

Weitere nennenswerte Entwicklungen auf dem Gebiet der Equivalent-Estimation-Versuchspläne wurden von Wang (2006) und Yuan (2011) geleistet: Wenn EEDs auf Basis eines zentral zusammengesetzten Versuchsplans entwickelt werden, dann besitzt (wie im vollständig randomisierten Fall) der Abstand der Sternpunkte zum Zentrum einen Einfluss auf die Vorhersageeigenschaften des Versuchsplans. Wang ermittelte separate Distanzen für WP- und SP-Faktoren, sodass eine Eigenschaft analog zur Drehbarkeit⁴⁶ vollständig randomisierter Versuchspläne erreicht wird.

EEDs für mehr als zwei Randomisierungsebenen (also Split-Split-Plot-EEDs) wurden von Yuan entwickelt. Diese erlauben eine weiter verfeinerte Unterteilung der Faktoren bezüglich ihres Änderungsaufwands und eine darauf angepasste Versuchsreihenfolge.

2.5.5.3 D-optimale Split-Plot-Versuchspläne

Goos und Vandebroek (2001) entwickelten Konstruktionstechniken für D-optimale Split-Plot-Designs. Sie leiten die Erstellung von Versuchsplänen für lineare und quadratische Modelle her. Für quadratische Modelle erstellen sie einen Computeralgorithmus. Diesem muss die Anzahl der WPs und SPs, die Form des zu erstellenden Modells, sowie das Varianzverhältnis d vorgegeben werden. d bezeichnet das Verhältnis zwischen WP- und SP-Varianz gemäß Gleichung 2.56.

$$d = \frac{\sigma_\delta^2}{\sigma^2} \quad 2.56$$

Zusätzlich benötigt der von ihnen erstellte Algorithmus eine vorgegebene Menge an möglichen Einstellungskombinationen aller Faktoren, aus der dann diejenigen ausgewählt werden, die eine möglichst effiziente Anpassung der vorgegebenen Modellform an die Versuchsdaten erlauben. Als Maß für die Effizienz und damit als Optimierungskriterium kommt die Informationsmatrix \mathbf{M} des Versuchsplans gemäß

⁴⁶ Siehe S. 46

Gleichung 2.57 zum Einsatz (Jones & Goos, 2012, S. 5). Diese Informationsmatrix stellt den Kehrwert der Kovarianzmatrix der Koeffizienten für GLS aus Gleichung 2.41 dar.⁴⁷ Das Maximieren von \mathbf{M} führt damit zur Minimierung von $\text{Cov}(\hat{\beta})$ und dadurch zu einem D-optimalen Versuchsplan.

$$\mathbf{M} = \mathbf{X}'\Sigma\mathbf{X} \quad 2.57$$

Ein alternativer Algorithmus, der keine Vorgabe von möglichen Einstellungskombinationen benötigt, wurde von Jones und Goos (2007) erstellt. Weitere Algorithmen, mit deren Hilfe D-effiziente Equivalent-Estimation-Versuchspläne generiert werden können, wurden von Macharia und Goos (2010), sowie Jones und Goos (2012) entwickelt. Die Erstellung von D-optimalen Split-Plot-Versuchsplänen mit drei Randomisierungsebenen beschreiben Jones und Goos (2009).

2.6 Statistische Versuchsplanung für nichtparametrische Regressionsmodelle

Die in Abschnitt 2.5 beschriebenen Versuchspläne und Auswertemethoden sind darauf optimiert, die Koeffizienten der linearen Regressionsmodelle aus demselben Abschnitt zu bestimmen. In den meisten Fällen werden als lineare Modelle Polynome maximal zweiten Grades eingesetzt. In einigen Fällen ist das zu modellierende Systemverhalten zu stark nichtlinear, um mit solchen Modellen abgebildet werden zu können, da es ansonsten zu Lack-of-fit kommt.⁴⁸ Erhöht man in diesen Fällen den Grad des verwendeten Polynoms, dann steigt die Gefahr eines Over-Fit.⁴⁹ Es steigt bei Erhöhung des Grades des verwendeten Polynoms also nicht unbedingt die Genauigkeit des Modells. Der Einsatz von Polynomen höheren Grades als Modell ist deshalb für Systeme mit stark nichtlinearem Verhalten meist nicht sinnvoll. (Barton, 1994, S. 239; Voss & Dean, 1999, S. 246)

Für solche Anwendungsfälle werden deshalb stattdessen üblicherweise nichtparametrische Regressionsmodelle eingesetzt, die in Abschnitt 2.6.2 beschrieben werden. Versuchspläne zur Bestimmung nichtparametrischer Regressionsmodelle sind nicht darauf ausgerichtet, eine vorgegebene Gleichung möglichst effizient parametrieren zu können. Stattdessen werden sie nach modellunabhängigen Kriterien ge-

⁴⁷ Für den Zusammenhang zwischen Ω und Σ siehe Gleichung 2.32

⁴⁸ Vgl. Abbildung 2.9

⁴⁹ Vgl. Abbildung 2.10

staltet. Bei ihnen wird versucht, den Versuchsraum möglichst gleichmäßig mit Versuchspunkten zu füllen. Sie werden deshalb auch *space-filling designs*, also raumfüllende Versuchspläne genannt. In Abschnitt 2.6.1 werden sie näher beschrieben.

2.6.1 Versuchspläne

Da raumfüllende Versuchspläne für stark nichtlineare Zusammenhänge entwickelt worden sind und auch meist hierfür verwendet werden, werden sie von Siebertz et al. (2010) *Versuchspläne für komplexe Zusammenhänge* genannt. Einige andere Autoren (z.B. Morris, 2012; Pronzato & Müller, 2012) bezeichnen sie auch als *Versuchspläne für Computer-Experimente*, da sehr komplexe Zusammenhänge üblicherweise nur in Computer-Experimenten untersucht werden. Dies hat folgenden Grund: Bei stark nichtlinearen Zusammenhängen steigt die Anzahl der zu untersuchenden Faktorstufen schnell an. Wenn mehrere Faktoren untersucht werden sollen, erhöht dies den Versuchsumfang bei raumfüllenden Versuchsplänen drastisch. Für physische Experimente ist der entstehende Aufwand zur Durchführung der Versuchsreihe dann meist zu groß. Im Gegensatz dazu können viele Computer-Experimente deutlich schneller durchgeführt werden, was sehr viel höhere Versuchsumfänge erlaubt. Mit Hilfe von Computer-Experimenten werden meist Metamodelle generiert, wenn die Rechengeschwindigkeit der originalen Simulationsmodelle für einen speziellen Anwendungsfall unzureichend ist.

Computer-Experimente sind im Gegensatz zu physischen Experimenten deterministisch. Sie besitzen keine stochastischen Störgrößeneinflüsse oder Messfehler. Es kann dadurch auf die in Abschnitt 2.4.2.1 beschriebenen Grundprinzipien zur Erstellung von Versuchsplänen von physischen Experimenten verzichtet werden (Siebertz et al., 2010, S. 153 f.). Des Weiteren sind aus demselben Grund die Einflüsse von Störgrößen und Messfehlern, sowie die daraus abgeleiteten Bewertungsgrößen, wie beispielsweise Vertrauensintervalle, bei Computer-Experimenten kein Bestandteil der Untersuchung. Da modellunabhängige Versuchspläne meist für Computer-Experimente eingesetzt werden, werden bei ihrer Entwicklung normalerweise keine Methoden zur Vermeidung oder Quantifizierung der stochastischen Einflüsse von Störgrößen und Messfehlern berücksichtigt.

2.6.1.1 Versuchspläne auf Basis von Gittern

Eine einfache Methode, einen vorgegebenen mehrdimensionalen Versuchsraum gleichmäßig mit Versuchspunkten zu füllen, ist die Verwendung von Gittern. Wenn diese Gitter rechtwinklig sind, wird von *grid designs* gesprochen. Nicht-rechtwinklige Gitter werden dagegen *lattice designs* genannt (Pronzato & Müller, 2012). Beide Varianten füllen den Versuchsraum ausreichend gleichmäßig. Einer ihrer Nachteile

ist, dass der Versuchsumfang immer dem Produkt der Stufenanzahlen aller Faktoren entspricht. Dadurch ergibt sich bei mehreren Faktoren eine sehr grobe Abstufung der möglichen Versuchsumfänge. Ihr größter Nachteil zeigt sich, wenn sich nach Durchführung der Versuchsreihe herausstellt, dass einer der untersuchten Faktoren keinen Einfluss auf die Zielgröße hatte. Die Dimension dieses Faktors kann dann aus dem Versuchsplan entfernt werden und alle Versuchspunkte werden auf die verbleibenden Dimensionen projiziert. Das führt vor allem bei rechtwinkligen Gittern dazu, dass mehrere Versuchspunkte aufeinander projiziert werden, was faktisch Versuchswiederholungen darstellt (Jurecka, 2007, S. 120). In Computer-Experimenten kann aus Versuchswiederholungen keine Zusatzinformation gezogen werden. Bei manchen Auswertemethoden können Wiederholungen sogar zu Problemen führen (Barton, 1994). Deshalb werden Versuchspläne auf Basis von Gittern nur selten für Computer-Experimente eingesetzt.

2.6.1.2 Monte-Carlo-Methode

Bei der Monte-Carlo-Methode werden die Faktorstufenkombinationen aller Versuche innerhalb ihrer Grenzen zufällig festgelegt. Algorithmen, die zur Erzeugung der dafür benötigten Zufallszahlen zum Einsatz kommen, wurden beispielsweise von Matsumoto & Saito (2008), Panneton et.al. (2006) oder Marsaglia & Tsang (2000) beschrieben. Die Lage der bereits erstellten Versuchspunkte im Versuchsraum wird zur Festlegung der Faktorstufenkombinationen der nachfolgenden Versuche nicht berücksichtigt. Dadurch kann eine gleichmäßige Verteilung der Versuchspunkte im Versuchsraum nur näherungsweise sichergestellt werden. Es wurden verschiedene Algorithmen entwickelt, die eine gleichmäßigere Verteilung der Versuchspunkte über den Versuchsraum gewährleisten sollen. Da diese Algorithmen einen deterministischen Anteil zur ansonsten zufälligen Generierung der Faktorstufenkombinationen liefern, werden sie als Quasi-Monte-Carlo-Methoden bezeichnet (Siebertz et al., 2010, S. 168 f.). Einen Überblick über diese Methoden liefert Lemieux (2010).

2.6.1.3 Latin Hypercube Samples

Die Methode Latin Hypercube Sampling basiert auf mehrdimensionalen lateinischen Quadraten und wurde von McKay et al. (1979) vorgestellt.

Bei dieser Methode wird der Wertebereich jedes Faktors in eine Anzahl gleich großer Abschnitte unterteilt. Die Anzahl der Abschnitte entspricht dem gewünschten Umfang der Versuchsreihe. Für jeden Faktor wird nun aus allen Abschnitten jeweils ein Zahlenwert zufällig ausgewählt. Die so ausgewählten Zahlenwerte werden für jeden Faktor getrennt in eine zufällige Reihenfolge gebracht. Die so generierten Zahlenreihen werden nun als Einstellungslisten für die Faktoren verwendet.

Bei dieser Methode wird die Lage der Versuchspunkte im Versuchsraum zwar auch nach dem Zufallsprinzip ermittelt, allerdings wird durch den verwendeten Algorithmus sichergestellt, dass jeder Abschnitt des Wertebereichs jedes Faktors exakt einmal im Versuchsplan untersucht wird. Dadurch weisen per Latin Hypercube Sampling generierte Versuchspläne bei gleichem Versuchsumfang bessere Vorhersageeigenschaften auf als solche, die mittels der Monte-Carlo-Methode erstellt wurden. (Morris, 2012, S. 401)

Es kann dennoch vorkommen, dass die Einstellungsvektoren mehrerer Faktoren stark miteinander korrelieren. Dadurch ist dann keine eindeutige Zuordnung der gemessenen Effekte zu den korrelierten Faktoren mehr möglich (siehe Abschnitt 2.5.1, Absatz 3). Um starke Korrelationen zu verhindern, erweiterten Tang (1993) und Owen (1992) die Methode so, dass mit ihr Abwandlungen von orthogonalen Testfeldern erstellt werden können. Die Methode ist weiterhin Gegenstand vieler auch aktueller Untersuchungen (Cioppa & Lucas, 2007; Deutsch & Deutsch, 2012; Leary et al., 2003 und viele andere). Ziel dieser Untersuchungen ist beispielsweise, die Orthogonalität der Versuchsreihe oder die Gleichmäßigkeit der Abdeckung des Versuchsraumes weiter zu verbessern (vgl. Abschnitt 2.5.1.8).

2.6.1.4 Gleichverteilte Testfelder (Uniform Designs)

In den Jahren 1980 und 1981 wurde von Fang (1980) und Wang und Fang (1981) das Konzept der Uniform Designs vorgestellt. Das Prinzip der Erstellung von Uniform Designs ähnelt dem von Latin Hypercube Samples. Zuerst werden die Wertebereiche der Faktorstufen in Abschnitte unterteilt, aus denen dann die Mitten (im Gegensatz zu beliebigen Werten bei Latin Hypercube Samples) als Faktorstufen ausgewählt werden. Daraufhin werden die Einstellungen aller Faktoren getrennt voneinander in eine Reihenfolge gebracht, die eine möglichst gleichmäßige Verteilung der Versuchspunkte über den Versuchsraum ergibt (Fang et al., 2000). Dabei können verschiedene Optimierungs-Algorithmen, wie beispielsweise das *Threshold Accepting*⁵⁰, zum Einsatz kommen. Als Maß für die Gleichmäßigkeit der Verteilung werden verschiedene Definitionen der *Diskrepanz* verwendet (Siebertz et al., 2010, S. 164 ff.).

2.6.1.5 Ungleich verteilte Testfelder

Alle bisher vorgestellten Verfahren versuchen die Versuchspunkte möglichst gleichmäßig über den Versuchsraum zu verteilen. Es kann aber vorkommen, dass in manchen Regionen des Versuchsraums eine genauere Approximation benötigt wird, als

⁵⁰ siehe (Dueck & Scheuer, 1990; Winker & Fang, 1997)

in anderen Regionen. In diesem Fall können die in den vorherigen Abschnitten beschriebenen Verfahren so angepasst werden, dass die Verteilungsdichte der Versuchspunkte gebietsweise erhöht wird. Solche Verfahren werden beispielsweise von Siebertz et al. (2010, S. 188 f.) behandelt.

2.6.2 Nichtparametrische Regressionsmodelle

Wie auf Seite 75 bereits erläutert, ist es meist nicht sinnvoll, den Grad eines Polynoms, das mittels multipler linearer Regression an Messdaten angepasst wird, zu erhöhen, um komplexe Abhängigkeiten besser abbilden zu können. Deshalb kommen zur Modellbildung aus Daten, die mittels der in Abschnitt 2.6.1 beschriebenen Versuchspläne generiert wurden, stattdessen meist nichtparametrische Verfahren zum Einsatz. Es sei darauf hingewiesen, dass der für diese Verfahren übliche Begriff „*nichtparametrisch*“ irreführend sein kann, da diese Verfahren durchaus auf Parametern basieren. Die Parameter sind bei diesen Verfahren lediglich aufgrund ihrer Anzahl oder der Modellstruktur einzeln nicht mehr sinnvoll vom Menschen interpretierbar. (Fahrmeir et al., 2009, S. 293)

Nichtparametrische Regressionsmodelle werden momentan zu einem großen Anteil für das maschinelle Lernen, also im Umfeld der Informatik, eingesetzt und weiterentwickelt. Stochastisches Verhalten spielt in der Informatik eine deutlich geringere Rolle als in der Statistik. Aus diesen Gründen stehen für nichtparametrische Regressionsmodelle Verfahren, mit der die statistische Signifikanz der abgebildeten Effekte bewertet werden kann, nicht im Fokus. Dasselbe gilt für die Quantifizierung der in den Residuen enthaltenen Streuung und damit auch für Bewertungsgrößen zur Vorhersagegenauigkeit des Modells (vgl. Abschnitt 2.5.2.2).

Einen Überblick über die verschiedenen nichtparametrischen Regressionsmodelle geben beispielsweise Frochte (2018) oder de Mello (2018). Nichtparametrische Regressionsmodelle stehen in dieser Arbeit nicht im Zentrum. Lediglich Polynomsplines, die in Abschnitt 2.6.2.1 vorgestellt werden, kommen in späteren Kapiteln zum Einsatz. Für alle anderen nichtparametrischen Regressionsmodelle werden stellvertretend nur die Künstlichen Neuronalen Netze in Abschnitt 2.6.2.2 kurz vorgestellt, die aktuell den mit Abstand populärsten Modelltyp darstellen.

2.6.2.1 Polynomsplines

Mittels Polynomsplines können komplexere Zusammenhänge auf Basis von Polynomen modelliert werden, ohne dass der Grad der Polynome hoch sein muss. Bei Polynomsplines wird der Wertebereich der Einflussfaktoren in mehrere Abschnitte unterteilt, in denen dann jeweils separate Polynome auf die Messdaten angepasst werden. Die Grenzen dieser Abschnitte werden Knoten genannt. Als zusätzliche

Bedingung wird gefordert, dass die Polynomsplines auch in den Knoten stetig und $p - 1$ mal stetig differenzierbar sind, wenn p der Grad der verwendeten Polynome ist (Fahrmeir et al., 2009, S. 295). Dies lässt sich durch verschiedene Verfahren erreichen.

Eine einfache Möglichkeit besteht darin, ein global gültiges mehrdimensionales Polynom p -ten Grades so zu erweitern, dass die Koeffizienten der Terme mit der höchsten Potenz an den Knoten angepasst werden. Dadurch, dass die Terme mit niedrigeren Potenzen an den Knoten nicht verändert werden, wird die geforderte Stetigkeit und stetige Differenzierbarkeit automatisch erfüllt. Die Anpassung der Terme der mit der höchsten Potenz erfolgt durch *trunkierte Potenzen*. Die so generierte Splines werden deshalb *TP-Splines* genannt.

Die trunkierte Potenz für den j -ten Knoten k_j ist gemäß Gleichung 2.58 definiert.

$$(x - k_j)_+^p = \begin{cases} (x - k_j)^p & x \geq k_j \\ 0 & x < k_j \end{cases} \quad 2.58$$

Für einen TP-Spline mit zwei Knoten, der auf einem Polynom zweiten Grades basiert, ergibt sich damit das Regressionsmodell nach Gleichung 2.59.

$$y_i = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 (x - k_1)_+^2 + \beta_5 (x - k_2)_+^2 + \varepsilon \quad 2.59$$

Ein mögliches Aussehen eines TP-Splines gemäß Gleichung 2.59 ist in Abbildung 2.26 dargestellt.

TP-Splines besitzen den Vorteil, dass sowohl ihre Berechnung, als auch die Interpretierbarkeit der Modellgleichung einfach ist. Zudem können ihre Koeffizienten β mit der gewöhnlichen Methode der kleinsten Fehlerquadrate⁵¹ bestimmt werden. Wenn jedoch viele Knoten in die Modellgleichung aufgenommen werden, entsteht das Problem, dass für große x -Werte viele trunkierte Potenzen gleichzeitig aktiv sind. Dadurch entsteht ein verhältnismäßig großer Rechenaufwand. In solchen Fällen sind *B-Splines* zu bevorzugen. Details zu der Erstellung dieser Splines und einen Überblick über weitere Verfahren liefern Fahrmeir et al. (2009, S. 293–325, 371–378), de Boor (1978) und Dierckx (1995).

⁵¹ Siehe Abschnitt 2.5.2.1

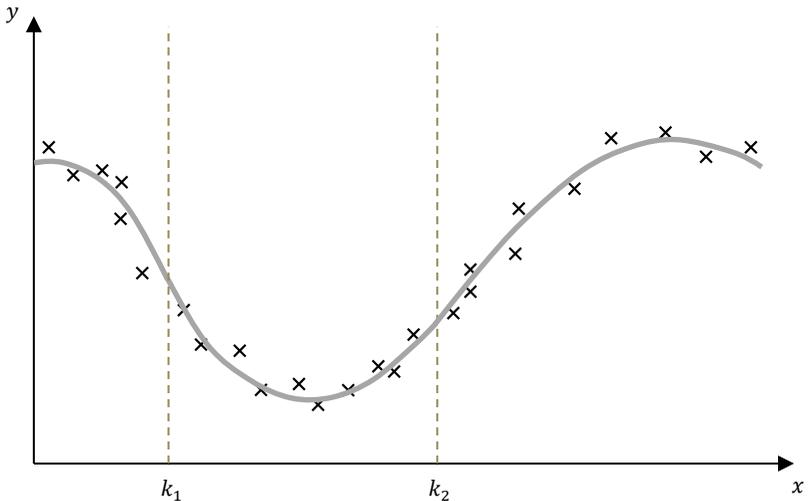


Abbildung 2.26: TP-Spline mit zwei Knoten

Die Abbildung von mehr als einem Faktor ist in Polynomsplines ebenfalls möglich. Die Wahl der Knotenpunkte wird dann allerdings so komplex, dass sie mittels eines Algorithmus vorgenommen werden sollte. Ein solcher Algorithmus, der sehr häufig eingesetzt wird (*Multivariate Adaptive Regression Splines, MARS*), wird von Friedman (1991) beschrieben.

2.6.2.2 Künstliche neuronale Netze

Künstliche neuronale Netze (KNN) sind nichtparametrische Modelle, die aus mehreren Faktoren die Werte einer oder mehrerer Zielgrößen berechnen. Einen großen Anwendungsbereich finden sie im Gebiet der künstlichen Intelligenz. Ihr Aufbau wurde durch das biologische Nervensystem inspiriert. Die grundlegende Idee künstlicher neuronaler Netze wurde bereits 1943 von McCulloch und Pitts (1943) entwickelt. Eine größere Verbreitung und Verwendung fanden sie allerdings erst mit dem Aufkommen moderner Computersysteme. Ein einfaches KNN stellt das *Multilayer Perceptron (MLP)* dar, anhand dessen die Funktionsweise künstlicher neuronaler Netze gut erklärt werden kann.

In einem MLP sind die *Neuronen* in mehreren Ebenen gemäß Abbildung 2.27 angeordnet. Neuronen sind in dieser Abbildung als Kreise dargestellt. Jedes Neuron n besitzt einen Zahlenwert a , der *Aktivierung* genannt wird. Die Neuronen der *Eingangsebene* beinhalten als Aktivierung die Einstellungen der Faktoren x , in deren Abhängigkeit die Zielgröße y berechnet werden soll. Die Aktivierung des Neurons der *Ausgangsebene* repräsentiert die berechnete Zielgröße \hat{y} . Üblicherweise kann die Aktivierung a Zahlenwerte zwischen null und eins annehmen. Deshalb liegen sowohl die Faktoren, als auch die Zielgröße in einer auf diesen Bereich normierten Form vor.

Zwischen der Eingangs- und der Ausgangsebene befinden sich eine oder mehrere *versteckte Ebenen*. Die Aktivierung jedes Neurons einer Ebene (abgesehen von der Eingangsebene) berechnet sich aus den Aktivierungen aller Neuronen der jeweils vorhergehenden Ebene. Dies wird durch die Linien in Abbildung 2.27 symbolisiert. Dabei fließen die Aktivierungen aller Neuronen der vorhergehenden Ebene zunächst linear in die Aktivierung jedes Neurons der nachfolgenden Ebene ein. Allerdings werden die Einflüsse der verschiedenen Neuronen der vorhergehenden Ebene mit unterschiedlichen Gewichtungsfaktoren b multipliziert. Zusätzlich besitzt die Aktivierung jedes Neurons ein Offset b_0 , das seine Aktivierung unabhängig von den Aktivierungen der vorhergehenden Ebene verschiebt. Anschließend wird der so berechnete Wert mittels einer *sigmoidalen Funktion* $sigm()$ auf einen Wertebereich zwischen null und eins normiert.

Dies führt zu Gleichung 2.60, die die Aktivierung eines Neurons der Ebene $k + 1$ in Abhängigkeit der Aktivierungen aller Neuronen der Ebene k beschreibt.

$$a_{k+1} = sigm(b_{k,0} + b_{k,1}a_{k,1} + b_{k,1}a_{k,1} + \dots + b_{k,n_e}a_{k,n_e}) \quad 2.60$$

Als sigmoidale Funktion kann beispielsweise Gleichung 2.61 zum Einsatz kommen. (Siebertz et al., 2010, S. 210)

$$sigm(x) = \frac{1}{1 + e^{-x}} \quad 2.61$$

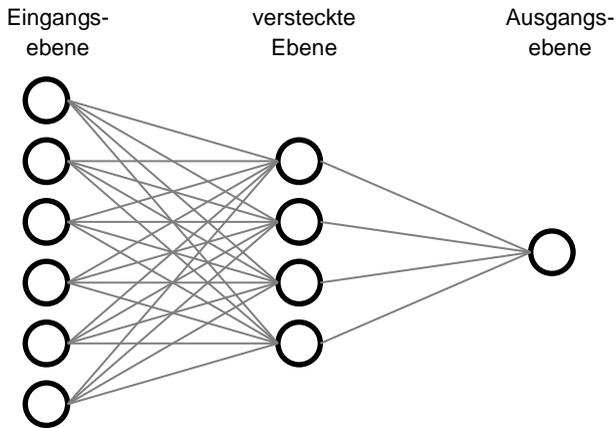


Abbildung 2.27: Schematische Darstellung eines neuronalen Netzes
 In Anlehnung an (Siebertz et al., 2010, Abb. 8.11)

Mittels dieser Formeln lassen sich zunächst aus den Aktivierungen der Neuronen der Eingangsebene die Aktivierungen der ersten versteckten Ebene berechnen. Für alle darauf folgenden Ebenen läuft die Berechnung nach demselben Schema ab. So erhält man letztlich die Aktivierung des Neurons der Ausgangsebene, die einen normierten Schätzwert für die Zielgröße liefert.

Falls künstliche neuronale Netze nicht zur Regression sondern zur Klassifizierung genutzt werden sollen, beinhaltet die Ausgangsebene für jede bestimmbare Klasse ein Neuron. Die Aktivierung der Neuronen der Ausgangsebene gibt dann wieder, für wie wahrscheinlich das KNN die Zugehörigkeit des zu bestimmenden Objekts zur jeweiligen Klasse hält.

Der Aufbau eines KNN (und damit die Form des funktionalen Zusammenhangs zwischen Eingangs- und Ausgangsebene) wird festgelegt, bevor es an die abzubildenden Datensätze angepasst wird. Das Anpassen des KNN wird *Trainieren* genannt. Dabei werden die Gewichtungsfaktoren und Offsets so optimiert, dass die berechnete Zielgröße \hat{y} für alle Elemente des Trainingsdatensatzes möglichst gut zu den darin enthaltenen realen Werten y passt. Eine Zusammenfassung des hierbei verwendeten Optimierungsverfahrens wird beispielsweise von Siebertz et al. (2010, Kap. 8.8) gegeben.

2.6.3 Semiparametrische Regressionsmodelle

Nichtparametrische Regressionsmodelle können mit linearen Regressionsmodellen kombiniert werden. Dadurch entstehen *semiparametrische Regressionsmodelle*. In solchen Modellen können stark nichtlineare Faktoren durch ein nichtparametrisches Teilmodell abgebildet werden, während weniger komplexe Abhängigkeiten durch ein lineares Teilmodell angenähert werden. Anwendung finden semiparametrische Regressionsmodelle, wenn für die Modellierung verschiedener Faktoren sehr unterschiedliche Auflösungsgrade benötigt werden.

Es existieren mehrere Ansätze, wie sich beide Modelltypen miteinander kombinieren lassen. Sehr häufig eingesetzt werden *additive Modelle*, bei denen beide Modellanteile separat modelliert und anschließend addiert werden. Beschreibt beispielsweise f ein nichtparametrisches Modell mit den darin abzubildenden Faktoren \mathbf{x}_{np} und $\mathbf{x}_{lin}\boldsymbol{\beta}$ ein lineares Modell mit den darin abzubildenden Faktoren \mathbf{x}_{lin} , dann kann ein additives Modell gemäß Gleichung 2.62 formuliert werden.⁵²

$$y = f(\mathbf{x}_{np}) + \mathbf{x}_{lin}\boldsymbol{\beta} + \varepsilon \quad 2.62$$

Alternativ können semiparametrische Modelle erstellt werden, in denen die Eingangsgrößen des nichtparametrischen Modells durch ein lineares Modell beschrieben werden. Diese Modelle werden *Index Modelle* genannt. Wenn f nur eine Eingangsgröße besitzt, die durch ein lineares Modell beschrieben wird, dann führt dies zu einem *Single Index Modell* nach Gleichung 2.63. Es sind auch flexiblere Ansätze gemäß Gleichung 2.64 möglich.⁵³

$$y = f(\mathbf{x}_{lin}\boldsymbol{\beta}) + \varepsilon \quad 2.63$$

$$y = f(\mathbf{x}_{np}, \mathbf{x}_{lin}\boldsymbol{\beta}) + \varepsilon \quad 2.64$$

In einer dritten Variante werden die Koeffizienten $\boldsymbol{\beta}$ des linearen Modells durch nichtparametrische Regressionsmodelle $\boldsymbol{\beta}(\mathbf{x}_{np})$ modelliert, die die Werte von $\boldsymbol{\beta}$ in

⁵² Für Details siehe (Fahrmeir et al., 2009, Kap. 8)

⁵³ Für Details siehe (Ichimura, 1993)

Abhängigkeit der nichtparametrisch modellierten Faktoren x_{np} beschreiben. Dies führt zu *Varying-Coefficient-Modellen* gemäß Gleichung 2.65.⁵⁴

$$y = x_{lin}\boldsymbol{\beta}(x_{np}) + \varepsilon \quad 2.65$$

Zur Bestimmung der Modellparameter semiparametrischer Regressionsmodelle können, in Abhängigkeit von der Art des semiparametrischen Regressionsmodells und von der Art des nichtparametrischen Modellteils, sehr viele verschiedene Verfahren zum Einsatz kommen. Aufgrund der Vielzahl und oft auch großen Komplexität der Verfahren wird auf diese hier nicht näher eingegangen, sondern auf weiterführende Literatur⁵⁵ verwiesen.

⁵⁴ Für Details siehe (Hastie & Tibshirani, 1993)

⁵⁵ Siehe z.B. (Chen & Tsay, 1993; Fan, 2018; Hastie, 2017; Hastie & Tibshirani, 1993; Ichimura, 1993; Ruppert et al., 2003; Speckman, 1988; Xia et al., 2002)

3 Motivation und Zielsetzung der Arbeit

Wie in Kapitel 1 erläutert, besitzen einfache Simulationsmodelle bei der Entwicklung mechatronischer Produkte eine große und weiter stetig steigende Bedeutung. In einigen Fällen können solche Simulationsmodelle nicht allein auf Basis physikalischer Gleichungen in Form eines White-Box-Modells erstellt werden. Oft liegt das daran, dass nicht alle physikalischen Zusammenhänge und Abläufe ausreichend genau bekannt sind, um sie so abbilden zu können. Es kann aber auch vorkommen, dass eine solche Modellierung zu aufwendig wäre oder zu rechenintensive Modelle als Ergebnis hätte. In all diesen Fällen kann die Zuhilfenahme von empirischer Modellbildung eine Lösung sein. Wenn die empirisch abzubildenden Verhaltensanteile in Abhängigkeit von mehreren Faktoren modelliert werden sollen, dann können zwei unterschiedliche Gruppen von Verfahren eingesetzt werden. Zum einen sind dies die linearen Regressionsmodelle aus den Abschnitten 2.5.2 bis 2.5.4 zusammen mit den Versuchsplänen aus Abschnitt 2.5.1. Diese stellen die „klassische“ statistische Versuchsplanung dar. In immer mehr Veröffentlichungen der letzten Jahre wird hingegen die zweite Gruppe von Verfahren eingesetzt. Diese besteht aus den nichtparametrischen Regressionsmodellen aus Abschnitt 2.6.2 kombiniert mit den Versuchsplänen aus Abschnitt 2.6.1.

Nichtparametrische Regressionsmodelle besitzen den großen Vorteil, dass sie auf sehr komplexe Zusammenhänge zwischen den Ein- und Ausgangsgrößen des zu modellierenden Systems angepasst werden können. Sie sind in der Lage, beliebig geformte Nichtlinearitäten im mehrdimensionalen Raum annähern zu können, während lineare Regressionsmodelle üblicherweise auf ein mehrdimensionales Polynom zweiten Grades beschränkt sind. Nichtparametrische Regressionsmodelle können deshalb für die Modellierung beliebiger technischer Systeme mit stationärem Verhalten eingesetzt werden. Wenn hingegen lineare Modelle auf Systeme mit zu komplexem Verhalten angepasst werden, so entsteht Lack-of-fit, woraus beträchtliche Abweichungen zwischen Modell und Realität resultieren können.¹ Wenn ungeeignete Versuchspläne und unzureichende Validierungsmaßnahmen² eingesetzt werden, können diese Abweichungen zudem unerkant bleiben.

Der scheinbar eindeutige Vorteil der höheren Auflösung nichtparametrischer Regressionsmodelle kommt allerdings für die Modellbildung in der Produktentwicklung in vielen Fällen nicht zum Tragen. Direkt mit dem Auflösungsvermögen eines Regressionsmodells verknüpft ist nämlich der Versuchsaufwand zur Durchführung der

¹ Vgl. S.28

² Siehe Abschnitt 2.5.2.3

Versuchsreihe, auf deren Ergebnisse das Modell angepasst wird. Dieser darf für einen Einsatz in der Produktentwicklung den Nutzen des erstellten empirischen Modells nicht übersteigen. Dadurch ist der mögliche Versuchsumfang einer Versuchsreihe begrenzt. Bei Versuchsplänen für nichtparametrische Regressionsmodelle steigt der Versuchsumfang im Allgemeinen exponentiell mit der Faktoranzahl. Im Gegensatz dazu wird in der statistischen Versuchsplanung für lineare Regressionsmodelle durch gezielte Vernachlässigung von Wechselwirkungen höherer Grade eine deutlich geringere Steigerung des Versuchsumfanges erreicht. Trotzdem wird auch hier für Versuchspläne für Polynome zweiten Grades bereits schnell die Grenze des praktikablen Versuchsumfanges erreicht. Dies bedeutet, dass beim Einsatz in der Produktentwicklung die Abbildungsleistung empirischer Modelle für höhere Anzahlen an Faktoren meist nicht durch die Fähigkeiten des Regressionsmodells, sondern durch den möglichen Versuchsumfang begrenzt wird.

Der größte Vorteil linearer Regressionsmodelle beim Einsatz in der Produktentwicklung ist, dass sie sehr gut verständlich sind. Dies resultiert daraus, dass lineare Modelle durch Polynome beschrieben werden. Dieser Gleichungstyp wird bereits in der Mathematik auf weiterführenden Schulen intensiv behandelt. Im Ingenieursstudium wird dieses Wissen in verschiedensten Zusammenhängen regelmäßig aufgefrischt, vertieft und angewandt. Ingenieure können dadurch die Effekte linearer und quadratischer Terme einzelner Einflussfaktoren nahezu intuitiv interpretieren. Ein Verständnis für Wechselwirkungsterme können sie in kurzer Zeit aufbauen, falls sie es nicht bereits besitzen. Damit lassen sich alle Bestandteile und Einflüsse eines linearen Modells schnell verstehen, unabhängig davon, wie viele Einflussfaktoren das Modell besitzt. Zudem kann beim Einsatz normierter Faktoren die Größe der Einflüsse aller Faktoren direkt aus ihren Koeffizienten abgelesen werden. Im Gegensatz dazu besitzen nichtparametrische Regressionsmodelle, wie beispielsweise künstliche neuronale Netze, keinen für Menschen verständlichen Zusammenhang zwischen ihren Parametern und den durch sie verursachten Effekte. Lineare Regressionsmodelle sind hingegen bei mehr als zwei Einflussfaktoren sogar einfacher verständlich als Kennfelder, da diese sich für mehr als zwei Faktoren nicht mehr sinnvoll graphisch darstellen lassen.

Durch die gute Verständlichkeit linearer Regressionsmodelle erhält der Entwickler beim ihrem Einsatz neben dem reinen Simulationsmodell des abgebildeten Systems auch einen Überblick über dessen Verhalten. Aus der Modellgleichung kann er direkt ablesen, welche Faktoren es in Richtung ungünstiger Betriebszustände (z.B. höhere Reibung oder niedrigere Effizienz) beeinflussen. Diese Informationen kann er beispielsweise dazu nutzen, die Grenzen der Einsatzbedingungen des Produkts festzulegen. Bei der Validierung des Produkts kann er hingegen auch gezielt diese ungünstigen Betriebszustände erzeugen und so die Zuverlässigkeit des Produkts

unter allen Einsatzbedingungen sicherstellen. Durch analytisches Denken können zudem aus den im Modell abgebildeten Effekten die Mechanismen abgeleitet werden, die zu diesen Effekten führen. Beispielsweise könnte aus einem großen Einfluss der Temperatur auf einen Reibungsvorgang darauf geschlossen werden, dass unterschiedliche thermische Ausdehnungen verschiedener Systemkomponenten die Normalkräfte im Reibkontakt beeinflussen. Aus solchen Informationen können wiederum Optimierungsmöglichkeiten für zukünftige Generationen des abgebildeten Produkts hergeleitet werden.

Lineare Regressionsmodelle besitzen zudem den Vorteil, dass sie einfach anzupassen sind. Ergeben sich Änderungen in neuen Produktgenerationen, die sich auf das Verhalten des Produkts auswirken, dann muss üblicherweise ein vollständig neues empirisches Modell erstellt werden. Wenn allerdings die Auswirkungen der Änderungen bekannt sind, lassen sich diese durch Parameteranpassungen einfach in ein lineares Regressionsmodell integrieren.

Die oben genannten positiven Effekte kommen daher, dass lineare Modelle für Ingenieure gut verständlich sind. Diese Verständlichkeit hat in der Produktentwicklung eine große Wirkung. Diese kommt daher, dass wie Albers et al. (2019c, Grundprinzip 1) treffend formuliert haben, der Mensch im Zentrum der Produktentwicklung steht:

„Durch sein kreatives Schaffen ist der Entwickler für die Entwicklung erfolgreicher Produkte verantwortlich. Für eine bestmögliche Unterstützung des Entwicklers bei der Durchführung seiner Aktivitäten müssen Prozesse und Methoden an seine Kreativität, Kompetenzen, Bedarfe und kognitiven Fähigkeiten angepasst werden. Zudem muss auch die Methodenauswahl hinsichtlich dieser Kriterien erfolgen.“

Aus dieser Erkenntnis folgt in Bezug auf die empirische Modellbildung in der Produktentwicklung die Forschungshypothese für diese Arbeit:

Forschungshypothese

Lineare Regressionsmodelle sind sehr gut an die Kompetenzen, Bedarfe und kognitiven Fähigkeiten von Produktentwicklern angepasst. Ihr Einsatz als Simulationsmodelle in der Produktentwicklung kann dadurch eine Vielzahl an Vorteilen gegenüber anderen Modelltypen bieten. Es ist deshalb sinnvoll, Lösungen für bestehende Einschränkungen ihrer Nutzbarkeit in diesem Kontext zu suchen.

Im Rahmen dieser Arbeit wurden drei grundlegende Einschränkungen für den Einsatz linearer Regressionsmodelle in der Produktentwicklung identifiziert.³ Wie diese gelöst werden können, ist Inhalt der drei nachfolgenden Forschungsfragen.

Ein großes Problem linearer Regressionsmodelle liegt in ihrer begrenzten Abbildungsleistung von Nichtlinearitäten. Deshalb muss ihre Abbildungsleistung gesteigert werden. Ein erster Schritt hierzu besteht darin, bestehendes Vorwissen über die physikalischen Zusammenhänge des abzubildenden Vorgangs in das Modell zu integrieren. Zwar kommt empirische Modellbildung in der Produktentwicklung üblicherweise dann zum Einsatz, wenn für die Erstellung eines reinen White-Box-Modells kein ausreichendes Wissen über den abzubildenden physikalischen Vorgang verfügbar ist. Meist ist aber auch in diesen Fällen zumindest Wissen über einzelne Teilaspekte des Systemverhaltens vorhanden. Wenn dieses Wissen mit linearen Modellen kombiniert wird, dann kann darin oft ein großer Teil der bestehenden Nichtlinearitäten abgebildet werden. Zudem wäre es sehr ineffizient, dieses Wissen ungenutzt zu lassen. Dadurch würde Versuchsaufwand zur Ermittlung bereits bekannter Zusammenhänge verschwendet. Daraus folgt die erste Forschungsfrage.

Forschungsfrage 1

Wie kann physikalisches Vorwissen in empirisch ermittelte lineare Modelle integriert werden?

Nichtlinearitäten, die nicht physikalisch basiert abgebildet werden können, müssen in den empirischen Modell-Anteil übernommen werden. Lineare Modelle sind hierzu jedoch, wie oben beschrieben, nur begrenzt geeignet. Dies führt zur zweiten Forschungsfrage.

Forschungsfrage 2

Wie kann die Abbildungsleistung linearer Modelle für starke Nichtlinearitäten einzelner Faktoren gesteigert werden?

Die dritte Forschungsfrage ergibt sich daraus, dass in Versuchsreihen oft einzelne Faktoren existieren, deren Einstellung sehr viel aufwendiger ist, als die der anderen

³ Siehe Kapitel 4 sowie Einleitungen der Kapitel 6, 7 und 8

Faktoren. Wenn eine solche Versuchsreihe so strukturiert wird, dass diese Faktoren möglichst selten geändert werden müssen, lässt sich der notwendige Versuchsaufwand deutlich reduzieren. Dies führt zu Split-Plot-Versuchsreihen, die oft nur einen Bruchteil des Versuchsaufwands vollständig randomisierter Versuchsreihen benötigen.⁴ Im industriellen Umfeld sind sehr oft Situationen anzutreffen, die den Einsatz von Split-Plot-Versuchsreihen erfordern. Cuthbert Daniel formulierte sogar etwas überspitzt:

„All industrial experiments are split-plot experiments“

Undatierte mündliche Aussage von Cuthbert Daniel,
zitiert von George E.P. Box (2005, S. 336)

In den Standardwerken (z.B. Kleppmann, 2011; Montgomery, 2009; Siebertz et al., 2010; Voss & Dean, 1999) zur statistischen Versuchsplanung werden Split-Plot-Versuchspläne trotzdem entweder gar nicht oder nur am Rande vorgestellt. Die wichtigsten Entwicklungen der letzten Jahre werden in diesen Büchern noch nicht berücksichtigt. Da sich unter den Entwicklungen der letzten Jahre unterschiedliche Verfahren befinden, folgt als dritte Forschungsfrage:

Forschungsfrage 3

Wie sollten Versuchsreihen zur Modellierung technischer Systeme in der Produktentwicklung strukturiert werden, um mit möglichst geringem Aufwand möglichst detaillierte und valide Ergebnisse zu liefern?

Die Beantwortung der drei oben genannten Forschungsfragen ist das Ziel dieser Arbeit.

⁴ Vgl. Abschnitt 2.5.5

4 Forschungsdesign

Jede der drei Forschungsfragen beinhaltet, wie ein spezieller Bedarf der Produktentwicklung an empirische Simulationsmodelle durch lineare Regressionsmodelle und die damit verknüpfte statistische Versuchsplanung gedeckt werden kann. Zur Beantwortung der Forschungsfragen werden Methodenbausteine benötigt, mit denen sich die bestehenden Verfahren erweitern lassen. Die Entwicklung dieser Methodenbausteine fällt in Kategorie II der entwicklungsmethodischen Forschungskategorien nach Cantamessa (2003, siehe Tabelle 4.1).

Tabelle 4.1: Fünf Kategorien von entwicklungsmethodischer Forschung nach Cantamessa (2003)

I	Empirical research, in which researchers analyze real-world design processes.
II	Development of new tools and methods for supporting the design process or elements of it.
III	Experimental research, in which researchers purposely set up design processes in a controlled environment.
IV	Implementation studies, in which researchers discuss the real-world deployment of innovative methods and tools.
V	Other, which includes papers dedicated to theory and education.

Marxen und Albers (2012) und Marxen (2014) bringen die fünf Kategorien nach Cantamessa in einen zeitlichen Zusammenhang und entwickeln so ein Modell, das die Entstehung von entwicklungsmethodischer Unterstützung in fünf Phasen beschreibt. Dieses Modell ist in Abbildung 4.1 graphisch dargestellt.

Den Anstoß für die vorliegende Arbeit gaben Schwierigkeiten bei der Modellbildung (vgl. Punkt „Difficulties in engineering design practice“ in Abbildung 4.1) für die im Praxisbeispiel aus Kapitel 5 beschriebenen Produkte. Dabei fiel auf, dass die aufgetretenen Schwierigkeiten nicht spezifisch für diese Komponenten sein können, sondern in vielen Fällen bei der empirischen Ermittlung von Simulationsmodellen technischer Systeme auftreten müssen. Dies führte zur in Abschnitt 3 näher beschriebenen Motivation und Zielsetzung dieser Arbeit.

Die Beantwortung der drei Forschungsfragen erfolgt in den Kapiteln 6 bis 8. Der Beantwortung jeder Frage wird jeweils ein Kapitel gewidmet. In jedem dieser Kapitel werden zunächst die Schwierigkeiten genauer beschrieben, die eine Beantwortung der zugehörigen Forschungsfrage erforderlich machen (Phase I in Abbildung 4.1).

Anschließend werden Methodenbausteine erarbeitet, die die Forschungsfrage zu beantworten versuchen (Phase II in Abbildung 4.1). Abschließend wird die Anwendung dieser Methodenbausteine am Praxisbeispiel aus Kapitel 5 demonstriert. Dies dient einerseits dazu, die Methodenbausteine und ihre Verwendung plastischer darzustellen. Andererseits werden damit aber auch die so erzielbaren Ergebnisse demonstriert (Phase III in Abbildung 4.1).

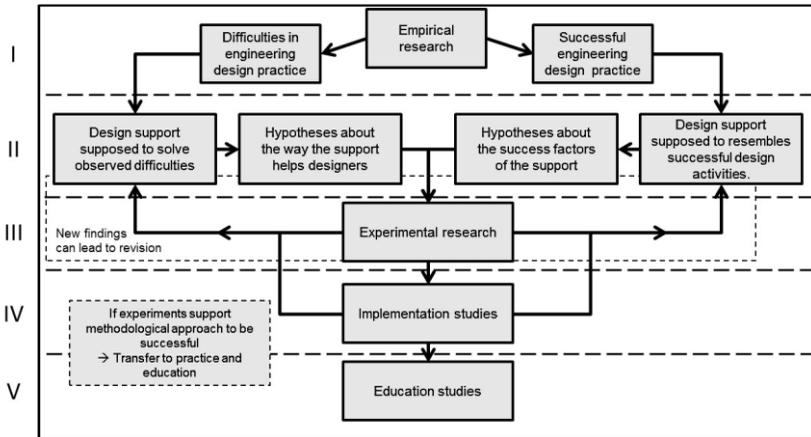


Abbildung 4.1: Entwicklungspfade entwicklungsmethodischer Unterstützung (Marxen, 2014)

Zur Beantwortung der Forschungsfragen werden in dieser Arbeit keine neuen statistischen Methoden entwickelt. Stattdessen wurde zur Beantwortung dieser Fragen eine breit angelegte Literaturrecherche durchgeführt. In dieser wurden Methoden der statistischen Versuchsplanung und Modellbildung mittels linearer Regressionsmodelle aus anderen Wissenschaftsbereichen untersucht. Bei potentieller Eignung zur Beantwortung der Forschungsfragen wurden diese in den in Kapitel 2 zusammengefassten relevanten Stand der Forschung aufgenommen. Die hierbei ermittelten Verfahren und Modelle werden in den Kapiteln 6 bis 8 auf die Anwendung zur Erstellung von Simulationsmodellen in der Produktentwicklung übertragen. Die mit ihnen erzielbaren Erkenntnisse werden anhand des Praxisbeispiels illustriert.

5 Praxisbeispiel: pneumatisch angetriebene Prozessarmaturen

Die Anwendung von DoE-Methoden im Produktentwicklungsprozess wird in dieser Arbeit anhand eines Beispiels aus der Prozessautomation demonstriert. In diesem Beispiel soll das dynamische Verhalten von pneumatisch angetriebenen Prozessarmaturen modelliert werden.

Prozessarmaturen sind Ventile, die zum Steuern und Regeln des Durchsatzes von Gas-, Flüssigkeits- oder Schüttgutströmen in Rohrleitungssystemen industrieller Anlagen eingesetzt werden. Als Bauformen kommen beispielsweise Kugelhähne, Klappen, Schieber oder verschiedene Sitzventile zum Einsatz. Prozessarmaturen werden oft pneumatisch angetrieben, da dieser Antriebstyp günstig, robust und sicher ist (Nesbitt, 2007, S. 398). Eine pneumatisch angetriebene Prozessarmatur für Schaltanwendungen besteht aus der Armatur, dem pneumatischen Antrieb, und einem Pilotventil, das die Luftzufuhr des Antriebs schaltet. Oft werden auf dem Antrieb zusätzlich Sensorboxen montiert, die die aktuelle Armaturstellung über Endlagenschalter erfassen. Wenn durch die Armatur der Massenstrom in einer Rohrleitung stufenlos geregelt werden soll, wird statt des Pilotventils und der Sensorbox ein Stellungsregler („Positioner“) auf dem pneumatischen Antrieb installiert. Dieser besteht aus einem Winkelsensor, pneumatischen Regelventilen und einer Regelelektronik. Eine Darstellung eines solchen Systems findet sich in Abbildung 5.1.

Für verschiedene Anwendungen in der Entwicklung solcher Systeme werden Simulationsmodelle benötigt. Details hierzu werden in Abschnitt 5.1 erläutert. Das in solchen Modellen abzubildende dynamische Verhalten wird sowohl durch den Antrieb, als auch durch die Prozessarmatur beeinflusst. Um ein funktionsfähiges Simulationsmodell des dynamischen Verhaltens des Gesamtsystems zu erhalten, müssen also beide Komponenten modelliert werden.

Es existieren sehr viele verschiedene Arten von Prozessarmaturen. In dieser Arbeit wird stellvertretend die Modellbildung von Kugelhähnen beschrieben. Der Aufbau und die Funktionsweise von Kugelhähnen wird in Abschnitt 5.2 beschrieben. Passend zu diesem Armaturtyp werden pneumatische Schwenkantriebe als zweite zu modellierende Komponente gewählt. Eine Beschreibung ihres Aufbaus und ihrer Funktionsweise findet sich in Abschnitt 5.3.

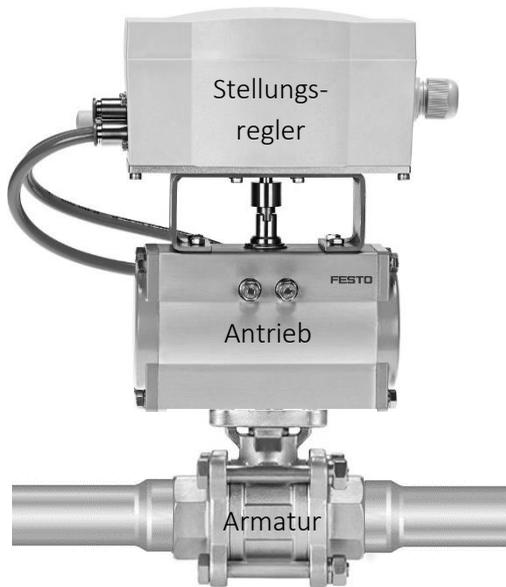


Abbildung 5.1: Pneumatisch angetriebener Kugelhahn mit Stellungsregler

5.1 Simulationsmodelle pneumatisch angetriebener Prozessarmaturen

Vor allem für die Entwicklung neuer Produktgenerationen von Positionern werden detaillierte Verhaltensmodelle der Prozessarmaturen inklusive ihres Antriebs benötigt. Die Modelle können einerseits zur modellbasierten Entwicklung und Auslegung von Regelungs- und Vorsteuerungsalgorithmen verwendet werden. Andererseits ermöglichen direkt auf dem Positioner lauffähige Modelle ganz neue Funktionalitäten, wie beispielsweise modellprädiktive Regelungen (MPC)¹ oder Condition Monitoring bzw. Predictive Maintenance Anwendungen. Unter Condition Monitoring wird die Überwachung des aktuellen Zustands eines technischen Systems verstanden. Mit diesen Informationen kann eine vom Anlagenalter unabhängige Aussage darüber getroffen werden, zu welchem Zeitpunkt welche Arten von Wartungsarbeiten notwendig werden (Predictive Maintenance, für Details siehe Thomas & Davies, 1993, S. 134). Da Stillstände beispielsweise von chemischen Produktionsanlagen sehr

¹ Aus dem Englischen für *model predictive control*

hohe Kosten nach sich ziehen, generieren diese Ansätze bei Produkten für die Prozessautomation einen sehr hohen Mehrwert. Mit ihrer Hilfe lassen sich Wartungsintervalle verlängern und gleichzeitig ausfallbedingte Anlagenstillstände vermeiden.

Dadurch, dass immer mehr Funktionen in Positioner integriert werden, wird auch ihre Validierung immer aufwendiger. Alle Funktionen sollen für verschiedenste Kombinationen aus Positionern, Antrieben und Armaturen unter allen relevanten Betriebsbedingungen gewährleistet sein. Um dies umfassend validieren zu können, sind sehr umfangreiche Versuchsreihen notwendig. Für diese Versuchsreihen müssen große Massenströme und hohen Druckdifferenzen erzeugt werden können, was die notwendigen Tests aufwendig, teuer und auch potentiell gefährlich macht. Um den zur Freigabe notwendigen Versuchsumfang wirtschaftlich und sicher bewältigen zu können, ist es notwendig, einen Teil der Versuche auf XiL-Prüfstände auszulagern, auf denen ein variabler Anteil des zu untersuchenden Systems auf einer Simulationshardware simuliert wird (vgl. Abschnitt 2.2.1). Dies trifft insbesondere auf Tests mit erhöhtem Gefährdungspotential (z.B. hohe Drücke und / oder Massenströme) und sehr zeitaufwendige Tests (z.B. Losbrechmoment nach langem Stillstand) zu.

In allen oben genannten Anwendungsfällen ist es hilfreich, wenn ein Modell nicht nur das durchschnittliche Verhalten des abgebildeten Systems wiedergeben kann, sondern auch Informationen darüber beinhaltet, wie stark dieses Verhalten unter normalen Bedingungen streuen kann. Nur so lassen sich ausreichend robuste Algorithmen auf Basis der Modelle entwickeln, bzw. prüfen, ob bestehende Algorithmen ausreichend robust sind. Sowohl Modelle, die auf dem Positioner lauffähig sein sollen, als auch solche für XiL-Prüfstände, müssen in Echtzeit berechnet werden können. Aufwendig zu berechnende Modelle, wie beispielsweise Finite Elemente Modelle, können hier deshalb nicht eingesetzt werden.

5.2 Kugelhähne

Kugelhähne sind sehr häufig vorkommende Prozessarmaturen. Sie sind im Vergleich zu Klappen und anderen Armaturtypen günstig und besitzen eine sehr gute Dichtheit in geschlossener Stellung. Im geöffneten Zustand verfügen Kugelhähne über den vollen Strömungsquerschnitt des angrenzenden Rohrleitungssystems ohne störende Elemente in der durchströmten Fläche. Damit erzeugen sie im geöffneten Zustand einen sehr geringen Druckabfall und sind zudem molchbar. Dies bedeutet, dass sie von sogenannten Molchen, also meist rotationssymmetrischen Laufkörpern, durchfahren werden können. Diese Molche werden meist von der Strömung durch Rohrleitungssysteme getrieben und können verschiedene Aufgaben

übernehmen. Dazu zählt beispielsweise das Reinigen oder das Inspizieren der Rohrleitung oder das Trennen verschiedener Flüssigkeiten (Skerra, 2000).

Der Aufbau eines Kugelhahns ist in Abbildung 5.2 dargestellt. Kugelhähne bestehen aus einer Ventilkugel, einer Ventilspindel, Primär- und Spindeldichtungen und einem Gehäuse. Die Ventilkugel besitzt eine Bohrung, die gerade durch die Kugel hindurch verläuft. An ihrer Oberseite ist ein Verbindungselement angebracht, über das sie mit der Ventilspindel verbunden wird. Meist ist dieses Verbindungselement als halbrunde Nut ausgeführt, in die ein entsprechend geformtes Gegenstück der Ventilspindel eingeführt wird. Über dieses Element sind Ventilspindel und Ventilkugel für Drehbewegungen formschlüssig verbunden. Die Ventilspindel wird aus dem Gehäuse herausgeführt und dabei von der Spindeldichtung geführt und abgedichtet. Die Ventilspindel wird aus dem Gehäuse herausgeführt und dabei von der Spindeldichtung geführt und abgedichtet.

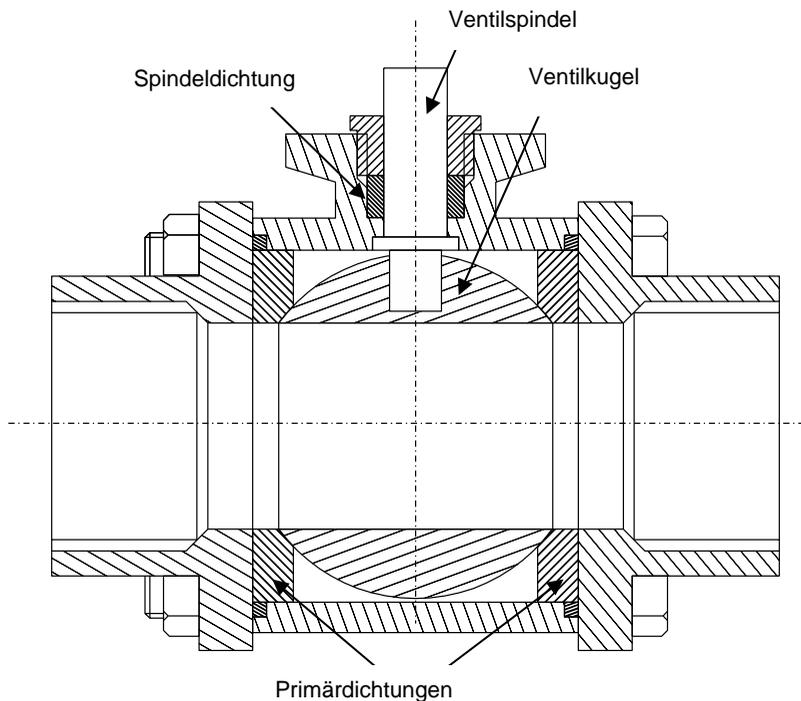


Abbildung 5.2: Aufbau eines Kugelhahns

Meist ist die Spindeldichtung als Dachmanschettenpaket oder Stopfbuchspackung ausgeführt. Über die Ventilspindel kann die Ventilkugel gedreht werden, wodurch der Weg für Fluide durch den Kugelhahn geöffnet oder geschlossen wird. Die Ventilkugel ist in Kugelhähnen meist schwimmend zwischen den beiden Primärdichtungen gelagert. Die Primärdichtungen besitzen in Rohrleitungsrichtung ein Übermaß, sodass der Kontakt zwischen Kugel und Dichtungen vorgespannt ist. Diese Vorspannung sorgt dafür, dass der Kugelhahn in geschlossener Stellung dicht ist, auch wenn keine Druckdifferenz anliegt.

Die Effekte beim Anlegen einer Druckdifferenz können anhand der C&C²-A Modellierung gemäß Abbildung 5.3 erklärt werden. In dieser Abbildung ist ein Schnitt senkrecht zur Drehachse der Ventilkugel durch einen geschlossenen Kugelhahn dargestellt. Zur Vereinfachung der Darstellung wird in dieser Darstellung der Fall betrachtet, dass nur einseitig Flüssigkeit im System vorhanden ist. Die Leitstützstruktur des Fluids überträgt den am Connector C1 anliegenden Druck auf das Wirkflächenpaar WFP1 zwischen Fluid und Ventilkugel. Die in diesem Wirkflächenpaar resultierende Kraft wird über die Leitstützstruktur der Kugel in das Wirkflächenpaar WFP2 zwischen Ventilkugel und hinterer Primärdichtung eingeleitet. Über den Connector C2 (wegen der Schnittdarstellung in Abbildung 5.3 wird dieser Connector dort in die zwei Teile C2.1 und C2.2 aufgeteilt) wird diese Kraft in das Gehäuse abgeleitet.

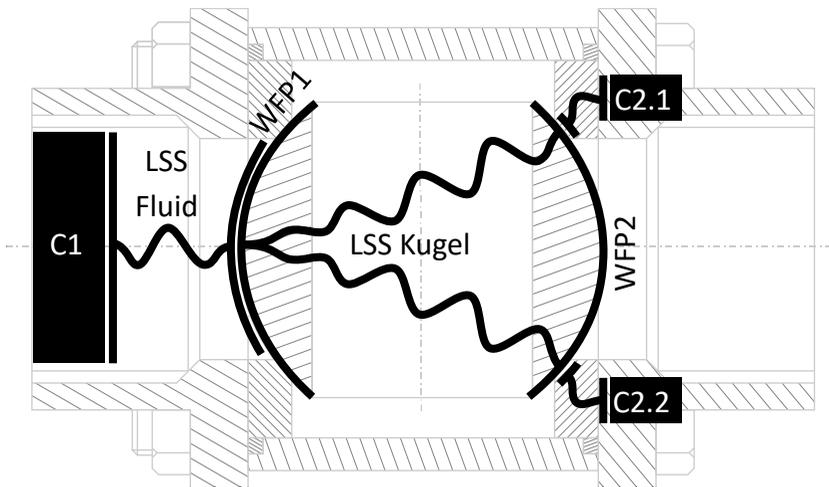


Abbildung 5.3: C&C²-A Modellierung eines Kugelhahns

Die Geometrie der Dichtung ist so gewählt, dass die Fläche des Wirkkontakts in WFP2 stets kleiner als die des Wirkkontakts in WFP1 ist. Dadurch ist die Normalspannung im Wirkkontakt des WFP2 stets größer als der stromaufwärts herrschende Fluiddruck. So kann die Dichtungsvorspannung und damit das Öffnungs- und Schließmoment niedrig eingestellt werden, ohne die Dichtheit bei großen Druckdifferenzen zu verlieren.

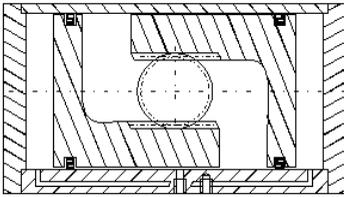
Das Material der Dichtung variiert je nach Anwendungszweck des Kugelhahns. Oft wird jedoch aufgrund seiner günstigen Reibeigenschaften und seiner hohen Temperatur- sowie Medienbeständigkeit Polytetrafluorethylen (PTFE) eingesetzt. Die Modellierung eines Kugelhahns dieser Bauart stellt das zweite Praxisbeispiel in dieser Arbeit dar.

5.3 Pneumatische Schwenkantriebe

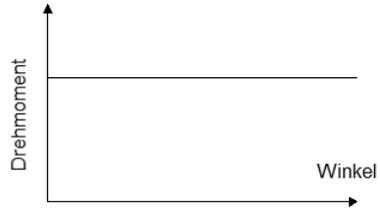
Pneumatische Schwenkantriebe für Prozessarmaturen besitzen eine hohe Robustheit und Zuverlässigkeit, sowie ein hohes Drehmoment bei geringer Baugröße. Durch ihren einfachen Aufbau sind sie zudem kostengünstig herzustellen und einfach zu warten. Des Weiteren sind sie auch für den Einsatz in explosionsgefährdeten Bereichen gut geeignet, da sie keine elektrische Energieversorgung benötigen.

Pneumatische Schwenkantriebe existieren in mehreren Bauformen. Die beiden in der Prozessautomation gebräuchlichsten Bauformen besitzen beide zwei gegenläufig im selben Zylinder angeordnete Kolben. Beide Antriebstypen besitzen dadurch eine innere und zwei äußere Zylinderkammern. Die beiden äußeren Zylinderkammern sind über einen Kanal im Gehäuse miteinander verbunden. Die beiden Antriebstypen unterscheiden sich in der Kinematik, mit deren Hilfe die gegenläufige Linearbewegung der Kolben in eine Rotation der Ausgangswelle übersetzt wird. Das Funktionsprinzip eines Antriebs mit Zahnstange-Ritzel-Mechanismus („Rack & Pinion“) ist in Abbildung 5.4 (a) dargestellt, während das eines Antriebs mit Joch-Kinematik („Scotch Yoke“) in Abbildung 5.4 (c) gezeigt ist. Der Zahnstange-Ritzel-Antrieb besitzt einen über den gesamten Hub konstanten Drehmomentverlauf (siehe Abbildung 5.4 (b)). Das Drehmoment von Antrieben mit Joch-Kinematik steigt hingegen in den Endlagen an (siehe Abbildung 5.4 (d)).

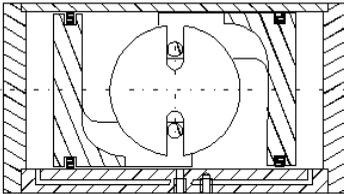
Als Praxisbeispiel wird in dieser Arbeit das Antriebsmoment von pneumatischen Schwenkantrieben mit Zahnstange-Ritzel-Mechanismus modelliert, da dies die am häufigsten eingesetzte Bauform pneumatischer Schwenkantriebe in der Prozessautomation ist (Klaer, 2003, S. 176).



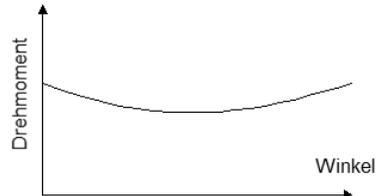
a) Aufbau eines Zahnstange-Ritzel-Antriebs



b) Drehmomentverlauf eines Zahnstange-Ritzel-Antriebs



c) Aufbau eines Joch-Kinematik-Antriebs



d) Drehmomentverlauf eines Joch-Kinematik-Antriebs

Abbildung 5.4: Gängige Bauweisen pneumatischer Schwenkantriebe

In Abbildung 5.5 ist die C&C²-A Modellierung eines solchen Antriebs zu sehen, bei dem die äußeren Zylinderkammern mit Druck beaufschlagt sind. Es wird darin nur das Teilsystem betrachtet, das das Verhalten des links dargestellten Kolbens beeinflusst. Die in der äußeren Zylinderkammer eingeschlossene Luft stützt sich links an WFP1 ab. Die dort erzeugte Kraft wird über den Connector C1.1 in das Gehäuse eingeleitet. Eine betragsmäßig identische aber entgegengerichtete Kraft erzeugt die Luft auf den Kolben in WFP2. Diese Kraft wird durch die Leitstützstruktur des Kolbens bis in WFP5 übertragen, über das das Ritzel der Abtriebswelle angetrieben wird. Das entstehende Moment wird im Ritzel über den Connector C2 in Richtung des Abtriebs abgeleitet. WFP5 ist im Gegensatz zu WFP2 nicht zentrisch am Kolben angebracht. Zudem ist es nicht parallel zu WFP2 und überträgt reibungsbedingt auch Kräfte in tangentialer Richtung. Dies führt in Summe zu einem Kippmoment auf den Kolben, das über Gleitelemente abgefangen wird, die zusammen mit der Zylinderwand die Wirkflächenpaare WFP3 und WFP4 bilden. Die Kräfte, die in diesen Wirkflächenpaaren übertragen werden, werden über die Connectoren C1.2 und C1.3 in das Gehäuse eingeleitet.

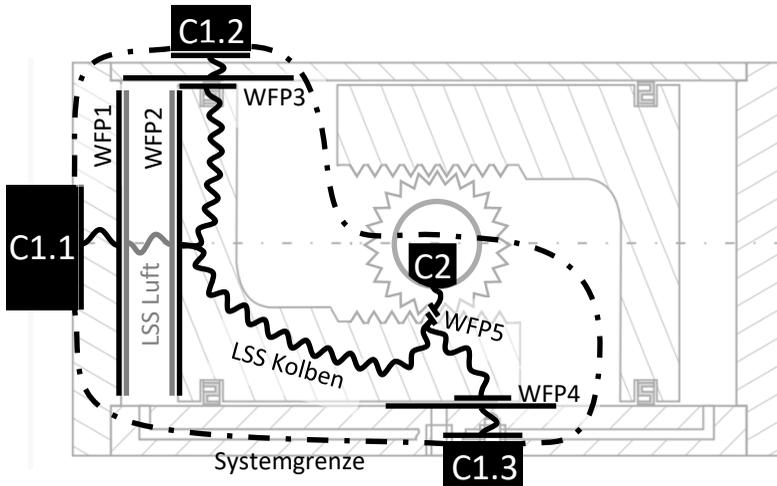


Abbildung 5.5: C&C²-A Modellierung eines Zahnstange-Ritzel-Antriebs

Wenn stattdessen die mittlere Zylinderkammer mit Druck beaufschlagt wird, dreht sich die Antriebsrichtung des Ritzels um. Dasselbe gilt für das Kippmoment, das auf den Kolben wirkt. Der Kolben stützt sich in diesem Fall auf der rechten Seite nach oben hin am Ritzel ab. Dadurch erweitert sich WFP5 auf die gegenüberliegende Zahnflanke. Linksseitig wird der Kolben nach unten hin über weitere Gleitelemente an der Innenwand des Zylinders abgestützt.

6 Integration von physikalischem Vorwissen in empirische Modelle

Im Rahmen der Entwicklung eines Produkts werden oft Simulationsmodelle des entwickelten Produkts, eines Teils seiner Subsysteme oder von mit dem Produkt verbundenen Systemen¹ benötigt. Diese Modelle können einerseits Teil des Handlungssystems zur Entwicklung des Produkts sein. Hierin werden sie beispielsweise genutzt, um bereits in frühen Entwicklungsstadien Validierungsaktivitäten starten zu können. Zum anderen können Simulationsmodelle auch Teil des Objektsystems des entwickelten Produkts sein, beispielsweise in Form von Regelungsalgorithmen.² In beiden Fällen muss mit der Erstellung der Simulationsmodelle parallel zur Entwicklung der Mechanik und Elektronik des Produkts oder sogar noch früher begonnen werden. Dass in dieser Entwicklungsphase noch kein fertiges Produkt als Prüfling für empirische Untersuchungen zur Verfügung steht, ist ein Problem für die empirische Modellbildung. Als Prüfling müssen dann Prototypen oder Elemente des Referenzsystems herangezogen werden, deren Verhalten aber meist nicht exakt dem des fertigen Produkts entspricht. Dies ist einer der Gründe dafür, dass die Erstellung von Simulationsmodellen in der Produktentwicklung in den meisten Fällen auf Basis physikalischer Grundgleichungen in Form von White-Box-Modellen³ erfolgt.

Für die Entwicklung von White-Box-Modellen wird eine breite Wissensbasis des Entwicklers über physikalische Vorgänge, die dazugehörenden Grundgleichungen und ihre Synthese zu Simulationsmodellen benötigt. Durch die hohe Granularität dieses Wissens kann es sehr flexibel für die Modellbildung verschiedenster technischer Systeme eingesetzt werden. Auch die auf diese Weise erstellten Simulationsmodelle besitzen eine hohe Granularität. Daraus folgt ein weiterer Vorteil für den Einsatz von White-Box-Modellen in Bezug auf das Modell der Produktgenerationsentwicklung nach Albers (2019): Wenn für neue Produktgenerationen Simulationsmodelle erstellt werden müssen, dann stellen Simulationsmodelle aus vorhergehenden Produktgenerationen Elemente des Referenzsystems dar und können für die Entwicklung der Modelle der neu entwickelten Produktgeneration genutzt werden. Je granularer die Modelle aufgebaut sind, desto einfacher, schneller und damit kostengünstiger lassen sie sich auf die neue Produktgeneration anpassen. Subsysteme des Produkts, die in die neue Produktgeneration durch Prinzipvariation übertragen werden, können im Modell durch seine hohe Granularität einfach ausgetauscht werden. Im Fall einer Gestaltvariation oder Übernahmevariation reicht bei

¹ „Connected Systems“, vgl. Abschnitt 2.2.1.2

² Vgl. Abschnitt 2.2

³ Vgl. Abschnitt 2.3.2

White-Box-Modellen sogar oft eine einfache Parameteranpassung, um das dazugehörige Modell auf die neue Produktgeneration anzupassen. Im Gegensatz hierzu ist eine Anpassung von empirisch ermittelten Black-Box-Modelle auf neue Produktgenerationen nur schwer oder gar nicht möglich. Oft muss das Verhalten der neuen Produktgeneration von Grund auf neu untersucht und modelliert werden.

Weitere Vorteile von physikalisch basierten White-Box-Modellen bestehen darin, dass sie problemlos stark nichtlineares und auch dynamisches Verhalten abbilden können. Die Form der Nichtlinearitäten ergibt sich dabei aus den Gleichungen, die die modellierten physikalischen Zusammenhänge beschreiben. Im Gegensatz hierzu müssen für die Erstellung von Black-Box-Modellen Nichtlinearitäten engmaschig untersucht und in meist nichtparametrischen Modellgleichungen abgebildet werden. Falls mehrere Faktoren nichtlineare Einflüsse auf das zu modellierende Verhalten haben, führt dies schnell zu einem sehr hohen Versuchsaufwand.

Aufgrund dieser und weiterer Vorteile wird bei der Modellbildung im industriellen Umfeld meist auf White-Box-Modelle zurückgegriffen. Es existieren aber auch Fälle, in denen technische Systeme nicht allein auf diese Weise modelliert werden können. Dies kann unter anderem an einer unvollständigen Wissensbasis des Entwicklers für einen speziellen Anwendungsfall liegen. Es kann aber auch sein, dass das abzubildende Verhalten im Detail so komplex ist, dass die Synthese des Wissens zu einem Modell nicht möglich bzw. nicht wirtschaftlich ist, oder dass das so entwickelte Modell zu komplex für seinen Einsatzzweck wäre. Beispiele hierfür können kontinuumsmechanische Vorgänge oder auch Reibungsphänomene sein.

In diesen Fällen muss, für die nicht physikalisch modellierbaren Subsysteme, auf empirische Black-Box-Modelle ausgewichen werden. Jedoch ist auch für Subsysteme, die empirisch modelliert werden müssen, fast immer zumindest Wissen über einzelne Teilaspekte des abzubildenden Systemverhaltens vorhanden. Dieses Wissen mit der Erstellung eines reinen Black-Box-Modells ungenutzt zu verwerfen, wäre sehr ineffizient. Zudem können durch Integration dieses Wissens einige der oben beschriebenen Vorteile von physikalisch basierten White-Box-Modellen nutzbar gemacht werden.

Dies führt zur Forschungsfrage, die in diesem Kapitel beantwortet werden soll:

Forschungsfrage 1

Wie kann physikalisches Vorwissen in empirisch ermittelte lineare Modelle integriert werden?

Zur Beantwortung dieser Frage, werden nachfolgend zunächst in Abschnitt 6.1 die speziellen Anforderungen in diesem Einsatzgebiet erklärt. Basierend auf diesen Anforderungen wird ein Lösungsansatz in Abschnitt 6.2 vorgestellt. Dessen Anwendung im Umfeld der Produktentwicklung wird in Abschnitt 6.3 erläutert. Zur Verdeutlichung wird anschließend die praktische Umsetzung in Abschnitt 6.4 anhand des Praxisbeispiels aus Kapitel 5.2 demonstriert. Die hierbei erzielten Ergebnisse führen zum Zwischenfazit in Abschnitt 6.5.

6.1 Spezielle Anforderungen in der Produktentwicklung

Es gibt eine Vielzahl an Anforderungen an Simulationsmodelle in der Produktentwicklung. Die wichtigsten Anforderungen, die die Kombination linearer Modelle mit physikalisch basierten Modellanteilen in der Produktentwicklung beeinflussen, werden nachfolgend erläutert.

1. Wie bereits in der Einleitung dieses Kapitels beschrieben, müssen Simulationsmodelle in der Produktentwicklung die Fähigkeit besitzen, *dynamisches und stark nichtlineares Verhalten abzubilden*.
2. In der Wissenschaft ist das Ziel einer Modellbildung meist, ein Modell zu erzielen, das die Realität mit der höchstmöglichen Genauigkeit abbildet. Der hierzu benötigte Aufwand spielt üblicherweise nur eine untergeordnete Rolle. Dies ist bei der Modellbildung eines technischen Systems im Rahmen der Produktentwicklung anders. Hier ist die Modellbildung an marktwirtschaftliche Randbedingungen gebunden. Deshalb muss hier ein Modell gefunden werden, das *ausreichende Genauigkeit für den gegebenen Anwendungsfall bei möglichst geringen Kosten (=Modellierungsaufwand)* bietet.
3. Für viele Anwendungen von Simulationen in der Produktentwicklung ist es zudem notwendig, dass die eingesetzten Simulationsmodelle *einen geringen Bedarf an Speichernutzung und Rechenleistung* aufweisen. Dies gilt sowohl für echtzeitfähige Anwendungen im Bereich der Funktionsentwicklung und Validierung, als auch für Anwendungen bei denen Modelle ein Teil des Produkts werden und somit auf dessen Hardware lauffähig sein müssen.⁴
4. Eine weitere Anforderung resultiert daraus, dass Systemverständnis zum entwickelten Produkt und damit verbundenen Systemen⁵ ein Element von

⁴ Vgl. Abschnitt 2.2

⁵ Vgl. „Connected Systems“ im IPEK-XiL Ansatz nach Albers, Abschnitt 2.2.1.2

zentraler Bedeutung für dessen Weiterentwicklung ist. Fehlendes Systemverständnis macht die zielführende Weiterentwicklung eines technischen Produkts unmöglich (Matthiesen et al., 2017, S. 2). Ein gutes Systemverständnis erleichtert hingegen nahezu alle Aktivitäten der Produktentwicklung und ist zudem auch beim Vertrieb des Produkts von sehr großem Wert. Durch die gut verständliche Form der Modellgleichung linearer Modelle sind diese in der Lage, ein Verständnis über die von ihnen modellierten Zusammenhänge zu erzeugen.⁶ Um dieses Potential nutzen zu können, sollten die *vom linearen Modell abgebildeten Zusammenhänge so gewählt werden, dass sie eine technische bzw. physikalische Bedeutung besitzen*. Nur so können Sie dabei helfen, ein verbessertes Systemverständnis zu gewinnen.

6.2 Lösungsansatz

Im Gegensatz zu üblichen empirischen (Black-Box-)Modellen erfüllen physikalische und phänomenologische (White-Box-)Modelle die erste oben genannte Anforderung problemlos. Sie können sowohl dynamisches Verhalten, als auch stark nichtlineares Verhalten umfassend und genau abbilden. In solchen Modellgleichungen sind die eingesetzten mathematischen Funktionen zudem optimal auf die Art der abzubildenden Zusammenhänge angepasst. Dadurch kann mit einer geringen Anzahl an Parametern eine gute Übereinstimmung zwischen Modell und Realität erreicht werden. Im Vergleich zu generischen Modellgleichungen linearer und nichtparametrischer Modelle führt dies zu einem geringen Bedarf an Speicherplatz. Dies ist in Bezug auf Anforderung 3 positiv zu sehen.

Es gibt aber auch Fälle, in denen White-Box-Modelle zu Gleichungssystemen führen, die zwar wenige Parameter besitzen, aber sehr viele Gleichungen umfassen. In diesen Fällen können White-Box-Modelle Anforderung 3 nicht erfüllen. Dies trifft insbesondere auf Vorgänge zu, die räumlich diskretisiert berechnet werden müssen, wie beispielsweise kontinuumsmechanische Vorgänge. Dies steigert den Rechenaufwand meist so stark, dass diese Modelle für echtzeitfähige Anwendungen in der Produktentwicklung⁴ nicht geeignet sind. Ebenso gibt es Fälle, in denen die detaillierte Abbildung von allen relevanten physikalischen Vorgängen, auf Grund von deren Komplexität schlicht zu aufwendig für den Einsatz in der Produktentwicklung ist. Dann ist Anforderung 2 verletzt. Dies trifft beispielsweise auf die Modellierung von

⁶ Vgl. S. 88

Reibung zu.⁷ Auch der alleinige Einsatz von White-Box-Modellen besitzt also Einschränkungen in der Produktentwicklung. Diese unterscheiden sich dabei von denen von Black-Box-Modellen.

Dies führt dazu, dass stets abgewogen werden muss, welcher Modelltyp für einen gegebenen Anwendungsfall zu bevorzugen ist. Anders ausgedrückt, müssen für jedes Teilsystem separat die am besten geeigneten Referenzsystemelemente für die Modellbildung ausgewählt werden. Dies sollte allerdings nicht für das gesamte Modell des zu modellierenden Systems getroffen werden. Vielmehr sollte das System zunächst gedanklich in seine Bestandteile zerlegt werden. Anschließend wird für jeden Bestandteil jeder physikalische Vorgang separat modelliert. So wird beispielsweise im Praxisbeispiel der pneumatischen Schwenkantriebe für jeden Kolben ein Reibungsmodell sowie ein separates Teilmodell für den Druckaufbau in der Zylinderkammer erstellt.⁸ Erst wenn keine weitere Unterteilung mehr möglich ist, sollte für jedes dieser Teilmodelle festgelegt werden, welcher Modelltyp eingesetzt wird.

Wenn ein solches Teilmodell empirisch ermittelt wird, muss überprüft werden, ob vorhandenes Vorwissen über die abzubildenden physikalischen Vorgänge in das Teilmodell integriert werden kann. Eine Kombination solcher White-Box-Anteile mit Black-Box-Anteilen zu Grey-Box-Modellen kann sehr sinnvoll sein, um die Einschränkungen beider Modelltypen zu umgehen. Wenn lineare Modelle als Teilmodell von Grey-Box-Modellen eingesetzt werden, dann können mit dem Gesamtmodell sowohl dynamische, als auch stark nichtlineare Vorgänge modelliert werden, obwohl lineare Modelle allein hierzu nicht in der Lage sind. Mit einem solchen Grey-Box-Modell lassen sich, wie nachfolgend gezeigt wird, alle oben genannten Anforderungen erfüllen. Damit liegt im Einsatz von Grey-Box-Modellen auch die Antwort auf die Forschungsfrage dieses Kapitels.

6.3 Anwendung des Lösungsansatzes in der Produktentwicklung

White-Box-Modelle besitzen, wie im vorhergehenden Abschnitt erläutert, in den meisten Bereichen sehr vorteilhafte Eigenschaften für die Modellbildung technischer Systeme. Auch bei der Erstellung von Grey-Box-Modellen im Umfeld der Produktentwicklung sollte deshalb zunächst versucht werden, möglichst große Anteile des zu modellierenden Verhaltens durch White-Box-Modellanteile zu beschreiben. Dies lässt sich, im Vergleich mit den anderen Grey-Box-Modelltypen aus Abschnitt 2.3.4,

⁷ Vgl. Abschnitt 6.4.2

⁸ Vgl. Abschnitt 6.4

am besten mit Hilfe hybrider Grey-Box-Modelle umsetzen. Deshalb wird an dieser Stelle ihr Einsatz empfohlen. Die Anwendung hybrider Grey-Box-Modelle wird nachfolgend erläutert.

6.3.1 Grundlegende Vorgehensweise

Die Erstellung eines hybriden Grey-Box-Modells beginnt stets mit der Erstellung und Parametrierung des White-Box-Teilmodells, welches durch eine physikalische oder phänomenologische Modellgleichung repräsentiert wird. Darin sollten, wie oben beschrieben, möglichst große Anteile des Systemverhaltens abgebildet werden.

Zur Verdeutlichung dieser Vorgehensweise wird zusätzlich zum Praxisbeispiel in Abschnitt 6.4 zunächst ein einfaches Beispiel eingeführt. In diesem Beispiel wird die Abkühlung eines gasumströmten Körpers modelliert. Als White-Box-Modell kann die Differentialgleichung 6.1 eingesetzt werden, die ein beschränktes Wachstum beschreibt. In ihr ist ϑ die Zustandsgröße (also in diesem Beispiel die Temperatur), c_1 die Wachstumskonstante und c_2 der Grenzwert des Wachstums. Gleichungen dieser Form sind zur Beschreibung einer Vielzahl an Vorgängen geeignet. Unter anderem für die zeitliche Entwicklung von Temperaturen, Stoffkonzentrationen oder Populationen von Lebewesen. Als Differentialgleichung ist diese Gleichung in der Lage auch dynamisches Verhalten und einen nichtlinearen zeitlichen Verlauf der Zustandsgröße zu modellieren.

$$\dot{\vartheta} = c_1 \cdot (c_2 - \vartheta) \tag{6.1}$$

In dieser Gleichung stellt der Temperaturgradient $\dot{\vartheta}$ die Zielgröße \hat{y}_{WB} des White-Box-Modells dar. Der einzige Faktor x_{WB} des White-Box-Modells ist durch die aktuelle Temperatur ϑ des Körpers gegeben. c_1 und c_2 sind zunächst unbekannte Parameter des White-Box-Modells, denen normalerweise durch theoretische Überlegungen oder experimentelle Untersuchungen ein fester Wert zugewiesen wird. Im Beispiel des gasumströmten Körpers kann der Parameter c_2 , der den Grenzwert des Abkühlvorgangs beschreibt, direkt als die Temperatur des umströmenden Gases festgelegt werden. Die Wachstumskonstante c_1 hängt hingegen von komplexen geometrischen und strömungsmechanischen Gegebenheiten ab und lässt sich deshalb nicht ähnlich einfach festlegen. Wenn Gleichung 6.1 den Abkühlvorgang bei stets konstanten Strömungsbedingungen beschreibt, dann ist c_1 eine Konstante, die experimentell ermittelt werden kann.

Wenn, wie hier, nur ein Parameter des White-Box-Modells experimentell bestimmt werden muss, dann ist hierzu nur ein einzelner Versuch notwendig. Bei der Versuchsdurchführung muss die Zielgröße \hat{y}_{WB} (hier ϑ), sowie alle Faktoren x_{WB} (hier ϑ) des White-Box-Modells aufgezeichnet werden. Anschließend kann Gleichung 6.1 nach dem zu bestimmenden Parameter (hier c_1) aufgelöst werden. Dadurch ergibt sich in diesem Beispiel Gleichung 6.2. Einsetzen der aufgezeichneten Werte der Zielgröße und der Faktoren des White-Box-Modells, sowie der auf anderem Weg bestimmten Parameter liefert den Wert des experimentell zu bestimmenden Parameters.

$$c_1 = \frac{\vartheta}{(c_2 - \vartheta)} \quad 6.2$$

Falls mehrere Parameter des White-Box-Modells auf diesem Weg bestimmt werden müssen, dann wird zur Bestimmung jedes Parameters jeweils mindestens ein Freiheitsgrad⁹ benötigt. Für dynamische Vorgänge, wie in diesem Beispiel, besteht eine Möglichkeit zur Erzeugung vieler Freiheitsgrade darin, einen zeitlichen Verlauf aufzuzeichnen (also im Beispiel einen Abkühlvorgang). Falls dies nicht möglich oder ausreichend ist, müssen die benötigten Freiheitsgrade durch das Durchführen mehrerer Versuche erzeugt werden. Hierbei ist darauf zu achten, dass alle Faktoren des White-Box-Modells so variiert werden, dass genügend Informationen über die im Modell enthaltenen Parameter gesammelt werden können.¹⁰

Anschließend können die Parameterwerte durch iterative Optimierungsverfahren so angepasst werden, dass das berechnete Modellverhalten möglichst genau dem gemessenen Verlauf entspricht. Dazu wird zunächst ein Fehler- oder Gütefunktional definiert, das die Güte der Anpassung des Modellverhaltens an das reale System quantifiziert. Anschließend werden die Werte der Parameter c_1 und c_2 durch ein Optimierungsverfahren so bestimmt, dass die Abweichungen zwischen Modell und Realität minimiert werden (Bohn & Unbehauen, 2016, S. 56). Eine Einführung in diese Verfahren gibt beispielsweise Ruckstuhl (2008).

Das oben beschriebene White-Box-Modell ist nur für die Strömungsbedingungen gültig, bei denen der Parameter c_1 bestimmt worden ist. Für andere Drücke, Strömungsgeschwindigkeiten oder Medien ist das Modell nicht aussagekräftig. Um die Einflüsse solcher Größen in ein White-Box-Modell aufzunehmen ist im Allgemeinen der Einsatz von mehrdimensionalen Strömungssimulationen

⁹ Vgl. Abschnitte 2.4.1 und 2.5.1

¹⁰ Vgl. Abschnitt 2.5.1

notwendig. Damit kann allerdings Anforderung 3 aus Abschnitt 6.1 nicht erfüllt werden. Deshalb ist es nicht sinnvoll, diese Einflüsse in einem reinen White-Box-Modell abzubilden. Eine Abbildung in einem reinen Black-Box-Modell wäre aber ebenso wenig zielführend, da dabei das Vorwissen in Form von Gleichung 6.1 nicht berücksichtigt würde. Stattdessen ist hier der Einsatz von Grey-Box-Modellen sinnvoll.

Hierzu dient das White-Box-Teilmodell aus Gleichung 6.1 als Grundlage. Auf dieser Basis wird das Verhalten des zu modellierenden Systems in Abhängigkeit der empirisch abzubildenden Faktoren in einer Versuchsreihe experimentell untersucht und in einem oder mehreren Black-Box-Teilmodellen abgebildet und mit dem White-Box-Teilmodell kombiniert. Die Vorgehensweise hierzu hängt davon ab, ob ein paralleler oder serieller Ansatz für die Erstellung des Gesamtmodells gewählt wird.

6.3.2 Paralleler Ansatz hybrider Grey-Box-Modelle

Beim parallelen Ansatz hybrider Grey-Box-Modelle ergibt sich gemäß Abbildung 2.7 der Schätzwert \hat{y} des Gesamtmodells aus der Summe der Vorhersagewerte \hat{y}_{WB} des White-Box-Modells und \hat{y}_{BB} des Black-Box-Teilmodells (Sohlberg & Jacobsen, 2008). Dies führt zu Gleichung 6.3. Darin beschreibt x die Gesamtmenge aller Faktoren, die einen Einfluss auf das Gesamtmodell besitzen. Entsprechend beschreibt x_{WB} die Faktoren, die auf das White-Box-Teilmodell wirken, und x_{BB} die, die auf das Black-Box-Teilmodell wirken.

$$\hat{y}(x) = \hat{y}_{WB}(x_{WB}) + \hat{y}_{BB}(x_{BB}) \quad 6.3$$

Zur Erstellung des Black-Box-Teilmodells wird das Verhalten des modellierten Systems nach Fertigstellung des White-Box-Modells in Abhängigkeit der Faktoren x_{BB} untersucht. Hierzu können beliebige Versuchspläne aus Abschnitt 2.5.1 zum Einsatz kommen. Zu bevorzugen sind allerdings Versuchspläne, die die Erstellung eines quadratischen Modells ermöglichen. Also Box-Behnken Pläne oder zentral zusammengesetzte Versuchspläne.¹¹

Nach Abschluss der Versuchsreihe wird das Black-Box-Teilmodell mittels eines Verfahrens aus Abschnitt 2.5 auf die Versuchsergebnisse angepasst. Dabei wird als Zielgröße für das Black-Box-Teilmodell die Differenz zwischen dem White-Box-Modell und den Versuchsergebnissen gemäß Gleichung 6.4 verwendet.

¹¹ Vgl. Abschnitt 7.3.3

$$\mathbf{y} - \hat{\mathbf{y}}_{WB}(\mathbf{X}_{WB}) = \mathbf{y}_{BB}(\mathbf{X}_{BB}) = \mathbf{X}_{BB}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad 6.4$$

In dieser Gleichung beschreibt \mathbf{y} die in der Versuchsreihe ermittelten Messwerte. $\hat{\mathbf{y}}_{WB}(\mathbf{X}_{WB})$ stellt die Vorhersagewerte des White-Box-Modells in Abhängigkeit der Einstellungen \mathbf{X}_{WB} der White-Box-Faktoren dar. $\mathbf{y}_{BB}(\mathbf{X}_{BB})$ beschreibt ein lineares Black-Box-Modell mit den Einstellungen \mathbf{X}_{BB} der Black-Box-Faktoren.¹² Der Residuenvektor $\boldsymbol{\varepsilon}$ enthält alle Abweichungen, die auch durch die Summe aus beiden Teilmodellen nicht abgebildet werden können.¹³ Nachdem auf diese Weise das Black-Box-Teilmodell erstellt ist, sind beide Teilmodelle und somit auch das Gesamtmodell gemäß Gleichung 6.3 fertiggestellt.

Beim oben dargestellten parallelen Ansatz gemäß der Definition von Sohlberg und Jacobsen werden die Vorhersagewerte beider Teilmodelle addiert. Damit ist der absolute Effekt des Black-Box-Modells unabhängig von den Vorhersagewerten des White-Box-Modells. In bestimmten Fällen wird stattdessen bevorzugt, dass alle Aspekte des White-Box-Modells durch das Black-Box-Modell proportional verstärkt oder reduziert werden. Dies führt zu einem multiplikativen parallelen Ansatz. Die Vorgehensweise hierbei entspricht nahezu vollständig der des additiven parallelen Ansatzes. Lediglich die Modellgleichung des Gesamtmodells wird zu Gleichung 6.5 geändert.

$$\hat{\mathbf{y}}(\mathbf{x}) = \hat{\mathbf{y}}_{WB}(\mathbf{x}_{WB}) \cdot \hat{\mathbf{y}}_{BB}(\mathbf{x}_{BB}) \quad 6.5$$

Beim multiplikativen Ansatz wird also anstatt der absoluten Differenz der relative Unterschied zwischen Messung und White-Box-Modell gemäß Gleichung 6.6 als Zielgröße für das Black-Box-Modell verwendet.

$$\frac{\mathbf{y}}{\hat{\mathbf{y}}_{WB}(\mathbf{X}_{WB})} = \mathbf{y}_{BB}(\mathbf{X}_{BB}) = \mathbf{X}_{BB}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad 6.6$$

Ob ein additiver oder multiplikativer paralleler Ansatz zu bevorzugen ist, hängt von der Art des zu modellierenden Zusammenhangs ab. Im Beispiel aus Abschnitt 6.3.1 würde ein additiver Ansatz beispielsweise zu Gleichung 6.7 führen. Wenn die Black-Box-Faktoren \mathbf{x}_{BB} die Einflüsse von Druck, Strömungsgeschwindigkeit und Art des

¹² Diese Gleichung beinhaltet alle Ergebnisse der Versuchsreihe. Dementsprechend unterscheiden sich die Formelzeichen zu denen in Gleichung 6.3. Vgl. auch Abschnitt 2.5.2

¹³ Vgl. Gleichung 2.16

Fluids abbilden sollen, ist dieser Ansatz sicherlich weniger gut geeignet. Dies liegt daran, dass diese Faktoren bei einem additiven Ansatz auch einen Einfluss auf die Grenztemperatur hätten, der in der Realität nicht gegeben ist. Ein multiplikativer Ansatz gemäß Gleichung 6.8 ist deshalb in diesem Fall zu bevorzugen.

$$\hat{\vartheta} = c_1 \cdot (c_2 - \vartheta) + y_{BB}(x_{BB}) \quad 6.7$$

$$\hat{\vartheta} = c_1 \cdot (c_2 - \vartheta) \cdot y_{BB}(x_{BB}) \quad 6.8$$

Falls nicht zuverlässig entschieden werden kann, welche Variante besser zum abgebildeten System passt, dann können probeweise auch beide Varianten an die Ergebnisse der Versuchsreihe angepasst werden. Anschließend kann die Vorhersagegenauigkeit der beiden so erstellten Modelle beispielsweise anhand des Bestimmtheitsmaßes¹⁴ verglichen und damit die besser geeignete Variante ausgewählt werden.

Bei dieser Entscheidung muss allerdings auch einbezogen werden, ob der Vorhersagewert des White-Box-Modells sehr klein oder sogar zu null werden kann. Dies kann beim multiplikativen Ansatz zu Singularitäten bei der Ermittlung des Black-Box-Modells führen: Wenn experimentell Ergebnisse ungleich null ermittelt werden, bei deren Versuchsbedingungen der Vorhersagewert des White-Box-Modells null ist, dann führt dies zu unendlich großen Vorhersagewerten des Black-Box-Modells. Deshalb sollte der multiplikative Ansatz nur gewählt werden, wenn dieses Problem ausgeschlossen werden kann.

Bei beiden Ansätzen muss, wie auch beim nachfolgend dargestellten seriellen Ansatz, berücksichtigt werden, dass die Regressionsgleichung auf transformierte Messwerte angepasst wird. Durch Transformationen ändern sich die Verteilungsdichtefunktionen der transformierten Größen (siehe Kleppmann, 2011, Kap. 6.4). Während unveränderte Messwerte meist (aber nicht immer) normalverteilt streuen, muss bei den durch Transformation bestimmten Parameterwerten davon ausgegangen werden, dass diese eine andere Verteilungsdichtefunktion besitzen. Dies kann jedoch leicht mithilfe der Methoden aus Abschnitt 2.5.2.3 überprüft werden. Falls sich hier deutliche Abweichungen von der Normalverteilung zeigen, dann werden

¹⁴ Siehe S. 52

die Effekte der Faktoren zwar korrekt geschätzt. Die Bewertungsgrößen gemäß Abschnitt 2.5.2.2 sowie die berechnete Vorhersagegenauigkeit des Modells besitzen dann aber keine Gültigkeit mehr.

Beide oben beschriebenen Ansätze sind einfach und universell einsetzbar. Ein Nachteil dieser Vorgehensweisen ist allerdings, dass mit ihnen nicht einzelne Aspekte des White-Box-Teilmodells gezielt beeinflusst werden können. Dies ist hingegen mit dem im nachfolgenden Abschnitt erläuterten seriellen Ansatz möglich.

6.3.3 Serieller Ansatz hybrider Grey-Box-Modelle

Um die gezielte Beeinflussung einzelner Aspekte des White-Box-Teilmodells zu ermöglichen, muss der Vorhersagewert des Black-Box-Teilmodells als Eingangsgröße (also als Faktor bzw. Funktionsargument) in die Modellgleichung des White-Box-Teilmodells aufgenommen werden. So kann das Black-Box-Teilmodell an beliebiger Stelle in die Modellgleichung des White-Box-Teilmodells integriert werden und so auch beliebige Elemente des White-Box-Teilmodells durch beliebige funktionale Zusammenhänge beeinflussen. Dadurch entsteht der serielle Ansatz hybrider Grey-Box-Modelle gemäß Gleichung 6.9 (Sohlberg & Jacobsen, 2008).

$$\hat{y}(x) = \hat{y}_{WB}(x_{WB}, \hat{y}_{BB}(x_{BB})) \quad 6.9$$

Zur Integration des Black-Box-Modells als unabhängige Variable in das White-Box-Modell bestehen zwei Möglichkeiten. Diese werden nachfolgend anhand des Beispiels aus Abschnitt 6.3.1 erläutert.

Die erste Möglichkeit besteht darin, die Ausgangsgröße des Black-Box-Modell als zusätzliche Variable an beliebiger Stelle der Funktionsgleichung des White-Box-Modells mit Hilfe einer beliebigen Operation zu integrieren. Gleichung 6.1 könnte so beispielsweise abgewandelt werden zu Gleichung 6.10.

$$\hat{y} = c_1 \cdot (c_2 - \vartheta)^{\hat{y}_{BB}(x_{BB})} \quad 6.10$$

Auch die beiden parallelen Ansätze des vorhergehenden Abschnitts gemäß den Gleichungen 6.7 und 6.8 können als Spezialfälle dieser Vorgehensweise verstanden werden. Diese Einbindungsmöglichkeit bietet größtmögliche Flexibilität. Allerdings besitzt hier die Zielgröße \hat{y}_{BB} des Black-Box-Modells meist keine unmittelbare physikalische Bedeutung, sondern muss als Korrekturparameter verstanden werden.

Die aus der Modellgleichung des Black-Box-Modells ablesbaren Informationen führen so zu keinem wertvollen Erkenntnisgewinn über das modellierte System.¹⁵

Dies ändert sich, wenn ein Black-Box-Modell nicht als zusätzliche Variable ins White-Box-Modell integriert, sondern stattdessen ein bestehender Parameter des White-Box-Modells durch das Black-Box-Modell ersetzt wird. In Bezug auf Forderung 4 aus Abschnitt 6.1 ist dies als vorteilhaft zu werten, da solche Parameter meist interpretierbare Systemeigenschaften beschreiben, wie beispielsweise die Zieltemperatur oder die Wachstumskonstante des obigen Beispiels. Im Beispiel wird also einer oder beide der Parameter c_1 und c_2 durch ein Black-Box-Modell \hat{y}_{BB} ersetzt. Die ersetzten Parameter besitzen damit keinen festen Wert mehr, sondern sind abhängig von den Faktoren x_{BB} , deren Einfluss empirisch abgebildet werden soll. Beispiele hierfür folgen weiter unten.

Beim Abkühlvorgang eines gasumströmten Körpers ist beispielsweise davon auszugehen, dass in Gleichung 6.1 hauptsächlich die Wachstumskonstante c_1 durch Strömungseigenschaften beeinflusst wird. Deshalb ist es hier sinnvoll, diesen Parameter durch ein Black-Box-Modell zu ersetzen, wenn das Gesamtmodell für unterschiedliche Strömungszustände gültig sein soll. Dies führt zu Gleichung 6.11.

$$\dot{\vartheta} = y_{BB}(x_{BB}) \cdot (c_2 - \vartheta) \quad 6.11$$

Wenn, wie hier, nur ein einzelner Parameter des White-Box-Modells durch ein Black-Box-Modell abgebildet werden soll, dann ergibt sich eine Vorgehensweise, die der des parallelen Ansatzes sehr ähnlich ist. Nach Fertigstellung des White-Box-Modells wird ein Versuchsplan aus Abschnitt 2.5 ausgewählt, mit dessen Hilfe die Einflüsse aller empirisch abzubildenden Faktoren auf das Systemverhalten untersucht werden. Gleichung 6.11 kann anschließend nach y_{BB} aufgelöst werden, wodurch sich die Zielgröße des Black-Box-Teilmodells berechnen lässt (solange nicht $\vartheta = c_2$ ist). So entsteht für dieses Beispiel Gleichung 6.12, anhand derer die lineare Regression für das Black-Box-Teilmodell durchgeführt werden kann. Durch Gleichung 6.11 ist anschließend das vollständig parametrisierte Gesamtmodell gegeben.

$$\frac{\dot{\vartheta}}{(c_2 - \vartheta)} = y_{BB}(x_{BB}) = X_{BB}\beta + \varepsilon \quad 6.12$$

¹⁵ Vgl. Anforderung 4 aus Abschnitt 6.1

Es kann allerdings auch vorkommen, dass mehrere Parameter existieren, die durch Black-Box-Modelle abgebildet werden sollen. Im Beispiel des gasumströmten Körpers könnte dies beispielsweise zutreffen, wenn dem Körper ein von mehreren Faktoren abhängiger Wärmestrom zugeführt wird. Wenn der Effekt dieses Wärmestroms genau wie die Wachstumskonstante des Abkühlvorgangs durch ein Black-Box-Modell abgebildet werden soll, ergibt sich als Gesamtmodell Gleichung 6.13. Darin bildet $y_{BB,1}(x_{BB,1})$ die Wachstumskonstante des Abkühlvorgangs und $y_{BB,2}(x_{BB,2})$ die durch den Wärmestrom verursachte Erwärmungsrate ab.

$$\dot{\vartheta} = y_{BB,1}(x_{BB,1}) \cdot (c_2 - \vartheta) + y_{BB,2}(x_{BB,2}) \quad 6.13$$

Die Vorgehensweise zur Erstellung eines solchen Modells hängt davon ab, ob es Einflussgrößen gibt, die als Faktoren mehrerer Black-Box-Teilmodelle vorkommen. Falls dies nicht zutrifft, es also keine Überschneidungen zwischen den Faktoren der verschiedenen Black-Box-Teilmodelle gibt, können diese unabhängig voneinander bestimmt werden. Die Faktoreinflüsse jedes Black-Box-Teilmodells werden dann in separaten Versuchsreihen untersucht, bei denen jeweils die Faktoreinstellungen aller anderen Black-Box-Teilmodelle konstant gehalten werden. Dadurch werden die anderen Black-Box-Teilmodelle innerhalb dieser Versuchsreihe zu konstanten Parametern, womit sich jedes Black-Box-Teilmodell gemäß der vorhergehend beschriebenen Vorgehensweise für ein einzelnes Black-Box-Teilmodell bestimmen lässt.

Falls allerdings manche Faktoren in mehreren Black-Box-Teilmodellen enthalten sind, dann müssen diese Teilmodelle simultan erstellt werden. Hierzu wird ein Versuchsplan aufgestellt, in dem die Einflüsse aller Faktoren der zu ermittelnden Black-Box-Teilmodelle auf das Systemverhalten untersucht werden. An jedem Punkt dieses Versuchsplans müssen nun die Werte der Zielgrößen aller Black-Box-Teilmodelle bestimmt werden. Die Vorgehensweise hierzu entspricht der Ermittlung von mehreren Parameterwerten aus Abschnitt 6.3.1. Es muss also jeweils ein Messwertverlauf ermittelt oder jeweils mehrere Einzelversuche durchgeführt werden, auf die anschließend die Werte der Zielgrößen aller Black-Box-Teilmodelle mittels eines iterativen Optimierungsverfahrens angepasst werden.

Falls an jedem Punkt des Versuchsplans mehrere Einzelversuche durchgeführt werden, entsteht eine Versuchsplan, der den gekreuzten Versuchsplänen aus Abschnitt 2.5.5.1 ähnelt. Entsprechend stellt sich hierbei die Frage, wie der so entstandene neue Versuchsplan randomisiert werden sollte. Es erscheint empfehlenswert, zwei separate Randomisierungsebenen einzusetzen: Zunächst sollten alle Versuchs-

punkte des ursprünglichen Versuchsplans untereinander randomisiert werden. Anschließend werden alle Einzelversuche für jeden Punkt des ursprünglichen Versuchsplans untereinander randomisiert und in zeitlich möglichst kurzer Abfolge durchgeführt. So können Änderungen von Störgrößeneinflüssen und damit die Fehlervarianz zwischen diesen Einzelversuchen gering gehalten werden. Dies erhöht die Genauigkeit der Parameterwerte, die so für jeden Punkt des ursprünglichen Versuchsplans durch ein iteratives Optimierungsverfahren ermittelt werden. Die Störgrößeneinflüsse wirken sich stattdessen hauptsächlich zwischen den Punkten des ursprünglichen Versuchsplans aus. Dies ist zu bevorzugen, da bei der Modellierung der Parameterunterschiede zwischen diesen Parametern mittels linearen Regressionsmodellen geeignete Verfahren zur Minimierung und Quantifizierung von deren Einflüssen existieren.¹⁶

6.4 Demonstration am Beispiel der Schwenkantriebe

Aufgrund ihres einfachen Aufbaus müssen pneumatische Schwenkantriebe für die Modellbildung nicht in Subsysteme zerlegt werden. Die Systemdynamik dieser Schwenkantriebe wird durch den Druckaufbau in den Zylinderkammern, Reibkräfte, Trägheitswirkungen und das von der angeschlossenen Armatur erzeugte Drehmoment beeinflusst. Der Druckaufbau und die Trägheitswirkungen lassen sich problemlos auf Basis physikalischer Grundgleichungen abbilden. Dies wird in Abschnitt 6.4.1 demonstriert. Die Reibung im System wird hingegen durch viele, nur schwierig in White-Box-Modellen abzubildende Effekte beeinflusst. Dies wird in Abschnitt 6.4.2 ausführlicher erläutert. Zur Umgehung dieses Problems ist deshalb der Einsatz eines Grey-Box-Modells sinnvoll. Die Vorgehensweise hierzu wird in Abschnitt 6.4.3 gezeigt. Die damit erzielten Ergebnisse finden sich in Abschnitt 6.4.4.

6.4.1 Modellierung der Kinematik und der Druckkräfte

Aus dem in Abbildung 6.1 dargestellten Aufbau eines Schwenkantriebes ergibt sich mit Hilfe des Kräfte- und Momentengleichgewichts Gleichung 6.14.

$$2 \cdot (\Delta p \cdot A - F_{K,fr} - m_K \ddot{x}) \cdot i_{rack} + M_{ext} - M_{W,fr} - J_W \cdot \ddot{\varphi} = 0 \quad 6.14$$

In dieser Gleichung sind die Zylinderquerschnittsfläche A , die Kolbenmasse m_K , das Massenträgheitsmoment J_W der Ausgangswelle und das Übersetzungsverhältnis

¹⁶ Vgl. Abschnitt 2.5.2

für schnelle Bewegungen des Antriebs kann in guter Näherung von einer isentropen Zustandsänderung ausgegangen werden. Für diese entspricht der Polytropenexponent dem Isentropenexponent. Der Isentropenexponent ist eine Stoffkonstante des eingesetzten Gases, die für Luft ungefähr den Wert 1,4 besitzt. Die Molare Masse M ist ebenfalls eine Stoffkonstante des eingesetzten Gases, die für Luft etwa den Wert 28,96 kg/kmol besitzt (Langeheinecke et al., 2017, S. 458).

Durch Umstellen von Gleichung 6.14 ergibt sich die Winkelbeschleunigung $\ddot{\varphi}$ des Schwenkantriebes zu Gleichung 6.17.

$$\ddot{\varphi} = \frac{2 \cdot (\Delta p \cdot A - F_{K,fr}) \cdot i_{rack} + M_{ext} - M_{W,fr}}{J_W + 2m_K \cdot i_{rack}^2} \quad 6.17$$

Zur Modellierung der in den Gleichungen 6.14 bis 6.17 beschriebenen Zusammenhangs ist der Einsatz von empirischer Modellbildung weder notwendig, noch sinnvoll, da hier der physikalisch basierter Modellierungsansatz, eine ausreichend genaue Modellierung unter Einbeziehung aller relevanten Einflüsse ermöglicht.

Neben dem Massenstrom, der in die beiden Zylinderkammern zu- oder abgeführt wird, sind die einzigen Variablen, die nun noch zu bestimmen sind, die Reibkräfte $F_{K,fr}$ sowie das Reibmoment $M_{W,fr}$. Ein Ansatz zu deren konventioneller Modellierung wird im nachfolgenden Abschnitt beschrieben.

6.4.2 White-Box Reibungsmodellierung

Einen Überblick über die zur Modellierung der Reibung in Pneumatikzylindern geeigneten Reibmodelle geben Heipl & Murrenhoff (2012). Bei physikalisch basierten White-Box-Modellierungsansätzen werden die gekoppelten strömungsdynamischen und kontinuumsmechanischen Vorgänge im elastohydrodynamischen Schmier-spalt räumlich und zeitlich diskretisiert modelliert (siehe z.B. Bartel, 2010, Kap. 7). Dabei müssen nicht nur die elastomeren Materialeigenschaften der Dichtung, sondern auch das nichtnewtonschen Fließverhaltens des Fettes berücksichtigt werden. Solche elastohydrodynamische Modelle sind für komplexere Schmier-spaltgeometrien, wie die eines pneumatischen Antriebs, nur numerisch lösbar (Heipl & Murrenhoff, 2012, S. 26–29). Sie sind sehr aufwendig in der Modellbildung und auch der Berechnung. Deshalb sind sie nicht in der Lage, die in Anforderung nach geringem Rechenaufwand von Simulationsmodellen in der Produktentwicklung zu erfüllen. Auch Heipl wählt zur Modellierung eines Pneumatikzylinders ein phänomenologisches Modell mit deutlich geringerer Komplexität.

Er setzt für sein Zylindermodell ein Reibungsmodell nach Carlson und Batista (1996) ein. Ein Modell desselben Typs verwendet auch Achenbach (2011) zur Modellierung des pneumatischen Schlagwerks einer Schlagbohrmaschine. Die Grundform des Modells ist durch Gleichung 6.18 gegeben.

$$F_{fr} = (F_{fr,s} - F_{fr,d})\theta + F_{fr,d} + \gamma\dot{x} \quad 6.18$$

Darin ist $F_{fr,s}$ die Haftreibungskraft und $F_{fr,d}$ die Gleitreibungskraft. $\gamma\dot{x}$ bildet einen viskosen Reibanteil nach, durch den die Reibkraft mit steigender Geschwindigkeit \dot{x} zunimmt. Der Schmierzustand wird durch die Zustandsgröße θ modelliert, die Werte zwischen null und eins annehmen kann. Dabei ist $\theta = 1$ einem ungeschmierten Kontakt und $\theta = 0$ einem vollständig ausgebildeten Schmierfilm gleichzusetzen. Die zeitliche Änderung des Schmierzustandes ist durch Gleichung 6.19 gegeben. In dieser Gleichung beschreibt der erste Term den von der Zeitkonstante τ abhängigen Abbau des Schmierfilms. Der Aufbau des Schmierfilms während Bewegungsvorgängen wird durch den zweiten Term beschrieben. Darin bezeichnet D die für diesen Aufbau charakteristische Wegstrecke.

$$\dot{\theta} = \frac{\theta}{\tau}(1 - \theta) - \frac{1}{D}\theta\dot{x} \quad 6.19$$

Das bis hier beschriebene Reibmodell berücksichtigt noch nicht, dass die Kontaktdrücke im Dichtkontakt variabel sind. Dies kommt einerseits von unterschiedlichen thermischen Ausdehnungskoeffizienten der verschiedenen Komponenten des Schwenkantriebes. Zudem ist die Dichtungsgeometrie selbstverstärkend gestaltet, sodass anliegende Druckdifferenzen eine Erhöhung der Dichtungsanpressung bewirken. So kann auch bei höheren Druckdifferenzen die Dichtheit des Systems gewährleistet werden, ohne bei niedrigen Druckdifferenzen die Reibung unnötig zu erhöhen. Erreicht wird dies durch die S-förmige Querschnittsgeometrie der Dichtungen (siehe Abbildung 6.1). Die Dichtung wird des Weiteren von den Reibkräften zwischen Dichtung und Zylinderwand verformt, was wiederum Änderungen der Anpresskräfte und damit eine Rückwirkung auf die Reibkräfte zur Folge hat. Diese Effekte werden zudem durch die Materialeigenschaften der Dichtung beeinflusst. Die Materialeigenschaften sind wiederum von der Temperatur, der Verformungsgeschwindigkeit (Mullins- und Payne-Effekt, Mullins (1969); Payne (1962)) und vielen weiteren Einflüssen abhängig.

Achenbach und Heipl berechnen diese Effekte über Finite-Elemente-Modelle (FE-Modelle). Aufgrund des rotationssymmetrischen Aufbaus der Dichtungen

berechnen sie jeweils nur die Verformungen und Spannungen eines Dichtungsquerschnittes. Das nichtlineare, hyperelastische Materialverhalten des Dichtungselastomers wird in ihren Modellen nach Mooney (1940) und Rivlin (1948) approximiert. Das viskoelastische Materialverhalten wird von Heipl durch eine Erweiterung des Cauchy-Greenschen Verzerrungstensors um zeitabhängige Terme ins Materialmodell integriert. Den Temperatureinfluss auf das Materialverhalten bildet er durch das Modell von William, Landel & Ferry (1955) ab. Achenbach beschreibt seine Vorgehensweise zur Abbildung viskoelastischer und thermischer Effekte nicht näher, erwähnt aber, dass er sie ausgehend von Relaxationsspektren (frequenzabhängige Messungen des komplexen Schubmoduls) ins Modell integriert. Um die sehr hohe Rechenzeit des FE-Modells auf ein für Systemsimulationen taugliches Niveau zu senken, erstellt Heipl vorab ein Kennfeld, in dem die Kontaktkräfte und –breiten in Abhängigkeit vom anliegenden Druck, der Temperatur und der Einwirkungszeit der Belastung abgelegt werden. Aus diesem dreidimensionalen Kennfeld werden die Kontaktkräfte und –breiten dann zur Laufzeit der Systemsimulation ausgelesen.

Da die Reibungs- und Materialmodelle von Achenbach und Heipl umfangreiche Reib- und Materialeigenschaften abbilden, benötigen sie viele Modellparameter, mit denen sie sich an das Verhalten in pneumatischen Antrieben anpassen lassen. Die Materialparameter beinhalten Parameter zur Beschreibung des hyperelastischen Verhaltens des Dichtungselastomers. Das Materialverhalten von Elastomeren im Druckbereich, bei Scherung und bei mehrachsiger Beanspruchung unterscheidet sich allerdings stark von dem bei einachsiger Zugbeanspruchung. Deshalb sollten zur Bestimmung dieser Parameter neben einem einachsigen Zugversuch immer auch die Ergebnisse mehrerer anderer Materialversuche, wie biaxialer oder planarer Zugversuche, Scherversuche oder Druckversuche herangezogen werden (MSC Software, 2010, S. 20). Des Weiteren sind Relaxationsversuche bei unterschiedlichen Temperaturen und Anregungsfrequenzen notwendig, um die Parameter zur Beschreibung des Relaxationsverhaltens und des darauf wirkenden Temperatureinflusses zu bestimmen (Heipl & Murrenhoff, 2012, S. 36). Abschließend muss für eine präzise Materialmodellierung noch der Wärmeausdehnungskoeffizient des Dichtungselastomers bestimmt werden. Mit Hilfe des parametrisierten Materialmodells können dann FE-Simulationen durchgeführt werden, um ein Kennfeld zu erzeugen, das die entstehenden Kontaktkräfte beinhaltet. Abschließend müssen dann die Parameter $F_{fr,s}$, $F_{fr,d}$, γ , τ und D des Reibmodells bestimmt werden.

Trotz des sehr hohen Aufwands, den beide Autoren bei der physikalisch basierten Modellierung betreiben, können sie nicht alle relevanten Effekte in ihre Modelle integrieren. Sie berücksichtigen beispielsweise nicht, dass alle von ihnen betrachteten

Systeme, genau wie pneumatische Schwenkantriebe, fettgeschmiert sind. Fette besitzen ein temperaturabhängiges nichtnewtonsches Fließverhalten. Diese Temperaturabhängigkeit der Schmierstoffeigenschaften wird in beiden Modellen vernachlässigt und nur die Eigenschaften des Dichtungswerkstoffes werden temperaturabhängig modelliert.

Bei Schwenkantrieben verläuft zudem die Wirklinie der an den Kolben angreifenden Druckkräfte und der durch die Kolbendichtung verursachten Reibkraft durch die Achse der Zylinderbohrung. Die daraus resultierende Kraft wird aber radial versetzt in die Verzahnung des Ritzels eingeleitet. Die dabei an dieser Stelle entstehende Gegenkraft bewirkt, dass ein Kippmoment auf die Kolben wirkt. Dieses wird an den Kontaktstellen des Kolbens zur Zylinderwand aufgenommen.¹⁷ Diese Verkippung der Kolben bewirkt eine asymmetrische Lastverteilung auf die Kolbendichtungen, die dadurch auch asymmetrisch verformt werden. Es wäre deshalb eine starke Vereinfachung, für pneumatische Schwenkantriebe nur die Verformungen einer Querschnittsfläche der Dichtung zu untersuchen und diese als konstant über den gesamten Kolbenumfang anzunehmen. Der Ansatz von Achenbach und Heipl müsste deshalb für pneumatische Schwenkantriebe geeignet erweitert werden. Dabei wäre es beispielsweise denkbar, das zweidimensionale FE-Modell des Dichtungsquerschnitts durch ein dreidimensionales Modell der gesamten Dichtung zu ersetzen.

Eine konventionelle Modellierung der Reibung von pneumatischen Schwenkantrieben ist also durchaus möglich. Trotz des sehr hohen Aufwands dieser Vorgehensweise ist es dabei allerdings extrem schwierig, alle relevanten Effekte realitätsnah ins Modell zu integrieren. Deshalb ist es hier sinnvoller, ein Grey-Box-Modell einzusetzen. Die Erstellung eines solchen Modells für die Reibung in pneumatischen Schwenkantrieben wird im nachfolgenden Abschnitt beschrieben.

6.4.3 Grey-Box Reibungsmodellierung

Es ist sehr schwierig, die Reibkräfte $F_{K,fr}$ der Kolben und das Reibmoment $M_{W,fr}$ der Welle separat experimentell zu untersuchen. Deshalb werden zunächst alle Reibkräfte und –momente zum Gesamtreibmoment M_{fr} gemäß Gleichung 6.20 zusammengefasst.

$$M_{fr} = 2 \cdot F_{K,fr} \cdot i_{rack} + M_{W,fr} \quad 6.20$$

¹⁷ Vgl. Abbildung 5.5 auf S. 102 (C&C²-A Modellierung) sowie Abbildung 6.1 auf S. 117 (Darstellung der wirkenden Kräfte und Momente)

Die direkte Modellierung dieses Reibmoments mittels eines reinen Black-Box-Modells ist nicht sinnvoll, da das Reibmoment weiterhin starke Nichtlinearitäten sowie dynamische Effekte beinhaltet. Um diese abzubilden wird ein hybrides Grey-Box-Modell eingesetzt, als dessen White-Box-Teilmodell ein Reibmodell nach Carlson und Batista dient. Durch Anpassung auf eine Drehbewegung ergeben sich dafür die Gleichungen 6.21 und 6.22.

$$M_{fr} = (M_{fr,s} - M_{fr,d})\theta + M_{fr,d} + \gamma\dot{\phi} \quad 6.21$$

$$\dot{\theta} = \frac{\theta}{\tau}(1 - \theta) - \frac{1}{D}\theta\dot{\phi} \quad 6.22$$

Wie im vorhergehenden Abschnitt beschrieben wurde, ist in dieser Form des Modells die thermische Abhängigkeit der Dichtungs- und Schmierstoffeigenschaften noch nicht berücksichtigt. Ebenso fehlen das nichtnewtonsche Fließverhalten des Fettes, die Materialeigenschaften der Dichtung sowie deren Verformung, die aus den anliegenden Drücken, sowie der Verkippung der Kolben resultiert. Die Auswirkungen dieser Vorgänge sollen durch Black-Box-Teilmodelle abgebildet werden. Faktoren die hierzu in diesen Teilmodellen berücksichtigt werden müssen sind die Temperatur T , die Winkelgeschwindigkeit $\dot{\phi}$ sowie die Druckdifferenz Δp zwischen den Zylinderkammern.

Mit einem seriellen Ansatz hybrider Grey-Box-Modelle kann lediglich das durch die Gleichungen 6.21 und 6.22 berechnete Reibmoment absolut (additiv) oder relativ (multiplikativ) durch das Black-Box-Teilmodell beeinflusst werden. Es muss jedoch davon ausgegangen werden, dass die Black-Box-Faktoren T , $\dot{\phi}$ und Δp unterschiedliche Einflüsse auf das Gleitreibmoment $M_{fr,d}$ und das Haftreibmoment $M_{fr,s}$ besitzen. Des Weiteren beeinflussen sie auch die Geschwindigkeiten des Auf- und Abbaus des Schmierfilms, die durch die Parameter τ und D beschrieben werden. Die Modellierung durch hybride Grey-Box-Modelle nach dem seriellen Ansatz ist somit nicht geeignet, alle Einflüsse der Black-Box-Faktoren auf das Systemverhalten realitätsgetreu abzubilden. Aufgrund der fehlenden physikalischen Bedeutung der Zielgröße des Black-Box-Teilmodells, wären zudem die Erkenntnisse über die Reibung in Schwenkantrieben, die aus den Versuchsergebnissen gezogen werden können, in beiden Fällen sehr begrenzt.

Aus diesen Gründen wird stattdessen ein paralleler Ansatz hybrider Grey-Box-Modelle eingesetzt. Darin werden alle Parameter des White-Box-Modells durch Black-

Box-Teilmodelle ersetzt. Damit ergeben sich für den White-Box-Anteil des Grey-Box-Modells die Gleichungen 6.23 und 6.24.

$$M_{fr} = (M_{fr,s}(x_{BB}) - M_{fr,d}(x_{BB}))\theta + M_{fr,d}(x_{BB}) \quad 6.23$$

$$\dot{\theta} = \frac{\theta}{\tau(x_{BB})}(1 - \theta) - \frac{1}{D(x_{BB})}\theta\dot{\phi} \quad 6.24$$

In Gleichung 6.23 wurde der Term $\gamma\dot{\phi}$ aus Gleichung 6.21 entfernt. Der Grund hierfür ist, dass bereits in $M_{fr,d}$ die Abhängigkeit des Reibmoments von der Geschwindigkeit $\dot{\phi}$ enthalten ist. Eine zusätzliche Berechnung durch den Term $\gamma\dot{\phi}$ ist deshalb weder notwendig noch sinnvoll.

Der Black-Box-Anteil des Grey-Box-Modells besteht aus vier separaten Black-Box-Teilmodellen, deren Zielgrößen in den Gleichungen 6.23 und 6.24 durch $M_{fr,s}(x_{BB})$, $M_{fr,d}(x_{BB})$, $\tau(x_{BB})$ und $D(x_{BB})$ repräsentiert werden. Für alle Black-Box-Teilmodelle werden Modellgleichungen derselben Form eingesetzt. Dadurch sind auch die Black-Box-Faktoren x_{BB} für alle Teilmodelle identisch. Um zumindest leichte Nicht-linearitäten abbilden zu können, die beispielsweise durch das nichtnewtonsche Fließverhalten des Fettes verursacht werden könnten, werden alle quadratischen Terme ins Modell aufgenommen. Dasselbe gilt für alle Wechselwirkungen zweiten Grades, mit deren Hilfe beispielsweise die Temperaturabhängigkeit der Dichtungseigenschaften beschrieben werden kann. Damit erhalten die Black-Box-Teilmodelle die Form von Gleichung 6.25.

$$\begin{aligned} y_{BB} = & \beta_0 + \beta_T T + \beta_\phi \dot{\phi} + \beta_p \Delta p \\ & + \beta_{TT} T^2 + \beta_{\phi\phi} \dot{\phi}^2 + \beta_{pp} \Delta p^2 \\ & + \beta_{T\phi} T \dot{\phi} + \beta_{Tp} T \Delta p + \beta_{\phi p} \dot{\phi} \Delta p \\ & + \varepsilon \end{aligned} \quad 6.25$$

Als Versuchsplan zur Bestimmung der Black-Box-Teilmodelle wird ein zentral zusammengesetzter Versuchsplan nach Abschnitt 2.5.1.7 eingesetzt. Dieser hat den

in Abbildung 2.20 dargestellten Aufbau.¹⁸ Als axiale Distanz α wird der normierte Wert eins eingesetzt. Im vollständig faktoriellen Teil des Versuchsplans besitzen die Faktoren die normierten Werte +0,5 bzw. -0,5. An jedem Punkt dieses Versuchsplans wird nicht nur ein einzelner Versuch, sondern jeweils elf Einzelversuche nach unterschiedlich langen Stillstandszeiten durchgeführt. Diese Einzelversuche werden für jeden Punkt des zentral zusammengesetzten Versuchsplans separat randomisiert. Anschließend werden die Punkte des zentral zusammengesetzten Versuchsplans untereinander randomisiert. So ergibt sich der Versuchsplan nach Tabelle A.5. Darin sind die Versuche in ihrer Durchführungsreihenfolge aufgelistet. Die Einstellungen aller Faktoren sind normiert angegeben, sodass ihre größte Einstellung dem normierten Wert eins und ihre kleinste Einstellung dem Wert minus eins entspricht.

Nach Durchführung der Versuchsreihe werden zunächst für jeden Punkt des zentral zusammengesetzten Versuchsplans separat die Parameter $M_{fr,s}$, $M_{fr,d}$, τ und D bestimmt. Hierzu werden iterative Optimierungsverfahren eingesetzt. Anschließend werden auf die so bestimmten Parameterwerte Regressionsmodelle nach Gleichung 6.25 angepasst. Die damit erzielten Ergebnisse werden im nachfolgenden Abschnitt präsentiert.

6.4.4 Ergebnisse

Exemplarisch sind in Abbildung 6.2 die gemessenen Reibmomentenverläufe für den Mittelpunkt des zentral zusammengesetzten Versuchsplans dargestellt. Aus Gründen der Geheimhaltung sind die untersuchten Stillstandszeiten darin durch t_i bis t_{11} bezeichnet. Je größer der Index dieser Bezeichnung ist, desto größer ist die untersuchte Stillstandszeit. Es ist zu erkennen, dass mit der Stillstandszeit das Losbrechmoment ansteigt. Ebenso ist sichtbar, wie das Reibmoment nach Erreichen des Maximums über einen gewissen Drehwinkel hinweg abfällt. Um dies besser sichtbar zu machen, ist in Abbildung 6.2 nur der Winkelbereich kurz nach dem Losbrechen dargestellt.

¹⁸ Es unterscheiden sich lediglich die Formelzeichen der Faktoren, die dort verallgemeinert durch x_1 , x_2 und x_3 gegeben sind. Im hier beschriebenen Beispiel sind diese durch T , ϕ und Δp zu ersetzen.

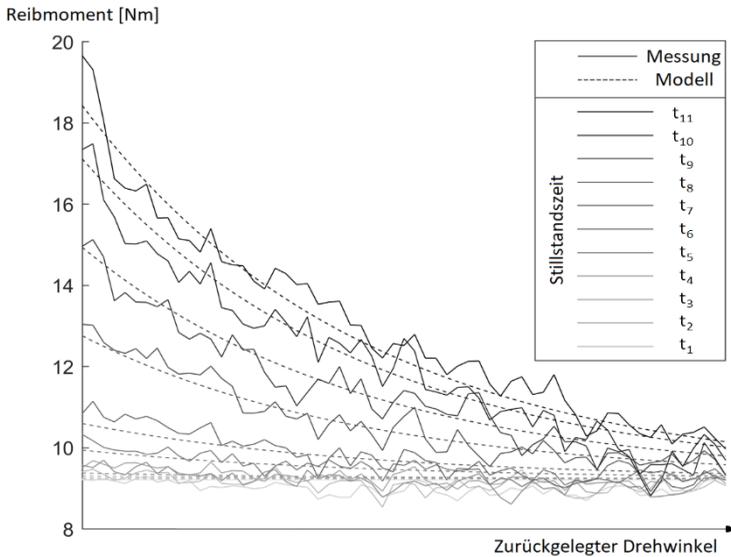


Abbildung 6.2: Reibmomentenverläufe nach verschiedenen Stillstandszeiten für den Zentrumspunkt des zentral zusammengesetzten Versuchsplans

Zusätzlich zu den gemessenen Reibmomentenverläufen sind auch die Vorhersagen des Reibungsmodells aus den Gleichungen 6.23 und 6.24 dargestellt. Die Parameter $M_{fr,s}(x_{BB})$, $M_{fr,d}(x_{BB})$, $\tau(x_{BB})$ und $D(x_{BB})$ dieses Reibungsmodells werden durch numerische Optimierung für jeden Punkt x_{BB} des zentral zusammengesetzten Versuchsplans separat an die gemessenen Reibmomentenverläufe angepasst. So erhält man für jeden Punkt verschiedene Schätzwerte dieser Parameter. Diese Schätzwerte sind in Tabelle 6.1 aufgelistet.

Auf die Schätzwerte der vier Parameter des (White-Box-)Reibungsmodells aus Tabelle 6.1 werden anschließend vier separate Black-Box-Teilmodelle angepasst. In den nachfolgenden Balkendiagrammen sind die Regressionskoeffizienten dieser

Black-Box-Teilmodelle dargestellt. Die Balken statistisch signifikanter Einflüsse¹⁹ sind darin dunkelgrau und die nicht signifikanter Einflüsse hellgrau eingefärbt.

Tabelle 6.1: Parameterwerte des Reibungsmodells in Abhängigkeit der Black-Box-Faktoren

Versuchsreihenfolge	Einstellungen der Faktoren			Werte der Zielgrößen			
	T	v	Δp	$M_{fr,s}$	$M_{fr,d}$	T_{nom}	D_{nom}
1	1,00	0,00	0,00	20,08	7,43	0,563	1,282
2	0,00	0,00	1,00	32,67	14,67	1,010	1,351
3	0,00	1,00	0,00	23,89	11,12	1,015	1,167
4	-0,50	0,50	-0,50	18,68	11,18	0,964	1,059
5	-0,50	0,50	0,50	26,21	14,59	1,040	1,183
6	0,50	0,50	-0,50	11,47	4,40	0,858	1,052
7	0,00	0,00	0,00	18,43	9,15	0,997	1,008
8	0,50	-0,50	-0,50	11,43	4,33	0,826	1,003
9	0,00	0,00	-1,00	10,81	6,62	0,919	1,075
10	0,50	0,50	0,50	17,02	6,44	0,876	1,337
11	-0,50	-0,50	-0,50	13,77	8,46	1,016	1,050
12	-1,00	0,00	0,00	21,86	15,05	0,925	1,146
13	0,50	-0,50	0,50	17,05	6,47	0,875	1,368
14	0,00	-1,00	0,00	12,72	4,69	1,050	1,117
15	-0,50	-0,50	0,50	21,82	11,81	1,055	1,026

Abbildung 6.3 zeigt die Regressionskoeffizienten für das Haftreibungsmoment $M_{fr,s}$. Signifikant ist hier nur die Konstante β_0 sowie der Effekt β_p der Druckdifferenz.²⁰ Das durch β_0 beschriebene durchschnittliche Losbrechmoment des untersuchten Schwenkantriebs lag während der Versuchsreihe bei etwa 16 Nm. Der Koeffizient für den Einfluss der Druckdifferenz hat einen Wert von etwa 9 Nm. Eine Druckänderung vom minimalen zum maximalen untersuchten Druck²¹ verursacht also einen Reibungsanstieg von etwa 18 Nm. Dieser wird einerseits dadurch verursacht, dass die Dichtung durch ihre selbstverstärkende Querschnittsgeometrie bei größeren

¹⁹ Einflüsse mit einem p-Wert von weniger als 0,05

²⁰ Details zu allen Koeffizienten finden sich im Anhang in Tabelle A.1

²¹ Normiert also von -1 nach +1

Druckdifferenzen stärker an die Zylinderwand angedrückt wird. Zum anderen erzeugen größere Druckdifferenzen auch größere Kräfte auf die Kolben und damit höhere Kontaktkräfte in den Führungen der Kolben, sowie der Verzahnung zum Ritzel.

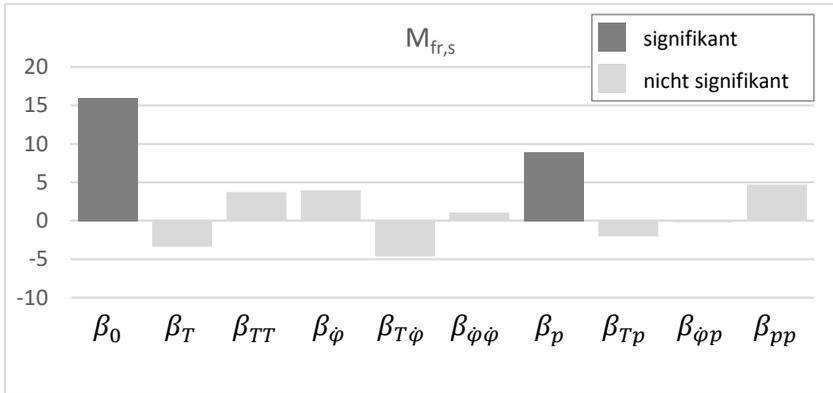


Abbildung 6.3: Regressionskoeffizienten für das Haftreibungmoment $M_{fr,s}$

Die Höhe der Balken zeigt die Größe des Einflusses. Berechnung von $M_{fr,s}$ gemäß Gleichung 6.25. Berechnung der Gesamtreibung gemäß Gleichungen 6.23 und 6.24.

In Abbildung 6.4 sind die Regressionskoeffizienten für das Gleitreibungmoment $M_{fr,d}$ dargestellt. Das Gleitreibungmoment besaß in der Versuchsreihe einen durchschnittlichen Wert von etwa 8 Nm. Es ist damit etwa halb so groß wie das durchschnittlich aufgetretene Losbrechmoment. Auch das Gleitreibungmoment wird durch höhere Druckdifferenzen gesteigert. Der Koeffizient β_p besitzt einen Wert von etwa 3,4 Nm. Zusätzlich sind beim Gleitreibungmoment auch die linearen Effekte der Temperatur und der Geschwindigkeit statistisch signifikant. Eine Erhöhung der Temperatur führt zu einer geringeren Schmierstoffviskosität und vermutlich dadurch zu einer verringerten Gleitreibung. Der Einfluss der Temperatur ist mit einem Koeffizienten β_T von etwa -5 Nm betragsmäßig größer als der aller anderen Faktoren.

Eine Erhöhung der Gleitgeschwindigkeit führt zu einer Erhöhung viskoser Reibanteile. Da diese gemäß Gleichung 6.23 im Gleitreibungmoment $M_{fr,d}$ mit abgebildet werden, zeigt sich hier ein positiver Koeffizient β_v mit einem Wert von etwa 2,3 Nm.

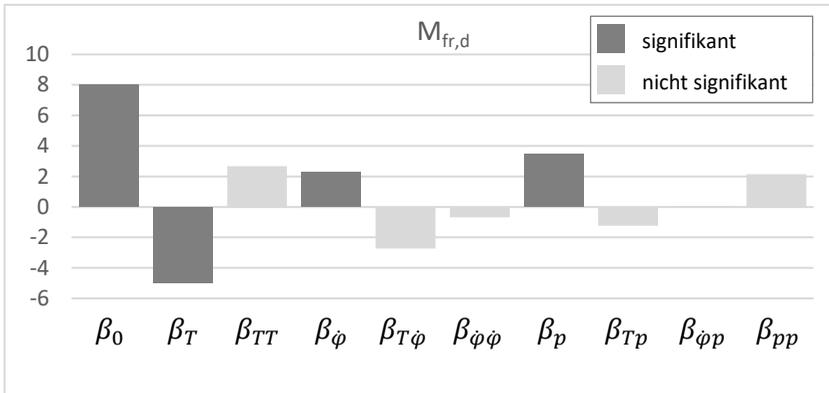


Abbildung 6.4: Regressionskoeffizienten für das Gleitreibungsmoment $M_{fr,d}$. Die Höhe der Balken zeigt die Größe des Einflusses. Berechnung von $M_{fr,d}$ gemäß Gleichung 6.25. Berechnung der Gesamtreibung gemäß Gleichungen 6.23 und 6.24.

Interessant ist auch, dass sich die Regressionskoeffizienten für das Gleitreibungsmoment und das Haftreibungsmoment sehr stark ähneln. Sie besitzen nahezu für alle Faktoren dasselbe Vorzeichen und auch zwischen den verschiedenen Faktoren ähnliche Verhältnisse. Es fallen nahezu alle Koeffizienten für das Gleitreibungsmoment betragsmäßig kleiner aus, als für das Haftreibungsmoment. Dennoch sind hier mehr Effekte statistisch signifikant. Dies liegt an einer geringeren Fehlervarianz. Diese beträgt für das Haftreibungsmoment etwa 3,9 Nm, während sie für das Gleitreibungsmoment nur etwa 1,8 Nm beträgt.²²

Die Zeitkonstante τ beschreibt die Geschwindigkeit des Anstiegs von der Gleitreibung hin zur Haftreibung. In Abbildung 6.5 sind die Regressionskoeffizienten für das Black-Box-Teilmodell dargestellt, das diesen Parameter beschreibt (aus Geheimhaltungsgründen normiert). Durch dieses Modell ist auch τ keine Konstante mehr, sondern abhängig von den Black-Box-Faktoren x_{BB} . Es ist erkennbar, dass lediglich die Temperatur einen nennenswerten Einfluss auf τ besitzt. Neben β_T und β_{TT} ist zwar auch β_p statistisch signifikant, allerdings sehr klein.

²² Details können Tabelle A.1 und Tabelle A.2 im Anhang entnommen werden.

β_T und β_{TT} besitzen negative Werte von etwa -0,17 und -0,25.²³ Dies bedeutet, dass gemäß diesem Modell τ bei der niedrigsten untersuchten Temperatur um etwa 0,08 größer als bei der mittleren untersuchten Temperatur ist. Bei weiterer Temperaturerhöhung sinkt τ dann deutlich ab und erreicht bei der höchsten untersuchten Temperatur sein Minimum von etwa 0,42. Da kleinere Werte von τ ein schnelleres Anwachsen des Reibmoments während Stillstandszeiten bedeuten, haben höhere Temperaturen also ein schnelleres Wachstum des Reibmoments zur Folge. Eine Ursache hierfür ist, dass beim Übergang von Gleitreibung zur Haftreibung das Schmiermedium aus dem Dichtspalt verdrängt wird. Da bei höheren Temperaturen dessen Viskosität abnimmt, läuft dieser Prozess hier schneller ab.

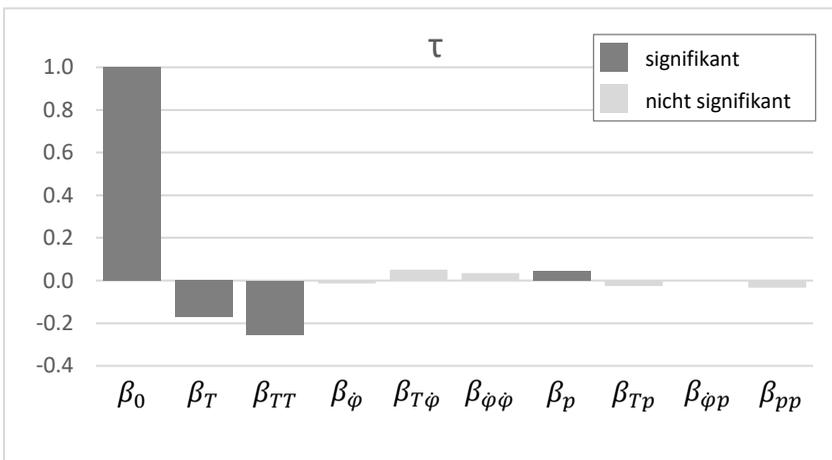


Abbildung 6.5: Regressionskoeffizienten für die Zeitkonstante τ (normiert)
 Die Höhe der Balken zeigt die Größe des Einflusses. Berechnung von τ gemäß Gleichung 6.25. Berechnung der Gesamtreibung gemäß Gleichungen 6.23 und 6.24.

In Abbildung 6.6 sind die Regressionskoeffizienten für die charakteristische Wegstrecke D normiert dargestellt. Statistisch signifikant sind hier die linearen und quadratischen Einflüsse der Temperatur und des Drucks, sowie die Wechselwirkung zwischen diesen beiden Faktoren. Alle Koeffizienten dieser Einflüsse besitzen positive Werte. Dies bedeutet, dass sowohl Temperaturerhöhungen, als auch Erhöhungen

²³ Für die genauen Werte und weitere Details siehe Tabelle A.3 im Anhang.

des Drucks eine verlängerte charakteristische Wegstrecke D zur Folge haben. Der Aufbau des Schmierfilms dauert dann also länger. Die Ursache hierfür liegt vermutlich in erhöhten Normalspannungen im Dichtkontakt sowie einer verringerten Schmierstoffviskosität.

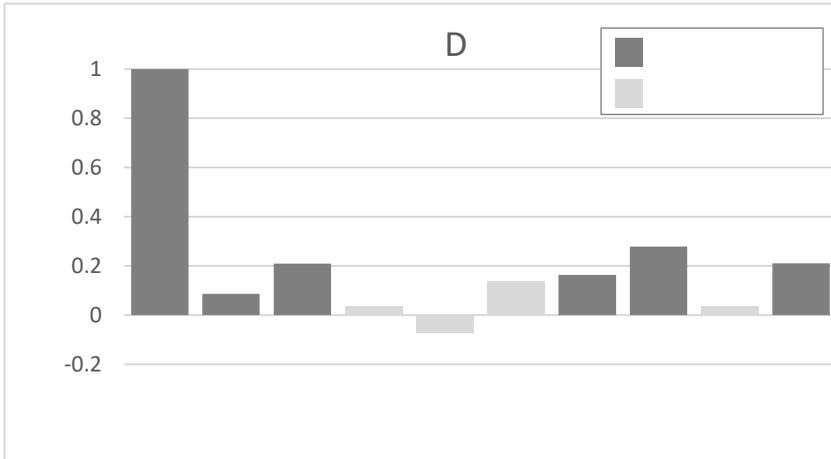


Abbildung 6.6: Regressionskoeffizienten für die charakteristische Wegstrecke D (normiert). Die Höhe der Balken zeigt die Größe des Einflusses. Berechnung von D gemäß Gleichung 6.25. Berechnung der Gesamtreibung gemäß Gleichungen 6.23 und 6.24.

Für das Black-Box-Teilmodell der charakteristischen Wegstrecke D werden deutlich mehr Koeffizienten als bei den anderen Teilmodellen als statistisch signifikant bewertet.²⁴ Dies liegt daran, dass für diese Black-Box-Teilmodelle gemäß Gleichung 2.25 ein größerer Schätzwert für die Fehlervarianz $\hat{\sigma}^2$ berechnet wird. Nach Gleichung 2.24 führt dies zu größeren Schätzwerten der Standardabweichungen $\hat{\sigma}_\beta$ der Koeffizienten. Diese haben wiederum zur Folge, dass die Signifikanz der Effekte niedriger bewertet wird.²⁵

Da die Fehlervarianz gemäß Gleichung 2.25 auf Basis der Residuen $\hat{\varepsilon}$ ermittelt wird, wird sie auch durch Lack-of-Fit²⁶ beeinflusst. In den in Abbildung 6.7 dargestellten

²⁴ Details können Tabelle A.4 im Anhang entnommen werden.

²⁵ Vgl. Abschnitt 2.5.2.2

²⁶ Vgl. S. 28

Residuendiagrammen zeigen sich hier auch auffällige Verläufe für die Teilmodelle von $M_{fr,s}$, $M_{fr,d}$ und τ .²⁷ Dies gilt vor allem für die in Zeile 3 dargestellte Temperaturabhängigkeit der Residuen. Um genauere Hinweise auf die Ursachen hierfür zu erhalten, müsste der Effekt der Temperatur in mehr verschiedenen Stufen untersucht werden.

Falls sich hierbei tatsächlich zeigte, dass die Temperaturabhängigkeit zu komplex für die Abbildung durch ein Polynom zweiten Grades ist, könnte für eine genauere Modellbildung das Verfahren nach Kapitel 7 eingesetzt werden. Die Unregelmäßigkeiten in den Residuendiagrammen spielen sich allerdings (für $M_{fr,s}$ und $M_{fr,d}$) im Bereich weniger Newtonmeter ab. Die dadurch verursachten Einbußen der Genauigkeit des Reibmodells sind in einem akzeptablen Bereich. Deshalb wird hier kein weiterer Versuchsaufwand zur Erhöhung der Genauigkeit des Modells getrieben.

6.5 Zwischenfazit

In den vorhergehenden Abschnitten dieses Kapitels konnte gezeigt werden, wie physikalisches Vorwissen über ein abzubildendes System in empirische Modelle integriert werden kann. Mit Hilfe der hierzu eingesetzten Grey-Box-Modelle lässt sich Vorwissen über Nichtlinearitäten und dynamische Verhaltensanteile in White-Box-Teilmodellen abbilden. Einflüsse, deren physikalisch basierte Modellierung sehr schwierig oder sehr aufwendig wäre, können dagegen in Black-Box-Teilmodelle aufgenommen werden, die auf empirischen Untersuchungen geeigneter Referenzsystemelemente basieren.

Ein besonders gut für den Einsatz in der Produktentwicklung geeigneter Typ von Grey-Box-Modellen sind die hybriden Grey-Box-Modelle. Ihre Erstellung wurde zunächst erklärt und anschließend anhand eines Praxisbeispiels demonstriert. Anhand der Ergebnisse dieses Beispiels konnte auch gezeigt werden, wie die gute Interpretierbarkeit, vor allem von seriellen hybriden Grey-Box-Modellen, zu einem verbesserten Systemverständnis führen kann.

Die zu Beginn dieses Kapitels gestellte Forschungsfrage kann damit als beantwortet betrachtet werden.

²⁷ Wie diese Diagramme im Normalfall aussehen sollten, wird in Abschnitt 2.5.2.3 beschrieben.

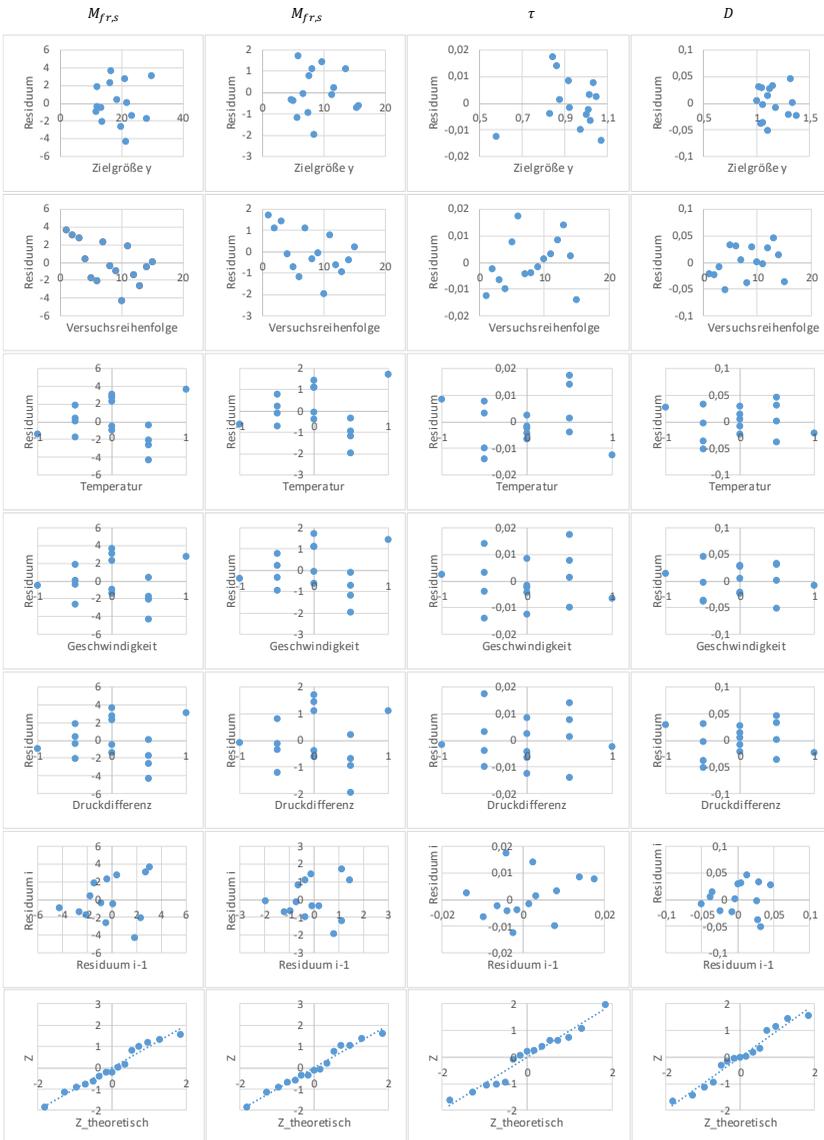


Abbildung 6.7: Residuendiagramme der vier Black-Box-Teilmodelle

7 Abbildung von Nichtlinearitäten

Technische Systeme besitzen, wie bereits im vorhergehenden Kapitel erläutert, oft nichtlineare Zusammenhänge zwischen den Ein- und Ausgangsgrößen, die in einem Modell abgebildet werden sollen. In vielen Fällen lassen sich diese in White-Box-Anteilen von Grey-Box-Modellen abbilden.

Dies trifft allerdings nicht immer zu, sodass oftmals nichtlineare Verhaltensanteile auch in Black-Box-Teilmodellen abgebildet werden müssen. Viele dieser Nichtlinearitäten sind wenig komplex. Damit ist gemeint, dass die Krümmung (zweite Ableitung) der dazugehörigen Funktionsverläufe näherungsweise konstant ist. Dies trifft beispielsweise auf die meisten Faktoren zu, die das Reibmoment von Kugelhähnen beeinflussen.¹ Die Einflüsse dieser Faktoren können problemlos durch lineare Modelle, in Form von Polynomen zweiten Grades, abgebildet werden.

Die Funktionsverläufe der Einflüsse mancher Faktoren weisen aber stark nichtlineare Verläufe auf, bei denen sich die Krümmung des Funktionsverlaufes deutlich und ggf. auch mehrfach ändert. Ein solcher Faktor ist die Armaturstellung von Kugelhähnen, in deren Abhängigkeit sich die Kontaktfläche der Dichtung und damit auch das Reibmoment mehrfach stark ändern. Eine Abbildung dieses Verlaufs durch ein Polynom zweiten Grades würde zu Lack-of-fit führen.² Auch eine Abbildung dieser Nichtlinearitäten in einem White-Box-Modell oder in White-Box-Teilmodellen eines Grey-Box-Modells gemäß Kapitel 6 ist hier nicht praktikabel.³

In solchen Fällen ist es wichtig, auch starke Nichtlinearitäten in Black-Box-(Teil-)Modellen abbilden zu können. Dies ist mit linearen Modellen allerdings nicht trivial.⁴ Daraus folgt die zweite Forschungsfrage, die in diesem Kapitel beantwortet werden soll.

Forschungsfrage 2

Wie kann die Abbildungsleistung linearer Modelle für starke Nichtlinearitäten einzelner Faktoren gesteigert werden?

¹ Siehe Abschnitt 5.2

² Vgl. Abbildung 2.9

³ Vgl. Abschnitt 7.4

⁴ Vgl. Einleitung von Abschnitt 2.6

In den nachfolgenden Abschnitten dieses Kapitels wird diese Frage beantwortet. Um ableiten zu können, welche Verfahren für den Einsatz in der Modellbildung für die Produktentwicklung am besten geeignet sind, werden nachfolgend zunächst in Abschnitt 7.1 die speziellen Anforderungen in diesem Einsatzgebiet erklärt. Anschließend wird in Abschnitt 7.2 beschrieben, welche Verfahren geeignet sind, diese Anforderungen zu erfüllen und in welchen Anwendungen diese bisher eingesetzt werden. In Abschnitt 7.3 wird erläutert, wie diese Verfahren für die Modellbildung in der Produktentwicklung eingesetzt werden können. Ihre praktische Anwendung wird in Abschnitt 7.4 anhand des Praxisbeispiels aus Kapitel 5.2 demonstriert. Dabei wird auch exemplarisch gezeigt, welche Ergebnisse damit erzielt werden können. Auf Basis dieser Ergebnisse wird in Abschnitt 7.5 ein kurzes Zwischenfazit gezogen.

7.1 Spezielle Anforderungen in der Produktentwicklung

Viele Einflüsse in technischen Systemen besitzen wenig komplexe Nichtlinearitäten, die sich durch lineare Modelle mit quadratischen Effekten und Wechselwirkungen zweiten Grades ausreichend genau annähern lassen. Nichtlineare Einflüsse, bei denen eine solche Näherung nicht ausreichend ist, lassen sich, wie im vorhergehenden Kapitel erläutert, oft durch White-Box-Anteile in Grey-Box-Modellen abbilden. Damit bleiben in den meisten Fällen entweder keine oder nur einzelne Faktoren übrig, deren Einflüsse in Black-Box-Modellteilen abgebildet werden müssen und bei denen eine Approximation durch lineare Modelle nicht ausreichend ist. Eine höhere Anzahl an solchen Faktoren wäre für die Modellbildung in der Produktentwicklung auch nicht tragbar: Eine empirische Abbildung komplexer Nichtlinearitäten verursacht einen sehr hohen Versuchsaufwand, der exponentiell mit der Anzahl der so abzubildenden Faktoren ansteigt. Daher übersteigt der Aufwand zur Modellerstellung für eine höhere Anzahl solcher Faktoren in der Produktentwicklung schnell den Nutzen der Modelle.

Eine häufige Besonderheit bei der empirischen Modellbildung technischer Systeme besteht darin, dass in einem Versuch ein abzubildender Einflussfaktor kontinuierlich variiert werden kann. Dieser eine Faktor lässt sich dann mit nahezu beliebigem Detaillierungsgrad abbilden, ohne die Anzahl der Versuche und damit den Versuchsaufwand zu steigern. Dies ist allerdings nur möglich, wenn dynamische Effekte für den so aufgezeichneten Messungsverlauf keine entscheidende Rolle spielen.

Nach Albers et al. (2019c, Grundprinzip 1) steht der Mensch im Zentrum der Produktentstehung. Deshalb *„müssen Prozesse und Methoden an seine Kreativität, Kompetenzen, Bedarfe und kognitiven Fähigkeiten angepasst werden. Zudem muss*

auch die Methodenauswahl hinsichtlich dieser Kriterien erfolgen.“: Es ist also essentiell, dass die erstellten Modelle für Ingenieure gut verständlich und handhabbar sind. So ist es beispielsweise für Validierungszwecke einfacher, die kritischsten Betriebszustände herauszufinden. Zudem kann das Systemverständnis oft deutlich gesteigert werden und die Anpassung der Modelle auf neue Randbedingungen (neue Produktgenerationen oder geänderte Einsatzbedingungen) wird vereinfacht.

Auch in Bezug auf die Anpassungsaufwände des erstellten Modells an neue Produktgenerationen ist eine gute Verständlichkeit des Modells sehr vorteilhaft: Vor allem wenn Subsysteme durch Gestalt- oder Übernahmevariation in die neue Produktgeneration übernommen werden, kann oft vorhergesagt werden, wie sich ihr Verhalten ändern wird. Wenn das Modell so aufgebaut ist, dass seine Parameter eine für den Entwickler verständliche Bedeutung besitzen, dann kann er sie entsprechend der prognostizierten Verhaltensänderung auf die neue Produktgeneration anpassen.

Die wichtigsten Randbedingungen zur Abbildung von Nichtlinearitäten in empirischen Simulationsmodellen für die Produktentwicklung lassen sich also wie folgt zusammenfassen:

1. Für alle Faktoren sollten quadratische Effekte und Wechselwirkungen zweiten Grades bestimmt werden können.
2. Nur für sehr wenige Faktoren sind noch höhere Abbildungsgenauigkeiten erforderlich.
3. Ein einzelner Faktor lässt sich oft mit geringem Aufwand in nahezu beliebiger Auflösung untersuchen.
4. Die erstellten Modelle sollten für Ingenieure gut verständlich sein.

7.2 Lösungsansatz

Auch abseits der Produktentwicklung gibt es viele Fälle, in denen einzelne Faktoren mit deutlich höherer Auflösung abgebildet werden müssen, als andere. Dies gilt insbesondere für Längsschnittstudien. Dies können beispielsweise medizinische Studien sein, in denen die Effekte verschiedener Behandlungen bei mehreren Patienten untersucht werden. Hier kann jeder Patient meist zu vielen verschiedenen Zeitpunkten der Behandlung, also in sehr hoher Auflösung, untersucht werden. Hingegen wird beispielsweise die Dosis einer Medikation üblicherweise in wenigen Stufen, also geringer Auflösung, untersucht. Ähnliche Anwendungen existieren im Umfeld der Ökonometrie zur Erstellung wirtschaftstheoretischer Modelle. Für solche und ähnliche Fälle wurden semiparametrische Regressionsmodelle entwickelt. Diese beinhalten einen nichtparametrischen Modellteil zur Abbildung komplex geformter

Nichtlinearitäten, die mit hoher Auflösung experimentell untersucht werden. Einfachere Nichtlinearitäten, die mit geringerer Auflösung untersucht werden, werden hingegen in einem linearen Modellteil abgebildet.⁵

Die Schwierigkeit beim Einsatz dieser Modelle liegt in der Bestimmung ihrer Modellparameter. Im Allgemeinen müssen die Parameter beider Anteile eines semiparametrischen Regressionsmodells gemeinsam bestimmt werden, was sehr aufwendig sein kann.⁵ Es gibt aber auch Fälle, in denen beide Modellanteile in zwei separaten Schritten bestimmt werden können. Ein solcher zweistufiger Ansatz vereinfacht die Bestimmung des Modells sehr deutlich. Welche Voraussetzungen dafür erfüllt sein müssen, wird in Abschnitt 7.3.3 genauer beleuchtet.

Die Anwendung eines solchen zweistufigen Verfahrens wurde bereits in mehreren Veröffentlichungen aus verschiedenen Disziplinen beschrieben. Aston et al. (2010) verwenden dieses Verfahren beispielsweise um den zeitlichen Verlauf der Grundfrequenz von gesprochenen Silben in Abhängigkeit von mehreren Faktoren in einem Modell abzubilden. Hierbei modellieren sie den zeitlichen Verlauf den nichtparametrischen Modellteil. Alle anderen Faktoren, wie beispielsweise die gesprochene Silbe oder das Geschlecht des Sprechers, werden durch ein gemischtes lineares Modell⁶ abgebildet.

Auch aus dem Ingenieursumfeld wurden bereits einzelne Anwendungen dieses Verfahrens publiziert. So bilden Woods et al. (2013) den Einfluss verschiedener Faktoren auf den Verlauf der Stribeck Kurve eines geschmierten Gleitkontakts in einem solchen Modell ab. Bei ihnen ist die Gleitgeschwindigkeit der einzige Faktor des nichtparametrischen Modellteils. Grove et al. (2004) beschreiben ein ähnliches Verfahren zur Abbildung von Motorenkennfeldern.

7.3 Anwendung des Lösungsansatzes in der Produktentwicklung

Nachfolgend wird erläutert, wie sich Nichtlinearitäten in empirischen Modellen für die Produktentwicklung durch semiparametrische Regressionsmodelle abbilden lassen. Die Vorgehensweise basiert dabei auf einem zweistufigen Lösungsansatz. Sie ist darauf ausgerichtet, die Anforderungen aus Abschnitt 7.1 bestmöglich zu erfüllen.

Zunächst wird in Abschnitt 7.3.1 kurz beschrieben, welche Faktoren im linearen und welche im nichtparametrischen Modellanteil abgebildet werden sollten. Welche

⁵ Siehe Abschnitt 2.6.3

⁶ Siehe Abschnitt 2.5.4

Form sich für diese beiden Modellteile empfiehlt, wird in Abschnitt 7.3.2 diskutiert. Anschließend wird in Abschnitt 7.3.3 die Frage geklärt, wie eine Versuchsreihe aufgebaut sein sollte, um eine zweistufige Regression eines semiparametrischen Modells zu ermöglichen. Anschließend wird in Abschnitt 7.3.4 erläutert, wie aus den so gewonnenen Messdaten ein semiparametrisches Regressionsmodell erstellt werden kann.

7.3.1 Auswahl der linear und nichtparametrisch abzubildenden Faktoren

Wie bereits erläutert, besitzen lineare Modelle für den Einsatz in der Produktentwicklung eine Reihe von Vorteilen gegenüber nichtparametrischen Regressionsmodellen.⁷ Dazu zählen vor allem eine bessere Verständlichkeit und ein geringerer Versuchsumfang. Dies sollte auch bei der Entscheidung, welche Faktoren im linearen und welche im nichtparametrischen Modellanteil eines semiparametrischen Modells abgebildet werden, berücksichtigt werden. Es sollten deshalb möglichst viele Faktoren im linearen Teilmodell abgebildet werden. Nur Faktoren deren Abbildung durch ein Polynom zweiten Grades nicht ausreichend ist, sollten in den nichtparametrischen Modellanteil aufgenommen werden. Dies sollten in keinem Fall mehr als zwei Faktoren sein. Eine größere Anzahl an Faktoren in diesem Modellteil führt neben dem stark vergrößerten Versuchsumfang auch zu einer schlechteren Interpretierbarkeit der erstellten Modelle. Details hierzu werden in Abschnitt 7.3.4 erläutert.

7.3.2 Auswahl eines geeigneten Regressionsmodells

Die Auswahl eines geeigneten semiparametrischen Regressionsmodells beinhaltet die Auswahl der Modelltypen für beide Teilmodelle, sowie der Art von deren Verknüpfung.

Beim linearen Teilmodell muss nur der Grad des zu verwendenden Polynoms festgelegt werden. Ein Polynom zweiten Grades stellt hierbei für die meisten Fälle die richtige Wahl dar. Der Einsatz von Polynomen des Grades drei oder höher birgt mehrere Probleme und sollte deshalb vermieden werden.⁸ Polynome des Grades eins können hingegen keinerlei Nichtlinearitäten abbilden. Deshalb sollte lediglich über den Einsatz eines Polynoms ersten Grades nachgedacht werden, wenn der mögliche Versuchsaufwand sehr stark eingeschränkt ist.

Für den nichtparametrischen Modellteil können prinzipiell beliebige nichtparametrische Regressionsmodelle eingesetzt werden. Eine besonders flexible Modellierung

⁷ Vgl. Kapitel 3

⁸ Vgl. Einleitung von Abschnitt 2.6

erlauben die künstlichen neuronalen Netze⁹. Vor allem wenn nur ein Faktor im nicht-parametrischen Modellteil abgebildet werden soll, sollte allerdings auch den TP-Splines¹⁰ besondere Aufmerksamkeit geschenkt werden. Diese besitzen den Vorteil, dass ihre Parameter sich mittels der gewöhnlichen Methode der kleinsten Fehlerquadrate¹¹ bestimmen lassen. Dadurch lassen sich beide Modellteile des semi-parametrischen Regressionsmodells mit derselben Methode schätzen. Da diese Methode lediglich auf wenigen gewöhnlichen Matrizenoperationen beruht, kann für die gesamte Erstellung des Modells ein sehr breites Spektrum an Software mit sehr geringem Aufwand eingesetzt werden.

Nachdem der Modelltyp für den nichtparametrischen Modellteil festgelegt ist, muss entschieden werden, wie beide Anteile miteinander verbunden werden. Dazu kommen additive Modelle, Index Modelle und Varying-Coefficient-Modelle in Frage.¹²

Additive Modelle gemäß Gleichung 2.62 sind dann zu empfehlen, wenn Wechselwirkungen zwischen nichtparametrischen und linearen Faktoren ausgeschlossen werden können. Dies ist allerdings ein Sonderfall, von dem nicht im Allgemeinen ausgegangen werden kann.

Alternativ können Index Modelle gemäß Gleichung 2.64 eingesetzt werden. Diese können so formuliert werden, dass die Parameter eines nichtparametrischen Modells durch lineare Regressionsmodelle beschrieben werden. Beispielsweise für ein künstliches neuronales Netz mit den Eingangsgrößen x_{np} werden also die Gewichtungsfaktoren¹³ \mathbf{b} durch jeweils ein lineares Modell der Form $x_{in}\mathbf{b}$ beschrieben. Damit ist durch diese linearen Teilmodelle zwar gut verständlich abgebildet, wie die Parameter \mathbf{b} des nichtparametrischen Modellteils sich in Abhängigkeit der linear modellierten Faktoren x_{in} verändern. Da die Parameter \mathbf{b} aber keine für Menschen verwertbaren Informationen beinhalten, ist damit kein Erkenntnisgewinn verbunden.

Sinnvoller ist hingegen die umgekehrte Vorgehensweise, die durch Varying-Coefficient-Modelle gemäß Gleichung 2.65 gegeben ist. Darin werden die Koeffizienten $\boldsymbol{\beta}$ eines linearen Modells durch jeweils ein nichtparametrisches Regressionsmodell beschrieben. Für bis zu zwei nichtparametrische Faktoren x_{np} können die in den nichtparametrischen Teilmodellen $\boldsymbol{\beta}(x_{np})$ abgebildeten Abhängigkeiten graphisch gut dargestellt werden. So erhält man graphische Darstellungen der Verläufe der

⁹ Siehe Abschnitt 2.6.2.2

¹⁰ Siehe Abschnitt 2.6.2.1

¹¹ Siehe Abschnitt 2.5.2.1

¹² Siehe Abschnitt 2.6.3

¹³ Siehe Abschnitt 2.6.2.2

Koeffizienten des linearen Regressionsmodells. Da diese Koeffizienten für Menschen interpretierbare Bedeutungen besitzen, kann das im Gesamtmodell abgebildete Verhalten mit wenigen Blicken aus diesen Darstellungen entnommen und verstanden werden. Anschaulicher wird dies anhand des Praxisbeispiels in Abschnitt 7.4.4 erklärt. Dort werden die so erzielten graphischen Darstellungen vorgestellt und auch erläutert, wie diese zu interpretieren sind.

7.3.3 Aufbau der Versuchsreihe

Zur Erstellung eines semiparametrischen Regressionsmodells mit Hilfe eines zweistufigen Ansatzes müssen für beide Teilmodelle, zunächst unabhängig voneinander, geeignete Versuchspläne ausgewählt werden. Für das lineare Teilmodell sind alle Versuchspläne aus Abschnitt 2.5.1 geeignet, während für das nichtparametrische Teilmodell die Versuchspläne aus Abschnitt 2.6.1 passend sind. Worauf bei der Auswahl dieser Versuchspläne geachtet werden sollte, wird nachfolgend beschrieben. Anschließend wird zum Ende dieses Abschnitts die Kombination beider Versuchspläne erläutert, um die Bestimmung eines semiparametrischen Regressionsmodells zu ermöglichen.

Zur Bestimmung des linearen Modellteils in Form eines Polynoms zweiten Grades stehen mehrere unterschiedliche Versuchsplantypen zur Auswahl.¹⁴ Der Einsatz von One-factor-at-a-time Versuchsreihen, sowie dreistufigen vollständigen oder fraktionell faktoriellen Versuchsplänen ist aufgrund mangelhafter Abbildungsleistung oder Effizienz nicht zu empfehlen.¹⁵ Damit bleiben Box-Behnken-Versuchspläne sowie zentral zusammengesetzte und optimale Versuchspläne¹⁶ als mögliche Kandidaten übrig. Auf die Ergebnisse aller drei Versuchspläne kann ein mehrdimensionales Polynom zweiten Grades inklusive aller Wechselwirkungen zweiten Grades angepasst werden. Der wesentliche Unterschied zwischen Box-Behnken Versuchsplänen und zentral zusammengesetzten Versuchsplänen ist, dass Box-Behnken-Versuchspläne mit einer kleineren Anzahl an Versuchen auskommen. Dafür untersuchen sie allerdings jeden Faktor nur in drei Stufen. Zentral zusammengesetzte Versuchspläne untersuchen hingegen jeden Faktor in fünf Stufen.¹⁷ Optimale Versuchspläne können ebenfalls so erstellt werden, dass die jeden Faktor in mehr als drei Stufen untersuchen. Falls die Annäherung des nichtlinearen Verhaltens durch ein Polynom zweiten Grades nicht ausreichend ist (Lack-of-fit), kann dies deshalb

¹⁴ Vgl. Abschnitt 2.5.1

¹⁵ Vgl. Abschnitte 2.5.1.1, 2.5.1.4 und 2.5.1.5

¹⁶ Vgl. Abschnitte 2.5.1.6, 2.5.1.7 und 2.5.1.8

¹⁷ für $\alpha \neq 1$, vgl. Abschnitt 2.5.1.7

bei Einsatz eines zentral zusammengesetzten oder optimalen Versuchsplans beispielsweise mittels Residuendiagrammen¹⁸ erkannt werden. Bei Einsatz eines Box-Behnken-Versuchsplans ist dies hingegen nicht möglich. Es besteht hier also die Gefahr, ein stark von der Realität abweichendes Modell zu erhalten, ohne dass dies aus den Versuchsergebnissen erkennbar wäre. Je nach Einsatz des Modelles in der Produktentwicklung resultieren hieraus große Risiken, die auch die Sicherheit des Produktes und seiner Anwender gefährden können. Aus diesem Grund empfiehlt es sich, zur Modellierung technischer Systeme für die Produktentwicklung zentral zusammengesetzte Versuchspläne oder optimale Versuchspläne mit mehr als drei Stufen einzusetzen. So entsteht zwar im Vergleich zu Box-Behnken Versuchsplänen ein geringer Mehraufwand. Dieser ist durch den Zugewinn an Zuverlässigkeit der Modelle aber in nahezu jedem Fall gerechtfertigt. Falls Lack-of-fit für einen Faktor erkennbar wird, dann können entweder Transformationen dieses Faktors oder der Zielgröße vorgenommen werden.¹⁹ Falls dies nicht möglich ist, muss dieser Faktor in den nichtparametrischen Modellteil verschoben werden.

Bei der Entscheidung zwischen zentral zusammengesetzten und optimalen Versuchsplänen sind in den meisten Fällen zentral zusammengesetzte Versuchspläne zu bevorzugen, da diese deutlich einfacher zu konstruieren sind. Zudem erlauben sie durch ihren geometrisch einfachen und symmetrischen Aufbau²⁰ die direkte Untersuchung von Effekten eines Faktors anhand von unbearbeiteten Messdaten. Hierfür können beispielsweise die Messdaten der beiden Sternpunkte eines Faktors und des Zentrumspunktes miteinander verglichen werden. Diese drei Punkte unterscheiden sich ausschließlich in der Einstellung dieses Faktors. Alle anderen Faktoren sind in allen drei Punkten identisch eingestellt. Damit kann direkt in den unbearbeiteten Rohdaten der Effekt dieses Faktors isoliert betrachtet werden. Bei optimalen Versuchsplänen ist dies aufgrund des unregelmäßigen Aufbaus nicht möglich. Dafür erlauben sie die optimale Ausfüllung geometrisch komplex geformter Faktorräume. Wenn also kompliziertere Randbedingungen²¹ für die untersuchbaren Faktorbereiche vorliegen, sollte der Einsatz von optimalen Versuchsplänen erwogen werden. Im Normalfall sind hingegen zentral zusammengesetzte Versuchspläne die bessere Wahl.²²

¹⁸ Vgl. Abschnitt 2.5.2.3, vor allem Abbildung 2.23

¹⁹ Siehe S. 88

²⁰ Vgl. Abbildung 2.20

²¹ Beispielsweise wenn Druck und Temperatur Faktoren sind und die zulässige Druckbelastung des Prüflings von der Temperatur abhängt

²² Siehe auch (Kleppmann, 2011, S. 217)

Für die Untersuchung der nichtparametrischen Modellanteile kommen alle Versuchspläne aus Abschnitt 2.6.1 in Frage. Versuchspläne auf Basis von Gittern²³ bieten allerdings den Vorteil, dass sie einfach zu konstruieren sind. Bei rechteckigen Gittern, die parallel zu den Koordinatenachsen ausgerichtet sind, lassen sich zudem stets benachbarte Versuchspunkte finden, die sich nur in der Einstellung eines Faktors unterscheiden. Falls im fertigen Regressionsmodell zunächst unerklärliche Effekte sichtbar werden, dann kann diese Eigenschaft die Ermittlung von deren Ursachen deutlich vereinfachen.

Eine alternative Vorgehensweise zur detaillierten Abbildung des Effekts eines Faktors besteht in der kontinuierlichen Variation dieses Faktors innerhalb eines Versuchs. So kann der Effekt dieses Faktors mit geringem Aufwand sehr detailliert untersucht und erfasst werden. Hierbei wird die mögliche Auflösung der Abbildung nur von der Abtastrate der verwendeten Messtechnik begrenzt. Dabei muss allerdings darauf geachtet werden, dass bei der Variation dieses Faktors dynamische Effekte keine zu große Rolle spielen, da diese ein so erstelltes Modell verfälschen würden. Um dynamische Effekte zumindest teilweise zu erfassen, kann die Änderungsgeschwindigkeit des so variierten Faktors als Faktor in den linearen Modellteil aufgenommen werden.

Wenn ein additives Modell ohne Wechselwirkungen zwischen linearen und nichtparametrischen Faktoren ausreichend ist, dann können beide Teilmodelle in unabhängigen Versuchsreihen untersucht und anschließend auch separat bestimmt werden. Um eine zweistufige Bestimmung der linearen und nichtparametrischen Modellanteile in einem Varying-Coefficient-Modell zu ermöglichen, muss hingegen für jede Faktorstufenkombination der nichtparametrischen Faktoren jeweils ein vollständiges lineares Modell erstellbar sein.

Es müssen deshalb gekreuzte Versuchspläne eingesetzt werden. Diese wurden bereits in Abschnitt 2.5.5.1 im Kontext von Split-Plot-Versuchen kurz vorgestellt. Um einen solchen Versuchsplan zu erstellen, werden zunächst separate Versuchspläne für das lineare und das nichtparametrische Teilmodell erstellt. Anschließend wird an jedem Punkt des Versuchsplans für den linearen Modellanteil ein vollständiger Versuchsplan für den nichtparametrischen Modellanteil eingesetzt. Dies wird in Abbildung 7.1 graphisch verdeutlicht. Umgekehrt, könnte auch an jedem Punkt des Versuchsplans des nichtparametrischen Modellanteils ein vollständiger Versuchsplan für den linearen Modellansatz eingesetzt werden. Die entstehenden Faktorstufenkombinationen des Gesamtversuchsplans sind für beide Vorgehensweisen iden-

²³ Vgl. Abschnitt 2.6.1.1

tisch. Was sich hingegen unterscheidet ist die entstehende Reihenfolge der Versuche. Im Falle einer vollständigen Randomisierung ist dies unerheblich, da die Versuchsreihenfolge ohnehin vollständig neu bestimmt wird.

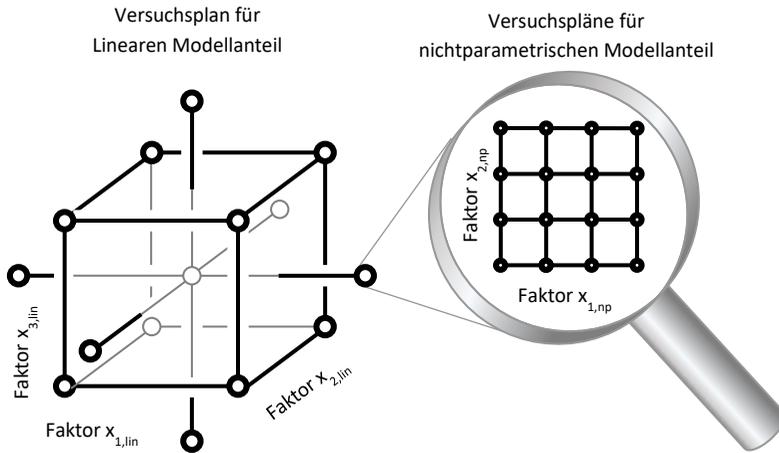


Abbildung 7.1: Gekreuzter Versuchsplan für semiparametrische Regressionsmodelle

Allerdings empfiehlt sich ein Randomisierungsschema analog zu den gekreuzten, separat randomisierten Versuchsplänen aus Abschnitt 2.5.5.1. Dabei sollte zunächst ein Versuchsplan für den linearen Modellanteil erstellt und randomisiert werden. Anschließend sollte für jeden Punkt dieses Versuchsplans ein vollständiger und jeweils separat randomisierter Versuchsplan für den nichtparametrischen Modellanteil eingesetzt werden. Durch dieses zweistufige Randomisierungsschema wird sichergestellt, dass jeder Versuchsplan für den nichtparametrischen Modellanteil jeweils in direkter Abfolge durchgeführt wird. Bei geringerer zeitlicher Distanz von zwei Versuchen kann davon ausgegangen werden, dass sich Fehlereinflüsse zwischen beiden Versuchen nur wenig verändern. Deshalb wird bei dieser Vorgehensweise die Fehlervarianz innerhalb der Versuchspläne für den nichtparametrischen Modellanteil minimiert. Die Fehlervarianz zwischen den Punkten des Versuchsplans für den linearen Modellanteil wird hingegen maximiert. Dies ist vorteilhaft, da nur für den linearen Modellanteil statistische Methoden zur Quantifizierung der Fehlervari-

anz und deren Auswirkungen existieren. Der Anteil der Fehlereinflüsse, der in statistischen Bewertungsmethoden²⁴ nicht berücksichtigten werden kann, und dadurch die mit ihnen erstellten Bewertungsgrößen verfälscht, wird so also minimiert.

Schwierig wird dieses zweistufige Randomisierungsschema allerdings, wenn Faktoren, deren Einfluss nichtparametrisch abgebildet werden soll, sehr aufwendig zu verändern sind. Das oben vorgeschlagene Randomisierungsschema steht dann einer effizienten Split-Plot Vorgehensweise im Wege.²⁵ In solchen Fällen muss dann zwischen der Reduzierung des Versuchsaufwand und der Genauigkeit der statistischen Bewertungsverfahren abgewogen werden.

7.3.4 Ermittlung des semiparametrischen Regressionsmodells

Nach Abschluss einer gemäß Abschnitt 7.3.3 aufgebauten Versuchsreihe existieren für jede Einstellungskombination der nichtparametrisch abzubildenden Faktoren Versuchsergebnisse eines vollständigen Versuchsplans der linear abzubildenden Faktoren. An jeder Einstellungskombination der nichtparametrisch abzubildenden Faktoren x_{np} wird nun zunächst jeweils ein lineares Regressionsmodell gemäß Gleichung 7.1 erstellt.

$$y(x_{np}) = X_{lin}\beta(x_{np}) + \varepsilon(x_{np}) \quad 7.1$$

Jedes dieser Modelle bildet dann die Einflüsse der linearen Faktoren für die jeweilige Einstellungskombination x_{np} der nichtparametrisch abzubildenden Faktoren ab. Die Koeffizienten $\beta(x_{np})$ dieser Regressionsmodelle weisen nun in Abhängigkeit der nichtparametrisch abzubildenden Faktoren unterschiedliche Werte auf. Für jedes Element dieser Koeffizienten wird nun separat jeweils ein nichtparametrisches Regressionsmodell erstellt, das diese Abhängigkeiten beschreibt. Das fertige Varying-Coefficient-Modell ist anschließend durch Gleichung 2.65 gegeben.

7.4 Demonstration am Beispiel von Kugelhähnen

Nachfolgend wird die Modellierung des Drehmoments beschrieben, das zum Öffnen von Kugelhähnen²⁶ benötigt wird, wenn der Kugelhahn zum Entlüften eines kleinen pneumatischen Druckspeichers eingesetzt wird. Das auf den Kugelhahn wirkenden

²⁴ Siehe Abschnitte 2.5.2.2, 2.5.3.3 und 2.5.4.2

²⁵ Vgl. Kapitel 8

²⁶ Siehe Abschnitt 5.2

Drehmoment setzt sich dabei im Wesentlichen aus Strömungs- und Reibmomenten zusammen.

Gemäß der Empfehlung aus Kapitel 6 sollte zunächst versucht werden, möglichst große Anteile des zu modellierenden Verhaltens eines technischen Systems durch White-Box-Modelle abzubilden. Dies erfolgt für Kugelhähne in Abschnitt 7.4.1. Anschließend wird in Abschnitt 7.4.2 gezeigt, warum die Modellierung des Betätigungsmoments von Kugelhähnen als reines White-Box-Modell nicht zielführend ist. Als alternative Vorgehensweise wird deshalb empirische Modellierung durch ein semi-parametrisches Grey-Box-Modell eingesetzt und in Abschnitt 7.4.3 beschrieben. Die Ergebnisse hiervon werden in Abschnitt 7.4.4 vorgestellt.

7.4.1 Modellierung der Kinematik und der Strömungskräfte

Die Kinematik von Kugelhähnen ist einfach, da die Ventilkugel und die mit ihr verbundene Ventilspindel die einzigen beweglichen Komponenten sind. Diese lassen sich gemeinsam um die in Abbildung 5.2 vertikal eingezeichnete Spindelachse drehen. Zur Modellierung des zum Öffnen eines Kugelhahnes benötigten Drehmoments müssen deshalb lediglich die Massenträgheit der Ventilkugel und Ventilspindel, sowie die an ihnen angreifenden Kräfte und Momente abgebildet werden.

Die auf die Ventilkugel wirkenden Kräfte und Momente sind in Abbildung 7.2 dargestellt. Während Abbildung 5.2 einen Schnitt durch die Spindelachse eines Kugelhahns zeigt, ist der Schnitt in Abbildung 7.2 senkrecht zur Spindelachse. Das an der Spindel anliegende Drehmoment M_{Sp} lässt sich aus dem Momentengleichgewicht entsprechend Abbildung 7.2 zu Gleichung 7.2 ableiten. Darin ist M_{Fl} das durch die Strömung verursachte Drehmoment parallel zur Spindelachse. r_K ist der Radius und J_K das kombinierte Massenträgheitsmoment der Ventilkugel und der Spindel um die Spindelachse. Die Reibkräfte $F_{R,1}$ und $F_{R,2}$ zwischen der Ventilkugel und den beiden Primärdichtungen hängen direkt von den im Reibkontakt wirkenden Normalkräften $F_{N,1}$ und $F_{N,2}$ ab. Diese sind wiederum abhängig von der Dichtungsvorspannung und den Strömungskräften $F_{Fl,x}$ und $F_{Fl,y}$.

$$M_{Sp} = M_{Fl} + (F_{R,1} + F_{R,2}) \cdot r_K + J_K \ddot{\varphi} \quad 7.2$$

Die oben dargestellte Betrachtungsweise ist allerdings stark vereinfacht, da die Reibkräfte $F_{R,1}$ und $F_{R,2}$ nicht punktuell an der Ventilkugel angreifen, sondern flächig über die beiden Dichtkontakte. Hierauf wird in Abschnitt 7.4.2 genauer eingegangen.

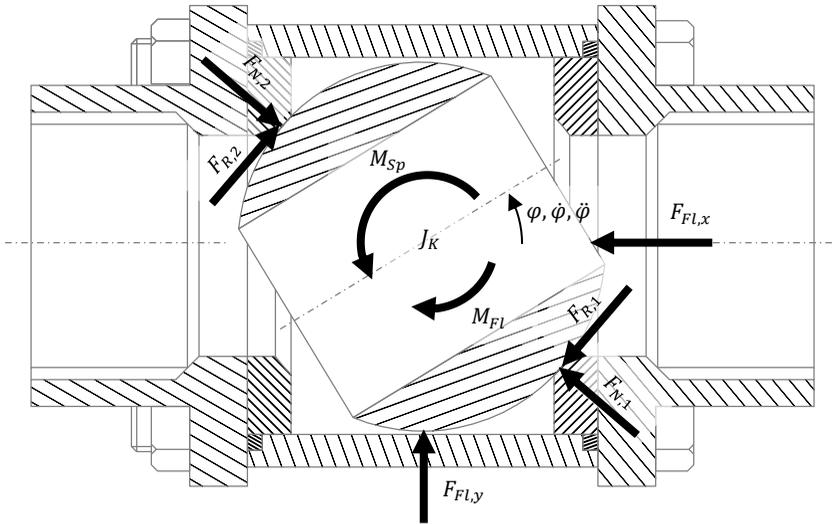


Abbildung 7.2: An der Ventilkugel angreifende Kräfte

Zur Modellierung des durch die Strömung verursachten Drehmoments M_{Fl} wird dieses zunächst durch mehrere dynamische CFD-Simulationen²⁷ bei unterschiedlichen Drücken bestimmt. Die Ergebnisse dieser Simulationen werden anschließend in einem Metamodell abgebildet, das dann zur Berechnung von M_{Fl} verwendet wird. Da das Strömungsmodell die im Rahmen dieser Arbeit zu erarbeitende Vorgehensweise nicht beeinflusst, wird in dieser Arbeit nicht genauer darauf eingegangen.

Wenn das Massenträgheitsmoment J_K bekannt ist, sind die Reibkräfte $F_{R,1}$ und $F_{R,2}$ die einzigen verbleibenden Unbekannten in Gleichung 7.2. Ihre Modellierung wird nachfolgend in Abschnitt 7.4.3 erläutert.

7.4.2 White-Box Reibungsmodellierung

Die in Abbildung 7.2 dargestellten Reibkräfte $F_{R,1}$ und $F_{R,2}$ greifen in Wirklichkeit nicht, wie dort vereinfacht dargestellt, punktuell an der Ventilkugel an. Vielmehr besteht durch die Reibung eine Tangentialspannung im gesamten Wirkkontakt der Dichtung. Die Größe und Lage des Wirkkontakts ist dabei variabel, da nicht immer

²⁷ „Computational Fluid Dynamics“, also numerische, meist zwei- oder dreidimensionale Strömungssimulationen

die gesamte Wirkfläche der Dichtung in Kontakt mit der Ventilkugel steht (siehe Abbildung 7.3). Und auch innerhalb des Wirkkontakts sind die Normalspannungen und der Schmierzustand nicht überall gleich, da sie unter anderem durch verschiedene Gleitrichtungen beeinflusst werden.

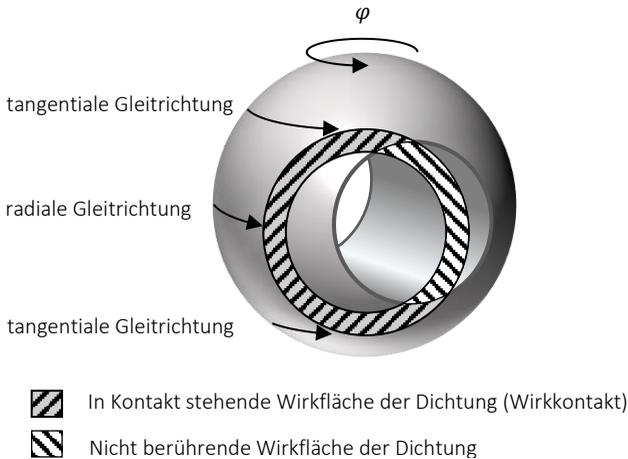


Abbildung 7.3: Wirkfläche zwischen Primärdichtung und Ventilkugel

Die Normalspannungen im Dichtkontakt werden durch die Dichtungsvorspannung, die Strömungskräfte $F_{FL,x}$ und $F_{FL,y}$, Differenzen zwischen den Reibkräften $F_{R,1}$ und $F_{R,2}$, sowie Kriech- und Relaxationsvorgänge des Dichtungswerkstoff beeinflusst. Diese entstehen aufgrund der viskoplastischen²⁸ Eigenschaften des als Dichtungswerkstoff eingesetzten PTFEs. Wenn beim Öffnen oder Schließen des Kugelhahnes Dichtungsanteile in Kontakt mit der Ventilkugel kommen, die zuvor im Bereich der Bohrung der Kugel lagen, dann baut sich die anfangs hohe Flächenpressung erst langsam durch Spannungsrelaxation ab. Umgekehrt baut sich in Bereichen, in denen die Dichtung nicht im Kontakt mit der Ventilkugel ist, eine aufgrund vorherigen Kontakts zurückbleibende Verformung nur langsam und unvollständig durch Kriechen ab. Die Normalspannungen im Dichtkontakt hängen zusätzlich von

²⁸ Viskoplastisches Materialverhalten beinhaltet elastische, viskose und plastische Anteile. Details zu diesem Verhalten und seiner Modellierung liefern beispielsweise Kletschkowski et al. (2001)

der Temperatur ab. Da PTFE einen deutlich höheren Wärmeausdehnungskoeffizient als Stahl hat, erhöht sich bei steigender Temperatur die Dichtungsverspannung. Temperaturänderungen wirken sich zudem auf die Materialeigenschaften der Dichtung aus.

Wie schon in Abschnitt 6.4 für Schwenkantriebe beschrieben, hat auch bei Kugelhähnen der Schmierzustand im Reibkontakt entscheidenden Einfluss auf die entstehenden Reibkräfte. Dieser wird unter anderem durch Stillstandszeiten nach vorangegangenen Bewegungen beeinflusst. Zudem variiert die Gleitrichtung der Kugeloberfläche auf der Dichtung je nach Position entlang des Dichtungsumfangs zwischen tangential und radial (siehe Abbildung 7.3). Dies hat ebenfalls Auswirkungen auf den Aufbau des Schmierfilms.

Während bei pneumatischen Schwenkantrieben über die Lebensdauer hinweg keine relevanten Auswirkungen von Verschleiß auf die entstehenden Reibkräfte erkennbar sind, sind bei Kugelhähnen schon nach wenigen Schaltspielen Verschleißerscheinungen sichtbar. Diese reichen vom reinen Materialabtrag an der Dichtungsoberfläche, mit damit einhergehendem Verlust an Dichtungsverspannung, bis hin zum Materialübertrag auf die Oberfläche der Ventilkugel.

Ein konventionelles Reibmodell für Kugelhähne könnte prinzipiell ähnlich zur in Abschnitt 6.4.2 für pneumatische Schwenkantriebe beschriebenen Vorgehensweise als White-Box-Modell erstellt werden. Diese ohnehin schon komplexe Vorgehensweise müsste allerdings um die zusätzlich abzubildenden Einflüsse von Armaturstellung, Verschleiß und Strömungskräften ergänzt werden. Zur Modellierung des Effekts des Verschleißes wäre die Ergänzung des Kennfeldes aus FE-Simulationsergebnissen um eine weitere Dimension erforderlich, sodass die Normalspannungen auch in Abhängigkeit des Verschleißzustandes beschrieben werden können.²⁹ Des Weiteren wäre ein zusätzliches Teilmodell zur Berechnung des entstehenden Verschleißes notwendig. Zudem müsste auch eine Möglichkeit gefunden werden, die unterschiedlichen Schmierzustände im Dichtspalt abzubilden. Ein Ansatz wäre eine diskretisierte Berechnung des Reibmodells für unterschiedliche Dichtungsbereiche. Dies erhöht allerdings den Aufwand zur Berechnung des Modells und erfordert eine Erweiterung, die das in Kontakt tretende Dichtungsflächen beschreibt.

Auch wenn diese Schwierigkeiten prinzipiell lösbar sind, wird dadurch die Erstellung eines solchen Reibungsmodells so aufwendig, dass sie für die Modellbildung im Rahmen einer normalen Produktentwicklung nicht praktikabel ist. Aus diesem Grund

²⁹ Vgl. S. 119

wird auch in diesem Beispiel zur Reibungsmodellierung ein Grey-Box-Modell eingesetzt. Aufgrund des stark nichtlinearen Einflusses der Armaturstellung ist hierbei allerdings, im Gegensatz zum Praxisbeispiel aus Kapitel 6, die Verwendung eines linearen Modells als Black-Box-Teilmodell nicht ausreichend. Stattdessen kommt ein semiparametrisches Regressionsmodell zum Einsatz. Dessen Erstellung wird im nachfolgenden Abschnitt beschrieben.

7.4.3 Reibungsmodellierung durch ein semiparametrisches Regressionsmodell

Aus den zuvor beschriebenen Vorgängen ergibt sich, dass, wie auch bei den Schwenkantrieben, die Temperatur, die anliegende Drücke, die Stellgeschwindigkeit und Stillstandszeiten einen Einfluss auf das Reibmoment von Kugelhähnen besitzen. Zusätzlich wird bei Kugelhähnen das Reibmoment auch durch die Armaturstellung sowie den Verschleißzustand beeinflusst. Diese Einflüsse gilt es in einem Modell geeignet abzubilden.

Analog zur in Abschnitt 6.4.2 für pneumatische Schwenkantriebe beschriebenen Vorgehensweise kann auch zur Modellierung des Reibmoments von Kugelhähnen ein Reibmodell nach Carlson und Batista (1996) als Grundlage dienen. Bei Kugelhähnen verläuft der Übergang von Gleitreibung zur Haftreibung allerdings sehr viel langsamer, als bei pneumatischen Schwenkantrieben. Eine Bestimmung der Abhängigkeit der Parameter τ und $M_{fr,s}$ von den Black-Box-Faktoren x_{BB} wäre deshalb extrem zeitaufwendig. Aus diesem Grund wird für Kugelhähne ein vereinfachtes Reibmodell eingesetzt.

Darin wird zur Reduzierung des Versuchsaufwands für Kugelhähne lediglich das Gleitreibmoment $M_{fr,d}$ in Abhängigkeit der Black-Box-Faktoren empirisch modelliert. Die restlichen Parameter in den Gleichungen 6.18 und 6.19 werden durch Konstanten angenähert, die bei Raumtemperatur im druckfreien Zustand ermittelt werden. Bei den Schwenkantrieben sind die Effekte der Black-Box-Faktoren auf das Haftreibmoment $M_{fr,s}$ und das Gleitreibmoment $M_{fr,d}$ sehr ähnlich.³⁰ Deshalb wird das Reibungsmodell für die Kugelhähne so angepasst, dass die Gleichungen 7.3 und 7.4 entstehen. Darin sind $M_{fr,s,NB}$, $M_{fr,d,NB}$, τ_{NB} und D_{NB} die Parameter des Reibungsmodells nach Carlson und Batista im druckfreien Zustand bei Raumtemperatur. Unter diesen Bedingungen entsprechen die Gleichungen 7.3 und 7.4 den Gleichungen 6.18 und 6.19. Durch die Umstellung von Gleichung 6.18 zu Gleichung 7.3

³⁰ Vgl. Abbildung 6.3 und Abbildung 6.4

wirken allerdings alle in $M_{fr,d}(x_{BB})$ enthaltenen Effekte in gleichem Maße auch auf das Haftreibungsmoment.

$$M_{fr} = (1 + \Theta \cdot \Delta M_{fr,rel,NB}) \cdot M_{fr,d}(x_{BB})$$

$$\text{mit } \Delta M_{fr,rel,NB} = \frac{M_{fr,s,NB} - M_{fr,d,NB}}{M_{fr,d,NB}} \quad 7.3$$

$$\Theta = \frac{\Theta}{\tau_{NB}} (1 - \Theta) - \frac{1}{D_{NB}} \Theta \dot{\varphi} \quad 7.4$$

Das einzige in diesen Gleichungen enthaltene Black-Box-Teilmodell ist also $M_{fr,d}(x_{BB})$. In diesem werden die Einflüsse der folgenden Black-Box-Faktoren auf das Gleitreibungsmoment abgebildet:

- Armaturstellung φ
- Stellgeschwindigkeit $\dot{\varphi}$
- Temperatur T
- Druckdifferenz Δp
- Absolvierte Schaltspielzahl n (Verschleiß)

Für jeden dieser fünf Faktoren muss nun abgewogen werden, ob eine Abbildung durch ein lineares Modell ausreichend ist. Für Stellgeschwindigkeit, Schaltspielzahl, Temperatur und Druckdifferenz wird hiervon ausgegangen. Um kleinere Nichtlinearitäten abbilden zu können, wird ein Polynom zweiten Grades inklusive aller Wechselwirkungen zweiten Grades eingesetzt. Der Einfluss der Armaturstellung ist hingegen so stark nichtlinear, dass dieser Faktor durch ein nichtparametrisches Modell abgebildet werden muss. Da nur ein einzelner Faktor nichtparametrisch abgebildet werden soll, kommt hierfür, gemäß der Empfehlung aus Abschnitt 7.3.2, ein TP-Spline zum Einsatz.

Als Gesamtmodell wird ein semiparametrisches Regressionsmodell eingesetzt. Bei dessen Wahl muss abgeschätzt werden, ob ein additives Modell ohne Wechselwirkungen zwischen linearen und nichtparametrischen Faktoren ausreicht. Da in Abhängigkeit der Armaturstellung unterschiedliche physikalische Vorgänge auf das Reibmoment wirken,³¹ muss davon ausgegangen werden, dass auch der Effekt der

³¹ Vgl. Abschnitt 7.4.2

linear abzubildenden Faktoren von der Armaturstellung abhängig ist. Ein wechselwirkungsfreies additives Modell reicht hier also nicht aus. Stattdessen wird ein Varying-Coefficient-Modell nach Gleichung 2.65 eingesetzt. Der Aufbau aller Variablen dieses Modells wird, für den hier vorgestellten Fall, in Gleichung 7.5 gezeigt.

$$\begin{aligned}
 y &= M_{fr,d} \\
 \mathbf{x}_{BB,np} &= \varphi \\
 \mathbf{x}_{BB,lin} &= (1 \quad \dot{\varphi} \quad T \quad \Delta p \quad n \quad \dot{\varphi}^2 \quad T^2 \quad \Delta p^2 \quad n^2 \quad \dot{\varphi}T \quad \dot{\varphi}\Delta p \quad \dot{\varphi}n \quad T\Delta p \quad Tn \quad \Delta pn) \\
 \boldsymbol{\beta}(\mathbf{x}_{BB,np}) &= \boldsymbol{\beta}(\varphi) = (\beta_0(\varphi) \quad \beta_{\dot{\varphi}}(\varphi) \quad \beta_T(\varphi) \quad \beta_{\Delta p}(\varphi) \quad \beta_n(\varphi) \quad \dots \\
 &\quad \beta_{\dot{\varphi}^2}(\varphi) \quad \beta_{T^2}(\varphi) \quad \beta_{\Delta p^2}(\varphi) \quad \beta_{n^2}(\varphi) \quad \dots \\
 &\quad \beta_{\dot{\varphi}\Delta p}(\varphi) \quad \beta_{\dot{\varphi}n}(\varphi) \quad \beta_{T\Delta p}(\varphi) \quad \beta_{Tn}(\varphi) \quad \beta_{\Delta pn}(\varphi))'
 \end{aligned} \tag{7.5}$$

Als Versuchsplan für den linearen Modellteil wird ein Equivalent-Estimation-Versuchsplan auf Basis eines zentral zusammengesetzten Versuchsplans eingesetzt. Dieser beinhaltet die linear abzubildenden Black-Box-Faktoren Stellgeschwindigkeit, Temperatur, Druckdifferenz und die Schaltspielzahl.

Die Armaturstellung, als einziger nichtparametrisch abzubildender Black-Box-Faktor, kann kontinuierlich variiert werden. Es ist also nicht notwendig, dass zur Untersuchung dieses Faktors ein separater Versuchsplan aufgestellt und anschließend mit dem Equivalent-Estimation-Versuchsplan zu einem gekreuzten Versuchsplan³² zusammengesetzt wird. Stattdessen werden in jedem Einzelversuch des Equivalent-Estimation-Versuchsplans Schwenkbewegungen über den gesamten interessierenden Winkelbereich durchgeführt. Dadurch, dass die Schwenkgeschwindigkeit $\dot{\varphi}$ als Faktor im linearen Teilmodell enthalten ist, können die aus ihr resultierenden dynamischen Effekte zumindest näherungsweise im Modell erfasst werden.

Das Ergebnis der so durchgeführten Versuchsreihe ist für jeden Versuch jeweils ein aufgezeichneter Messwerteverlauf für die Armaturstellung φ und das Reibmoment $M_{fr,d}$. Zur Erstellung des Regressionsmodells nach Gleichung 7.5 wird zunächst durch Interpolation das Reibmoment für jede Armaturstellung von $\varphi = 0^\circ$ bis $\varphi = 90^\circ$ in Schritten von $0,1^\circ$ bestimmt. Das so ermittelte Reibmoment wird nun für jede dieser Armaturstellungen durch jeweils ein lineares Modell gemäß Gleichung

³² Vgl. S. 71

7.6 approximiert. Durch ein geeignetes Schätzverfahren³³ wird dabei jeweils ein Schätzwert für den Vektor $\hat{\beta}$ bestimmt.

$$y(\varphi) = \mathbf{x}_{BB,lin}\boldsymbol{\beta}(\varphi) + \varepsilon \quad 7.6$$

So erhält man für jede Armaturstellung einen Vektor $\hat{\boldsymbol{\beta}}(\varphi)$. Die Abhängigkeit dessen Komponenten $\hat{\beta}_*$ von der Armaturstellung kann nun durch jeweils ein nichtparametrisches Regressionsmodell abgebildet werden. Hierbei kommen in diesem Beispiel, wie weiter oben beschrieben, TP-Splines gemäß Gleichung 2.59 zum Einsatz. Zu deren Bestimmung wird das gewöhnliche Verfahren der kleinsten Fehlerquadrate³⁴ verwendet.

7.4.4 Ergebnisse

Die bei Reibungsuntersuchungen erhaltenen Ergebnisse werden nachfolgend dargestellt und erläutert. Jede der nachfolgenden Grafiken besteht aus zwei Diagrammen. Das jeweils obere Diagramm zeigt die Verläufe der geschätzten Koeffizienten $\hat{\boldsymbol{\beta}}(\varphi)$ über die Armaturstellung hinweg. Die aus den Messwerten direkt abgeleiteten Kennlinien sind in den nachfolgenden Grafiken mit durchgehender Linie eingezeichnet. Die im Modell verwendeten Splines, die auf diese Verläufe angepasst sind, werden in den nachfolgenden Grafiken gestrichelt dargestellt. In allen Diagrammen sind stets die Kennlinien aller Faktoren enthalten. Die des aktuell diskutierten Faktors ist dabei schwarz und die aller anderen Faktoren sind hellgrau dargestellt. So kann die Effektgröße jedes Faktors besser mit der der restlichen Faktoren verglichen werden.

Im jeweils unteren Diagramm ist der p-Wert des betrachteten Koeffizienten in Abhängigkeit der Armaturstellung dargestellt. Auch hier ist der Verlauf des aktuell betrachteten Koeffizienten schwarz und die Verläufe aller anderen Koeffizienten sind hellgrau eingezeichnet. Schwarz gestrichelt ist in diesem Diagramm die Grenze von 0,05 eingezeichnet, bis zu der der Effekt des zugehörigen Faktors als statistisch signifikant betrachtet wird. Da der p-Wert nur zur Bewertung der statistischen Signifikanz der Effekte dient, aber im Reibungsmodell keine Verwendung findet, werden die hier gezeigten Kurvenverläufe nicht durch Splines angenähert.

³³ Siehe Abschnitte 2.5.2.1, 2.5.3.1 und 2.5.3.2

³⁴ Siehe Abschnitt 2.5.2.1

In den nachfolgenden Abschnitten werden nur statistisch signifikante Einflüsse erläutert. Statistisch nicht signifikante Einflüsse finden sich ohne weitere Erläuterungen im Anhang.

Mittleres Reibmoment

Das bei der Versuchsreihe bestimmte mittlere Reibmoment des Kugelhahns ist in Abbildung 7.4 dargestellt. Dieses Diagramm zeigt den Reibmomentenverlauf für Versuchsbedingungen, bei denen alle Faktoren auf ihren Mittelwert eingestellt sind. Alle Faktoren besitzen hier also normiert den Wert Null.

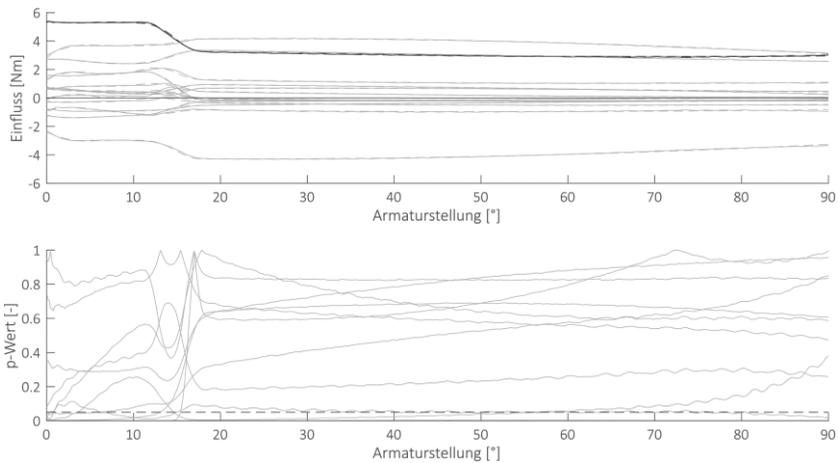


Abbildung 7.4: Mittleres Reibmoment

Anhand dieses Diagramms kann der Versuchsablauf gut nachvollzogen werden. Jeder Versuch beginnt in geschlossener Stellung (0°). In diesem Zustand liegt die eingestellte Druckdifferenz am Kugelhahn an. Bei einer Armaturstellung von etwa 12° öffnet sich ein erster Spalt durch die Bohrung der Ventilkugel. Durch diesen Spalt beginnt sich der Druckspeicher zu entleeren. Dadurch sinkt die Druckdifferenz zwischen den beiden Anschlüssen des Kugelhahns. Dies hat ein Absinken der Strömungskräfte $F_{FL,x}$ und $F_{FL,y}$ zur Folge,³⁵ was wiederum ein Absinken des Reibmoments bewirkt. Der mögliche Durchfluss durch diesen Spalt steigt mit größerer Öffnungsweite schnell an. Bei Erreichen einer Armaturstellung von etwa 18° ist der

³⁵ Vgl. Abbildung 7.2

Druckausgleich bereits abgeschlossen. Für noch weiter geöffnete Armaturstellungen bleibt das Reibmoment anschließend nahezu konstant.

In geschlossener Stellung beträgt das mittlere Reibmoment etwa 5,3 Nm. Beim Abfallen der anliegenden Druckdifferenz sinkt dieser Wert auf etwa 3,3 Nm. Im Diagramm nur schwer zu erkennen ist, dass das Reibmoment über den weiteren Verlauf weiter abnimmt, bis es bei etwa 75° sein Minimum von etwa 2,8 Nm erreicht. Dies könnte einerseits damit zusammenhängen, dass die in Kontakt stehende Dichtungsfläche mit steigender Öffnungsweite immer kleiner wird. Kurz vor Erreichen der vollständig geöffneten Stellung (90°) nimmt die in Kontakt stehende Dichtungsfläche allerdings wieder schnell zu.³⁶ Da hier nur ein minimaler Anstieg des Reibmoments auf etwa 3,0 Nm erkennbar ist, kann der Reibmomentenverlauf eventuell auch auf eine leicht ovale Form der Ventilkugel hinweisen.

Der Verlauf des p-Werts in Abbildung 7.4 liegt so genau auf der x-Achse, dass er von dieser nicht unterschieden werden kann. Der Mittelwert des Reibmoments ist also für alle Armaturstellungen eindeutig signifikant von null verschieden.

Einfluss der Druckdifferenz

In Abbildung 7.5 ist der lineare Einfluss der Druckdifferenz dargestellt. Auch hier ist gut zu erkennen, dass nur innerhalb der ersten 12° des Öffnungsvorgangs wirklich die volle Druckdifferenz anliegt. Hier beträgt der lineare Einfluss der Druckdifferenz etwa 1,8 Nm. Die kleinste untersuchte Druckdifferenz in der Versuchsreihe betrug 0 bar. Diese besitzt den normierten Wert von -1. Im druckfreien Zustand ist in diesem Bereich also das Reibmoment um etwa 1,8 Nm kleiner, als das in Abbildung 7.4 dargestellte mittlere Reibmoment. Die dort sichtbare Erhöhung des Reibmoments innerhalb dieses Bereichs wird somit im druckfreien Zustand fast vollständig ausgeglichen.

Für Armaturstellungen von größer als 12° sinkt die Kennlinie für den linearen Einfluss der Druckdifferenz auf einen Wert von etwa -0,2 Nm ab. Eine hohe Druckdifferenz hat hier also eine Verringerung des Reibmoments zur Folge. Dies kann dadurch erklärt werden, dass im geschlossenen Zustand, in dem die Druckdifferenz tatsächlich anliegt, die Ventilkugel durch die Strömungskraft $F_{Fl,x}$ in die gegenüberliegende Dichtung gedrückt wird. Hierdurch wird diese Dichtung komprimiert und kriecht beim Öffnen des Kugelhahns nach Abfallen der Druckdifferenz nur langsam in ihre ursprüngliche Form zurück. Dies führt zu einer kurzzeitig verringerten Dichtungsvorspannung und dadurch zu einem geringeren Reibmoment. Wie am Verlauf

³⁶ Vgl. Abbildung 7.3

des p-Werts unten in Abbildung 7.5 erkannt werden kann, ist dieser Einfluss allerdings nicht statistisch signifikant nachzuweisen.

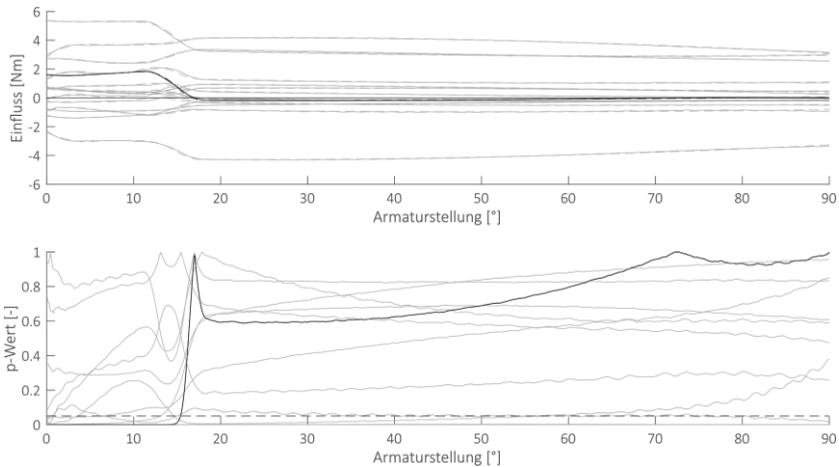


Abbildung 7.5: Einfluss der Druckdifferenz

Der quadratische Einfluss der Druckdifferenz ist im Anhang in Abbildung A.1 dargestellt. Dieser ist über alle Armaturstellungen hinweg sehr klein und auch nahezu durchgängig statistisch nicht signifikant.

Einfluss der Temperatur

In Abbildung 7.6 ist der lineare Einfluss der Temperatur auf die Reibkraft zu sehen. Über den gesamten Öffnungsweg hinweg ist dieser Einfluss positiv und klar statistisch signifikant. Eine Temperaturerhöhung bewirkt also auch eine Erhöhung des Reibmoments. Dies kann durch die unterschiedlichen thermischen Ausdehnungskoeffizienten der Dichtung und des Gehäuses erklärt werden. Die aus PTFE³⁷ bestehende Dichtung besitzt einen deutlich höheren thermischen Ausdehnungskoeffizienten als das aus rostfreiem Stahl gefertigte Gehäuse. Bei Erwärmung dehnt sich die Dichtung also stärker aus als das Gehäuse, was zu einer erhöhten Dichtungsverspannung führt. Dieser Effekt ist im geschlossenen Zustand mit etwa 2,5 Nm

³⁷ Polytetrafluorethylen, oft als „Teflon“ bezeichnet (Handelsname der Firma DuPont)

etwas kleiner, als im geöffneten Zustand mit etwa 3 Nm. Dies könnte daher kommen, dass die Ventilkugel im geschlossenen Zustand durch die Strömungskraft $F_{FL,x}$ in eine der beiden Dichtungen gedrückt wird. Wenn dadurch der Kontakt zur entlasteten Dichtung vollständig verloren geht, hat hier eine Temperaturerhöhung zunächst keine Erhöhung der Dichtungsvorspannung mehr zur Folge. Erst sobald der Spalt in dieser Dichtfläche durch thermische Ausdehnung überbrückt ist, führt eine weitere Temperaturerhöhung zu einem Anstieg der Dichtungsvorspannung und damit auch zu einer Erhöhung des Reibmoments.

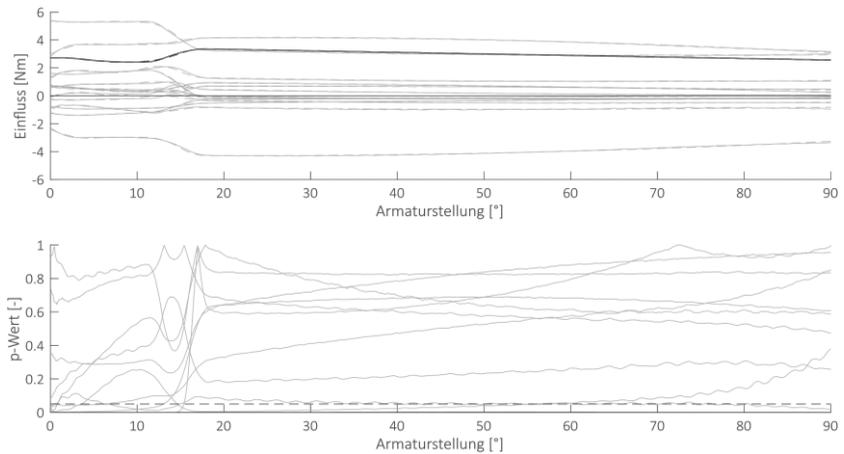


Abbildung 7.6: Einfluss der Temperatur

Diese These kann auch durch Abbildung 7.7 gestützt werden, in der der Effekt der Wechselwirkung zwischen Druckdifferenz und Temperatur auf das Reibmoment dargestellt ist. Diese Wechselwirkung besitzt ausschließlich für geschlossene Armaturstellungen einen signifikanten negativen Effekt. Das bedeutet, dass hier bei hohen anliegenden Druckdifferenzen der an sich positive Effekt einer Temperaturerhöhung³⁸ verkleinert wird.

³⁸ vgl. Abbildung 7.6

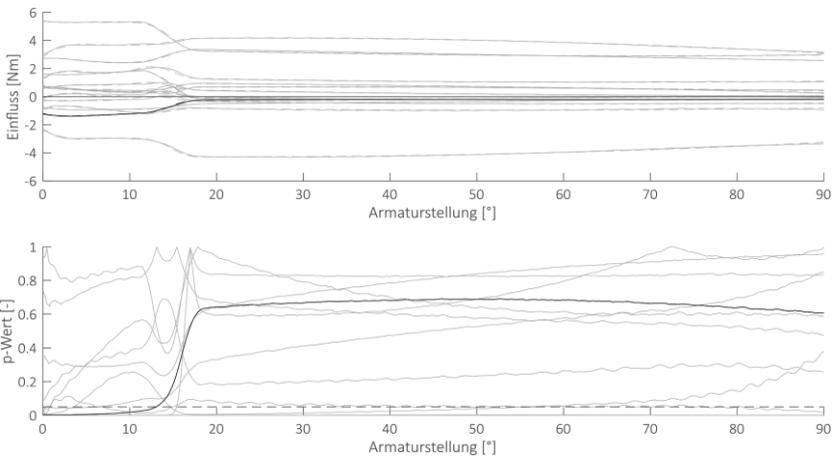


Abbildung 7.7: Wechselwirkung zwischen Druckdifferenz und Temperatur

Umgekehrt kann diese Wechselwirkung auch als Einfluss der Temperatur auf den Effekt der Druckdifferenz gedeutet werden. Im geschlossenen Zustand hat eine höhere Druckdifferenz eine Erhöhung des Reibmoments zur Folge.³⁹ Diese Erhöhung fällt bei hohen Temperaturen, aufgrund der hier größeren Dichtungsvorspannung, kleiner aus, als bei niedrigen Temperaturen.

Der in Abbildung 7.8 dargestellte quadratische Einfluss der Temperatur ist positiv, aber nicht für alle Armaturstellungen signifikant. Dies bedeutet, dass der Anstieg des Reibmoments über die Temperatur bei hohen Temperaturen größer ist, als bei niedrigen Temperaturen. Auch dies passt zur oben erläuterten These, da gemäß ihr Temperaturerhöhungen erst nach Überbrückung aller Spalte überhaupt eine Auswirkung besitzen. Mit dem vorhandenen Systemverständnis nicht erklärbar ist hingegen, dass der quadratische Einfluss im geschlossenen Zustand kleiner ist, als im geöffneten Zustand. Um eine Erklärung hierfür zu finden, wären weitere Untersuchungen notwendig.

³⁹ Vgl. Abbildung 7.5

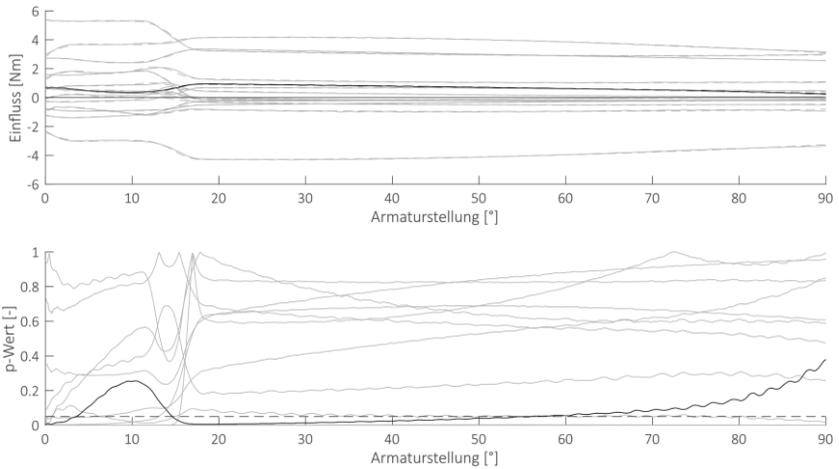


Abbildung 7.8: Quadratischer Einfluss der Temperatur

Einfluss der Öffnungsgeschwindigkeit

In Abbildung 7.9 ist der lineare Einfluss der Öffnungsgeschwindigkeit auf das Reibmoment zu sehen. Dieser ist über den gesamten Öffnungsvorgang hinweg positiv und statistisch signifikant. Höhere Öffnungsgeschwindigkeiten führen also zu erhöhten Reibwerten. Ähnlich der durchschnittlichen Reibkraft ist auch der Einfluss der Öffnungsgeschwindigkeit in geschlossenen Armaturstellungen leicht erhöht.

Der in Abbildung 7.10 dargestellte quadratische Effekt der Geschwindigkeit ist durchgehend negativ und etwa halb so groß wie deren linearer Effekt. Dies bedeutet, dass sich der Anstieg des Reibmoments über die Geschwindigkeit für hohe Geschwindigkeiten verlangsamt. Der zum quadratischen Effekt gehörende p-Wert schwankt um die Signifikanzgrenze von 0,05.

Abbildung von Nichtlinearitäten

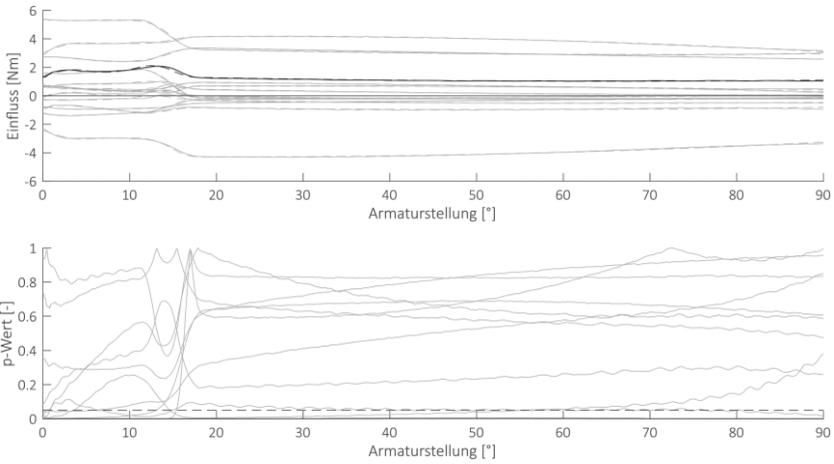


Abbildung 7.9: Einfluss der Öffnungsgeschwindigkeit

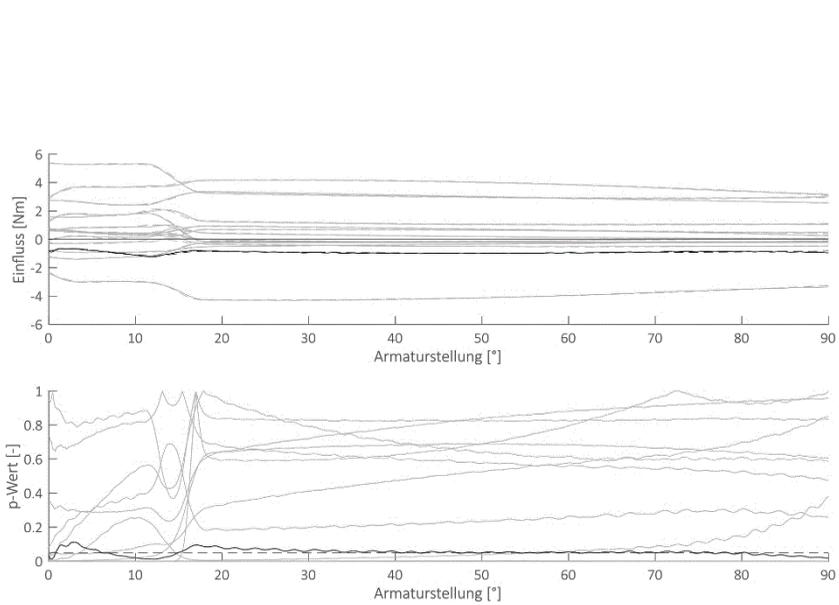


Abbildung 7.10: Quadratischer Einfluss der Öffnungsgeschwindigkeit

Eine interessante, wenn auch nicht statistisch signifikante Wechselwirkung ist in Abbildung 7.11 dargestellt. Diese besitzt durchgängig nahezu keinen Einfluss. Lediglich zwischen etwa 12° und 18° ist der Kurvenverlauf deutlich erhöht, obgleich auch diese Werte keine statistische Signifikanz erreichen. Dies ist der Bereich, in dem der Kugelhahn einen ersten Strömungsquerschnitt freigibt und darüber den Druckspeicher auf der Versorgungsseite entleert. Bei hohen Öffnungsgeschwindigkeiten erreicht der Kugelhahn eine weiter geöffnete Stellung, bevor der Druckspeicher vollständig entleert ist. Dadurch liegt über einen größeren Winkelbereich eine Druckdifferenz und damit ein erhöhtes Reibmoment an. Diese Erhöhung des Druckeinflusses für größere Geschwindigkeiten führt zum in Abbildung 7.11 sichtbaren Verlauf dieser Wechselwirkung.

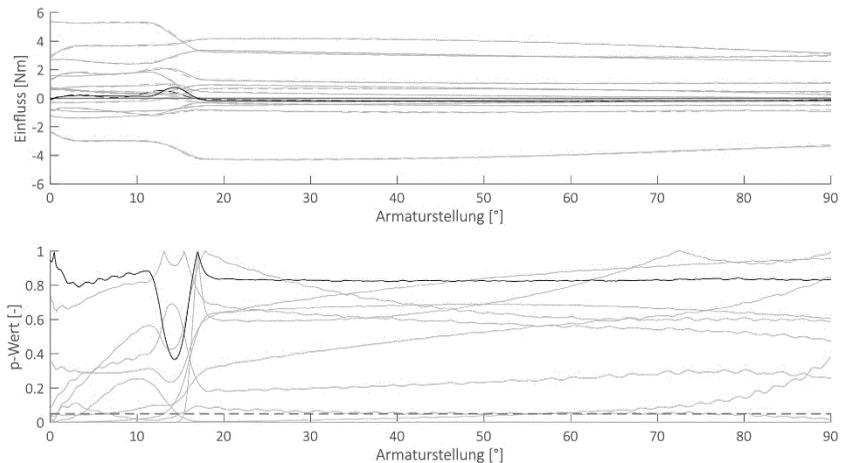


Abbildung 7.11: Wechselwirkung zwischen Öffnungsgeschwindigkeit und Druck

Einfluss des Verschleißes

In Abbildung 7.12 ist der lineare Einfluss des Verschleißes auf das Reibmoment zu sehen. Dieser ist durchgehend negativ. Der in Abbildung 7.13 gezeigte quadratische Einfluss ist hingegen durchgehend positiv und besitzt einen nahezu exakt spiegelbildlichen Verlauf zum linearen Einfluss. Im geschlossenen Zustand besitzen beide Kurvenverläufe einen Betrag von etwa 3 Nm, der bei der Öffnung des ersten Strömungsquerschnitts (etwa bei 12°) auf 4 Nm ansteigt. Im weiteren Verlauf nehmen die Beträge beider Kurven wieder leicht ab und erreichen etwa 3,2 Nm. Beide Einflüsse sind über den gesamten untersuchten Bereich hinweg klar signifikant.

Abbildung von Nichtlinearitäten

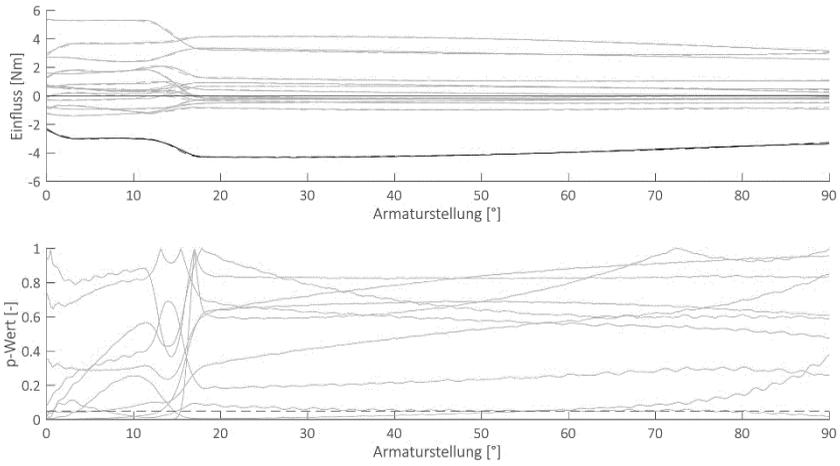


Abbildung 7.12: Einfluss des Verschleißes

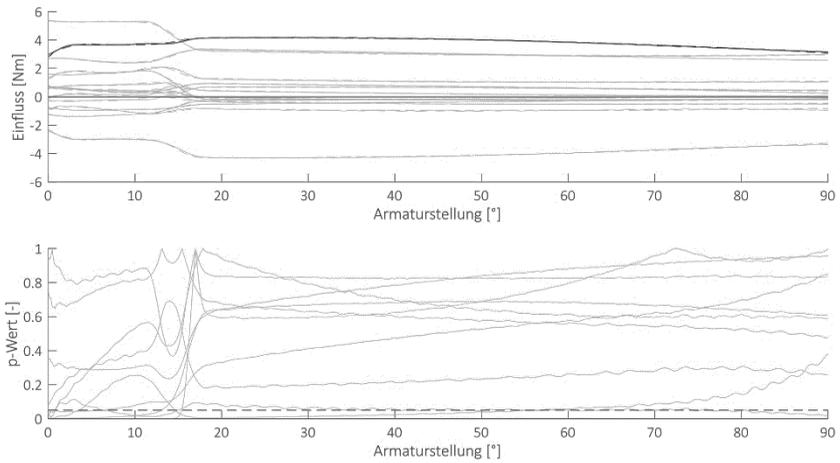


Abbildung 7.13: Quadratischer Einfluss des Verschleißes

Die negativen Werte des linearen Einflusses bedeuten, dass das Reibmoment für höhere Schaltspielzahlen abnimmt. Die positiven Werte desselben Betrags für den quadratischen Einfluss sind so zu deuten, dass nahezu das gesamte Absinken des Reibmoments in der ersten Hälfte der untersuchten Schaltspielzahlen abläuft: Der Neuzustand des Kugelhahns besitzt die normierte Schaltspielzahl -1 . Um das Reibmoment in diesem Zustand zu erhalten, muss der lineare Einfluss der Schaltspielzahl mit -1 und der quadratische Einfluss mit $(-1)^2=1$ multipliziert und anschließend zum durchschnittlichen Reibmoment aus Abbildung 7.4 hinzuaddiert werden. Somit ist im Neuzustand das Reibmoment beispielsweise bei einer Armaturstellung von 20° um etwa 8 Nm im Vergleich zum Reibmoment bei der Schaltspielzahl mit dem normierten Wert 0 erhöht. Für die größte untersuchte Schaltspielzahl mit einem normierten Wert von 1 gibt es hingegen keine Änderung im Vergleich zur mittleren untersuchten Schaltspielzahl.

Die Abnahme des Reibmoments in geschlossenen Armaturstellungen ist etwas geringer als im geöffneten Bereich. Dies liegt daran, dass hier die Normalspannung in einem der beiden Dichtkontakte durch die Strömungskraft $F_{Fl,x}$ aufrechterhalten wird, auch nachdem die Dichtungsvorspannung vollständig durch Verschleiß abgebaut ist.

7.5 Zwischenfazit

Die Abbildungsleistung für Nichtlinearitäten ist in linearen Modellen sehr begrenzt. In diesem Kapitel konnte gezeigt werden, wie diese Abbildungsleistung für Nichtlinearitäten einzelner Faktoren deutlich gesteigert werden kann.

Hierzu kommen semiparametrische Regressionsmodelle zum Einsatz. Besonders gut geeignet für den Einsatz in der Produktentwicklung sind Varying-Coefficient-Modelle, da sie sehr flexibel sind und trotzdem gut verständliche Parameter besitzen. Dadurch besitzen diese Modelle einerseits eine gute Abbildungsleistung und können in einer Produktentwicklung zudem zur Steigerung des Systemverständnisses führen.

Wenn nur ein Faktor starke Nichtlinearitäten aufweist und dynamische Effekte eine untergeordnete Rolle spielen, kann der Einfluss dieses Faktors durch kontinuierliche Variation seiner Einstellung untersucht werden. Dies ermöglicht eine sehr schnelle und trotzdem sehr hochauflösende Untersuchung seines Einflusses. Die Einflüsse aller anderen Faktoren können durch einen beliebigen Versuchsplan aus Abschnitt 2.5.1 untersucht werden. Die Untersuchung eines Variationsverlaufs des stark nicht-linearen Faktors an jedem Punkt dieses Versuchsplans generiert mit geringem Aufwand ausreichend umfangreiche Versuchsergebnisse, um ein Varying-Coefficient-

Modell erstellen zu können. Falls mehrere Faktoren stark nichtlineare Einflüsse aufweisen, oder dynamische Effekte eine große Rolle spielen, muss hierfür stattdessen auf gekreuzte Versuchspläne zurückgegriffen werden. Die Erstellung eines Varying-Coefficient-Modells aus den Ergebnissen beider Arten von Versuchsreihen ist durch ein relativ einfaches zweistufiges Verfahren möglich.

Alle Schritte von der Versuchsplanung bis zur Erstellung solchen Varying-Coefficient-Modells wurden in diesem Kapitel zunächst theoretisch beschrieben und abschließend mittels eines Praxisbeispiels demonstriert. Anhand der Ergebnisse dieses Beispiels wurde auch gezeigt, wie diese zu interpretieren sind und wie sie zu einem gesteigerten Verständnis über die im System ablaufenden physikalischen Vorgänge beitragen können.

Die zu Beginn dieses Kapitels gestellte Forschungsfrage kann damit als beantwortet betrachtet werden.

8 Steigerung der Effizienz von Versuchsreihen

In sehr vielen Fällen, wie auch in den Praxisbeispielen dieser Arbeit, benötigt die Durchführung eines Einzelversuchs wenige Sekunden. Die Änderung vieler Faktoren, wie beispielsweise eines Drucks, kann innerhalb einer ähnlichen Zeitspanne erfolgen. Es gibt aber auch Faktoren, deren Änderung mehrere Stunden Zeit benötigt. Hierzu zählt beispielsweise oft die Temperatur, da diese meist nur sehr langsam geändert werden kann. Wenn solche schwierig zu ändernden Faktoren in einer Versuchsreihe enthalten sind, dann hängt der Aufwand zu deren Durchführung nur zu einem geringen Teil von der Gesamtzahl der durchzuführenden Versuche ab. Vielmehr sind hier die Änderungen schwierig einzustellender Faktoren entscheidend. Wenn Versuchsreihen wie üblich vollständig randomisiert werden,¹ dann wird allerdings jeder Faktor bei nahezu jedem Versuch neu eingestellt. Durch eine effiziente Umsortierung der Versuchsreihe lässt sich hier viel Aufwand einsparen. Hierbei müssen allerdings gewisse Regeln eingehalten werden, um die Validität der Versuchsergebnisse nicht zu gefährden.

Zudem müssen auch die Randbedingungen berücksichtigt werden, die sich durch die Modellbildung im Rahmen einer Produktentwicklung ergeben. Der Zwang zu wirtschaftlichem Handeln bedeutet hier, dass der Gesamtaufwand zur Modellbildung minimiert werden sollte. Und darin enthalten sind neben der reinen Versuchsdurchführung auch Aufwände zur Versuchsplanung und zur Erstellung von Prüfständen, mit deren Hilfe die empirische Untersuchung des zu modellierenden Objekts durchgeführt wird. Im Optimalfall werden bei der Bewertung dieser Aufwände und der daraus folgenden Wahl der Vorgehensweise zur Modellbildung bereits zukünftige Produktgenerationen bedacht. Für diese kann eine Anpassung der Modelle und eine Wiederholung der durchgeführten Versuchsreihen erforderlich sein.

Die dritte zu beantwortende Forschungsfrage dieser Arbeit lautet deshalb:

Forschungsfrage 3

Wie sollten Versuchsreihen zur Modellierung technischer Systeme strukturiert werden, um mit möglichst geringem Aufwand möglichst detaillierte und valide Ergebnisse zu liefern?

¹ Vgl. Abschnitt 2.4.2.1

Die Antwort zu dieser Frage liefern Split-Plot-Versuchspläne.² Um ableiten zu können, welche Verfahren für den Einsatz in der Modellbildung für die Produktentwicklung am besten geeignet sind, werden zunächst in Abschnitt 8.1 die speziellen Anforderungen in dieser Anwendung erarbeitet. In den Abschnitten 8.2 und 8.3 werden Empfehlungen zur Auswahl und zum Einsatz geeigneter Split-Plot-Versuchspläne gegeben, um diese Anforderungen bestmöglich zu erfüllen. Anschließend wird die Anwendung dieser Verfahren in Abschnitt 8.4 anhand eines Praxisbeispiels aus Kapitel 5 demonstriert und gezeigt, welche Ergebnisse damit erzielt werden können. Auf Basis der Ergebnisse folgt in Abschnitt 8.5 ein kurzes Zwischenfazit zum Einsatz von Split-Plot-Versuchsreihen zur Modellbildung in der Produktentwicklung.

8.1 Spezielle Anforderungen in der Produktentwicklung

Zur empirischen Modellierung von technischen Systemen, Teilsystemen und Komponenten wird üblicherweise zunächst ein Prüfstand erstellt. Dieser muss das zu modellierende Objekt den relevanten Einflussfaktoren aussetzen, Störgrößeneinflüsse möglichst klein oder zumindest konstant halten und zudem den interessierenden Teil des Systemverhaltens messen und aufzeichnen. Der Gesamtaufwand zur Durchführung einer Versuchsreihe besteht also im technischen Umfeld neben der reinen Versuchsdurchführung auch in der Konzeption, Konstruktion und Fertigung des für sie benötigten Prüfstandes. Diese Zusatzaufwände müssen ebenfalls möglichst gering gehalten werden.

Neben der Wahl effizienter Versuchspläne stellt sich damit auch die Frage nach dem Automatisierungsgrad des zur Modellbildung eingesetzten Prüfstands. Im Allgemeinen bedeutet ein höherer Automatisierungsgrad eine Erhöhung des Aufwands zur Erstellung des Prüfstandes aber eine Senkung des Aufwands zur Durchführung der Versuchsreihen. Beide Aufwände müssen stets gegeneinander abgewogen werden. Um hierbei einen guten Kompromiss finden zu können, muss die Entwicklung des Prüfstandes parallel und in enger Abstimmung zur Entwicklung des Versuchsplanes für die durchzuführenden Versuchsreihen stattfinden. Es ist zudem zu prüfen, ob Prüfstände, die Teil des Validierungssystems des entwickelten Produkts sind, auch für Zwecke der Modellbildung herangezogen werden können. Bei den Abwägungen zum Automatisierungsgrad sollte berücksichtigt werden, dass empiri-

² Siehe Abschnitt 2.5.5

sche Modelle ausschließlich das untersuchte Objekt abbilden. Für Nachfolgenerationen des modellierten Objekts, bei denen die Gestalt oder das Prinzip geändert wurde³, müssen die Versuchsreihen also meist wiederholt werden.

In vielen Fällen ist das Ergebnis dieser Abwägungen, dass Prüfstände entwickelt werden, die einige Faktoren automatisiert einstellen können, während zur Änderung anderer Faktoren ein manueller Eingriff erforderlich ist. Automatisiert einstellbare Faktoren können zu jeder Tageszeit verändert werden, während manuell einzustellende Faktoren üblicherweise nur an Werktagen und dann auch nur an etwa 8 Stunden pro Tag verändert werden können. Manuelle Eingriffe haben also in Versuchsreihen oft Verzögerungen von vielen Stunden bis hin zu mehreren Tagen zur Folge. Zudem binden manuelle Eingriffe auch menschliche Arbeitskraft. Aus diesem Grund ist der durchschnittliche Aufwand zur manuellen Änderung von Faktoreinstellungen fast immer höher, als der zur automatisierten Einstellung. In beiden Fällen lassen sich die Faktoren zudem oft nach dem zu ihrer Einstellung benötigten Zeit- und Ressourcenaufwand unterteilen. So kann im Praxisbeispiel der Kugelhähne die automatisierte Änderung eines Drucks üblicherweise um mehrere Größenordnungen schneller ablaufen, als die einer Temperatur. Automatisiert durchführbare Temperaturänderungen verursachen dagegen deutlich weniger Aufwand, als der manuelle Wechsel eines Prüflings. Dieser ist wiederum deutlich weniger aufwendig, als der ebenfalls manuelle Wechsel des Prüfmediums, bei dem der Prüfstand aufwendig gereinigt werden muss. Eine Unterteilung des Einstellungsaufwands der Faktoren in nur zwei Stufen ist deshalb bei der Untersuchung technischer Systeme nicht immer ausreichend. In diesen Fällen werden Split-Plot-Versuchsreihen mit mehr als zwei Randomisierungsebenen benötigt, um eine wirklich effiziente Versuchsdurchführung zu ermöglichen.

Neben des Aufwands zur Erstellung eines Prüfstands und zur Versuchsdurchführung sollte allerdings auch der Aufwand zur Planung und Auswertung der Versuchsreihe nicht unterschätzt werden. Diese beiden Schritte werden durch die komplexere Versuchsstruktur von Split-Plot-Versuchsreihen und die daraus resultierende Fehlerstruktur in den Versuchsergebnissen aufwendiger und schwieriger als bei vollständig randomisierten Versuchsplänen. Dies wird vor allem dann problematisch, wenn weder entsprechend geschultes Personal, noch geeignete Software zur Planung und Auswertung von Split-Plot-Versuchsreihen zur Verfügung stehen. Um die hierdurch entstehenden Zusatzaufwände niedrig zu halten, und dennoch valide Ergebnisse erhalten zu können, sind einfach anzuwendende Verfahren zu bevorzugen.

³ Vgl. Abschnitt 2.1.1

Wie bereits in Kapitel 7 erläutert, ist es zudem in jedem Fall wichtig, dass die eingesetzten Versuchspläne eine Bestimmung von quadratischen Effekten und Wechselwirkungen zweiten Grades erlauben. Weiterhin ist es vorteilhaft, wenn alle Faktoren in mehr als drei Stufen untersucht werden, um Lack-of-fit Tests für solche Modelle zu ermöglichen.

Die speziellen Anforderungen für den Einsatz von Split-Plot-Versuchsreihen für die Modellbildung in der Produktentwicklung lassen sich also wie folgt zusammenfassen:

1. In einigen Fällen ist eine Unterteilung des Versuchsplans in mehr als zwei Ebenen nötig.
2. Um Einstiegshürden niedrig zu halten, sind einfach anzuwendende Verfahren zu bevorzugen, die keine spezielle Software erfordern.
3. Für alle Faktoren sollten quadratische Effekte und Wechselwirkungen zweiten Grades bestimmt werden können.

8.2 Lösungsansatz

Wie bereits in der Einleitung des Kapitels erläutert, liegt der Schlüssel zu einer effizienten Strukturierung von Versuchsreihen im industriellen Umfeld in Split-Plot-Versuchsreihen. Prinzipiell können alle in Abschnitt 2.5.5 vorgestellten Split-Plot-Versuchsplantypen auch für die Modellbildung in der Produktentwicklung eingesetzt werden. Durch die Forderung eines Versuchplantyps, der quadratische Effekte und Wechselwirkungen zweiten Grades abbilden kann, wird diese Auswahl bereits sehr stark eingeschränkt. Übrig bleiben dann Equivalent-Estimation-Versuchspläne⁴ und D-optimale Split-Plot-Versuchspläne⁵. Auf den Gebieten beider Versuchsplantypen gab es in den letzten Jahren nennenswerte Weiterentwicklungen in einer Reihe von Veröffentlichungen. Nachfolgend werden Einsatzempfehlungen für die darin beschriebenen Methoden gegeben.

8.3 Anwendung von Split-Plot-Versuchsreihen in der Produktentwicklung

Die Auswahl des einzusetzenden Versuchplantyps sowie die genaue Vorgehensweise zur Erstellung und Auswertung der Split-Plot-Versuchsreihe hängen von der

⁴ Siehe Abschnitt 2.5.5.2

⁵ Siehe Abschnitt 2.5.5.3

Anzahl der benötigten Randomisierungsebenen ab. In Abschnitt 8.3.1 werden Empfehlungen für den Einsatz von Split-Plot-Versuchsreihen gegeben, bei denen eine Unterteilung des Einstellungsaufwands der Faktoren in bis zu drei Gruppen ausreichend ist. Für alle anderen Fälle finden sich entsprechende Empfehlungen in Abschnitt 8.3.2.

8.3.1 Empfehlungen für Split-Plot-Versuchsreihen mit bis zu drei Randomisierungsebenen

D-optimalen Split-Plot-Versuchspläne basieren auf den in Abschnitt 2.5.1.8 vorgestellten Versuchsplänen. Die dort erwähnten Algorithmen können so angepasst werden, dass sie Split-Plot-Versuchspläne generieren. Ein Algorithmus dem eine vorab erstellte Kandidatenliste vorgegeben werden muss und der anschließend die effizientesten Versuchspunkte aus dieser Liste auswählt, wird von Goos und Vandebroek (2001) vorgestellt. Einen Algorithmus, der eine solche Liste nicht benötigt, zeigen Jones und Goos (2007). Beide Algorithmen benötigen eine Vorabschätzung des Verhältnisses $d = \sigma_{\delta}^2 / \sigma^2$ der Varianzkomponenten⁶ und generieren Versuchspläne mit zwei Randomisierungsebenen. Einen Algorithmus zur Erstellung von Versuchsplänen mit drei Randomisierungsebenen beschreiben Jones und Goos (2009). Im dreistufigen Fall gibt es neben der Varianz σ^2 der Einzelversuche für beide höheren Randomisierungsebenen separate Varianzen $\sigma_{\delta_1}^2$ und $\sigma_{\delta_2}^2$. Die Verhältnisse beider Varianzen zur Varianz der Einzelversuche (also $d_1 = \sigma_{\delta_1}^2 / \sigma^2$ und $d_2 = \sigma_{\delta_2}^2 / \sigma^2$) werden vom Algorithmus als Eingangsgrößen benötigt.

Ein großer Vorteil von D-optimalen Split-Plot-Versuchsplänen ist, dass sie eine sehr flexible Anordnung der Versuchspunkte im Faktorraum ermöglichen. Dies ist sehr hilfreich, wenn kompliziert begrenzte Faktorräume vorliegen. Auftreten kann dies beispielsweise, wenn manche Komponenten des Prüfstands oder des Prüflings eine Druckfestigkeit aufweisen, die eine Funktion der Temperatur ist. Durch eine solche oder ähnliche Einschränkung verliert der zu untersuchende Faktorraum seine (Hyper-)Quaderform. Bei starken Abweichungen von dieser Form können Versuchspläne mit vorgegebener Geometrie, wie beispielsweise faktorielle, zentral zusammengesetzte oder Box-Behnken-Versuchspläne, den Faktorraum nur noch unzureichend ausfüllen und somit nicht umfassend untersuchen. Optimale Versuchspläne sind hingegen ausreichend flexibel, um auch solche Faktorräume umfassend und dennoch effizient untersuchen zu können.

Equivalent-Estimation-Versuchspläne sind speziell strukturierte zentral zusammengesetzte Versuchspläne. Zu ihrer Erstellung wird ein zentral zusammengesetzter

⁶ Vgl. Gleichung 2.56

Versuchsplan zunächst so in Whole-Plots aufgeteilt, dass innerhalb der Whole-Plots alle aufwendig einzustellenden Faktoren konstante Einstellungen besitzen. Anschließend werden nach einem festen Schema Wiederholungen von Einzelversuchen und teilweise auch kompletten Whole-Plots dem Versuchsplan hinzugefügt. Abschließend werden alle Einzelversuche innerhalb jedes Whole-Plots und danach alle Whole-Plots untereinander randomisiert. Einen sehr gut verständlichen Einstieg in die Erstellung und Auswertung von Equivalent-Estimation-Versuchsplänen liefern Kowalski et al. (2007). Die genauen Verfahrensweisen zu ihrer Erstellung beschreiben Vining und Kowalski (2008), Vining et al. (2005) und Parker et al. (2007a, 2007b). Die Verfahrensweise von Vining und Kowalski (2008) erzeugt zwar leicht umfangreichere Versuchspläne als die anderen Verfahrensweisen. Dafür liefert sie aber modellunabhängige Schätzwerte für beide Varianzkomponenten σ^2 und σ_δ^2 , die die Grundlage für die statistische Bewertung des Modells sind.⁷ Die anderen Konstruktionsmethoden erzeugen hingegen Versuchspläne, bei denen zur Bestimmung der Varianzkomponenten weiterhin REML oder ähnliche Verfahren benötigt werden. Bei Versuchsplänen, die nach der Verfahrensweise von Vining und Kowalski erzeugt worden sind, lassen sich die Varianzkomponenten gemäß den Gleichungen 2.54 und 2.55 ermitteln. Anschließend kann die Kovarianzmatrix der Regressionskoeffizienten mit Hilfe der Gleichungen 2.32, 2.41, 2.50 und 2.51 geschätzt werden. Da in diesen Gleichungen keine Modellannahmen über feste Effekte enthalten sind, werden diese Schätzwerte und die mit ihrer Hilfe abgeleiteten statistischen Bewertungskennzahlen⁸ weder durch Lack-of-fit, noch durch Over-fit beeinträchtigt (vgl. Bochterle et al., 2015, S. 2). Eine Methode zur Erstellung von Versuchsplänen mit drei separaten Randomisierungsebenen wird von Yuan (2011) vorgestellt.

Bereits in Abschnitt 7.3.3 wurde festgestellt, dass für den Einsatz in der Produktentwicklung zentral zusammengesetzte Versuchspläne den optimalen Versuchsplänen im vollständig randomisierten Fall meist vorzuziehen sind. Dies liegt vor allem an ihrer einfacheren Erstellung, aber auch an der Möglichkeit, die bei der Regression ermittelten Effekte anhand von Vergleichen von unverarbeiteten Messdaten nachzuvollziehen. Bei Split-Plot-Versuchsreihen kommt hinzu, dass zur Erstellung von D-optimalen Versuchsplänen eine Vorabschätzung des Verhältnisses der Varianzkomponenten benötigt wird. Um hier eine gute Abschätzung treffen zu können, wird viel Erfahrung benötigt. Zur Erstellung von Equivalent-Estimation-Versuchsplänen ist keine derartige Abschätzung notwendig. Einen weiteren Vorteil besitzen die auf zentral zusammen gesetzten Versuchsplänen basierenden Equivalent-Estimation-

⁷ Vgl. S. 70 f.

⁸ Siehe Abschnitt 2.5.2.2

Versuchspläne bei ihrer Auswertung: Durch ihre Eigenschaft der OLS-GLS Äquivalenz⁹ lassen sich diese Versuchspläne durch OLS¹⁰ auswerten. Dies ist vor allem dann wichtig, wenn die verwendete Auswertungssoftware weder GLS¹¹ noch REML¹² unterstützt. Da inzwischen auch im Ingenieursumfeld gängige Softwarepakete, wie beispielsweise MathWorks MATLAB, diese Auswertungsverfahren unterstützen, verliert dieser Umstand für den Einsatz in der Produktentwicklung allerdings mehr und mehr an Bedeutung. In Summe ist dennoch die Erstellung und Auswertung von D-optimalen Split-Plot-Versuchsplänen aufwendiger und erfordert mehr Vorkenntnisse als die von Equivalent-Estimation-Versuchsplänen.

Damit lassen sich die folgenden Empfehlungen für die Auswahl von Split-Plot-Versuchsplänen treffen: In den meisten Fällen sind Equivalent-Estimation-Versuchspläne für die Modellbildung in der Produktentwicklung den D-optimalen Split-Plot-Versuchsplänen vorzuziehen. Dabei spielt es keine Rolle, ob der Versuchsplan in zwei oder drei Randomisierungsebenen unterteilt werden soll. Lediglich für kompliziert begrenzte Faktorräume muss manchmal von Equivalent-Estimation-Versuchsplänen auf D-optimale Split-Plot-Versuchspläne ausgewichen werden. Bei der Konstruktion von Equivalent-Estimation-Versuchsplänen mit zwei Randomisierungsebenen ist der Einsatz der Methode von Vining und Kowalski (2008) empfehlenswert.

8.3.2 Empfehlungen für Split-Plot-Versuchsreihen mit mehr als drei Randomisierungsebenen

Für mehr als drei Randomisierungsebenen bestehen aktuell weder Verfahren zur Konstruktion von Equivalent-Estimation-Versuchsplänen, noch zur Konstruktion von D-optimalen Versuchsplänen. Falls ein solcher Versuchsplan notwendig wird, muss er also vom Anwender selbst erstellt werden. Um eine effiziente Anordnung der Versuchspunkte im Faktorraum zu gewährleisten, sollte als Grundlage ein Versuchsplan aus den Abschnitten 2.5.1.6 bis 2.5.1.8 dienen. Empfehlenswert ist auch hier der Einsatz eines zentral zusammengesetzten Versuchsplans.

Der gewählte Versuchsplan wird zunächst so in Whole-Plots unterteilt, dass innerhalb jedes Whole-Plots die Faktoren mit dem größten Änderungsaufwand konstant sind. Jeder dieser Whole-Plots kann anschließend weiter in Split-Plots aufgeteilt werden, sodass in diesen die Einstellungen der etwas weniger aufwendig einzustel-

⁹ Siehe Abschnitt 2.5.5.2

¹⁰ Siehe Abschnitt 2.5.2.1

¹¹ Siehe Abschnitt 2.5.3.1

¹² Siehe Abschnitt 2.5.3.2

lenden Faktoren konstant sind. Nach diesem Prinzip kann weiter vorgegangen werden, bis der Versuchsplan in eine nahezu beliebig viele Randomisierungsebenen aufgeteilt ist. Zu berücksichtigen ist jedoch, dass jede weitere Randomisierungsebene Freiheitsgrade bei der Versuchsauswertung kostet.¹³ Die Versuchsreihe sollte also auch nicht feiner strukturiert werden, als notwendig. Die Reihenfolge der Versuchsdurchführung ergibt sich anschließend durch folgende Vorgehensweise: Zunächst werden alle Whole-Plots in eine zufällige Reihenfolge gebracht. Anschließend wird die Reihenfolge der Split-Plots innerhalb jedes Whole-Plots randomisiert. Nach diesem Muster wird weiter verfahren, bis auch die Einzelversuche innerhalb der niedrigsten Split-Split-...-Plot Ebene randomisiert sind. So erhält man einen Versuchsplan, dessen Reihenfolge so randomisiert ist, dass möglichst wenig Einstellungsaufwand für schwierig zu ändernde Faktoren erzeugt wird.

Die Struktur einer solchen Versuchsreihe wirkt sich auf Fehlerstruktur der Versuchsergebnisse aus, weshalb von einer heteroskedastischen Fehlerstruktur der Versuchsergebnisse ausgegangen werden muss.¹⁴ Der Einsatz eines klassischen linearen Modells und damit OLS zur Auswertung einer solchen Versuchsreihe scheidet deshalb aus. Am besten lässt sich die heteroskedastische Fehlerstruktur in Form von zufälligen Effekten eines gemischten linearen Modells abbilden. Zur Auswertung einer solchen Versuchsreihe wird deshalb Software benötigt, die gemischte lineare Modelle unterstützt.¹⁵

Der Anwender muss einer solchen Software üblicherweise den Ergebnisvektor \mathbf{y} des gemischten linearen Modells gemäß Gleichung 2.42, sowie dessen Einstellungsmatrizen \mathbf{X} und \mathbf{Z} der festen und zufälligen Faktoren zur Verfügung stellen. Zudem muss er Annahmen über den Aufbau der Kovarianzmatrizen \mathbf{G} und \mathbf{R} treffen.¹⁶ Hierzu wird nachfolgend eine kurze Hilfestellung gegeben.

Zunächst wird nachfolgend der einfache Fall einer Split-Plot-Versuchsreihe mit lediglich zwei Randomisierungsebenen betrachtet. Die $i = 1 \dots m$ Whole-Plots einer

¹³ Vgl. Abschnitt 2.5.4.1

¹⁴ Siehe S. 69 f. für Details

¹⁵ Im Ingenieursumfeld kann dies beispielsweise MathWorks MATLAB sein, in dessen „Statistics and Machine Learning Toolbox“ geeignete Auswertungsverfahren implementiert sind. Der Befehl zur Erstellung gemischter linearer Modelle aus Daten in Matrixform lautet hier `fitlmematrix`.

¹⁶ Siehe Gleichungen 2.45 und 2.46

solchen Versuchsreihe werden im gemischten linearen Modell als Cluster abgebildet. Wenn die Modellgleichung analog zu Gleichung 2.44¹⁷ in Einzelmatrizen für jeden Whole-Plot (bzw. jedes Cluster) aufgeteilt wird, dann lassen sich diese Einzelmatrizen gemäß Gleichung 2.43¹⁷ formulieren.

Bei Split-Plot-Versuchsplänen wird im Allgemeinen allerdings angenommen, dass die Effekte aller Faktoren über die gesamte Versuchsreihe hinweg konstant sind und zwischen den Whole-Plots lediglich Unterschiede im Mittelwert bestehen. Diese Unterschiede werden in Gleichung 2.49 durch den Whole-Plot Fehlerterm δ beschrieben, der innerhalb jedes Whole-Plots konstant bleibt, aber zwischen verschiedenen Whole-Plots variiert. Um dies in der Formulierung als gemischtes lineares Modell abzubilden, bleibt in \mathbf{Z}_i nur die erste Spalte (bestehend aus Einsen) erhalten und alle anderen Spalten fallen weg. Analog dazu reduziert sich \mathbf{b}_i zu seinem ersten Element $b_{i,0}$. \mathbf{Z}_i und \mathbf{b}_i sind für Split-Plot-Versuchspläne also durch Gleichung 8.1 gegeben. Die Anzahl der Zeilen in \mathbf{Z}_i entspricht dabei der Anzahl der Versuche im i -ten Whole-Plot.

$$\mathbf{Z}_i = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \mathbf{b}_i = b_{i,0} \quad 8.1$$

Aufbauend hierauf kann auch die Abbildung von mehr als zwei Randomisierungsebenen erläutert werden. In einem solchen Fall ändert sich lediglich die Form von Gleichung 2.42. Die Definitionen der in ihr enthaltenen Komponenten bleiben unverändert gültig. Für einen Versuchsplan mit p Randomisierungsebenen ergibt sich ein gemischtes lineares Modell gemäß Gleichung 8.2.

¹⁷ In MATLAB muss \mathbf{Z} abweichend von Gleichung 2.44 folgendermaßen formuliert werden: $\mathbf{Z} = (\mathbf{Z}_1 \ \mathbf{Z}_2 \ \dots \ 0)^T$. Die daraus resultierende Struktur von \mathbf{Z} erlaubt keine Rückschlüsse auf die Unterteilung der Versuchsreihe in Whole-Plots. Diese Information wird in MATLAB stattdessen über eine Matrix \mathbf{G} definiert. Diese darf nicht mit der Kovarianzmatrix \mathbf{G} der zufälligen Effekte verwechselt werden. Der Aufbau dieser Kovarianzmatrix kann in MATLAB über das Funktionsargument „CovariancePattern“ definiert werden. Für Details siehe Mathworks (2019).

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sum_{j=1}^{p-1} \mathbf{Z}_j \mathbf{b}_j + \boldsymbol{\varepsilon} \quad 8.2$$

Darin beschreibt $\mathbf{Z}_j \mathbf{b}_j$ die zufälligen Effekte für die j -te Randomisierungsebene. Der Fehlerterm jeder Randomisierungsebene wird also in separaten Matrizen abgebildet, die anschließend aufsummiert werden. In den Matrizen \mathbf{Z}_j und \mathbf{b}_j wird die Unterteilung der j -ten Randomisierungsebene analog zur Definition von \mathbf{Z} und \mathbf{b} in Gleichung 2.44 abgebildet.

Unabhängig von der Anzahl der Randomisierungsebenen sind die zu treffenden Annahmen über die Kovarianzmatrizen \mathbf{G} und \mathbf{R} der zufälligen Effekte und der Residuen: Für \mathbf{R} ist fast immer die Annahme $\mathbf{R} = \sigma^2 \mathbf{I}$ ausreichend.¹⁸ Und da für jede Randomisierungsebene nur ein zufälliger Faktor enthalten ist, besitzt auch die Kovarianzmatrix der zufälligen Effekte lediglich ein Element. Diese Matrix besitzt also keine Kovarianzstruktur, die näher definiert werden müsste.

8.4 Demonstration am Beispiel der Kugelhähne

Das nachfolgende Praxisbeispiel behandelt die Modellierung des Losbrechmoments von Kugelhähnen¹⁹. Es basiert auf der Arbeit von Bochterle et al. (2015), in der versucht wurde, die Anwendung von Split-Plot-Versuchsplänen und die damit möglichen Effizienzgewinne möglichst anschaulich zu demonstrieren. Im Gegensatz zu den vorhergehenden beiden Kapiteln, soll hier also lediglich das Losbrechmoment von Kugelhähnen und kein vollständiges Reibungsmodell ermittelt werden.

8.4.1 Auswahl und Aufbau des Versuchsplans

Die Faktoren, in deren Abhängigkeit das Losbrechmoment abgebildet werden soll, sind:

- Die Temperatur T
- Die Stillstandszeit t_s seit der letzten Bewegung des Kugelhahns
- Die Druckdifferenz Δp zwischen beiden Anschlüssen
- Die Stellgeschwindigkeit $\dot{\varphi}$

¹⁸ Beispielsweise MATLAB bietet auch keine Option diese Kovarianzen abweichend von dieser Annahme zu spezifizieren.

¹⁹ Vgl. Abschnitt 5.2

Alle diese Faktoren können auf dem zur Untersuchung eingesetzten Prüfstand automatisiert eingestellt werden. Die Einstellungsänderung der Stillstandszeit, der Druckdifferenz und der Stellgeschwindigkeit können dabei innerhalb weniger Sekunden erfolgen. Temperaturänderungen benötigen hingegen etwa eine Stunde bis der Prüfling ausreichend gleichmäßig durchwärmt ist. Die Temperatur ist in diesem Beispiel also der einzige, aufwendig einzustellende Faktor. Eine Unterscheidung der Faktoren nach ihrem Einstellungsaufwand in zwei Stufen ist deshalb ausreichend. Aus diesem Grund kann für dieses Beispiel das in Abschnitt 8.3.1 beschriebene Verfahren angewendet werden.

Wie dort empfohlen, wird als Versuchsplan ein Equivalent-Estimation-Versuchsplan eingesetzt, der nach der Methode von Vining und Kowalski (2008) aufgebaut wird. Der so erstellte Versuchsplan wird in Tabelle 8.1 gezeigt. Die Darstellungsreihenfolge in dieser Tabelle entspricht der realen Versuchsreihenfolge. Sie beinhaltet also bereits das Ergebnis der Randomisierung. Zudem sind auch die gemessenen Werte des Losbrechmoments $M_{fr,s}$ für alle Versuche enthalten.

8.4.2 Aufbau des gemischten linearen Modells und Regression

Als Modellgleichung wird ein gemischtes lineares Modell gemäß Gleichung 2.42 eingesetzt. Der Aufbau der darin enthaltenen Matrizen für den Versuchsplan aus Tabelle 8.1 wird nachfolgend erläutert. \mathbf{y} ist der Ergebnisvektor, der gemäß Gleichung 8.3 alle Versuchsergebnisse enthält.

$$\mathbf{y} = (M_{fr,s,1} \quad M_{fr,s,2} \quad \dots \quad M_{fr,s,80})' \quad 8.3$$

Der Koeffizientenvektor $\boldsymbol{\beta}$ der festen Effekte ist ein Spaltenvektor, der alle Faktoren enthält. Passend zu einem zentral zusammengesetzten Versuchsplan sind dies alle Terme eines mehrdimensionalen Polynoms bis zum zweiten Grad. Diese sind zunächst noch unbekannt. Ihre Bestimmung ist das primäre Ziel der Modellbildung. Der Aufbau des Koeffizientenvektors ist in Gleichung 8.4 dargestellt.

$$\boldsymbol{\beta} = (\beta_0 \quad \beta_T \quad \beta_{T^2} \quad \beta_{t_s} \quad \beta_{t_s^2} \quad \beta_{\Delta p} \quad \beta_{\Delta p^2} \quad \beta_{\varphi} \quad \beta_{\varphi^2} \quad \beta_{T t_s} \quad \beta_{T \Delta p} \quad \beta_{T \varphi} \quad \beta_{t_s \Delta p} \quad \beta_{t_s \varphi} \quad \beta_{\Delta p \varphi})' \quad 8.4$$

Tabelle 8.1: Equivalent-Estimation-Versuchsplan inklusive der Versuchsergebnisse (Bochterle et al., 2015, S. 5)

	T	t_s	Δp	$\dot{\varphi}$	M_{Frs}		T	t_s	Δp	$\dot{\varphi}$	M_{Frs}
WP1 (centre)	0	0	0	0	16.04	WP6 (axial)	1.19	0	0	0	21.46
	0	0	0	0	16.55		1.19	0	0	0	22.97
	0	0	0	0	16.43		1.19	0	0	0	23.22
	0	0	0	0	16.36		1.19	0	0	0	23.24
	0	0	0	0	16.28		1.19	0	0	0	23.32
	0	0	0	0	16.28		1.19	0	0	0	23.32
	0	0	0	0	16.31		1.19	0	0	0	23.32
WP2 (axial)	0	0	2	0	18.16	WP7 (axial)	0	0	0	-2	9.81
	0	0	2	0	18.29		0	0	0	2	18.53
	0	0	-2	0	16.02		0	0	0	-2	9.72
	0	0	2	0	18.04		0	0	0	2	18.70
	0	0	2	0	18.31		0	0	0	-2	9.64
	0	0	-2	0	16.02		0	0	0	2	18.70
	0	0	-2	0	14.11		0	0	0	2	18.38
WP3 (factorial)	0	0	-2	0	14.14	0	0	0	-2	9.64	
	-1	1	1	1	14.18	WP8 (factorial)	1	1	-1	1	22.12
	-1	-1	1	-1	12.26		1	-1	-1	-1	18.77
	-1	1	-1	1	11.04		1	1	-1	-1	18.80
	-1	1	1	-1	12.13		1	-1	1	-1	19.51
	-1	1	-1	-1	9.30		1	1	1	-1	20.34
	-1	-1	-1	-1	9.25		1	-1	-1	1	24.39
-1	-1	-1	1	10.99	1		1	1	1	24.00	
WP4 (axial)	-1	-1	1	1	14.18	1	-1	1	1	25.02	
	-1.19	0	0	0	11.35	WP9 (axial)	0	2	0	0	15.92
	-1.19	0	0	0	11.62		0	-2	0	0	16.21
	-1.19	0	0	0	11.60		0	-2	0	0	16.19
	-1.19	0	0	0	11.62		0	2	0	0	16.09
	-1.19	0	0	0	11.65		0	2	0	0	16.26
	-1.19	0	0	0	11.60		0	2	0	0	16.33
-1.19	0	0	0	11.60	0		-2	0	0	16.38	
WP5 (centre)	-1.19	0	0	0	11.57	0	-2	0	0	16.38	
	0.00	0	0	0	16.09	WP10 (centre)	0	0	0	0	16.24
	0.00	0	0	0	16.63		0	0	0	0	16.24
	0.00	0	0	0	16.53		0	0	0	0	16.19
	0.00	0	0	0	16.48		0	0	0	0	16.19
	0.00	0	0	0	16.46		0	0	0	0	16.21
	0.00	0	0	0	16.43		0	0	0	0	16.21
0.00	0	0	0	16.41	0		0	0	0	16.24	
0.00	0	0	0	16.38	0	0	0	0	16.24		

Δp : pressure difference; t_s : idle time; T : temperature; WP: whole plot; M_{Frs} : breakaway torque; $\dot{\varphi}$: actuating speed.

Die Designmatrix X der festen Effekte enthält alle Einstellungen der Faktoren des Koeffizientenvektors β . Da die vollständige Designmatrix 80 Zeilen besitzt, ist in Gleichung 8.5 nur ein Ausschnitt von ihr dargestellt. Dieser enthält, neben dem ersten und letzten Versuch der Versuchsreihe, alle Zeilen, die zum dritten Whole-Plot gehören. Zur besseren Verständlichkeit sind die Formelzeichen der in den Spalten enthaltenen Faktoreinstellungen über den Spalten angegeben.

$$\mathbf{X} = \begin{pmatrix}
 1 & T & T^2 & t_s & t_s^2 & \Delta p & \Delta p^2 & \dot{\varphi} & \dot{\varphi}^2 & T t_s & T \Delta p & T \dot{\varphi} & t_s \Delta p & t_s \dot{\varphi} & \Delta p \dot{\varphi} \\
 \vdots & \vdots \\
 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 \vdots & \vdots \\
 1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 \\
 1 & -1 & 1 & -1 & 1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & 1 & -1 \\
 1 & -1 & 1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 \\
 1 & -1 & 1 & 1 & 1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\
 1 & -1 & 1 & 1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 \\
 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
 1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 & 1 & 1 & 1 & -1 & 1 & -1 & -1 \\
 1 & -1 & 1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 \\
 \vdots & \vdots \\
 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
 \end{pmatrix} \quad 8.5$$

Der Koeffizientenvektor \mathbf{b} der zufälligen Effekte ist ein Spaltenvektor mit zehn Zeilen. Er enthält die vorab unbekanntene Werte des Whole-Plot-Fehlers δ für die zehn Whole-Plots. Für das hier vorgestellte Beispiel besitzt er die Form von Gleichung 8.6.

$$\mathbf{b} = (b_{1,0} \quad b_{2,0} \quad \dots \quad b_{10,0})' \quad 8.6$$

Die Designmatrix \mathbf{Z} der zufälligen Effekte besitzt die in Gleichung 8.7 angegebene Form.

$$\mathbf{Z} = \begin{pmatrix}
 \mathbf{Z}_1 & \mathbf{0} & \dots & \mathbf{0} \\
 \mathbf{0} & \mathbf{Z}_2 & \dots & \mathbf{0} \\
 \vdots & \vdots & \ddots & \vdots \\
 \mathbf{0} & \mathbf{0} & \dots & \mathbf{Z}_{10}
 \end{pmatrix} \text{ mit } \mathbf{Z}_1 = \mathbf{Z}_2 = \dots = \mathbf{Z}_{10} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}^{10 \times 1} \quad 8.7$$

$$\text{und } \mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}^{10 \times 1}$$

Der Residuenvektor $\boldsymbol{\varepsilon}$ enthält gemäß Gleichung 8.8 die Fehlerterme aller Einzelversuche des Versuchsplans.

$$\boldsymbol{\varepsilon} = (\varepsilon_1 \quad \varepsilon_2 \quad \dots \quad \varepsilon_{80})' \quad 8.8$$

Durch die OLS-GLS-Äquivalenz des eingesetzten Versuchsplans können die Koeffizienten der festen Effekte direkt durch Einsetzen von \mathbf{y} und \mathbf{X} in Gleichung 2.21 bestimmt werden. Falls zusätzlich auch die statistischen Bewertungsgrößen aus Ab-

schnitt 2.5.2.2 ermittelt werden sollen, ist dies durch den Einsatz der Konstruktionsmethode von Equivalent-Estimation-Versuchsplänen nach Vining und Kowalski ebenfalls einfach und ohne Einsatz von spezieller Software möglich: Schätzwerte für die Varianzkomponenten σ^2 und σ_δ^2 lassen sich gemäß den Gleichungen 2.54 und 2.55 ermitteln. Anschließend kann mit Hilfe der Gleichungen 2.32, 2.41, 2.50 und 2.51 die Kovarianzmatrix der Koeffizienten geschätzt werden. Auf deren Basis lassen sich anschließend, wie in Abschnitt 2.5.2.2 beschrieben, verschiedene Zuverlässigkeitskennzahlen für die geschätzten Werte der Koeffizienten β bestimmen.

8.4.3 Ergebnisse

Das Ziel von Split-Plot-Versuchsplänen ist die Steigerung der Effizienz der durchzuführenden Versuchsreihe. Es soll also das Verhältnis aus der Qualität der Versuchsergebnisse und des Versuchsaufwands verbessert werden. Um die Effizienz des oben gezeigten Equivalent-Estimation-Versuchsplans greifbar zu machen, werden nachfolgend sowohl der Versuchsaufwand, als auch die Genauigkeit des damit erstellten Modells, mit denen von zwei vollständig randomisierten Versuchsreihen verglichen.

Eine dieser Versuchsreihen ist eine dreistufige vollständig faktorielle Versuchsreihe, die in vollständig randomisierter Reihenfolge durchgeführt worden ist. Diese Versuchsreihe umfasst $3^4=81$ Versuche. Auf ihre Ergebnisse wurde ein mehrdimensionales Polynom angepasst, das alle Terme ersten und zweiten Grades enthält. Die zweite Versuchsreihe ist eine One-factor-at-a-time-Versuchsreihe, bei der jeder Faktor gesondert in 21 Stufen untersucht worden ist. Die Teilversuchsreihen für die Faktoren sind separat randomisiert durchgeführt worden. Der Versuchsumfang beträgt hier $4 \cdot 21=84$ Versuche. Die drei verglichenen Versuchsreihen besitzen also alle einen ähnlichen Versuchsumfang. Auf die Ergebnisse der One-factor-at-a-time-Versuchsreihe wurde ein mehrdimensionales Polynom fünften Grades angepasst. Aufgrund des Aufbaus der Versuchsreihe, kann dieses Polynom jedoch keinerlei Wechselwirkungen enthalten. In Abbildung 8.1 ist der Zeitaufwand dargestellt, der zur Durchführung der drei Versuchsreihen benötigt worden ist.

Die Equivalent-Estimation-Versuchsreihe besaß in diesem Beispiel mit Abstand den geringsten Zeitaufwand. Die faktorielle Versuchsreihe hat mit noch größerem Abstand den höchsten Zeitaufwand verursacht. Es ist erkennbar, dass der Zeitaufwand deutlich mit der Anzahl der im Versuchsplan enthaltenen Temperaturänderungen korreliert: In der faktoriellen Versuchsreihe musste die Temperatur 73-mal neu eingestellt werden, in der Equivalent-Estimation-Versuchsreihe 8-mal und in der One-factor-at-a-time-Versuchsreihe 22-mal. Ähnliche Verhältnisse zeigen sich auch im benötigten Zeitaufwand.

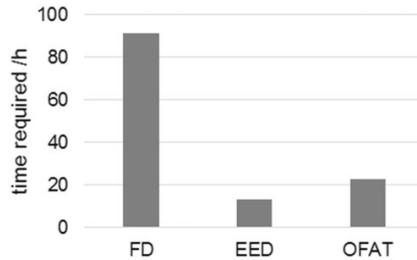


Abbildung 8.1: Zeitaufwand für die Durchführung der drei Versuchsreihen. FD steht für die faktorielle Versuchsreihe, EED für die Equivalent-Estimation-Versuchsreihe und OFAT für die One-factor-at-a-time-Versuchsreihe. (Bochterle et al., 2015, S. 6)

Die Genauigkeit der auf die Versuchsergebnisse angepassten linearen Modelle wurde mit Hilfe eines Testdatensatzes überprüft. Dieser besteht aus 15 Versuchen mit zufälligen Einstellungskombinationen der vier Faktoren (T , t_s , Δp und ϕ). Die Ergebnisse dieser Versuche werden in Abbildung 8.2 den Schätzwerten der drei Modelle gegenübergestellt.

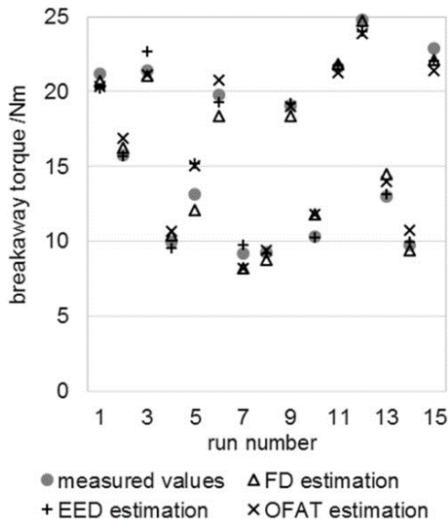


Abbildung 8.2: Vergleich der Schätzwerte mit den Messwerten des Testdatensatzes (Bochterle et al., 2015, S. 7)

Um die Vorhersagequalität der drei Modelle bezogen auf die Ergebnisse des Testdatensatzes besser beurteilen zu können, wurde das Bestimmtheitsmaß²⁰ für die Schätzwerte jedes Modells ermittelt. Die Bestimmtheitsmaße der drei Modelle werden in Abbildung 8.3 einander gegenübergestellt. Hier zeigt sich, dass die Equivalent-Estimation-Versuchsreihe, bzw. das auf sie angepasste Modell, die besten Ergebnisse liefert.

Bei der Interpretation dieser Ergebnisse für die Vorhersagequalität der drei Modelle muss allerdings berücksichtigt werden, dass diese im Wesentlichen nicht durch die Split-Plot-Vorgehensweise verursacht werden. Vielmehr sind hier der Versuchstyp, sowie das auf ihn angepasste Regressionsmodell entscheidend. Durch die Split-Plot-Vorgehensweise wird die Vorhersagegenauigkeit des erzielten Modells aber auch nicht wesentlich verschlechtert. Durch die mit ihr erzielbaren Zeiteinsparungen verbessert sich deshalb das Verhältnis aus Vorhersagequalität und Versuchsaufwand sehr deutlich.

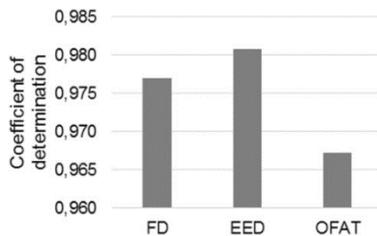


Abbildung 8.3: Bestimmtheitsmaße der drei Modelle für den Testdatensatz (Bochterle et al., 2015, S. 7)

8.5 Zwischenfazit

Wie bereits Daniel (Box 2005) erkannte²¹, sind im industriellen Umfeld fast ausschließlich Experimente anzutreffen, bei denen der Aufwand zur Einstellung der enthaltenen Faktoren stark variiert. In solchen Fällen kann der benötigte Versuchsaufwand durch die Split-Plot-Vorgehensweise auf einen Bruchteil reduziert werden. Dies konnte auch im Praxisbeispiel dieses Kapitels gezeigt werden. In solchen Fällen ist der Einsatz von Split-Plot-Versuchsplänen also essentiell, um die Versuchsreihe effizient gestalten zu können.

²⁰ Siehe Gleichung 2.23

²¹ Vgl. S. 91

Zudem konnte erklärt werden, welche Arten von Split-Plot-Versuchsplänen in welchen Anwendungsfällen im Bereich der Modellbildung für die Produktentwicklung am besten geeignet sind: Wenn eine Unterteilung des Einstellungsaufwands in bis zu drei Stufen ausreichend ist, dann ist in den meisten Fällen ein Equivalent-Estimation-Versuchsplan zu bevorzugen. Wenn Beschränkungen für durchführbare Faktorbereiche vorliegen, die zu kompliziert geformten Faktorräumen führen, dann sollte stattdessen der Einsatz von D-optimalen Split-Plot-Versuchsplänen erwogen werden. Wenn hingegen der Einstellungsaufwand der Faktoren in mehr als drei Stufen erfolgen muss, um eine effiziente Versuchsdurchführung zu ermöglichen, müssen vom Anwender Split-Plot-Versuchspläne frei erstellt werden. Die dabei entstehende Fehlerstruktur muss bei der Versuchsauswertung in einem gemischten linearen Modell abgebildet werden.

Für alle drei Fälle wurden die durchzuführenden Schritte beschrieben und, wo nötig, auf geeignete weiterführende Literatur verwiesen. Zusammen mit den in den Abschnitten 2.4 und 2.5 beschriebenen Grundlagen der statistischen Versuchsplanung für lineare Regressionsmodelle sollte dies interessierten Produktentwicklern und Versuchsingenieuren ermöglichen, anwendungsgerechte Split-Plot-Versuchsreihen für den Einsatz in der Modellbildung für die Produktentwicklung zu planen, durchzuführen und auszuwerten.

Die zu Beginn dieses Kapitels gestellte Forschungsfrage kann damit als beantwortet betrachtet werden.

9 Zusammenfassung und Ausblick

In dieser Arbeit wurden Methodenbausteine entwickelt, die den Einsatz linearer Regressionsmodelle als Simulationsmodelle in der Produktentwicklung ermöglichen. Die dabei erzielten Ergebnisse werden in Abschnitt 9.1 zusammengefasst. Abschließend wird in Abschnitt 9.2 ein Ausblick auf potentielle weitere Forschungsarbeiten gegeben, mit deren Hilfe die Einsatzmöglichkeiten und der Nutzen linearer Regressionsmodelle in der Produktentwicklung weiter gesteigert werden könnte.

9.1 Zusammenfassung

Lineare Regressionsmodelle besitzen viele positive Eigenschaften, die ihre Verwendung als Simulationsmodelle in der Produktentwicklung attraktiv machen. Hierzu gehört die Verfügbarkeit von effizienten Versuchsplänen, ein geringer Bedarf an Rechenleistung und Speicherplatz, sowie die Möglichkeit, aus ihren Parametern Einblicke in das abgebildete Systemverhalten zu erlangen.

Ihr größter Vorteil besteht darin, dass das von ihnen abgebildete Systemverhalten aus ihrer Modellgleichung und den darin enthaltenen Parametern von Entwicklungsingenieuren abgelesen und verstanden werden kann. Diese Eigenschaft ist essentiell, da laut Albers et al. (2019c) der Mensch im Zentrum der Produktentstehung steht, weshalb *„Prozesse und Methoden an seine Kreativität, Kompetenzen, Bedarfe und kognitiven Fähigkeiten angepasst“* werden müssen. Die gute Verständlichkeit hat damit zur Folge, dass lineare Regressionsmodelle nicht nur als reines Simulationsmodell im Rahmen einer Produktentwicklung dienen können, sondern auch das Systemverständnis steigern. Dies hat eine ganze Reihe von positiven Auswirkungen. Nicht zuletzt verbessert es die Möglichkeiten bestehende Modelle auf Nachfolgenerationen des aktuellen Produkts anpassen zu können. Ihr Nutzen als Referenzelement zukünftiger Produktgenerationen gemäß dem Modell der Produktgenerationsentwicklung nach Albers (2019) wird dadurch erheblich gesteigert.

Lineare Regressionsmodelle besitzen in ihrer Grundform allerdings auch Eigenschaften, die beim Einsatz als Simulationsmodell in der Produktentwicklung problematisch sind. Das Ziel dieser Arbeit war die Ermittlung von Methodenbausteinen, mit deren Hilfe der Einsatz linearer Regressionsmodelle für die Modellbildung in der Produktentwicklung ermöglicht wird. Hierbei sollten die oben erwähnten positiven Eigenschaften linearer Regressionsmodelle möglichst wenig beeinträchtigt werden.

Der größte Nachteil linearer Regressionsmodelle für den Einsatz in der Produktentwicklung besteht darin, dass sie weder dynamisches Verhalten, noch starke Nichtlinearitäten abbilden können. Oft können solche Verhaltensanteile allerdings auf Basis physikalischer Grundgleichungen in White-Box-Modellen abgebildet werden. Dies führt zur ersten Forschungsfrage, die davon handelt, wie solche White-Box-Modelle mit linearen (Black-Box-)Regressionsmodellen kombiniert werden können. Als Antwort hierauf wurde in Kapitel 6 der Einsatz hybrider Grey-Box-Modelle empfohlen. Es wurde dort beschrieben, wie solche Modelle aufgebaut sind, wie Versuchsreihen zu ihrer Ermittlung gestaltet werden sollten und wie die Modelle anhand der Versuchsergebnisse parametrisiert werden können. Diese Schritte wurden anschließend auch anhand eines Praxisbeispiels demonstriert. Anhand von dessen Ergebnissen konnte auch gezeigt werden, wie diese zu interpretieren sind und wie sie zu einem verbesserten Systemverständnis führen können.

Auch bei Nutzung von Grey-Box-Modellen kann es vorkommen, dass stark nichtlineare Verhaltensanteile verbleiben, die vom Black-Box-Modellanteil abgebildet werden müssen. Wie dies gelöst werden kann, ist Inhalt der zweiten Forschungsfrage. Diese wurde in Kapitel 7 beantwortet. Die hierfür gefundene Lösung beruht auf dem Einsatz von semiparametrischen Regressionsmodellen, bei denen lineare Regressionsmodelle mit nichtparametrischen Modellanteilen kombiniert werden. Besonders gut geeignet für den Einsatz in der Produktentwicklung sind hierbei Varying-Coefficient-Modelle, da sie sehr flexibel einsetzbar sind und trotzdem gut verständliche Parameter besitzen. Auch für die Erstellung dieser Modelle wurden alle Schritte von der Versuchsplanung bis zur Versuchsauswertung und Modellbildung zunächst theoretisch beschrieben und anschließend anhand eines Praxisbeispiels demonstriert.

Die dritte und letzte Forschungsfrage wurde daraus motiviert, dass in industriellen Versuchen sehr häufig einzelne Faktoren einen sehr großen Einstellungsaufwand besitzen. Dies tritt bei der empirischen Modellbildung für die Produktentwicklung beispielsweise auf, wenn die Versuchsreihe auf einem Prüfstand durchgeführt wird, bei dem nicht alle Faktoren automatisiert eingestellt werden können. In solchen Fällen ermöglicht die in der statistischen Versuchsplanung übliche vollständige Randomisierung keine effiziente Durchführung einer Versuchsreihe. In Kapitel 8 wurde beantwortet, wie dieses Problem mittels Split-Plot-Versuchsreihen gelöst werden kann. In Abhängigkeit der Randbedingungen der durchzuführenden Versuchsreihe ist in der Produktentwicklung der Einsatz von Equivalent-Estimation-Versuchsplänen, D-optimalen Split-Plot-Versuchsplänen oder vom Anwender frei erstellten Split-Plot-Versuchsplänen zu empfehlen. Deren Erstellung, Durchführung und Auswertung wurde in Kapitel 8 erläutert. Für Equivalent-Estimation-Versuchspläne wurden diese Schritte zudem anhand eines Praxisbeispiels demonstriert. Anhand dieses

Beispiels wurde zudem die mögliche Effizienzsteigerung im Vergleich zu vollständig randomisierten Versuchsreihen aufgezeigt.

Mit den oben beschriebenen Methodenbausteinen konnten lineare Regressionsmodelle im Rahmen der untersuchten Praxisbeispiele die Anforderungen der empirischen Modellbildung in der Produktentwicklung erfüllen. Sie scheinen damit für den Einsatz in der Produktentwicklung eine Alternative zu nichtparametrischen Regressionsmodellen, wie beispielsweise künstlichen neuronalen Netzen zu sein. Es konnte ebenfalls demonstriert werden, wie die genannten Vorteile zum Tragen kommen und beispielsweise zu einem gesteigerten Systemverständnis führen können.

9.2 Ausblick

Im Rahmen dieser Arbeit wurden Methodenbausteine entwickelt und in einem kontrollierten Umfeld anhand der beiden Praxisbeispiele aus Kapitel 5 untersucht. Damit bewegt sich diese Arbeit im Bereich entwicklungsmethodischer Forschung in den Kategorien zwei und drei nach Cantamessa (2003). Nachfolgende Untersuchungen können sich beispielsweise mit der Anwendung der Methodik im industriellen Umfeld befassen. Hierbei wäre eine Kosten-Nutzen-Analyse interessant, die alle Aufwände dem erzielbaren Nutzen gegenüberstellt. Hierbei sollten nach Möglichkeit sowohl Trainingsaufwände als auch konkrete wirtschaftliche Auswirkungen der beschriebenen positiven Eigenschaften linearer Regressionsmodelle berücksichtigt werden. Eine solche Arbeit ließe sich in Kategorie vier der entwicklungsmethodischen Forschung nach Cantamessa einordnen.

Neben solchen Anwendungsstudien wäre eine Untersuchung der Übertragbarkeit der Methode auf andere physikalische Zusammenhänge interessant, da sich beide Praxisbeispiele dieser Arbeit sich mit Reibungsvorgängen befassen. Alternative Anwendungsfälle könnten beispielsweise im Bereich der Kontinuumsmechanik oder der Verfahrenstechnik bestehen.

Eine weitere Untersuchung der Übertragbarkeit der Methode könnte sich auf die Anwendung der Methode in anderen Disziplinen als der Produktentwicklung fokussieren. Da die Methode in der Lage ist, verschiedene Faktoren in unterschiedlichen Detaillierungsgraden zu untersuchen und abzubilden, scheint sie für den Einsatz in Longitudinal-Studien geeignet zu sein. Die Möglichkeit zur Durchführung von Split-Plot-Versuchsreihen könnte beispielsweise in der Landwirtschaft interessant sein. Somit böten sich z.B. Versuchsreihen zur Untersuchung des Wachstumsverhaltens verschiedener Pflanzen in Abhängigkeit der Einflüsse mehrerer Faktoren an.

Eine logische Möglichkeit zur Weiterentwicklung der Methodik resultiert daraus, dass lineare Modelle in Form von gemischten linearen Modellen auch umfangreiche Aussagen über die Streuung des modellierten Systemverhaltens liefern können. Wenn dieses Potential durch geeignete Methodenbausteine nutzbar gemacht würde, wäre es möglich, damit anwendungsspezifische Streubreiten¹ für das prognostizierte Systemverhalten zukünftiger Produkte zu ermitteln. So könnte dann beispielsweise die Robustheit von Regelalgorithmen gegenüber realitätsnahen Schwankungen des Systemverhaltens getestet werden oder es könnte in Condition Monitoring Algorithmen geprüft werden, ob das beobachtete Systemverhalten innerhalb der normalen Toleranzen liegt. Wie der Einsatz der Modelle in solchen Anwendungen ablaufen kann und welche Möglichkeiten dadurch für die Validierung und andere Nutzungsmöglichkeiten im Rahmen einer Produktentwicklung ermöglicht werden, wäre ebenfalls eine interessante Fragestellung für nachfolgende Arbeiten.

¹ Konfidenz- und Prädiktionsintervalle, siehe Abschnitt 2.5.2.2

Literaturverzeichnis

- Achenbach, M. (2011). A piston seal as an important subsystem in the driving mechanism of an electro-pneumatic hammer. In *NAFEMS Seminar: „The Role of CAE in System Simulation“* (pp. 1-9), Wiesbaden
- Albers, A.; (2010a). Five hypotheses about engineering processes and their consequences. In *Proceedings of the Eighth International Symposium on Tools and Methods of Competitive Engineering, TMCE 2010* (pp. 343-356), Ancona, Italy
- Albers, A.; Behrendt, M.; Klingler, S.; Matros, K. (2016a). Verifikation und Validierung im Produktentstehungsprozess. In U. Lindemann (Hrsg.), *Handbuch Produktentwicklung* (S. 541–569), München: Hanser
- Albers, A.; Braun, A. (2011a). A generalised framework to compass and to support complex product engineering processes. *International Journal of Product Development*, 15 (1-3), 6–25
- Albers, A.; Braun, A. (2011b). Der Prozess der Produktentstehung. In F. Henning; E. Moeller (Hrsg.), *Handbuch Leichtbau* (S. 5–30), München: Hanser Verlag
- Albers, A.; Braun, A.; Muschik, S. (2010). Ein Beitrag zum Verständnis des Aktivitätsbegriffs im System der Produktentstehung. In *Tag des Systems Engineering* (S. 87-96), München
- Albers, A.; Burkhardt, N.; Meboldt, M.; Saak, M. (2005). SPALTEN problem solving methodology in the product development. In *Engineering Design and the Global Economy: 15th International Conference on Engineering Design – ICED* (pp. 553-554), Melbourne, Australia
- Albers, A.; Bursac, N.; Rapp, S. (2017b). PGE – Produktgenerationsentwicklung am Beispiel des Zweimassenschwungrads. *Forschung im Ingenieurwesen*, 81 (1), 13-31
- Albers, A.; Bursac, N.; Wintergerst, E. (2015). Produktgenerationsentwicklung - Bedeutung und Herausforderungen aus einer entwicklungsmethodischen Perspektive. In *Stuttgarter Symposium für Produktentwicklung (SSP)* (o. S.), Stuttgart
- Albers, A.; Deigendesch, T. (2010). Patterns in product development: Background, application and exemplification. In *Proceedings of the Eighth International Symposium on Tools and Methods of Competitive Engineering, TMCE 2010* (pp. 1299–1300), Ancona, Italy
- Albers, A.; Fischer, J.; Klingler, S.; Behrendt, M. (2014a). Durchgängige Validierung und Verifizierung am Beispiel der akustischen Eigenschaften eines Elektrofahrzeugs. In *Interdisziplinäre Entwicklung von Straßen- und Schienenfahrzeugen: Prozesse, Methoden, Tools, Best Practices; 7. Grazer Symposium Virtuelles Fahrzeug (GSVF)* (o. S.), Graz, Österreich

- Albers, A.; Heimicke, J.; Spadinger, M.; Reiß, N.; Breitschuh, J.; Richter, T.; Bursac, N.; Marthaler (2019c). *Eine Systematik zur situationsadäquaten Mechatroniksystementwicklung durch ASD - Agile Systems Design* (KIT Scientific Working Papers Nr. 113). Karlsruhe: Karlsruher Institut für Technologie (KIT), IPEK – Institut für Produktentwicklung
- Albers, A.; Gladysz, B.; Kniel, J.; Aschoff, M.; Meyer, A. (2016b). Integration von Versuchsergebnissen in C&C² - Modellen zur Wiederverwendung in der Produktgenerationsentwicklung am Beispiel eines trockenlaufenden Kupplungssystems. In *14. Gemeinsames Kolloquium Konstruktionstechnik: Traditio et Innovatio - Entwicklung und Konstruktion* (S. 10–20), Rostock
- Albers, A.; Meboldt, M. (2006). A new approach in product development based on systems engineering and systematic problem solving. In *Proceedings of AEDS Workshop. 15th Workshop on Applied Engineering Design Science* (pp. 5–12), Pilsen, Czech Republic
- Albers, A.; Meboldt, M. (2007). IPEMM -- Integrated product development process management model, based on systems engineering and systematic problem solving. In *16th International Conference on Engineering Design* (pp. 611-612), Paris
- Albers, A.; Rapp, S.; Spadinger, M.; Richter, T.; Birk; Marthaler, F.; Heimicke, J.; Kurtz, V.; Wessels, H. (2019a). The reference system in the model of PGE: Proposing a generalized description of reference products and their interrelations. In *Proceedings of the 22nd International Conference on Engineering Design (ICED19)* (pp. 1693–1702), Delft, The Netherlands
- Albers, A.; Rapp, S.; Birk; Bursac, N. (2017a). Die Frühe Phase der PGE – Produktgenerationsentwicklung. In *4. Stuttgarter Symposium für Produktentwicklung 2017 (SSP)* (o. S.), Stuttgart
- Albers, A.; Reiß, N.; Bursac, N.; Breitschuh, J. (2016c). 15 years of SPALTEN problem solving methodology in product development. In *12th Biennial Norddesign 2016 Conference „Highlighting the Nordic Approach“* (pp. 411–420), Trondheim, Norway
- Albers, A.; Reiß, N.; Bursac, N.; Richter, T. (2016d). IPeM-integrated product engineering model in context of product generation engineering. In *26th CIRP Design Conference* (pp. 100–105), Stockholm
- Albers, A.; Reiß, N.; Bursac, N.; Urbanec, J.; Lüdcke, R. (2014b). Situation-appropriate method selection in product development process - empirical study of method application. In *Proceedings of NordDesign 2014 Conference* (pp. 550–559), Espoo, Finland
- Albers, A.; Turki, T.; Lohmeyer, Q. (2012). Transfer of engineering experience by shared mental models. In *Design education for future wellbeing: Proceedings of the 14th International Conference on Engineering and Product Design Education* (pp. 77–82), Antwerp, Belgium

- Albers, A.; Wilmsen, M.; Dühr, K.; Heimicke, J. (2019b). The first steps towards innovation: A reference process model for developing product profiles. In *International Conference on Engineering Design (ICED19)* (pp. 1673–1682), Delft, The Netherlands
- Albers, M. (2010b). *Statistic Seminar, 6th talk, Prediction of New Observations*, Zürich, Eidgenössische Technische Hochschule (ETH) Zürich, Seminar for Statistics
- Arenz, A.; Potrykus, S. (2010). Modellbasierte Funktionsentwicklung. *ATZ-Automobil-technische Zeitschrift*, 112 (1), 28–32
- Aston, J.; Chiou, J.; Evans, J. (2010). Linguistic pitch analysis using functional principal component mixed effect models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59 (2), 297–317
- Baehr, H.; Kabelac, S. (2016). *Thermodynamik: Grundlagen und technische Anwendungen*, Berlin, Heidelberg: Springer
- Bartel, D. (2010). *Simulation von Tribosystemen*, Wiesbaden: Vieweg+Teubner
- Barton, R. (1994). Metamodeling: a state of the art review. In *Proceedings of the 26th conference on Winter simulation* (pp. 237–244), Orlando, FL
- Bingham, D.; Sitter, R. (1999). Minimum-aberration two-level fractional factorial split-plot designs. *Technometrics*, 41 (1), 62–70
- Bochterle, M.; Gauchel, W.; Lorentz, B. (2015). The use of equivalent estimation designs in tribological studies. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, 229 (8), 1022-1029
- Bohn, C.; Unbehauen, H. (2016). *Identifikation dynamischer Systeme: Methoden zur experimentellen Modellbildung aus Messdaten*, Wiesbaden: Springer Vieweg
- De Boor, C. (1978). *A practical guide to splines*, New York: Springer
- Box, G.; Hunter, J.; Hunter, W. (2005). *Statistics for experimenters: design, innovation, and discovery*, Hoboken, NJ: Wiley-Interscience
- Box, G.; Wilson, K. (1951). On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society. Series B*, 13 (1), 1–45
- Braun, A. (2013). Modellbasierte Unterstützung der Produktentwicklung - Potentiale der Modellierung von Produktentstehungsprozessen am Beispiel des integrierten Produktentstehungsmodells (iPeM). Dissertation. In A. Albers (Hrsg.), *Forschungsberichte des IPEK – Institut für Produktentwicklung. Systeme, Methoden, Prozesse* (Bd. 72). Karlsruhe: Karlsruher Institut für Technologie (KIT)
- Breitbach, T. (2015). *Effizienzsteigerung der Zellenmodellierung für Industrieroboter durch hybride Simulationsmodelle*. Aachen, Apprimus Verlag

- Bursac, N. (2016). Model Based Systems Engineering zur Unterstützung der Baukastenentwicklung im Kontext der Frühen Phase der Produktgenerationsentwicklung. Dissertation. In A. Albers (Hrsg.), *Forschungsberichte des IPEK – Institut für Produktentwicklung. Systeme, Methoden, Prozesse* (Bd. 93). Karlsruhe: Karlsruher Institut für Technologie (KIT)
- Cantamessa, M. (2003). An empirical perspective upon design research. *Journal of Engineering Design*, 14 (1), 1–15
- Carlson, J.; Batista, A. (1996). Constitutive relation for the friction between lubricated surfaces. *Physical Review E*, 53 (4), 4153–4165
- Chen, R.; Tsay, R. (1993). Functional-coefficient autoregressive models. *Journal of the American Statistical Association*, 88 (421), 298–308
- Cioppa, T.; Lucas, T. (2007). Efficient nearly orthogonal and space-filling Latin hypercubes. *Technometrics*, 49 (1), 45–55
- Cooper, R. (1990). Stage-gate systems: A new tool for managing new products. *Business Horizons*, 33 (3), 44–54
- Czado, C.; Schmidt, T. (2011). *Mathematische Statistik, Statistik und ihre Anwendungen*, Berlin, Heidelberg: Springer
- Daenzer, W.; Huber, F. (1997). *Systems engineering: Methodik und Praxis*, Zürich: Industrielle Organisation Zürich
- Deutsch, J.; Deutsch, C. (2012). Latin hypercube sampling with multidimensional uniformity. *Journal of Statistical Planning and Inference*, 142 (3), 763–772
- Dierckx, P. (1995). *Curve and surface fitting with splines*, Oxford: Oxford University Press
- Draper, N.; John, J. (1998). Theory & methods: response surface designs where levels of some factors are difficult to change. *Australian & New Zealand Journal of Statistics*, 40 (4), 487–495
- Draper, N.; Lin, D. (1990). Small response-surface designs. *Technometrics*, 32 (2), 187–194
- Dueck, G.; Scheuer, T. (1990). Threshold accepting: a general purpose optimization algorithm appearing superior to simulated annealing. *Journal of computational physics*, 90 (1), 161–175
- Eigner, M.; Koch, W.; Muggeo, C. (2017). *Modellbasierter Entwicklungsprozess cybertronischer Systeme*, Berlin, Heidelberg: Springer Vieweg
- English, T.; Simpson, J.; Landman, D.; Parker, P. (2012). An efficient split-plot approach for modeling nonlinear aerodynamic effects. *Quality Engineering*, 24 (4), 522–530

- Fahrmeir, L.; Kneib, T.; Lang, S. (2009). *Regression - Modelle, Methoden und Anwendungen*, Berlin, Heidelberg: Springer
- Fan, J.; Gijbels, I. (2018). *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, New York: Taylor & Francis
- Fang, K. (1980). The uniform design: application of number-theoretic methods in experimental design. *Acta Mathematicae Applicatae Sinica*, 3 (4), 363–372
- Fang, K.; Lin, D.; Winker, P.; Zhang, Y. (2000). Uniform design: Theory and application. *Technometrics*, 42 (3), 237–248
- Fisher, R. (1925). *Statistical methods for research workers*, Edinburgh: Oliver and Boyd
- Fisher, R. (1935). *The design of experiments*, Edinburgh: Oliver and Boyd
- Freudenmann, T. (2014). Ontologien zur Validierung von Produkten basierend auf dem Contact & Channel - Ansatz (C&C²-Ansatz). Dissertation. In A. Albers (Hrsg.), *Forschungsberichte des IPEK – Institut für Produktentwicklung. Systeme, Methoden, Prozesse* (Bd. 78). Karlsruhe: Karlsruher Institut für Technologie (KIT)
- Friedman, J. (1991). Multivariate adaptive regression splines. *Annals of Statistics*, 19 (1), 1–67
- Frochte, J. (2018). *Maschinelles Lernen: Grundlagen und Algorithmen in Python*, München: Hanser
- Galecki, A. (2013). *Linear mixed-effects models using R: A step-by-step approach*, New York: Springer
- Gausemeier, J.; Plass, C.; Wenzelmann, C. (2009). *Zukunftsorientierte Unternehmensgestaltung: Strategien, Geschäftsprozesse und IT-Systeme für die Produktion von morgen*, München: Hanser
- Goos, P.; Vandebroek, M. (2001). Optimal split-plot designs. *Journal of Quality Technology*, 33 (4), 436–450
- Greene, W. (2008). *Econometric analysis*, Upper Saddle River, NJ: Pearson Prentice Hall
- Grove, D.; Woods, D.; Lewis, S. (2004). Multifactor B-spline mixed models in designed experiments for the engine mapping problem. *Journal of Quality Technology*, 36 (4), 380–391
- Grundlach, C. (2004). *Entwicklung eines ganzheitlichen Vorgehensmodells zur problemorientierten Anwendung der statistischen Versuchsplanung*. Kassel: Kassel University Press
- Harville, D. (2008). Accounting for the estimation of variances and covariances in prediction under a general linear model: an overview. *Tatra Mountains Mathematical Publications*, 39 (1), 1–15

- Harville, D.; Jeske, D. (1992). Mean squared error of estimation or prediction under a general linear model. *Journal of the American Statistical Association*, 87 (419), 724–731
- Hasegawa, Y.; Ikeda, S.; Matsuura, S.; Suzuki, H. (2010). A study on methodology for total design management (the 4th report). A study on the response surface method for split-plot designs using the generalized least squares. In *Proceedings of The 92nd Japanese Society for Quality Control (JSQC) Technical Conference* (pp. 235-238), Tokio
- Hastie, T. (2017). Generalized additive models. In J. Chambers, T. Hastie (Eds.), *Statistical models in S* (pp. 249–307), Boca Raton, FL: Routledge
- Hastie, T.; Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 55 (4), 757–796
- Heipl, O.; Murrenhoff, H. (2012). *Schlussbericht zum Vorhaben 16229 N/1: Reibkraftmodellierung an Dichtkontakten in der Pneumatik*. Aachen: Rheinisch-Westfälische Technische Hochschule (RWTH) Aachen, Institut für fluidtechnische Antriebe und Systeme (IFAS)
- Huang, P.; Chen, D.; Voelkel, J. (1998). Minimum-aberration two-level split-plot designs. *Technometrics*, 40 (4), 314–326
- Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics*, 58 (1), 71–120
- Ikeda, S.; Matsuura, S.; Suzuki, H. (2014). Two-step residual-based estimation of error variances for generalized least squares in split-plot experiments. *Communications in Statistics - Simulation and Computation*, 43 (2), 342–358
- Isermann, R. (2014). *Engine modeling and control : modeling and electronic management of internal combustion engines*, Berlin, Heidelberg: Springer
- Jensen, W.; Kowalski, S. (2012). Response surfaces, blocking and split plots: An industrial experiment case study. *Quality Engineering*, 24 (4), 531–542
- Jones, B.; Goos, P. (2007). A candidate-set-free algorithm for generating D-optimal split-plot designs. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 56 (3), 347–364
- Jones, B.; Goos, P. (2009). D-optimal design of split-split-plot experiments. *Biometrika*, 96 (1), 67–82
- Jones, B.; Goos, P. (2012). An algorithm for finding D-efficient equivalent-estimation second-order split-plot designs. *Journal of Quality Technology*, 44 (4), 363-374
- Jones, B.; Nachtsheim, C. (2009). Split-plot designs : what , why , and how. *Journal of Quality Technology*, 41 (4), 340–361

- Jurecka, F. (2007). *Robust Design Optimization Based on Metamodeling Techniques* (Schriftenreihe des Lehrstuhls für Statik TU München. Band 8). München, Technische Universität München
- Klaer, R. (2003) (Hrsg.). *Praxis-Handbuch Industriearmaturen*, Essen: Vulkan-Verlag
- Kleppmann, W. (2011). *Taschenbuch Versuchsplanung: Produkte und Prozesse optimieren*, München: Hanser
- Kletschkowski, T.; Schomburg, U.; Bertram, A. (2001). Viskoplastische Materialmodellierung am Beispiel des Dichtungswerkstoffes Polytetrafluorethylen. *Technische Mechanik*, 21 (3), 227–241
- Klingler, S. (2017). Eine Methode zur effizienten und effektiven Unterstützung der kontinuierlichen Validierung im Kontext der PGE – Produktgenerationsentwicklung. Dissertation. In A. Albers (Hrsg.), *Forschungsberichte des IPEK – Institut für Produktentwicklung. Systeme, Methoden, Prozesse* (Bd. 101). Karlsruhe: Karlsruher Institut für Technologie (KIT)
- Kowalski, S.; Parker, P.; Vining, G. (2007). Tutorial: industrial split-plot experiments. *Quality Engineering*, 19 (1), 1–15
- Kroll, A. (2000). Grey-box models: Concepts and application. *New Frontiers in Computational Intelligence and its Applications*, 57 (1), 42–51
- Langeheinecke, K.; Kaufmann, A.; Thieleke, G. (2017). *Thermodynamik für Ingenieure: Ein Lehr- und Arbeitsbuch für das Studium*, Wiesbaden: Springer Vieweg
- Leary, S.; Bhaskar, A.; Keane, A. (2003). Optimal orthogonal-array-based latin hypercubes. *Journal of Applied Statistics*, 30 (5), 585–598
- Lemieux, C. (2010). *Monte carlo and quasi-monte carlo sampling*, New York: Springer
- Letsinger, J.; Myers, R.; Lentner, M. (1996). Response surface methods for bi-randomization structures. *Journal of quality technology*, 28 (4), 381–397
- Lindemann, U. (2009). *Methodische Entwicklung technischer Produkte: Methoden flexibel und situationsgerecht anwenden*, Berlin, Heidelberg: Springer
- Macharia, H.; Goos, P. (2010). D-optimal and D-efficient equivalent-estimation second-order split-plot designs. *Journal of Quality Technology*, 42 (4), 358–372
- Maier, M. (1998). Architecting principles for system of systems. *Systems Engineering*, 1 (4), 267-284
- Marsaglia, G.; Tsang, W. (2000). The ziggurat method for generating random variables. *Journal of Statistical Software*, 5 (8), 1–7

- Marxen, L. (2014). A Framework for Design Support Development based on the integrated Product Engineering Model iPeM. Dissertation. In A. Albers (Hrsg.), *Forschungsberichte des IPEK – Institut für Produktentwicklung. Systeme, Methoden, Prozesse* (Bd. 74). Karlsruhe: Karlsruher Institut für Technologie (KIT)
- Marxen, L.; Albers, A. (2012). Supporting validation in the development of design methods. In *Proceedings of DESIGN 2012 - the 12th International Design Conference* (pp. 1009–1018), Dubrovnik, Croatia
- Mathworks: *MATLAB Hilfe zum Befehl „fitlmematrix“*. Zugriff am 09.02.2022. Verfügbar unter <https://de.mathworks.com/help/stats/fitlmematrix.html>
- Matthiesen, S.; Grauberger, P.; Hölz, K.; Nelius, T.; Bremer, F.; Wettstein, A.; Gessinger, A.; Pflegler, B.; Nowoseltschenko, K.; Voß, K. (2018). *Modellbildung mit dem C&C²-Ansatz in der Gestaltung - Techniken zur Analyse und Synthese* (KIT Scientific Working Papers Nr. 58). Karlsruhe: Karlsruher Institut für Technologie (KIT), IPEK – Institut für Produktentwicklung
- Matthiesen, S.; Hölz, K.; Grauberger, P. (2017). Systemverständnis durch Analysemethoden. In *Proceedings of the 28th Symposium Design for X* (S. 263–274), Bamberg
- McCulloch, W.; Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of mathematical biophysics*, 5 (4), 115–133
- McKay, M.; Beckman, R.; Conover, W. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21 (2), 239–245
- Meboldt, M. (2008). Mentale und formale Modellbildung in der Produktentstehung-als Beitrag zum integrierten Produktentstehung-Modell (iPeM). Dissertation. In A. Albers (Hrsg.), *Forschungsberichte des IPEK – Institut für Produktentwicklung. Systeme, Methoden, Prozesse* (Bd. 29). Karlsruhe: Universität Karlsruhe (TH)
- Meissner, F.; Shirokinskiy, K.; Alexander, M. (2020). Computer on wheels / Disruption in automotive electronics and semiconductors. *Focus – Roland Berger*, 01.2020, 1-24
- de Mello, R. (2018). *Machine learning: a practical approach on the statistical learning theory*, Cham, Schweiz: Springer Nature
- Meywerk, M. (2007). *CAE-Methoden in der Fahrzeugtechnik*, Berlin, Heidelberg: Springer
- Montgomery, D. (2009). *Design and analysis of experiments*, Hoboken, NJ: Wiley
- Mooney, M. (1940). A theory of large elastic deformation. *Journal of Applied Physics*, 11 (9), 582-592

- Morris, M. (2012). Computer experiments. In K. Hinkelmann (Ed.), *Design and analysis of experiments: Volume 3; Special designs and applications* (pp. 379–411), Hoboken, NJ: Wiley
- MSC Software: *Nonlinear Finite Element Analysis of Elastomers*. Zugriff am 09.02.2022. Verfügbar unter http://www.axelproducts.com/downloads/WP_Nonlinear_FEA-Elastomers.pdf
- Mullins, L. (1969). Softening of rubber by deformation. *Rubber Chemistry and Technology*, 42 (1), 339–362
- Murakami, K.; Seborg, D (2000). Constrained parameter estimation with applications to blending operations. *Journal of Process Control*, 10 (2), 195–202
- Nesbitt, B. (2007). *Handbook of valves and actuators*, Oxford: Butterworth-Heinemann
- NIST/SEMATECH (2012). *Engineering Statistics Handbook*. Washington, DC: U.S. Department of Commerce
- Nollau, R. (2009). *Modellierung und Simulation technischer Systeme: Eine praxisnahe Einführung*, Berlin, Heidelberg: Springer
- Owen, A. (1992). Orthogonal arrays for computer experiments, integration and visualization. *Statistica Sinica*, 2 (1), 435–452
- Pahl, G.; Beitz, W.; Feldhusen, J.; Grote, K. (2007). *Konstruktionslehre: Grundlagen erfolgreicher Produktentwicklung Methoden und Anwendung*, Berlin, Heidelberg: Springer
- Panneton, F.; L'Ecuyer, P.; Matsumoto, M. (2006). Improved long-period generators based on linear recurrences modulo 2. *ACM Transactions on Mathematical Software*, 32 (1), 1–16
- Panreck, K. (2002). *Rechnergestützte Modellbildung physikalisch-technischer Systeme*, Düsseldorf: VDI-Verlag
- Parker, P.; Kowalski, S.; Vining, G. (2007a). Unbalanced and minimal point equivalent estimation second-order split-plot designs. *Journal of quality technology*, 39 (4), 376–388
- Parker, P.; Kowalski, S.; Vining, G. (2007b). Construction of balanced equivalent estimation second-order split-plot designs. *Technometrics*, 49 (1), 56–65
- Payne, A. (1962). The dynamic properties of carbon black-loaded natural rubber vulcanizates. Part I. *Journal of Applied Polymer Science*, 6 (19), 57–63
- Ponn, J.; Lindemann, U. (2011). *Konzeptentwicklung und Gestaltung technischer Produkte*, Heidelberg: Springer

- Potcner, K.; Kowalski, S. (2004). How to analyze a split-plot experiment. *Quality Progress*, 37 (1), 67–74
- Pronzato, L.; Müller, W. (2012). Design of computer experiments: Space filling and beyond. *Statistics and Computing*, 22 (3), 681–701
- Raković, S.; Levine, W. (2019). *Handbook of model predictive control*, Cham, Schweiz: Birkhäuser
- Raudenbush, S.; Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods*, New York: Sage
- Rhinehart, R. (2016). *Nonlinear regression modeling for engineering applications: modeling, model validation, and enabling design of experiments*, Hoboken, NJ: Wiley
- Rivlin, R. (1948). Large elastic deformations of isotropic materials. IV. Further developments of the general theory. *Philosophical Transactions of the Royal Society of London Series A*, 241 (835), 379–397
- Rogers, T.; Holmes, F.; Cross, E.; Worden, K. (2017). On a grey box modelling framework for nonlinear system identification. In *Proceedings of the 35th IMAC, Volume 6: Special Topics in Structural Dynamics* (pp. 167–178), Garden Grove, CA
- Ropohl, G. (2009). *Allgemeine Technologie: eine Systemtheorie der Technik*. Karlsruhe: Universitätsverlag Karlsruhe
- Ruckstuhl, A. (2008). *Einführung in die nichtlineare Regression*. Zürich: Zürcher Hochschule für Angewandte Wissenschaften (ZHAW), Institut für Datenanalyse und Prozessdesign (IDP)
- Ruppert, D.; Wand, M.; Carroll, R. (2003). *Semiparametric regression*, Cambridge: Cambridge University Press
- Saak, M. (2007). Entwicklung eines Konzeptes und eines Prototypen für ein rechnergestütztes Werkzeug zum effizienten Einsatz der Problemlösungsmethodik SPALTEN. Dissertation. In A. Albers (Hrsg.), *Forschungsberichte des IPEK – Institut für Produktentwicklung. Systeme, Methoden, Prozesse* (Bd. 23). Karlsruhe: Universität Karlsruhe (TH)
- Saito, M.; Matsumoto, M. (2008). SIMD-oriented fast mersenne twister: a 128-bit pseudorandom number generator. In *Proceedings of the Seventh International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing* (pp. 607–622), Ulm
- Shao, L.; Mahajan, A.; Schreck, T. (2017). Interactive regression lens for exploring scatter plots. *Computer Graphics Forum*, 36 (3), 157-166
- Sharma, R.; Nori, A.; Aiken, A. (2014). Bias-variance tradeoffs in program analysis. *ACM SIGPLAN Notices*, 49 (1), 127-137

- Siebertz, K.; Bebber, D.; Hochkirchen, T. (2010). *Statistische Versuchsplanung: Design of Experiments (DoE)*, Berlin, Heidelberg: Springer
- Skerra, B. (2000). *Handbuch Molchtechnik*, Essen: Vulkan-Verlag
- Sohlberg, B.; Jacobsen, E. (2008). Grey box modelling – branches and experiences. In *Proceedings of the 17th IFAC World Congress* (pp. 11415–11420), Seoul
- Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 50 (3), 413–436
- Stachowiak, H. (1973). *Allgemeine Modelltheorie*, Wien, New York: Springer
- Strahilov, A. (2015). *Simulation des physikalischen Verhaltens bei der digitalen Absicherung von automatisierten Montageanlagen*, Berlin: VDE Verlag
- Tang, B. (1993). Orthogonal array-based Latin hypercubes. *Journal of the American Statistical Association*, 88 (424), 1392–1397
- Thomas, R.; Davies, C. (1993). Efficient power generation through predictive maintenance. In B. Rao (Ed.), *Profitable condition monitoring* (pp. 133–143), Dordrecht, The Netherlands: Springer
- Trinca, L.; Gilmour, S. (2001). Multistratum response surface designs. *Technometrics*, 43 (1), 25–33
- Tulleken, H. (1991). Application of the grey-box approach to parameter estimation in physicochemical models. In *Proceedings of the 30th IEEE Conference on Decision and Control* (pp. 1177–1183), Brighton, UK
- VDI 2206 (2004). *VDI-Richtlinie: VDI 2206 Entwicklungsmethodik für mechatronische Systeme*. Düsseldorf: VDI-Verlag
- VDI 2221 (2019). *VDI-Richtlinie: VDI 2221 Entwicklung technischer Produkte und Systeme*. Düsseldorf: VDI-Verlag
- Verbeke, G.; Molenberghs, G. (2001). *Linear mixed models for longitudinal data*, New York: Springer
- Vining, G. (2012). Split-plot response surface designs. In Hinkelmann (Ed.), *Design and analysis of experiments: Volume 3; Special designs and applications* (pp. 471–500), Hoboken, NJ: Wiley-Interscience
- Vining, G.; Kowalski, S. (2008). Exact inference for response surface designs within a split-plot structure. *Journal of quality technology*, 40 (4), 394–406
- Vining, G.; Kowalski, S.; Montgomery, D. (2005). Response surface designs within a split-plot structure. *Journal of quality technology*, 37 (2), 115–129
- Voss, D.; Dean, A. (1999). *Design and analysis of experiments*, New York: Springer

- Wang, L. (2006). *Recommendations for Design Parameters for Central Composite Designs with Restricted Randomization*. Blacksburgh, VA: Virginia Polytechnic Institute and State University
- Wang, Y.; Fang, K. (1981). A note on uniform distribution and experimental design. *Kexue Tongbao*, 26 (1), 485–489
- Weerakkody, G.; Johnson, D. (1992). Estimation of within model parameters in regression models with a nested error structure. *Journal of the American Statistical Association*, 87 (419), 708–713
- Wentura, D.; Pospeschill, M. (2015). *Multivariate Datenanalyse: Eine kompakte Einführung*, Wiesbaden: Springer Fachmedien
- Williams, M.; Landel, R.; Ferry, J. (1955). The temperature dependence of relaxation mechanisms in amorphous polymers and other glass-forming liquids. *Journal of the American Chemical Society*, 77 (14), 3701–3707
- Winker, P.; Fang, K. (1997). Application of threshold-accepting to the evaluation of the discrepancy of a set of points. *SIAM Journal on Numerical Analysis*, 34 (5), 2028–2042
- Woods, D.; Marley, C.; Lewis, S. (2013). Designed experiments for semi-parametric models and functional data with a case-study in tribology. In *Proceedings of the 59th ISI World Statistics Congress* (pp.735-740), Hong Kong
- Xia, Y.; Tong, H.; Li, W.; Zhu, L. (2002). An adaptive estimation of dimension reduction space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64 (3), 363–410
- Yuan, F. (2011). *Construction of Estimation-Equivalent Second-Order Split-Split-Plot Designs*. Tuscaloosa, AL: The University of Alabama
- Zeller, P. (2012). *Handbuch Fahrzeugakustik: Grundlagen, Auslegung, Berechnung, Versuch*, Wiesbaden: Vieweg+Teubner
- Zingel, C. (2013). Basisdefinition einer gemeinsamen Sprache der Produktentwicklung im Kontext der Modellbildung technischer Systeme und einer Modellierungstechnik für Zielsystem und Objektsystem technischer Systeme in SysML auf Grundlage des ZHO-Prinzips. Dissertation. In A. Albers (Hrsg.), *Forschungsberichte des IPEK – Institut für Produktentwicklung. Systeme, Methoden, Prozesse* (Bd. 70). Karlsruhe: Karlsruher Institut für Technologie (KIT)

Anhang

Tabelle A.1: Regressionsmodell für das Haftreibungsmoment $M_{fr,s}$

Model fit statistics

AIC	BIC	R ²
54,62	50,32	0,859

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
beta_0	15,955	3,385	4,714	5	0,0053	7,255	24,656
beta_T	-3,363	1,937	-1,736	5	0,1431	-8,343	1,617
beta_TT	3,741	4,642	0,806	5	0,4570	-8,192	15,673
beta_v	3,958	1,937	2,043	5	0,0965	-1,022	8,938
beta_vT	-4,640	5,477	-0,847	5	0,4356	-18,720	9,440
beta_vv	1,076	4,642	0,232	5	0,8258	-10,856	13,009
beta_p	8,937	1,954	4,574	5	0,0060	3,914	13,960
beta_pT	-2,054	5,577	-0,368	5	0,7277	-16,390	12,282
beta_pv	-0,172	5,577	-0,031	5	0,9766	-14,508	14,164
beta_pp	4,715	4,636	1,017	5	0,3559	-7,204	16,633

Residuals

StdDev	3,873
--------	-------

Tabelle A.2: Regressionsmodell für das Gleitreibungsmoment $M_{fr,d}$

Model fit statistics

AIC	BIC	R ²
46,71	42,42	0,923

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
beta_0	7,980	1,535	5,198	5	0,0035	4,033	11,927
beta_T	-4,943	0,879	-5,625	5	0,0025	-7,202	-2,684
beta_TT	2,662	2,106	1,264	5	0,2618	-2,751	8,075
beta_v	2,298	0,879	2,615	5	0,0474	0,039	4,557
beta_vT	-2,724	2,485	-1,096	5	0,3229	-9,111	3,663
beta_vv	-0,673	2,106	-0,320	5	0,7621	-6,086	4,740
beta_p	3,432	0,886	3,871	5	0,0117	1,153	5,710
beta_pT	-1,232	2,530	-0,487	5	0,6470	-7,735	5,271
beta_pv	0,050	2,530	0,020	5	0,9850	-6,453	6,553
beta_pp	2,148	2,103	1,021	5	0,3540	-3,259	7,554

Residuals

StdDev	1,757
--------	-------

Tabelle A.3: Regressionsmodell für die Zeitkonstante τ (normiert)**Model fit statistics**

AIC	BIC	R ²
81,33	77,03	0,995

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
beta_0	1,0000	0,0134	74,827	5	8,08E-09	0,9656	1,0343
beta_T	-0,1702	0,0076	-22,252	5	3,40E-06	-0,1899	-0,1505
beta_TT	-0,2544	0,0183	-13,878	5	3,49E-05	-0,3015	-0,2073
beta_v	-0,0130	0,0076	-1,704	5	0,1492	-0,0327	0,0066
beta_vT	0,0508	0,0216	2,350	5	0,0655	-0,0048	0,1064
beta_vv	0,0340	0,0183	1,856	5	0,1227	-0,0131	0,0811
beta_p	0,0457	0,0077	5,926	5	0,0020	0,0259	0,0656
beta_pT	-0,0251	0,0220	-1,142	5	0,3054	-0,0817	0,0315
beta_pv	0,0027	0,0220	0,121	5	0,9084	-0,0539	0,0593
beta_pp	-0,0333	0,0183	-1,821	5	0,1282	-0,0804	0,0137

Residuals

StdDev	0,0153
--------	--------

Tabelle A.4: Regressionsmodell für die charakteristische Wegstrecke D (normiert)**Model fit statistics**

AIC	BIC	R ²
-19,71	-24,01	0,948

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
beta_0	1,0000	0,0432	23,172	5	2,79E-06	0,8891	1,1109
beta_T	0,0866	0,0247	3,505	5	0,0172	0,0231	0,1501
beta_TT	0,2086	0,0592	3,524	5	0,0169	0,0564	0,3607
beta_v	0,0352	0,0247	1,427	5	0,2131	-0,0283	0,0987
beta_vT	-0,0744	0,0698	-1,065	5	0,3357	-0,2539	0,1052
beta_vv	0,1369	0,0592	2,314	5	0,0686	-0,0152	0,2891
beta_p	0,1640	0,0249	6,581	5	0,0012	0,0999	0,2280
beta_pT	0,2785	0,0711	3,916	5	0,0112	0,0957	0,4613
beta_pv	0,0341	0,0711	0,480	5	0,6516	-0,1487	0,2169
beta_pp	0,2107	0,0591	3,564	5	0,0162	0,0587	0,3626

Residuals

StdDev	0,0494
--------	--------

Tabelle A.5: Zur Untersuchung des Schwenkantriebs eingesetzter Versuchsplan

V.Nr.	T	Δp	v	t_s	V.Nr.	T	Δp	v	t_s	V.Nr.	T	Δp	v	t_s
1	1	0	0	t_4	56	0,5	-0,5	0,5	t_1	111	-0,5	-0,5	-0,5	t_6
2	1	0	0	t_{11}	57	0,5	-0,5	0,5	t_3	112	-0,5	-0,5	-0,5	t_{10}
3	1	0	0	t_7	58	0,5	-0,5	0,5	t_{10}	113	-0,5	-0,5	-0,5	t_1
4	1	0	0	t_{10}	59	0,5	-0,5	0,5	t_5	114	-0,5	-0,5	-0,5	t_5
5	1	0	0	t_5	60	0,5	-0,5	0,5	t_2	115	-0,5	-0,5	-0,5	t_3
6	1	0	0	t_8	61	0,5	-0,5	0,5	t_4	116	-0,5	-0,5	-0,5	t_4
7	1	0	0	t_6	62	0,5	-0,5	0,5	t_6	117	-0,5	-0,5	-0,5	t_8
8	1	0	0	t_3	63	0,5	-0,5	0,5	t_9	118	-0,5	-0,5	-0,5	t_9
9	1	0	0	t_9	64	0,5	-0,5	0,5	t_7	119	-0,5	-0,5	-0,5	t_{11}
10	1	0	0	t_1	65	0,5	-0,5	0,5	t_8	120	-0,5	-0,5	-0,5	t_7
11	1	0	0	t_2	66	0,5	-0,5	0,5	t_{11}	121	-0,5	-0,5	-0,5	t_2
12	0	1	0	t_9	67	0	0	0	t_{10}	122	-1	0	0	t_8
13	0	1	0	t_4	68	0	0	0	t_5	123	-1	0	0	t_{10}
14	0	1	0	t_8	69	0	0	0	t_4	124	-1	0	0	t_3
15	0	1	0	t_6	70	0	0	0	t_8	125	-1	0	0	t_7
16	0	1	0	t_{11}	71	0	0	0	t_6	126	-1	0	0	t_6
17	0	1	0	t_2	72	0	0	0	t_{11}	127	-1	0	0	t_4
18	0	1	0	t_1	73	0	0	0	t_2	128	-1	0	0	t_9
19	0	1	0	t_7	74	0	0	0	t_1	129	-1	0	0	t_5
20	0	1	0	t_3	75	0	0	0	t_9	130	-1	0	0	t_2
21	0	1	0	t_5	76	0	0	0	t_3	131	-1	0	0	t_{11}
22	0	1	0	t_{10}	77	0	0	0	t_7	132	-1	0	0	t_1
23	0	0	1	t_2	78	0,5	-0,5	-0,5	t_8	133	0,5	0,5	-0,5	t_9
24	0	0	1	t_1	79	0,5	-0,5	-0,5	t_9	134	0,5	0,5	-0,5	t_2
25	0	0	1	t_7	80	0,5	-0,5	-0,5	t_1	135	0,5	0,5	-0,5	t_5
26	0	0	1	t_3	81	0,5	-0,5	-0,5	t_2	136	0,5	0,5	-0,5	t_6
27	0	0	1	t_5	82	0,5	-0,5	-0,5	t_6	137	0,5	0,5	-0,5	t_8
28	0	0	1	t_8	83	0,5	-0,5	-0,5	t_3	138	0,5	0,5	-0,5	t_1
29	0	0	1	t_{10}	84	0,5	-0,5	-0,5	t_{10}	139	0,5	0,5	-0,5	t_{11}
30	0	0	1	t_{11}	85	0,5	-0,5	-0,5	t_5	140	0,5	0,5	-0,5	t_4
31	0	0	1	t_4	86	0,5	-0,5	-0,5	t_{11}	141	0,5	0,5	-0,5	t_{10}
32	0	0	1	t_9	87	0,5	-0,5	-0,5	t_4	142	0,5	0,5	-0,5	t_7
33	0	0	1	t_6	88	0,5	-0,5	-0,5	t_7	143	0,5	0,5	-0,5	t_3
34	-0,5	-0,5	0,5	t_6	89	0	-1	0	t_{10}	144	0	0	-1	t_7
35	-0,5	-0,5	0,5	t_9	90	0	-1	0	t_4	145	0	0	-1	t_2
36	-0,5	-0,5	0,5	t_3	91	0	-1	0	t_7	146	0	0	-1	t_4
37	-0,5	-0,5	0,5	t_1	92	0	-1	0	t_6	147	0	0	-1	t_8
38	-0,5	-0,5	0,5	t_8	93	0	-1	0	t_3	148	0	0	-1	t_9
39	-0,5	-0,5	0,5	t_{11}	94	0	-1	0	t_9	149	0	0	-1	t_{10}
40	-0,5	-0,5	0,5	t_7	95	0	-1	0	t_5	150	0	0	-1	t_1
41	-0,5	-0,5	0,5	t_5	96	0	-1	0	t_{11}	151	0	0	-1	t_5
42	-0,5	-0,5	0,5	t_2	97	0	-1	0	t_2	152	0	0	-1	t_3
43	-0,5	-0,5	0,5	t_4	98	0	-1	0	t_8	153	0	0	-1	t_6
44	-0,5	-0,5	0,5	t_{10}	99	0	-1	0	t_1	154	0	0	-1	t_{11}
45	-0,5	0,5	0,5	t_2	100	0,5	0,5	0,5	t_6	155	-0,5	0,5	-0,5	t_{11}
46	-0,5	0,5	0,5	t_7	101	0,5	0,5	0,5	t_5	156	-0,5	0,5	-0,5	t_9
47	-0,5	0,5	0,5	t_5	102	0,5	0,5	0,5	t_{10}	157	-0,5	0,5	-0,5	t_2
48	-0,5	0,5	0,5	t_6	103	0,5	0,5	0,5	t_2	158	-0,5	0,5	-0,5	t_1
49	-0,5	0,5	0,5	t_3	104	0,5	0,5	0,5	t_{11}	159	-0,5	0,5	-0,5	t_7
50	-0,5	0,5	0,5	t_4	105	0,5	0,5	0,5	t_9	160	-0,5	0,5	-0,5	t_8
51	-0,5	0,5	0,5	t_1	106	0,5	0,5	0,5	t_1	161	-0,5	0,5	-0,5	t_{10}
52	-0,5	0,5	0,5	t_{10}	107	0,5	0,5	0,5	t_7	162	-0,5	0,5	-0,5	t_4
53	-0,5	0,5	0,5	t_9	108	0,5	0,5	0,5	t_3	163	-0,5	0,5	-0,5	t_5
54	-0,5	0,5	0,5	t_{11}	109	0,5	0,5	0,5	t_4	164	-0,5	0,5	-0,5	t_3
55	-0,5	0,5	0,5	t_8	110	0,5	0,5	0,5	t_8	165	-0,5	0,5	-0,5	t_6

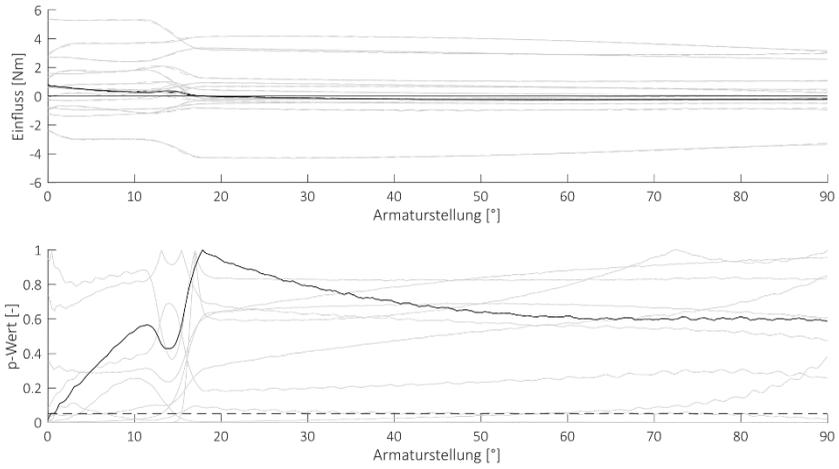


Abbildung A.1: Quadratischer Einfluss der Druckdifferenz

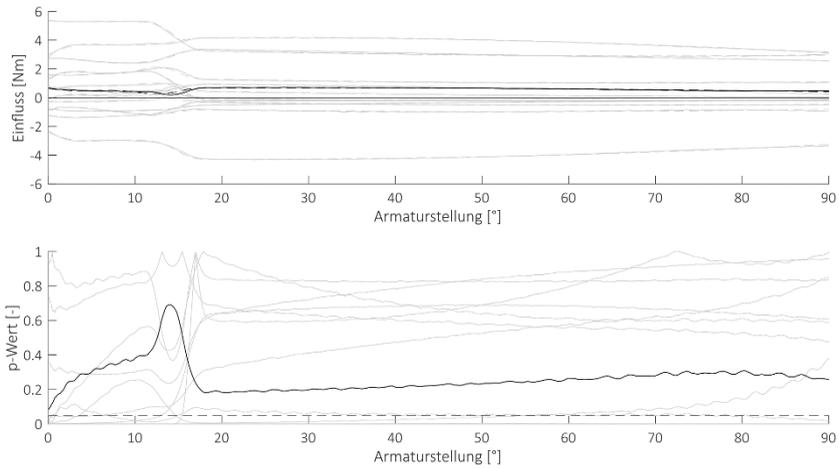


Abbildung A.2: Wechselwirkung zwischen Geschwindigkeit und Temperatur

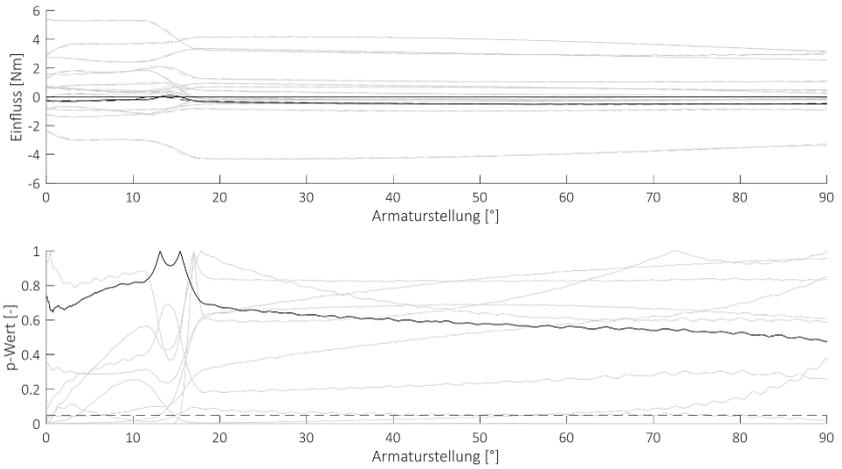


Abbildung A.3: Wechselwirkung zwischen Geschwindigkeit und Schaltspielzahl

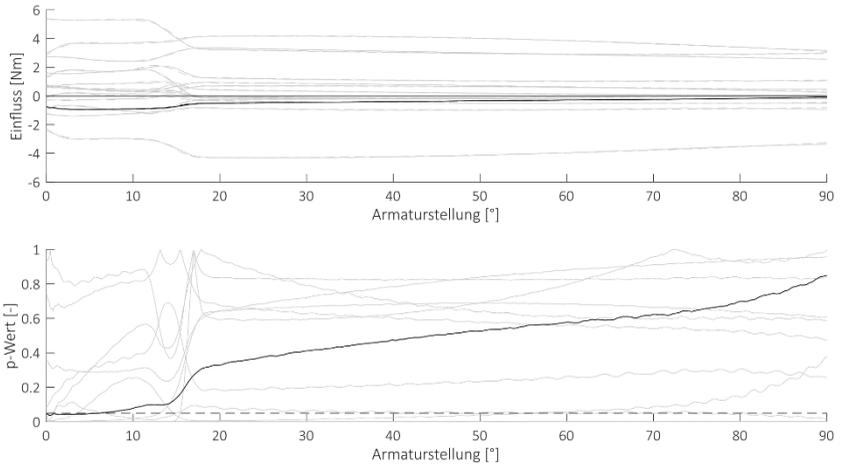


Abbildung A.4: Wechselwirkung zwischen Temperatur und Schaltspielzahl

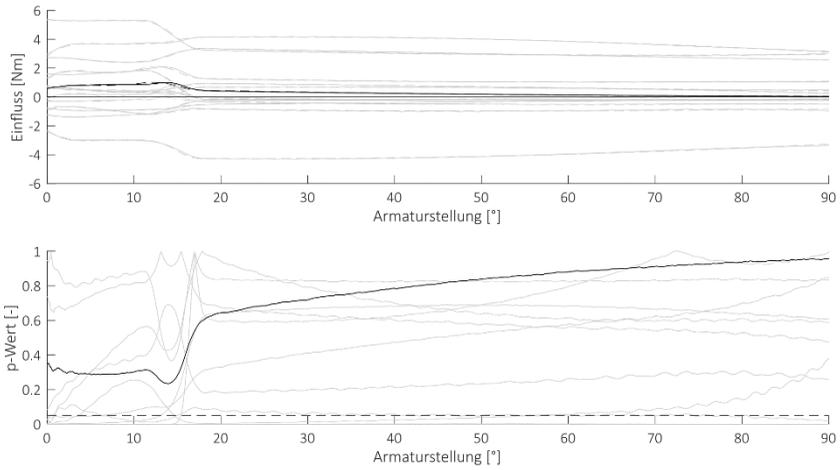


Abbildung A.5: Wechselwirkung zwischen Druck und Schaltspielzahl